



ÉCOLE DOCTORALE DE MANAGEMENT
PANTHÉON-SORBONNE

PÔLE DE RECHERCHES INTERDISCIPLINAIRES EN SCIENCES DU MANAGEMENT

Essays on Asset Price Dynamics Modeling

Par LÉO PARENT

Thèse de doctorat en SCIENCES DE GESTION

Dirigée par JEAN-PAUL LAURENT

Soutenue le 5 Décembre 2024 devant un jury composé de :

M. EDUARDO ABI JABER	École Polytechnique	Rapporteur
M. JEAN-PHILIPPE BOUCHAUD	Capital Fund Management	Rapporteur
M. JEAN-PAUL LAURENT	Université Paris 1 Panthéon-Sorbonne	Examinateur
M. CHARLES-ALBERT LEHALLE	École Polytechnique	Examinateur
M. YANNICK MALEVERGNE	Université Paris 1 Panthéon-Sorbonne	Examinateur
M. THIERRY RONCALLI	Amundi Asset Management	Examinateur

Remerciements

Mes premiers remerciements s'adressent naturellement à mon directeur de thèse, Jean-Paul Laurent, sans qui cette thèse n'aurait pas vu le jour. Ces remerciements sont d'autant plus profonds que Monsieur Laurent a accepté de me prendre sous sa direction alors même que rien ne le prédisposait a priori à le faire. En effet, je n'avais jamais été son élève ni ne l'avais rencontré auparavant, et de surcroît, ce projet de thèse constituait une reprise d'études deux ans après l'obtention de mon Master. Malgré cela, Monsieur Laurent a généreusement pris de son temps pour considérer ce projet, et m'a finalement fait l'honneur d'accéder à ma demande. Je le remercie aussi pour l'ensemble des conseils prodigués durant ces trois années, ainsi que pour la grande liberté qu'il m'a accordée dans mes recherches.

Je tiens également à exprimer ma profonde reconnaissance à Eduardo Abi Jaber et Jean-Philippe Bouchaud d'avoir accepté d'être mes rapporteurs de thèse, et je suis profondément honoré que mon travail soit évalué par des chercheurs de cette envergure. Je les remercie également pour leurs observations et suggestions constructives formulées lors de la pré-soutenance, qui ont grandement contribué à enrichir et affiner cette thèse.

Je remercie aussi Charles-Albert Lehalle et Yannick Malevergne d'avoir accepté d'examiner ma thèse et de participer au jury de soutenance.

Je souhaite également exprimer ma profonde gratitude à Thierry Roncalli, également membre du jury, qui m'a transmis le goût de la recherche en finance quantitative, un intérêt qui a été déterminant dans ma décision de poursuivre un doctorat. Thierry a aussi été un soutien précieux tout au long de cette thèse, de son état embryonnaire de projet, aux différentes étapes de sa réalisation.

Je remercie aussi Abida Saidyassine, qui a toujours fait preuve d'une grande disponibilité pour répondre et traiter les diverses requêtes administratives attenantes à la vie d'un chercheur doctorant.

Je remercie enfin mes parents ainsi que mon frère Romain pour leurs encouragements et leur présence durant ces trois années de doctorat.

Résumé

Cette thèse considère différentes problématiques de modélisation des dynamiques de prix d'actifs.

Après un chapitre introductif exposant l'approche générale dans laquelle la thèse s'inscrit, le deuxième chapitre propose une extension du modèle de Heston, qui surmonte certaines limites structurelles de sa forme standard. Il est alors montré que ce modèle reproduit une part majeure des propriétés des dynamiques jointes de prix et de volatilité.

Le chapitre suivant introduit une classe de modèles permettant de capturer à la fois le caractère *rough* de la volatilité et sa *path-dependence*. Au travers d'une étude empirique et d'expériences numériques, des éléments de réponse sont ensuite apportés à la question : quelles modélisations au sein de cette classe sont congruentes avec les données de marché ?

Le quatrième chapitre explore les implications de diverses spécifications d'un certain modèle de volatilité path-dependent. L'impact de la tendance du prix sur la dynamique de la volatilité dans ce cadre est alors mis en évidence.

Le cinquième chapitre introduit une nouvelle méthode d'estimation pour les modèles à volatilité stochastique, orientée vers un objectif de prédiction. L'efficacité de cette approche est évaluée, d'abord sur données synthétiques, puis sur données de marché.

Le sixième chapitre développe enfin un modèle multivarié de dynamique des prix d'actifs dans un cadre factoriel et de marché path-dependent. De ce modèle est dérivé un générateur de marché, dont l'efficacité à produire des séries financières réalistes est démontrée au travers de différentes évaluations.

Mots clés : *Dynamiques de prix d'actifs, Modèle de volatilité stochastique, Modèle de Heston, Effet Zumbach, Volatilité path-dependent, Volatilité rugueuse, Deep estimation, Générateur de marché.*

Abstract

This thesis considers different issues related to modeling asset price dynamics.

After an introductory chapter outlining the general approach in which the thesis is situated, the second chapter proposes an extension of the Heston model, which overcomes certain structural limitations of its standard form. It is then shown that this model reproduces a major part of the properties of the joint dynamics of price and volatility.

The following chapter introduces a class of models that captures both the *rough* nature of volatility and its *path-dependence*. Through an empirical study and numerical experiments, answers are subsequently provided to the question: which modelings within this class are consistent with market data?

The fourth chapter explores the implications of various specifications of a certain path-dependent volatility model. The impact of price trends on the dynamics of volatility within this framework is then highlighted.

The fifth chapter introduces a new estimation method for stochastic volatility models, aimed at a prediction objective. The effectiveness of this approach is evaluated first on synthetic data and then on market data.

Finally, the sixth chapter develops a multivariate model of asset price dynamics within a factor and path-dependent market framework. From this model, a market generator is derived, whose effectiveness in producing realistic financial series is demonstrated through various evaluations.

Keywords: *Asset price dynamics, Stochastic volatility model, Heston model, Zumbach effect, Path-dependent volatility, Rough volatility, Deep estimation, Market generator.*

Contents

1	Introduction	1
1.1	Introduction générale	1
1.2	Propriétés empiriques et modélisation des prix d’actifs	2
1.2.1	Phénoménologie des dynamiques de prix d’actifs	2
1.2.2	Cadre général de la modélisation en temps continu	4
1.2.3	Volatilité et dynamiques de prix	8
1.3	Structuration de la thèse	11
1.3.1	The EWMA Heston Model	11
1.3.2	Rough Path-Dependent Volatility Models	12
1.3.3	Price and Volatility Dynamics Under Path-Dependent Models	14
1.3.4	Deep Estimation for Volatility Forecasting	15
1.3.5	The Factorial Path-Dependent Market Model	17
2	The EWMA Heston Model	20
2.1	Introduction	20
2.2	The EWMA Heston model	22
2.2.1	Presentation of the model	22
2.2.2	The thresholded version of the model as a limit-case	25
2.2.3	Comparison with volatility models based on quadratic Hawkes processes	27
2.2.4	Model discretization scheme	29
2.3	Consistency of the model with market data	30
2.3.1	The relationship between EWMA of returns and the volatility	31
2.3.2	The dynamics of price and volatility	32
2.3.3	Volatility and log-returns distributions for different time horizons	39
2.3.4	The time-reversal asymmetry	45
2.4	Conclusion	46
2.5	Acknowledgments	47
2.A	Convergence of the attraction volatility under given specification	48
2.B	The model calibration procedure	49
2.C	Standard deviation of an EWMA of a Brownian motion	50
2.D	Complementary results	51

3	Rough Path-Dependent Volatility Models	53
3.1	Introduction	53
3.2	Rough path-dependent volatility models	55
3.2.1	A general form of path-dependent volatility model	55
3.2.2	Rough path-dependent volatility models	58
3.2.3	Markovian multi-factor approximation of the RPDV model	59
3.3	Comparison of model specifications	62
3.3.1	The tested RPDV models	62
3.3.2	Comparative regression results	64
3.4	Numerical experiments	70
3.4.1	Experimental setup	70
3.4.2	Results of numerical experiments	71
3.5	Conclusion	76
3.6	Acknowledgments	78
3.A	Stochastic differential equations for the Markovian approximation of the RPDV model	79
3.B	Numerical quadratures	79
3.C	The ARPDV model as a simulation device	80
3.D	Estimation of the minimum value of the standard deviation of the log-returns of volatility	82
3.E	Exponential-OU process	83
3.F	The model used for numerical experiments	83
4	Price and Volatility Dynamics Under Path-Dependent Models	84
4.1	Introduction	84
4.2	The function θ as a constant	87
4.2.1	The case $\lambda_2 = 0$	88
4.2.2	The case $\lambda_2 > 0$	89
4.3	The function θ as a path-dependent process	92
4.4	The function θ as a long-run coherence operator between price and value	94
4.4.1	The attraction volatility as an adjustment variable	96
4.4.2	The risk premium as an adjustment variable	98
4.5	Conclusion	99
4.A	Highlighting the mean-reverting component of the volatility process	101
4.B	Zero of the function f in polynomial drift hypothesis	101
5	Deep Estimation for Volatility Forecasting	103
5.1	Introduction	103
5.2	Exposition of the estimation problem	105
5.2.1	The model to be estimated	105

5.2.2	The Bayesian estimation problem to solve: a forecasting objective-based estimation problem	107
5.2.3	The loss function: a sum of proxy divergence measures	109
5.3	Construction of the Bayesian estimator function	111
5.3.1	The neural network as a proxy for function M	111
5.3.2	The estimator function	116
5.3.3	The estimation procedure	120
5.4	Assessment of the estimation method	126
5.4.1	Evaluation of estimation method using synthetic data	126
5.4.2	Evaluation of estimation procedure using market data	132
5.5	Conclusion	138
5.6	Acknowledgments	139
5.A	Approximation of the considered RPDV model	140
5.A.1	Stochastic differential equations for the Markovian approximation of the considered RPDV model	140
5.A.2	Approximation of the power law kernel	141
5.B	Parameters of the log-normal approximation to the conditional volatility distribution	143
5.C	Proofs of convergence results	143
5.C.1	Convergence of the estimation procedure of the function \mathcal{M}	143
5.C.2	Convergence of the estimation procedure of the estimator function Θ	144
5.D	Annex results	146
5.D.1	Value of the integral of \hat{K} over $\mathbb{R}_{\geq 0}$	146
5.D.2	Standard deviation of a BSS process	146
5.D.3	The variance of the volatility process	146
6	The Factorial Path-Dependent Market Model	147
6.1	Introduction	147
6.2	General framework of the factorial Path-Dependent Market Model	149
6.2.1	General framework of the FPDM model	149
6.2.2	Volatilities and drifts of elementary factors	153
6.3	The Factorial Path-Dependent Market Generator	159
6.3.1	The market generator framework	159
6.3.2	Calibration of the market generator	164
6.4	Empirical assessment of the FPDM generator performance	171
6.4.1	Modalities of the conducted assessments	171
6.4.2	General properties of generated datasets	173
6.4.3	Evaluation of strategy features replication	179
6.5	Conclusion	191
6.A	List of notations	194
6.B	Stochastic differential equations involved in the FPDM model	195
6.B.1	Calculation of the asset price vector solution	195

6.B.2	EWMA estimators and their stochastic differential equations	195
6.C	Important specifications of the FPDM model	196
6.C.1	The 4-factor PDV as a specific case of the FPDM model	196
6.C.2	Definition of the factorial drift vector in the context of the CAPM	197
6.D	Approximation of the Wasserstein distance for eigenvalues	198
6.E	Normal log-normal mixture	198
6.E.1	First four moments of a normal log-normal mixture of the form $We^{s(B-s)}$. .	198
6.E.2	First four moments of an i.i.d. sample from a normal log-normal mixture . .	198
6.E.3	Estimation of the vector S using the method of moments	199
6.F	Estimation of the market sensitivity operator by maximum likelihood	200
6.G	Additional results of the numerical experiment	201
6.G.1	Fitted kernels	201
6.G.2	The market factor	202
6.G.3	The market sensitivity operator	203
6.G.4	Some distributions of individual stock returns	204
6.G.5	Cumulative returns of the considered investment strategies	207

7 Conclusion

*"Valuation itself is the treasure and jewel of the valued things.
Through valuation only is there value; and without valuation
the nut of existence would be hollow."*

Thus Spoke Zarathustra, Friedrich Nietzsche

CHAPTER 1

Introduction

1.1 Introduction générale

L'importance des choix de modélisation financière a été rendue manifeste ces dernières décennies par un certain nombre d'épisodes marquants. Parmi les plus fameux figure la chute de Long-Term Capital Management (LTCM), qui manqua d'engendrer en 1998 une crise systémique majeure. À l'origine de cette débâcle, des décisions basées sur l'utilisation de modèles structurellement inaptes à saisir les risques réellement encourus ([128], [187]). De manière analogue, les défaillances des modèles d'estimation des probabilités de défaut ont joué un rôle significatif dans les octrois de crédits ayant abouti en 2008 à la crise financière des subprimes ([175]).

Si ces événements historiques mettent en évidence les dangers d'une formalisation inadaptée, la modélisation financière n'en reste pas moins une nécessité impérieuse. Tant la mesure des différents risques financiers que la couverture d'options, en passant par la construction de portefeuilles, requièrent des outils quantitatifs d'objectivation du réel et de ses potentiels devenir. Aussi, de par la nature aléatoire des marchés financiers - a minima au regard du référentiel d'information des agents¹ - cette objectivation s'effectue de manière générale à partir de modèles probabilistes qui permettent d'affecter à un ensemble d'états du monde une mesure de probabilité. Si les problématiques de modélisation financière auxquelles viennent répondre ces modèles sont multiples, quelques-unes se distinguent par leur centralité, parmi lesquelles figure la modélisation des dynamiques de prix d'actifs. En effet, du calcul de mesures de risque ou de performance de portefeuilles, au traitement des problématiques d'évaluation ou de couverture d'options, l'ensemble de ces opérations dépendent, de manière directe ou implicite, d'hypothèses sur les mouvements de prix d'actifs. Par conséquent, la question de la formalisation de ces dynamiques surdétermine les réponses à apporter à tout un ensemble de problématiques de finance quantitative connexes.

¹La précision apportée souligne qu'il ne s'agit pas ici de trancher entre l'hypothèse objectiviste d'un hasard ontologique des marchés financiers, et l'hypothèse subjectiviste qui suppose que le système financier constitue, en dernière analyse, rigoureusement déterministe. Il est simplement entendu ici par "nature aléatoire des marchés financiers", qu'à partir du niveau dont disposent les agents, l'état du futur des marchés financiers ne peut être prédit de manière certaine.

Fort de cette importance, la présente thèse vise à proposer des modélisations des dynamiques des prix d'actifs, capables de rendre compte des caractéristiques empiriques des séries financières dans toute leur complexité. En cela, l'approche adoptée dans cette thèse diffère d'une part majeure de la littérature financière, qui subordonne les choix de modélisation à la possibilité d'en dériver des formules analytiques fermées ou semi-fermées destinées à répondre à des problématiques d'ingénierie financière spécifiques. Ici, la primauté sera accordée à la capture des différentes propriétés empiriques des historiques de prix d'actifs plutôt qu'à la commodité mathématique des outils quantitatifs mobilisés. Pour cette même raison, les modèles introduits seront généralement évalués par comparaison entre données simulées et données historiques. Point important, ces différentes évaluations menées tout au long de la thèse portent uniquement sur des actifs du marché action (indice et action individuelle). Cependant, de par les propriétés communes des dynamiques de prix du marché actions à celles d'autres classes d'actifs (commodities, obligations, indices, etc.), les résultats relatifs aux modélisations considérées seront remobilisables au-delà de ce cadre ([70]).

Afin d'appréhender son objet, la présente thèse adopte une approche inductive. Il s'agit donc de partir des réalités phénoménales² des dynamiques de prix afin d'en identifier les traits les plus saillants, pour en induire des choix de modélisation. En conséquence, ce chapitre commencera par exposer les principales propriétés empiriques des dynamiques de prix d'actifs, puis les choix de modélisation adoptés tout au long de la thèse qui en découleront. La section suivante présentera ensuite la structure générale de la thèse et le contenu de chacun de ses chapitres.

1.2 Propriétés empiriques et modélisation des prix d'actifs

1.2.1 Phénoménologie des dynamiques de prix d'actifs

1.2.1.1 Les distributions univariées de rendements

Un premier ensemble de propriétés caractéristiques des séries financières porte sur les distributions de rendements. Pour les horizons de court à moyen terme (haute-fréquence à fréquence mensuelle), celles-ci sont fondamentalement non gaussiennes et présentent en particulier des queues lourdes. Si leur forme exacte reste difficile à déterminer de manière précise, ces queues de distributions sont proches d'un comportement parétien ou en loi puissance pour les horizons haute fréquence à quelques jours ([70], [144]) et sont assez bien capturées par une loi de Student ([45]). Cont ([70]) montre en outre que les coefficients de queue observés empiriquement tendent à invalider l'hypothèse initialement formulée par Mandelbrot ([146], [147]) selon laquelle les rendements suivent un processus de variance infinie. Par ailleurs, pour une part significative des actifs, les distributions

²Par opposition aux approches nouménales des dynamiques de prix d'actifs, telle que celles prenant pour point de départ la notion de "valeur fondamentale".

de rendements sont asymétriques à différentes échelles de temps. Pour les indices boursiers par exemple, le skewness de la distribution des rendements tend à devenir de plus en plus négatif en fonction de l'échelle de temps jusqu'à atteindre un pic (négatif), après quoi le skewness converge lentement vers zéro. Un phénomène analogue de convergence s'observe pour la leptokurticité des distributions de rendements, à ceci près que ce mouvement semble pour sa part suivre une relation décroissante homogène. D'un niveau très élevé pour les rendements de haute fréquence, les queues de distribution tendent à se normaliser avec l'augmentation de l'échelle de temps. De manière plus générale, les distributions de rendement semblent converger très lentement vers un comportement gaussien ([169], [47]). Concrètement donc, les distributions de rendements annuels sont plus proches d'un comportement gaussien que ne le sont les distributions des rendements journaliers.

1.2.1.2 Les caractéristiques trajectorielles des prix d'actifs

Les variations de prix d'actifs considérées sous l'angle trajectoriel sont caractérisées là encore par un certain nombre de propriétés remarquables. Premièrement, l'autocorrélation linéaire des rendements est globalement non significative ([70]), ce qui implique concrètement que les mouvements de prix sont difficilement anticipables à court terme. Cette propriété n'est cependant pas synonyme d'indépendance des rendements. En effet, leur valeur absolue (ou leur carré) présente une autocorrélation forte et persistente. Ce phénomène connu sous le nom de *clusters de volatilité* ou "*syndrome de Joseph*", qui fut mis en évidence par Mandelbrot ([146], [148]), constitue l'une des caractéristiques des séries financières les plus importantes, puisqu'elle explique dans une large mesure la non-gaussianité des distributions de rendements soulignée en section 1.2.1.1. Ainsi, les distributions de rendements journaliers ajustés de cet effet tendent à se rapprocher d'une distribution normale ([41]). Il faut cependant noter que cet ajustement opéré à partir de méthodes standards³ n'annihile pas complètement la présence de queues épaisses de distributions ([70]) du fait de la survenue brutale d'événements extrêmes, généralement baissiers. Ce type d'événements, qualifié par Mandelbrot de "*syndrome de Noé*" ([148]), semble également appartenir aux invariants des dynamiques financières. Ainsi, dès 1891, Zola décrivait par le menu une occurrence de ce phénomène, narrant la chute de la valeur de l'action de la banque Universelle, dont le cours de 2400 francs à l'ouverture termina à 830 francs à la clôture⁴ ([202]). Si ce type d'événements en constitue la manifestation la plus spectaculaire, l'augmentation du niveau de variabilité des prix conjointe aux tendances baissières reste également vérifiée dans des situations de marché plus conventionnelles. En effet, en plus d'être positivement autocorrélés, les carrés des rendements sont également négativement corrélés aux rendements simples pour la plupart des actifs. Ce phénomène, généralement désigné par *leverage effect*⁵, implique donc que le niveau de volatilité tend à augmenter lorsque le prix diminue ([35], [44]). En termes distributifs, cette relation permet d'expliquer - au moins partiellement - le skewness négatif des distributions de rendements évoqué en section 1.2.1.1 ([121], [114]). Un phénomène empirique connexe au leverage effect est l'effet Zumbach ([203], [204], [205]).

³Typiquement par un modèle de type GARCH à processus d'innovation gaussien.

⁴Si la banque dont fait l'objet ce récit (tiré du roman *L'Argent*) est fictive, celui-ci s'inspire dans une large mesure du krach boursier de Paris de 1882 ([176]).

⁵L'appellation *leverage effect* ne se réfère pas ici à son sens restreint qui explique ce phénomène par un effet de modification de structure du capital ([97]).

La première dimension de ce phénomène, qualifiée par Gatheral *et al.* ([106]) d'*effet Zumbach faible*, désigne le phénomène empirique selon lequel "les carrés des rendements passés prévoient mieux la volatilité intégrée futures que la volatilité intégrée passée ne prévoit les carrés des rendements futurs". Cette propriété constitue une caractéristique notable des dynamiques de prix du fait de sa non trivialité ([39]). La seconde dimension de l'effet Zumbach, l'*effet Zumbach fort*, consiste en ce que "la dynamique de la volatilité conditionnelle au passé dépend non seulement de la trajectoire passée de la volatilité, mais aussi de la trajectoire historique du prix". Par conséquent, la relation entre dynamiques des prix et dynamiques de la volatilité excède le simple leverage effect ([44]).

1.2.1.3 Propriétés des dynamiques multivariées de prix d'actifs

À l'instar des dynamiques individuelles, les dynamiques multivariées de prix d'actifs présentent un certain nombre de propriétés notables. Tout d'abord, les rendements individuels des titres au sein d'une même classe d'actifs sont significativement corrélés, rendant impropre l'argument du théorème central limite appliqué aux distributions de rendements de portefeuilles. Cette propriété, associée aux caractéristiques des rendements individuels, conduit effectivement à des distributions de rendements fortement non gaussiennes pour les portefeuilles uniclasse, typiquement les indices boursiers ([169]). Par ailleurs, l'analyse spectrale des matrices de rendements révèle que cette structure de corrélation est engendrée dans une très large mesure par un petit nombre de facteurs ([135]). Le plus important, qui peut être appréhendé comme le facteur de marché, occupe une place prépondérante et permet à lui seul d'expliquer une part très significative des covariations de prix ([68]). En outre, au-delà même des propriétés attenantes à leurs marges, les distributions multivariées de rendements admettent des caractéristiques remarquables relatives à la structure de corrélation. Chicheportiche et Bouchaud ([64]) ont notamment mis en évidence la non-ellipticité des rendements d'actions. En particulier, les structures de corrélation elliptiques induites par la copule normale ou la copule de Student échouent à capturer de manière satisfaisante les caractéristiques empiriques des paires de rendements faiblement corrélés. Autre propriété majeure : les structures de corrélation des univers d'investissement sont dynamiques. Cet état de fait se matérialise par la coexistence de périodes où certaines paires d'actifs sont faiblement corrélées, avec d'autres périodes où ces mêmes paires présentent des corrélations proches de 1. Typiquement, pour le marché action, les corrélations (linéaires) moyennes entre actifs sont modérées lors des phases d'expansion, et subissent un très fort mouvement de recorrélation lorsque se matérialise un risque systémique ([68]). De ce fait, la matrice de corrélation des rendements est insuffisante pour rendre compte de la complexité de la structure de dépendance des actifs.

1.2.2 Cadre général de la modélisation en temps continu

1.2.2.1 Le mouvement brownien comme brique élémentaire de la modélisation

Depuis l'émergence de la formalisation financière initiée par les travaux précurseurs de Louis Bachelier ([13]), la modélisation des dynamiques de prix d'actifs en temps continu s'est majoritairement construite autour d'un objet mathématique particulier : le mouvement brownien. Formellement, un

mouvement brownien standard B_t est un processus stochastique défini par les propriétés suivantes :

1. $B_0 = 0$ presque sûrement.
2. les incréments de B sont indépendants : pour tout $0 \leq s < t$, l'accroissement $B_t - B_s$ est indépendant de $\{B_u : 0 \leq u \leq s\}$.
3. les incréments de B sont stationnaires et gaussiens : pour tout $0 \leq s < t$, $B_t - B_s \sim \mathcal{N}(0, t-s)$, où \mathcal{N} désigne la distribution normale.
4. les trajectoires de B_t sont continues presque sûrement.

De par ces caractéristiques, le mouvement brownien constitue un outil commode pour rendre compte des dynamiques aléatoires des variations de prix. Tout d'abord, la propriété d'indépendance des incréments a permis une formalisation en temps continu de la notion d'efficience informationnelle des marchés selon laquelle le prix actuel d'un actif est le meilleur prédictor de son prix futur actualisé ([185], [96]). De manière plus pratique, le caractère martingale du mouvement brownien constitue un outil puissant pour modéliser l'absence de corrélation linéaire des rendements évoqués en section 1.2.1.2. Par ailleurs, la propriété d'indépendance des incréments permet de construire des modèles de dynamiques de prix distinguant clairement la composante martingale de la composante déterministe (décomposition de Doob-Meyer).

L'utilisation du mouvement brownien dans la formalisation des dynamiques de prix d'actifs a cependant fait l'objet d'un certain nombre de critiques. Parmi les plus fameuses figurent celles adressées par Benoit Mandelbrot, pointant notamment l'inadéquation entre les propriétés d'indépendance et de gaussiannité des incréments avec les propriétés des séries empiriques de variations de prix exposées précédemment (voir section 1.2.1). Il faut cependant noter que ces critiques portent sur un certain emploi du mouvement brownien associé aux modèles de dynamiques de prix de première génération, décrivant le prix comme un brownien arithmétique ([13]) ou géométrique ([184], [38]). Dans ce type de modélisation, les rendements - standards dans le cas du brownien arithmétique, logarithmiques dans le cas du brownien géométrique - sont en effet gaussiens, et les propriétés d'indépendance et de stationnarité des incréments produisent une invariance du niveau de variabilité des prix. Cependant, ces modèles n'épuisent évidemment pas l'usage du mouvement brownien dans un cadre de modélisation financière. Ainsi, de nombreuses modélisations ultérieures de dynamiques de prix d'actifs également construites à partir de cet objet ont montré leur capacité à capturer les phénomènes empiriques de variations non-gaussiennes des cours et de concentration de la volatilité ([86], [114]). Dans ces modèles cependant, le mouvement brownien ne constitue que l'une des briques élémentaires du modèle articulée à d'autres composantes. De ce fait, par exemple, les propriétés d'indépendance des incréments du brownien ne se traduisent pas nécessairement par une indépendance des incréments du prix (ou du logarithme du prix). C'est dans ce cadre d'utilisation du mouvement brownien que s'inscrit la modélisation adoptée dans cette thèse, présentée dans la section suivante.

1.2.2.2 Les équations différentielles stochastiques d'Itô pour décrire les dynamiques de prix

Le cadre de modélisation en temps continu adopté dans cette thèse reprend l'approche la plus courante de la modélisation financière dans laquelle les dynamiques de prix d'actifs sont décrites à partir d'équations différentielles stochastiques (EDS). Bien que la majeure partie de la thèse se centre sur des modèles univariés, la formalisation introduite dans cette section part du cas général dans lequel on considère un vecteur de prix n -dimensionnel \mathbf{P} pour dériver ensuite le cas particulier unidimensionnel.

Pour commencer, dans l'ensemble des modèles considérés, la dynamique d'un vecteur de prix d'actifs peut être décrite par l'EDS suivante :

$$d\mathbf{P}_t = \mathbf{P}_t \odot (\mathbf{A}d\mathbf{F}_t), \quad (1.1)$$

où \mathbf{A} est une matrice $m \times n$, $\sqrt{\boldsymbol{\Omega}}$ une matrice diagonale $n \times n$, et \mathbf{F} est un processus Itô m -dimensionnel défini par :

$$d\mathbf{F}_t = \boldsymbol{\mu}_t dt + \sqrt{\boldsymbol{\Omega}_t} d\mathbf{B}_t,$$

avec \mathbf{B} un mouvement brownien m -dimensionnel. Les marges du processus \mathbf{F} modélisent les différents facteurs qui engendrent les variations des prix d'actifs, la matrice \mathbf{A} représente les sensibilités des actifs à ces différents facteurs. Dans ce cadre, la dynamique du prix du i -ème actif est donnée par l'expression suivante :

$$dP_{i,t} = (\mathbf{P}_t)_i \cdot (\mathbf{A}d\mathbf{F}_t)_i = P_{i,t} \cdot \sum_{j=1}^m (\mathbf{A})_{i,j} \cdot (\boldsymbol{\mu}_t dt + \sqrt{\boldsymbol{\Omega}_t} d\mathbf{B}_t)_j, \quad (1.2)$$

en posant $\mu_{i,t} = \sum_{j=1}^m (\mathbf{A})_{i,j} \cdot \mu_{j,t}$ et $\sigma_{i,t} = \sqrt{\sum_{j=1}^m ((\mathbf{A})_{i,j} \cdot \sigma_{j,t})^2}$, puis en introduisant le brownien W dont la dynamique est définie par⁶

$$dW_t = \frac{\sum_{j=1}^m (\mathbf{A})_{i,j} \cdot \sigma_{j,t} dB_{j,t}}{\sqrt{\sum_{j=1}^m ((\mathbf{A})_{i,j} \cdot \sigma_{j,t})^2}},$$

cette dynamique peut se réécrire

$$dP_{i,t} = P_{i,t} (\mu_{i,t} dt + \sigma_{i,t} dW_t). \quad (1.3)$$

⁶ W est défini comme une combinaison linéaire de browniens indépendants. Par conséquent, dès lors que l'on suppose les processus σ_j adaptés et de carré intégrable, W est une martingale. De plus :

$$(dW_t)^2 = \frac{\sum_{j=1}^m ((\mathbf{A})_{i,j} \cdot \sigma_{j,t})^2 dt}{\sum_{j=1}^m ((\mathbf{A})_{i,j} \cdot \sigma_{j,t})^2} = dt.$$

De ce fait, par le théorème de caractérisation de Lévy d'un mouvement brownien, W est bien un mouvement brownien.

L'expression du prix au temps t est donc définie par

$$P_{i,t} = P_{i,0} + \int_0^t P_u (\mu_{i,u} du + \sigma_{i,u} dB_u).$$

Du fait que les facteurs sont définis comme des processus d'Itô, les processus de drift et de coefficients de diffusion respectent certaines propriétés générales.

Les drifts tout d'abord, sont des processus \mathcal{F}_t -adaptés, où \mathcal{F}_t correspond à la filtration naturelle du processus, et intégrables tels que $\int_0^t |\mu_{j,u}| du < \infty$. Ceux-ci définissent l'espérance du prix futur. En effet, par propriété de l'intégrale stochastique :

$$\mathbb{E}[P_{i,t}] = P_{i,0} + \sum_{j=1}^m (\mathbf{A})_{i,j} \cdot \int_0^t P_{i,u} \mu_{j,u} du = P_{i,0} + \int_0^t P_{i,u} \mu_{i,u} du.$$

Il s'ensuit donc que si μ est nul pour tout t , le prix est alors une martingale. Les coefficients de diffusion σ , pour leur part, sont également des processus \mathcal{F}_t -adaptés, vérifiant $\int_0^t \sigma_u^2 du < \infty$. Ils déterminent la variance du prix futur, puisque :

$$\text{Var}[P_{i,t}] = \sum_{j=1}^m (\mathbf{A})_{i,j}^2 \cdot \int_0^t \mathbb{E}[(P_{i,u} \sigma_{j,u})^2] du = \int_0^t \mathbb{E}[(P_{i,u} \sigma_{i,u})^2] du.$$

Dans ce cadre de modélisation, la qualité de la modélisation des dynamiques du prix P dépend donc de la spécification de μ et de σ .

L'approche de μ et de σ comme processus dépendant du temps implique que le prix P ou son logarithme perd certaines des propriétés attribuées au mouvement brownien. En particulier, les propriétés d'indépendance et de stationnarité des incréments ne sont généralement plus vérifiées ni pour P , ni pour $\log(P)$. Cependant, les conditions de mesurabilité et d'intégrabilité posées sur μ et σ garantissent la markovianité du processus de prix.

Point notable, le vecteur des prix d'actifs est défini comme un processus purement diffusif, ce qui pourrait paraître constitué un choix de modélisation restrictif. En particulier, la capture des ajustements abrupts des prix est souvent associée à l'incorporation d'une composante de sauts⁷. Si cet ajout constitue une manière cohérente de modéliser ces phénomènes de marché ([60]), la forme purement diffusive (1.1) permet également la modélisation de ce type de mouvements extrêmes. Par exemple, certaines spécifications de (1.3) convergent vers des processus à sauts ([152], [2]). Cette très grande plasticité est permise en pratique par la sous-détermination des deux ingrédients clés dont dépend la dynamique des prix dans le cadre considéré : le drift et le processus de volatilité.

⁷Typiquement, une modélisation telle que proposée par Pelger ([171]), où les facteurs suivent un processus de la forme :

$$dF_{j,t} = \mu_{j,t} dt + \sigma_{j,t} dB_{j,t} + dJ_{j,t},$$

où J_j correspond à un processus à sauts.

1.2.3 Volatilité et dynamiques de prix

1.2.3.1 De la centralité du processus de volatilité dans les dynamiques de prix

Si le drift et le processus de volatilité occupent chacun une fonction spécifique dans la modélisation des dynamiques de prix, leur importance respective ne peut être mise dans une relation de stricte équivalence.

À cet égard, il convient de commencer par noter que la majeure partie de la littérature académique dédiée aux modèles de prix en temps continu porte presque exclusivement son attention sur la modélisation du processus de volatilité. Cet état de fait s'explique en première instance par les objectifs poursuivis par cette littérature, relatifs au traitement des problèmes de tarification et de couverture d'options. Or, ces travaux s'inscrivent pour la plupart dans le cadre de la théorie moderne des options qui évacue la question de la modélisation du drift sous mesure physique par la mobilisation du concept charnière de diffusion du prix sous probabilité risque-neutre. Cependant, l'utilisation même de cet artefact mathématique par les praticiens de la finance n'est en réalité pas totalement étrangère aux propriétés du drift sous mesure physique. En effet, hors de l'espace théorique du hedging en temps continu, les rebalancements d'un portefeuille de réplication s'effectuent dans le monde réel en temps discret. De ce fait, l'impact du drift sur la couverture de l'option n'est en pratique pas complètement éliminé ([126]). Malgré cela, cet impact est généralement ignoré pour une raison qui, elle, est indépendante du cadre d'utilisation du modèle de diffusion : la faiblesse du ratio espérance sur écart-type des rendements à court terme.

Ce phénomène peut être facilement appréhendé à partir du cas simple où μ et σ sont constants. Le ratio espérance sur écart-type des rendements logarithmiques est alors donné par :

$$\frac{\mathbb{E}[\log(P_t/P_0)]}{\text{std}[\log(P_t/P_0)]} = \frac{\mu - 0.5\sigma^2}{\sigma} \sqrt{t}.$$

Sous cette hypothèse donc, ce ratio croît en racine de t dès lors que le drift du prix est positif. Ainsi, en considérant $\mu = 0.06$ et $\sigma = 0.15$, une spécification raisonnable pour le S&P500, ce ratio est d'environ 1.7% à l'horizon d'un jour, et d'environ 9.4% à l'horizon d'un mois. Hors de ce cadre analytique, la faiblesse du ratio signal-sur-bruit à court-moyen terme reste vérifiée empiriquement ([85]). Pour ces horizons temporels, les dynamiques de prix sont donc davantage fonction du terme $P_t\sigma_t dB_t$ que de $P_t\mu_t dt$. Ainsi, dans le cadre de la gestion des risques financiers, σ reste l'élément central de la modélisation.

Outre cet aspect quantitatif, l'importance première du processus de volatilité tient également à son rôle direct dans les phénomènes caractéristiques des données financières exposés en section 1.2.1. Le plus évident est l'effet Joseph, qui par sa définition même comme phénomène de persistance de la volatilité, définit son origine dans la dynamique de σ . Sa capture dépend donc directement de la bonne modélisation de l'autocorrélation de la volatilité. De manière plus indirecte, les caractéris-

tiques des distributions de rendements, éloignées du cadre gaussien, sont également la résultante des dynamiques de volatilités. Ainsi, la relation négative entre les rendements de l'actif et le niveau de volatilité, qui constitue déjà en lui-même un phénomène caractéristique des séries financières (le leverage effect), engendre le skewness négatif des distributions de rendements. De manière analogue, les queues de ces distributions sont la contrepartie des propriétés trajectorielles de la volatilité.

La primauté de l'importance de la modélisation de la volatilité tient également aux hypothèses usuelles du drift sous mesure historique qui, de par leurs formes mêmes, tendent à subordonner la capture de cette composante de tendance à la bonne modélisation du coefficient de diffusion ([108]). En effet, la majeure partie de la littérature portant sur cette question suppose un drift fonction du processus de volatilité ([188], [118]), typiquement de la forme :

$$\mu_t = \alpha_t + \sum_{k=1}^p \lambda_{t,p} \cdot (\sigma_t)^p.$$

Généralement, α_t est défini par le taux sans risque et p est égal à 1 ou 2, de telle sorte que le drift ajusté du taux sans risque suit une relation linéaire ou quadratique du processus de volatilité. Cette relation de μ est la plupart du temps supposée croissante par rapport à σ , venant modéliser le *there is no free lunch* des marchés financiers : une plus haute espérance de rendement implique un plus haut niveau de risque.

En dehors des éléments relatifs aux dynamiques marginales de prix, la composante volatilité est également centrale dans la capture des phénomènes propres à la modélisation multi-actifs. En particulier, le cadre exposé en section 1.2.2.2 indique que la corrélation instantanée des rendements entre un actif i et j est définie par⁸ :

$$(\mathbf{C}_t)_{i,j} = \frac{\sum_{k=1}^m (\mathbf{A}_t)_{i,k} \cdot (\mathbf{A}_t)_{j,k} \cdot (\sqrt{\Omega}_t)_{k,k}}{\sqrt{\sum_{k=1}^m (\mathbf{A}_t)_{i,k}^2 \cdot (\sqrt{\Omega}_t)_{k,k}} \sqrt{\sum_{k=1}^m (\mathbf{A}_t)_{j,k}^2 \cdot (\sqrt{\Omega}_t)_{k,k}}}.$$

Les volatilités factorielles déterminent ainsi la structure de corrélation de l'univers d'investissement considéré ainsi que sa dynamique. Par corollaire, dans le cadre d'une modélisation du drift de type CAPM dans laquelle les espérances de rendements sont des fonctions des covariances des actifs avec certains portefeuilles d'actifs, les dynamiques de tendance des prix sont également conditionnées par $\sqrt{\Omega}$.

Pour ces différentes raisons, bien que la question de la modélisation du drift soit considérée, l'accent est principalement mis dans cette thèse sur la modélisation du processus de volatilité.

⁸Puisque $dt dt = 0$, $dt d\mathbf{B}_t = \mathbf{0}_m$ et $d\mathbf{B}_t d\mathbf{B}_t^\top = \mathbf{I}_m dt$, la matrice de covariance des rendements instantanés est donnée par:

$$(d\mathbf{P}_t \otimes \mathbf{P}_t) (d\mathbf{P}_t \otimes \mathbf{P}_t)^\top = \left(\mathbf{A} \mu_t dt + \mathbf{A} \sqrt{\Omega}_t d\mathbf{B}_t \right) \left(\mathbf{A} \mu_t dt + \mathbf{A} \sqrt{\Omega}_t d\mathbf{B}_t \right)^\top = \mathbf{A} \Omega_t \mathbf{A}^\top dt.$$

1.2.3.2 Volatilité et path-dependence

De par sa centralité dans la dynamique des prix soulignée dans la section précédente, la modélisation de la volatilité a donné lieu à une littérature académique abondante. La multiplicité des approches proposées se différencie sur la forme de l'expression mathématique par laquelle est défini ce processus en général, ainsi que par la relation qu'entretiennent les dynamiques de prix et de volatilité en particulier.

Sur ce dernier point, Guyon et Lekeufack ([117]) identifient dans le cadre univarié deux cas polaires : les modèles strictement indépendants de la trajectoire passée et les modèles purement path-dependents. Pour la première catégorie de modèles, la trajectoire passée du prix n'a rigoureusement aucune influence sur le niveau de la volatilité. Dans ce cadre, les dynamiques de volatilités sont de nature purement exogène. À l'autre extrême, les modèles purement path-dependents définissent la volatilité comme une fonction déterministe de la trajectoire passée du prix, telle que

$$\sigma_t = \sigma(\{P_u\}_{u \leq t}).$$

Cette formalisation suppose donc que tout mouvement de volatilité peut être appréhendé comme un phénomène endogène. Entre ces pôles se situent les modèles partiellement path-dependents dans lesquels la volatilité est à la fois fonction des dynamiques du prix et d'origine exogène.

Dans une perspective de modélisation réaliste des dynamiques de prix d'actif, notamment sur le marché des actions, les modèles strictement indépendants de la trajectoire passée du prix doivent être écartés d'emblée. En effet, le leverage effect ou l'effet Zumbach mentionnés en section 1.2.1 sont des phénomènes largement documentés dans la littérature académique ([35], [44],[203], [204], [205]) qui contredisent frontalement l'hypothèse d'indépendance entre les dynamiques du prix et de la volatilité. Par conséquent, la modélisation de la volatilité doit s'inscrire au minimum dans une approche partiellement path-dependent. La question de la nécessité d'intégrer une composante de frictions exogènes se pose alors. À cet égard, les résultats des études empiriques tendent à montrer que si les dynamiques de volatilité sont explicables majoritairement de manière endogène, une part résiduelle, mais significative, est indépendante des variations de prix ([117]). Les modélisations purement path-dependent et purement path-independent constituent donc toutes deux des approches hémiplegiques, mais leurs lacunes respectives ne sont pas symétriques. D'un côté, des modèles de volatilité purement path-dependent sont en mesure de capturer la majeure partie des caractéristiques des dynamiques de prix d'actifs, de l'autre, l'hypothèse de path-independence annihile les conditions de possibilité de leur capture.

Sur la base de ces éléments, la volatilité est modélisée dans les différents chapitres de la thèse comme un processus essentiellement path-dependent, intégrant une composante exogène mais principalement déterminée par les trajectoires passées des prix ou des facteurs déterminant les dynamiques de prix.

1.3 Structuration de la thèse

Hors des chapitres introductif et conclusif, cette thèse se compose de cinq chapitres, chacun correspondant à un article de recherche pouvant être considéré de manière indépendante. Les chapitres 2, 3 et 4 introduisent et examinent de nouveaux modèles univariés de dynamique des prix d'actifs, dans lesquels le processus de volatilité est essentiellement path-dependent. Le chapitre 5 propose une méthode d'estimation innovante pour les modèles de dynamique de prix univariés, dans laquelle l'estimation est subordonnée à un objectif de prédiction de la volatilité. Enfin, le chapitre 6 développe un modèle de dynamique des prix multivariés dans un cadre factoriel et de marché path-dependent.

1.3.1 The EWMA Heston Model

Les trois premiers chapitres de la thèse s'inscrivent dans une démarche commune consistant à proposer des modèles univariés de prix d'actifs dans lesquels le processus de volatilité est essentiellement path-dependent. Ce choix de modélisation vise à saisir différentes propriétés empiriques des séries temporelles de prix d'actifs que les modèles standards (comme le modèle de Heston [121] ou le modèle CEV [76]) ne parviennent pas à capturer, notamment la modélisation de l'effet Zumbach dans ses différentes dimensions, ainsi que les caractéristiques fines des distributions de rendements à différents horizons.

Dans cette perspective, le premier chapitre introduit une extension du modèle de Heston : le EWMA Heston model. Le paramètre de variance de long terme du modèle de Heston dans sa forme standard est ici substitué par un processus de "variance d'attraction" purement path-dependent défini comme une fonction déterministe d'une ou de plusieurs moyennes mobiles exponentielles des rendements passés de l'actif. De manière analogue, le paramètre de volatilité de la volatilité du modèle de Heston est pondéré dans le EWMA Heston model par la racine carrée de cette variance d'attraction. Cette formalisation intègre donc le modèle de Heston standard comme cas particulier dans lequel la variance d'attraction est une fonction constante. En dehors de ce cas spécifique, les modifications apportées conduisent à des changements majeurs des dynamiques de prix et de volatilité modélisées. Premièrement, le niveau vers lequel tend la variance étant fonction de la trajectoire passée du prix, le modèle est structurellement apte à rendre compte de l'effet Zumbach au sens fort. Ainsi, là où dans le modèle de Heston standard le leverage effect ne pouvait être capturé qu'à travers une corrélation négative entre les browniens respectivement associés au prix et au processus de variance, la relation négative entre la tendance passée du prix et le niveau de volatilité est également causée par la variance d'attraction dans le EWMA Heston model. Cette double détermination permet par ailleurs un possible découplage entre les déterminants des mouvements de volatilité à court terme de ceux de moyen-long terme. En particulier, lorsque la corrélation entre les processus browniens du prix et de la variance est nulle ou faible, les variations de la volatilité à court terme sont principalement de nature exogène (c'est-à-dire indépendantes ou quasi-indépendantes des variations du prix de l'actif). À l'inverse, les mouvements de la volatilité à moyen terme et son niveau sont

en première instance engendrés par la dynamique de la variance d'attraction, et sont de ce fait une résultante de la tendance passée du prix. Par ailleurs, la spécification de la variance d'attraction comme fonction décroissante d'un ou de plusieurs EWMA des rendements passés peut également induire des effets sur la tendance des prix. En effet, sous l'hypothèse standard d'un drift croissant du processus de volatilité, la baisse de la volatilité engendre une hausse du drift du prix et tend donc à produire un retour à la moyenne du prix par un effet prime de risque.

Après avoir exposé ces propriétés théoriques, le chapitre démontre, sur la base d'expériences numériques réalisées à partir d'une version discrétisée d'une certaine spécification du EWMA Heston model, que celui-ci reproduit une part importante des caractéristiques empiriques de l'indice S&P500 et de sa volatilité. Premièrement, les phénomènes relatifs aux dynamiques conjointes du prix et de la volatilité, tels que la coïncidence entre les pics de volatilité et les chutes brutales de l'indice, ou l'existence de périodes de faible volatilité conjointes à une tendance haussière de l'indice, sont bien modélisés. Le comportement erratique de la volatilité à court terme observable sur les données de marché est également assez bien capturé grâce aux valeurs élevées prises à la fois par le paramètre de volatilité de la volatilité et par le paramètre de retour vers le niveau d'attraction. Cette configuration de paramètres produit ainsi un processus de volatilité sujet à de forts mouvements aléatoires intra-journaliers, contenus par une importante force de retour vers le niveau de volatilité d'attraction. Au-delà de ces caractéristiques trajectorielles, les distributions de rendement et de volatilité générées par le modèle sont globalement proches de leurs contreparties empiriques. En particulier, le modèle est capable de capturer les déformations respectives de ces distributions en fonction de l'horizon temporel.

Ce chapitre souligne néanmoins que, malgré ces qualités, la version du EWMA Heston model considérée présente quelques limitations. En particulier, bien que l'autocorrélation du processus de volatilité soit assez bien reproduite par le modèle pour les horizons inférieurs à 2 mois, celle-ci décroît trop rapidement pour les horizons plus longs au regard des données empiriques. En outre, la relation liant la valeur absolue attendue des incréments de la log-volatilité et l'échelle de temps considérée diffère entre les données générées par le modèle et les données de marché. Ces limitations semblent découler de la spécification considérée davantage que du EWMA Heston modèle dans sa forme générale, qui fait dépendre la variance d'attraction d'une seule EWMA des rendements passés. A la place, une combinaison de plusieurs EWMA pourrait s'avérer davantage à même de la structure d'autocorrélation de la volatilité empirique.

1.3.2 Rough Path-Dependent Volatility Models

Le troisième chapitre de la thèse poursuit la démarche entreprise dans le chapitre 2 consistant à proposer de nouvelles modélisations des dynamiques univariées de prix d'actifs cohérentes avec les données de marché, dans lesquelles le processus de volatilité est essentiellement voire entièrement path-dependent.

Le premier apport de ce chapitre consiste à introduire une classe générale de modèles dans laquelle la volatilité dépend de deux processus sous forme d'équations de convolution stochastique : le premier représente essentiellement un facteur de tendance passée du prix, le second un facteur de volatilité (ou variance) historique. Cette structure comprend divers modèles majeurs de la littérature, parmi lesquels figurent le modèle rough et le modèle quadratic rough Heston, ainsi que le modèle PDV introduit par Guyon et Lekeufack. Cette forme très flexible est ensuite utilisée comme outil comparatif pour examiner différentes manières de modéliser la volatilité et de rendre compte de sa propriété de path-dependence.

À partir de ce cadre général, le chapitre définit alors une sous-classe de modèles : les modèles *rough path-dependent volatility* (RPDV). Celle-ci englobe l'ensemble des modèles inclus dans la classe générale présentée, pour lesquels les noyaux de convolution associés au processus de volatilité sont en loi puissance.

Une étude comparative sur des données de marché évalue ensuite la capacité de différentes variantes de modèles RPDV, utilisés comme modèles de régression, à expliquer la volatilité réalisée de cinq indices boursiers. Plusieurs résultats clés en ressortent. Premièrement, comme l'ont démontré Guyon et Lekeufack ([117]), dans le cadre d'une modélisation purement path-dependent, les modèles qui intègrent un facteur de volatilité historique ont un fort pouvoir prédictif et sont nettement plus performants que les modèles qui reposent uniquement sur un processus de tendance passée des prix. Par ailleurs, pour l'ensemble des modèles considérés, une approche purement path-dependent ne permet pas d'expliquer entièrement les variations de la volatilité réalisée, suggérant ainsi la nécessité d'adopter une approche partiellement path-dependent. Cependant, les propriétés des résidus des régressions impliquent que la composante exogène au processus de volatilité ne peut être appréhendée de la même manière entre les modèles intégrant un facteur de volatilité historique et ceux qui n'en intègrent pas. Pour les premiers, les mouvements exogènes de la volatilité représentent un phénomène intrajournalier. Il est alors cohérent de modéliser le processus de volatilité comme le produit d'un modèle purement path-dependent et d'un processus de retour à la moyenne très rapide dont la composante stochastique est indépendante du brownien du prix. Dans le cadre de modélisations n'intégrant pas de facteur de volatilité historique, il semble davantage cohérent de considérer les mouvements exogènes de la volatilité comme une composante de même nature que la source endogène d'aléa associée à la dynamique des prix, à l'instar des modèles de volatilité stochastique classiques.

Le chapitre considère enfin l'hypothèse selon laquelle l'apparente dépendance structurelle de la volatilité par rapport à sa trajectoire passée serait en fait causée par une composante exogène tierce, plutôt que par un réel phénomène structurel de rétroaction de la volatilité. La validité ou non de cette hypothèse a des implications fortes. Si celle-ci est rejetée, les modèles n'intégrant pas une dépendance structurelle de la volatilité à sa propre trajectoire passée sont structurellement inaptes à modéliser correctement les dynamiques réelles de prix. Une expérience numérique est donc mise en place afin d'évaluer la vraisemblance de cette hypothèse. Il est alors démontré qu'à

l'instar de ce qui est observé sur les données de marché, une apparente autorégressivité du processus de volatilité apparaît, bien que les données soient générées à partir d'un modèle qui ne comprend pas de facteur de volatilité historique. Ces résultats suggèrent ainsi que l'apparent feedback positif de la volatilité peut être expliqué, au moins partiellement, par la dépendance de la volatilité à des variables exogènes.

1.3.3 Price and Volatility Dynamics Under Path-Dependent Models

Le chapitre 4 explore les implications de diverses spécifications d'un modèle PDV appartenant à la classe générale des modèles introduite au chapitre 3. Dans ce modèle, la volatilité est définie par une équation convolutive stochastique à un seul noyau, tandis que le drift du prix suit une relation quadratique du processus de volatilité intégrant à la fois une prime de volatilité et une prime de variance.

Dans un premier temps, le chapitre examine comment, dans le cadre considéré, la jonction de la propriété de path-dépendance de la volatilité avec les hypothèses sur le drift du prix conditionne à la fois le niveau vers lequel la volatilité tend - qualifiée de *volatilité d'attraction* - et la force d'attraction vers cette valeur. Il est notamment montré que le niveau de la volatilité d'attraction décroît avec le niveau de la prime de risque, tandis que la vitesse de retour à la moyenne de la volatilité augmente avec celui-ci. De plus, la structure de la prime de risque - c'est-à-dire, les valeurs respectives des primes de risque de volatilité et de variance - conditionne les caractéristiques des forces de retour à la moyenne. En particulier, la présence d'une prime de variance strictement positive entraîne une force de retour à la moyenne plus importante lorsque la différence entre la volatilité d'attraction et la volatilité actuelle est négative que lorsque celle-ci est positive, pour un même différentiel absolu. A l'inverse, si la prime de variance est nulle et la prime de volatilité positive, la force de rappel de la volatilité vers sa valeur d'attraction est symétrique.

Ce chapitre formule ensuite des hypothèses concurrentes relatives aux mécanismes de formation du prix et de la volatilité à travers différentes spécifications du modèle général considéré.

La première suppose que le niveau de volatilité d'attraction est lui-même un processus path-dépendent, entraînant une intrication de path-dépendance. Cette spécification permet de renforcer l'effet Zumbach du modèle, l'une des caractéristiques empiriques majeures des données financières ([203], [204], [205], [39]) évoquée en section 1.2.1.2. Elle constitue de ce fait une réponse possible au problème de symétrie excessive (par rapport aux données empiriques) entre les mouvements baissiers et haussiers de la volatilité, que tendent à générer les modèles PDV n'intégrant pas de facteur de volatilité historique.

Le dernier type de spécification considéré dans ce chapitre vise à proposer des formalisations qui concilient la nature hautement endogène des dynamiques de prix, résultant de la propriété de path-dépendance de la volatilité, avec l'hypothèse selon laquelle le prix tend à long terme à revenir au

voisinage de sa "valeur fondamentale". Deux modélisations concurrentes sont alors proposées.

Dans la première, l'ajustement entre le prix et la valeur fondamentale s'effectue par une variation du niveau de volatilité d'attraction. Sous le jeu d'hypothèses considéré, la volatilité d'attraction est croissante à la fois de l'écart entre le taux de croissance de la valeur fondamentale et le taux sans risque, ainsi que de l'écart entre le prix et la valeur fondamentale. Ainsi, les variations de la volatilité d'attraction entraînent des modifications du niveau de volatilité qui, à leur tour, par un effet de prime de risque (dont la structure reste inchangée), engendrent un ajustement du niveau du drift.

La seconde spécification inverse la relation causale, en consacrant la prime de risque comme opérateur d'ajustement entre le prix et la valeur fondamentale. Ainsi, dans cette seconde hypothèse, alors que la volatilité d'attraction reste constante, les primes de volatilité et de variance augmentent avec le différentiel entre le taux de croissance de la valeur fondamentale et le taux sans risque, ainsi qu'avec celui entre le prix et la valeur fondamentale. De plus, la relation positive entre la prime de risque et la force de retour à la moyenne engendre une relation croissante entre cette force et les écarts entre le taux de croissance de la valeur fondamentale et le taux sans risque, ainsi qu'entre la valeur fondamentale et le prix.

1.3.4 Deep Estimation for Volatility Forecasting

Le cinquième chapitre occupe une place quelque peu singulière dans la thèse. Contrairement aux chapitres 2, 3, 4 et 6 axés sur la question de la modélisation des dynamiques de prix et de volatilité, le chapitre 5 s'attache à traiter une problématique connexe. L'objectif est d'introduire une nouvelle méthode d'estimation sur données historiques de modèles de volatilité stochastique tels que ceux considérés dans les chapitres 2, 3, et 4, dans une perspective de prédiction des deux premiers moments de la volatilité à différents horizons. Ce chapitre se focalise plus spécifiquement sur l'estimation d'un RPDV model particulier. Toutefois, le principe général de l'approche proposée peut être appliqué à l'estimation de tout autre modèle de dynamique de prix univarié.

Le chapitre commence par formaliser le problème d'estimation considéré dans les coordonnées de la théorie bayésienne de la décision. L'adoption de ce cadre théorique aboutit à traduire le problème d'estimation en un programme d'optimisation à résoudre, dans lequel la fonction objectif est une espérance de coût. Ce coût est déterminé par la différence entre les moments de volatilité associés aux paramètres réels et ceux associés aux paramètres estimés. L'espérance, quant à elle, est définie sous une mesure postérieure induite par une mesure a priori sur les paramètres à estimer mise à jour par les données disponibles à partir desquelles on cherche à estimer le modèle.

La résolution du programme d'optimisation considéré pose cependant deux défis majeurs : d'une part, le calcul direct de la mesure postérieure associée à la fonction objectif est impraticable ; d'autre part, la fonction des moments du processus de volatilité n'admet pas de formule fermée connue. Le

premier obstacle est contourné en considérant un programme d'optimisation connexe sous la mesure a priori, qui elle est connue. Il est ainsi démontré que, sous certaines hypothèses raisonnables, la résolution de ce nouveau problème d'optimisation permet asymptotiquement de résoudre le problème initial sous la mesure postérieure. Le chapitre introduit donc un réseau de neurones (NN) estimateur qui vise à déterminer, à partir d'une matrice de données, les paramètres solutions de ce nouveau programme d'optimisation. Cependant, l'entraînement de ce NN requiert toujours la fonction des moments, pour laquelle on ne dispose pas de formule analytique. Cette absence est palliée par l'introduction d'un second NN chargé d'approximer cette fonction des moments.

Ces éléments aboutissent ainsi à formaliser la procédure d'estimation qui se compose de trois grandes étapes. La première consiste à générer les données d'entraînement à partir du modèle de dynamique des prix et de la volatilité que l'on cherche à estimer, en utilisant des vecteurs de paramètres tirés de la mesure a priori. Dans un second temps, le NN servant de proxy de la fonction des moments est entraîné à partir de ces données. Enfin, le NN estimateur est entraîné par interaction avec le second NN. Ce dernier sert à calculer le coût des jeux de paramètres proposés par le NN estimateur. Grâce à ce calcul, le NN estimateur peut alors s'ajuster en conséquence. À l'issue de ces étapes, le NN estimateur peut alors être utilisé pour estimer les paramètres du modèle de volatilité considéré.

Cette procédure d'estimation est ensuite soumise à une évaluation approfondie sur données synthétiques, puis sur données de marché. Sur données synthétiques d'abord, les résultats des évaluations menées démontrent que la procédure atteint bien son objectif. Ainsi, les moments prédits associés aux paramètres estimés sont très proches des moments réels. Sur données de marché, les résultats obtenus sont là encore encourageants bien que plus contrastés. Globalement, le modèle de volatilité entraîné par la méthode proposée surclasse significativement les benchmarks de la littérature pour les prédictions de volatilité sur les horizons supérieurs à une semaine. Cependant, ses performances prédictives sont moindres pour les horizons plus courts, typiquement d'une journée. Deux explications sont alors avancées pour expliquer ces performances hétérogènes selon l'horizon temporel considéré. La première tient au choix du modèle de volatilité qui induit certaines hypothèses implicites sur le comportement réel de la volatilité. Dans le cadre de cette première explication, l'adoption d'un modèle de volatilité plus flexible pourrait accroître la capacité prédictive de l'approche pour les horizons courts. La seconde explication avancée repose sur la méthode d'estimation elle-même. Plus particulièrement, le fait que le NN estimateur soit uniquement entraîné sur des données synthétiques non bruitées avec un pas de temps constant, là où les données réelles présentent des irrégularités telles que la variabilité de la fréquence d'observation, pourrait introduire un biais dans la prédiction des variables d'état, biais dont l'importance décroît avec l'horizon de prédiction. Des amendements de la méthode d'estimation proposée sont donc envisagés afin de rendre l'approche plus robuste, comme le bruitage volontaire des données d'entraînement ou un apprentissage partiel du NN estimateur à partir de données réelles.

1.3.5 The Factorial Path-Dependent Market Model

Le dernier chapitre de la thèse s'attaque à la modélisation des dynamiques de prix multi-actifs. L'objectif est ici de proposer un cadre de modélisation capable de rendre compte conjointement des différentes caractéristiques empiriques des dynamiques de prix répertoriées en section 1.2.1. Pour ce faire, un certain nombre d'idées-forces, originellement conçues dans un contexte de modélisation univariée, sont redéployées dans un cadre étendu de modélisation multi-factorielle et multi-actifs.

Le cadre général de modélisation adopté ici correspond à la formalisation définie en section 1.2.2.2, dans laquelle les dynamiques des actifs sont décrites par une combinaison linéaire de dynamiques de facteurs suivant des processus d'Itô de browniens indépendants. Ces facteurs, qualifiés de *facteurs élémentaires*, se scindent en deux groupes : les facteurs communs aux différents actifs, et les facteurs idiosyncratiques. Si cette formalisation est relativement standard dans la littérature financière, le modèle proposé se singularise par sa spécification des drifts et des volatilités factorielles, qui fait de ces composantes des processus essentiellement fonction des trajectoires passées d'un ensemble de portefeuilles de facteurs élémentaires. Pour chacun de ces portefeuilles, l'information dont dépendent les drifts et les volatilités factorielles est agrégée dans deux types de variables d'état : la première est un produit convolutif de la dynamique passée du portefeuille, la seconde, une variable de volatilité historique de ce portefeuille. Ces variables déterminent le niveau des drifts et des volatilités factorielles, produisant ainsi une path-dependence factorielle du vecteur des prix.

Le vecteur des volatilités factorielles d'abord est défini comme le produit (au sens multiplicatif) de trois composantes. La première, qui constitue la partie purement path-dependent des volatilités, est une fonction multilinéaire des variables d'état engendrées par les portefeuilles factoriels. C'est par cette composante que sont capturés les différents effets découlant de la relation entre les trajectoires des prix et les volatilités (effet de levier, effet Zumbach). La deuxième est un vecteur stochastique qui modélise les différents facteurs exogènes de variation des volatilités factorielles. Cette composante a pour fonction principale de capturer les mouvements de volatilité à court terme, voire très court terme, à l'origine des queues lourdes des distributions conditionnelles de rendements⁹ ([70]). La troisième est un opérateur univarié de sensibilité du marché aux frictions aléatoires qui modélise les variations homothétiques du système des volatilités factorielles.

Le drift d'un facteur élémentaire dépend pour sa part des coefficients de régression instantanés de ce facteur avec les portefeuilles factoriels dont dépendent les variables d'état du modèle. Plus exactement, chacun de ces coefficients de régression est pondéré par la prime associée au portefeuille factoriel correspondant. Ces primes résultent de la combinaison additive de quatre composantes distinctes. Les deux premières sont des primes de risque, respectivement fonction de la volatilité instantanée et de la variable de volatilité historique du portefeuille factoriel attendant. La troisième dépend de la variable définie par le produit convolutif de la dynamique passée du portefeuille, et

⁹Il est entendu ici par "conditionnelles" les rendements corrigés de l'effet de regroupement de volatilité à l'échelle journalière.

permet en pratique de modéliser des phénomènes de rétroaction comme l'effet momentum ou l'effet reversal. Enfin, la quatrième composante peut être appréhendée comme un résidu potentiellement fonction d'informations exogènes.

Le chapitre considère ensuite une spécification particulière de ce modèle. Dans celle-ci, la composante purement path-dépendent des volatilités factorielles n'est fonction que de la trajectoire passée du facteur élémentaire auquel celles-ci sont respectivement associées ainsi que de la composante path-dépendent de la volatilité du facteur élémentaire correspondant au facteur de marché. Ce dernier est par ailleurs l'unique facteur élémentaire disposant d'un drift non nul, produisant une hypothèse sur les drifts de type CAPM ([77]). Le facteur de marché joue donc un rôle central dans cette spécification. D'une part, il conditionne le niveau et les dynamiques des volatilités, ainsi que la structure de corrélation entre les différents actifs et sa dynamique, ce par le biais de la composante purement path-dépendent des volatilités factorielles. D'autre part, les drifts des prix sont entièrement déterminés par ce facteur de marché.

Le chapitre dérive alors de cette spécification un market generator, puis expose une méthode d'estimation de ce modèle à partir d'un historique de rendement de l'univers d'investissement considéré. La première étape consiste à déterminer les facteurs élémentaires et les coefficients de sensibilité des actifs à ces facteurs en appliquant une décomposition en valeurs singulières sur la matrice des rendements préalablement standardisée. Un algorithme opère alors une séparation entre les m plus grandes valeurs singulières rattachées au signal, des autres valeurs singulières associées (supposément) au bruit contenu dans les données. Les éléments rattachés aux m plus grandes valeurs singulières sont utilisés pour déterminer les composantes associées aux facteurs élémentaires communs, les éléments rattachés aux autres valeurs singulières pour déterminer les composantes associées aux facteurs élémentaires idiosyncratiques. De plus, les éléments se rapportant à la plus grande valeur singulière sont appréhendés comme les composantes du facteur de marché. Sur la base de cette décomposition, une méthode dérivée du maximum de vraisemblance est utilisée pour déterminer le reste des paramètres du modèle associés aux drifts et aux volatilités factorielles.

Le market generator estimé à partir de cette méthode est ensuite évalué via différents tests opérés hors échantillon, en considérant comme univers d'investissement 436 actions du S&P500.

En premier lieu, les distributions marginales des rendements d'actifs à différentes échelles de temps issues des simulations se révèlent globalement très cohérentes avec leurs contreparties empiriques. De même, les différentes propriétés trajectorielles caractéristiques des dynamiques individuelles de prix exposées en section 1.2.2.2 sont également bien reproduites. Sur le plan de la modélisation des caractéristiques empiriques propres au cadre multivarié, le modèle se révèle également convaincant. En particulier, l'estimateur de la matrice de covariance calculé à partir des données simulées par le modèle présente des performances hors échantillon significativement meilleures que celles des estimateurs de la matrice de covariance calculés sur données historiques.

Il est enfin montré que le market generator proposé constitue un outil puissant pour le backtesting de stratégies. Ainsi, les propriétés empiriques des différentes stratégies considérées, notamment les moments des distributions de rendements à différentes échelles de temps, sont globalement bien reproduites. En outre, pour la plupart des paires de stratégies considérées, les corrélations prédites sont assez proches de leurs contreparties empiriques. Cependant, la différence entre les corrélations sur données simulées et celles sur données empiriques est assez significative pour les stratégies dynamiques dont la composition dépend des tendances passées des prix (stratégies de type trend-following ou reversal, rebalancées fréquemment). Ce fort différentiel semble donc indiquer que la spécification de la dérive, plaçant le facteur de marché comme unique déterminant des drifts des actifs, est probablement trop simpliste et mériterait d'être amendée.

CHAPTER 2

The EWMA Heston Model

Abstract

This chapter introduces the exponentially weighted moving average (EWMA) Heston model, a Markovian stochastic volatility model able to capture a wide range of empirical features related to volatility dynamics while being more tractable for simulations than rough volatility models based on fractional processes. After presenting the model and its principal characteristics, our analysis focuses on the use of its associated Euler-discretization scheme as a time-series generator for Monte-Carlo simulations. Using this discretization scheme, and on the basis of S&P500 empirical time series, we show that the EWMA Heston model is overall consistent with market data, making it a credible alternative to other existing stochastic volatility models.

2.1 Introduction

The modeling of asset price dynamics is one of the most important issues in quantitative finance and has resulted in a large body of academic research. It has for long been established that the modeling quality of a price dynamics model depends to a large extent, on its ability to accurately reproduce the empirical features of volatility dynamics. Different approaches have been applied to take account of this.

The ARCH (or autoregressive conditional heteroskedasticity) family of models is one of the most important approaches and includes a wide range of variants able to describe the stylized facts of financial time series. For instance, the famous generalized ARCH or GARCH model ([92]) in which the variance process depends on an exponential moving average of past squared returns provides a good representation of many of the empirical features of financial data, such as tail heaviness, volatility clustering, and feedback effect. More sophisticated extensions such as the EGARCH ([163]), NGARCH ([136]), FIGARCH ([17], [26]) and QGARCH ([43]) models, further improved modeling quality by reproducing other features of financial time series including the leverage effect, time-reversal asymmetry, and the feedback effect.

Although some of these models have continuous-time counterparts this family of models in its canonical form adopts a discrete-time approach. However, some financial issues require a continuous-time framework which led to the parallel development of other continuous-time stochastic volatility models.

Notable examples include the Stein-Stein ([191]), Heston ([121]), and Bergomi ([27]) models. While these models offer significant advantages in terms of modeling and tractability due to their Markovian nature, they fail to capture certain key empirical properties of volatility. These shortcomings have, to a large extent, motivated the introduction of rough volatility models, which in turn succeed in capturing a substantial portion of these properties ([23], [89], [90], [91]). However, despite their strengths, the application of rough volatility models in Monte Carlo experiments, particularly in risk or asset management contexts, faces certain limitations. First, these models are founded on non-Markovian and non-semimartingale processes, which makes unbiased simulations based on such models a tricky task ([91]). In addition to the strictly technical issues, some empirical features of the realized volatility are not entirely captured by this type of model ([39], [106]).

The present chapter introduces a new stochastic volatility model - the EWMA Heston model (HM) - which aims to address some aspects of these different issues. The aim is to propose a model able to reproduce a broad spectrum of empirical features related to volatility dynamics while being more tractable for simulations than rough volatility models based on fractional stochastic processes. We show that, while being Markovian, the EWMA HM can capture the different dimensions of the Zumbach effect and accurately reproduce the joint dynamics of the asset returns and volatility, as well as the empirical returns and volatility distributions for different time horizons. Unlike most of the academic literature on stochastic volatility models, the present chapter deals not with the model's application for derivative pricing issues but focuses instead on the model's use as a tool for Monte Carlo experiments.

The chapter is organized as follows. First, we present the EWMA HM and its principal characteristics. Second, we consider a given specification of this model and one of its limit-case with interesting properties. Third, we compare the EWMA HM with stochastic volatility models based on quadratic Hawkes processes ([39], [80], [105]), to highlight their similarities and their differences. We then focus on use of its associated Euler-discretization scheme as a time-series generator for Monte-Carlo simulations. Using this discretization, the consistency of this model is evaluated against empirical data from the S&P 500 time series. It is demonstrated that this version of the EWMA HM fairly well reproduces the empirical distributions of returns and volatilities while also capturing subtle properties of financial time series, such as time-reversal asymmetry. However, it is also shown that this version has shortcomings in capturing the autocorrelation structure of empirical volatility, suggesting that a version of the EWMA HM with multiple exponentials might be better suited for modeling the slow decay in the memory of empirical volatility.

2.2 The EWMA Heston model

2.2.1 Presentation of the model

Let us introduce the EWMA HM defined by the following stochastic differential equations (SDEs) system¹:

$$\begin{cases} \frac{dS_t}{S_t} &= \mu_t dt + \sqrt{V_t} dW_t \\ dV_t &= \frac{1}{\tau_1} \left(\nu(t, m_t)^2 - V_t \right) dt + \xi \nu(t, m_t) \sqrt{V_t} dB_t \\ dm_t &= \Lambda \mathbf{1} \cdot \left(\frac{dS_t}{S_t} - \eta \mu_t dt \right) - \Lambda m_t dt \end{cases} \quad (2.1)$$

where $\{W\}_{t \in T}$ and $\{B\}_{t \in T}$ are Brownian motions such as $\langle dW, dB \rangle_t = \rho \in [-1 : 1]$, $\tau_1, \xi \in \mathbb{R}_+$, $\eta \in \{0, 1\}$, $\mathbf{1}$ a $d \times 1$ vector of ones, $m_t \in \mathbb{R}^d$, Λ a $d \times d$ diagonal matrix as $\forall j \in \{1, \dots, d\}, \Lambda_{j,j} \in \mathbb{R}_+$, and with $\nu : \mathbb{R}_+ \times \mathbb{R}^d \rightarrow \mathbb{R}_+$.

The specificity of this extension of the HM lies in the process ν . Whereas in the standard HM ν^2 is a constant and corresponds to the "long variance" of price, here it is an attraction variance function of the d -dimensional random vector m_t solution of the third SDE of (2.1)

$$m_t = m_0 e^{-\Lambda t} + \int_0^t e^{-\Lambda(t-u)} \Lambda \left(\frac{dS_u}{S_u} - \eta \mu_u du \right).$$

Therefore, each coordinate of m_t corresponds to an exponential weighted moving average (EWMA) estimator of the price trend adjusted (i.e. $\eta = 1$) or not (i.e. $\eta = 0$) by its deterministic component. Due to the dependence between m_t and the attraction volatility ν_t , the model is able to capture the leverage effect, even if $\{W\}_{t \in T}$ and $\{B\}_{t \in T}$ are uncorrelated. More importantly, this property makes the EWMA HM structurally adapted to capturing the "strong Zumbach effect" which corresponds to the fact that "conditional dynamics of volatility with respect to the past depend not only on the past volatility trajectory but also on the historical price path" ([106] p.3). The properties of the EWMA HM imply that it is included in the family of path-dependent volatility models ([116]). Additionally, the generic nature of the EWMA HM allows it to capture a wide array of path dependencies through attraction volatility. For instance, exploiting the results in [40] or [2], makes it possible to use a specification of Λ and $\nu(\cdot)$ in a way that mimics the behavior of a process with a power-law kernel.

After setting out this general framework, we now focus on an EWMA HM where $\eta = 0$ and m_t

¹The questions regarding the conditions for the existence and uniqueness of solutions for the system in 2.1 will not be addressed in this work. Moreover, the exotic nature of the system generated by the process ν makes it difficult to relate to existing literature, except in the specific case where ν is a constant. Therefore, these technical issues warrant further investigation in future works.

depends on a unique EWMA parameter (i.e. m_t is a scalar), defined by the following SDE system:

$$\begin{cases} \frac{dS_t}{S_t} &= \mu_t dt + \sqrt{V_t} dW_t \\ dV_t &= \frac{1}{\tau_1} (\nu_t^2 - V_t) dt + \xi \nu_t \sqrt{V_t} dB_t \\ d\nu_t &= \left(\frac{\psi}{\nu_t - \underline{\nu}} - \nu_t + \underline{\nu} + \alpha \right) \frac{dt}{\tau_2} - \frac{\beta}{\tau_2} \frac{dS_t}{S_t} \end{cases} \quad (2.2)$$

where $V_0, \nu_0, \psi, \tau_1, \underline{\nu}, \beta \in \mathbb{R}_+$ and $\alpha \in \mathbb{R}$. We assume also that $\forall t \in T$ the Feller condition is respected²:

$$\frac{2\nu_t^2}{\tau_1} > (\xi \nu_t)^2 \iff \frac{2}{\tau_1} > \xi^2.$$

At first sight, because the attraction volatility is described here by its dynamic through an SDE, membership of this model in the family of EWMA HMs defined above is not obvious. To highlight this and fully understand the concrete effects of the historical price path on the volatility in this EWMA HM, we need to consider the SDE describing attraction volatility. First, notice that if $\beta = 0$, given $\nu_0 > \underline{\nu}$, the dynamic of the attraction volatility is described by an ordinary differential equation (ODE) that converges as follows³:

$$\lim_{t \rightarrow +\infty} \nu^* = \frac{\underline{\nu} + \alpha + \sqrt{\alpha^2 + 4\psi}}{2}.$$

Therefore, in this specific case (where $\beta = 0$), the EWMA HM converges toward a standard HM. To consider the dynamics of ν_t in the general case, we set $f(\nu_t, t) = \nu_t e^{\frac{t}{\tau_2}}$ and apply the Itô lemma:

$$df(\nu_t, t) = \frac{e^{\frac{t}{\tau_2}}}{\tau_2} \nu_t dt + e^{\frac{t}{\tau_2}} d\nu_t = \frac{e^{\frac{t}{\tau_2}}}{\tau_2} \left(\frac{\psi}{\nu_t - \underline{\nu}} + \underline{\nu} + \alpha \right) dt - \frac{\beta e^{\frac{t}{\tau_2}}}{\tau_2} \frac{dS_t}{S_t}.$$

Thus, the solution of the SDE is

$$\nu_t = \nu_0 e^{-\frac{t}{\tau_2}} + \frac{1}{\tau_2} \int_0^t e^{\frac{1}{\tau_2}(u-t)} \left(\frac{\psi}{\nu_u - \underline{\nu}} + \underline{\nu} + \alpha \right) du - \beta \underbrace{\frac{1}{\tau_2} \int_0^t e^{\frac{1}{\tau_2}(u-t)} \frac{dS_u}{S_u}}_{m_t}.$$

This relationship clarifies the membership of the model in (2.1), highlighting that the attraction variance is indeed a time-dependent process function of m_t , an EWMA of the past returns of $\{S\}_{t \in T}$, which is the solution of the following SDE when $m_0 = 0$:

$$dm_t = \frac{1}{\tau_2} \left(\frac{dS_t}{S_t} - m_t dt \right).$$

Further, from a statistical point of view, this process $\{m\}_{t \in T}$ has a theoretical foundation since it

²The fact that the Feller condition $\forall t \in T$ is respected implies that ν stays strictly positive (as in the standard Cox-Ingersoll-Ross (CIR) processes) but has yet to be demonstrated.

³See appendix 4.B.

can be considered the Kalman filter estimator of an unobservable trend ([119], [56], [129]). The sensitivity of the attraction volatility to this trend estimator is defined by the parameter β . Since β is positive, the model is able to capture the leverage effect, even when $\{W\}_{t \in T}$ and $\{B\}_{t \in T}$ are uncorrelated. This aspect of the EWMA HM makes it more appropriate to model stock market dynamics where the empirical leverage effect is stronger than in other markets such as foreign exchanges which have no leverage effects. Another important and related element is the asymmetry between the positive and negative trends in attraction volatility. The fact that the repulsion force approaches infinity when the attraction volatility tends toward $\underline{\nu}$, means attraction volatility cannot fall below $\underline{\nu}$. Therefore, $\underline{\nu}$ constitutes the floor attraction volatility value which means that a long period of a strong positive trend leads to a period where attraction volatility remains around $\underline{\nu}$. Conversely, long periods of a strong negative trend make $\underline{\nu}$ converge toward an affine relationship of m_t . Indeed, a negative trend increases attraction volatility which decreases the repulsion force produced by $\frac{\psi}{\nu_t - \underline{\nu}} dt$. Consequently, the more negative m_t , the closer $d\nu_t$ becomes to $-\beta dm_t$. The speed of the convergence depends on the parameter ψ . All other things being equal, the lower the value of ψ , the greater the rate of convergence toward this situation. Furthermore, similar to the parameter α but related less directly, the value of ψ shows a positive relation to the expected attraction volatility.

These features of the attraction volatility process allow a better specification of the variance process. Based on the above results, the spot variance can be written as:

$$\begin{aligned} V_t &= V_0 e^{\frac{-t}{\tau_1}} + \frac{1}{\tau_1} \int_0^t e^{\frac{1}{\tau_1}(u-t)} \nu_u^2 du + \xi \int_0^t e^{\frac{1}{\tau_1}(u-t)} \nu_u \sqrt{V_u} dB_u, \\ \nu_t &= \nu_0 e^{\frac{-t}{\tau_2}} + \frac{1}{\tau_2} \int_0^t e^{\frac{1}{\tau_2}(u-t)} \left(\frac{\psi}{\nu_u - \underline{\nu}} + \underline{\nu} + \alpha \right) du - \frac{\beta}{\tau_2} \int_0^t e^{\frac{1}{\tau_2}(u-t)} \frac{dS_u}{S_u}. \end{aligned}$$

Also, the variance process can be written such that it more clearly demonstrates the impact of the past trend on the variance:

$$\begin{aligned} V_t &= V_0 e^{\frac{-t}{\tau_1}} + \frac{1}{\tau_1} \int_0^t e^{\frac{1}{\tau_1}(u-t)} (z_u - \beta m_u)^2 du + \xi \int_0^t e^{\frac{1}{\tau_1}(u-t)} (z_u - \beta m_u) \sqrt{V_u} dB_u, \\ z_t &= \nu_0 e^{\frac{-t}{\tau_1}} + \frac{1}{\tau_2} \int_0^t e^{\frac{1}{\tau_2}(u-t)} \left(\frac{\psi}{\nu_u - \underline{\nu}} + \underline{\nu} + \alpha \right) du. \end{aligned}$$

The above expression emphasizes that the strong Zumbach effect is encoded by the model since the volatility is a function of past price trends through m_t . It shows also that the model reproduces the quadratic feedback of the price trends on volatility. Additionally, note that $\{v\}_{t \in T}$ exhibits a very interesting mathematical property related to its dependence on the past. We see that in contrast to rough volatility models ([91]) the variance process in the EWMA HM is Markovian with the state vector (V_t, ν_t) . This Markovian property makes the model very tractable for Monte-Carlo simulations based on a discretization scheme of the associated SDE system.

2.2.2 The thresholded version of the model as a limit-case

Following our presentation of the EWMA HM, in the remainder of the chapter we focus on the case 2.2 where ψ tends toward zero while remaining strictly positive. In this limit case, attraction volatility is given by:

$$\nu_t = \underline{\nu} + (\alpha - \beta m_t)_+. \quad (2.3)$$

Thus, the SDE system associated with the EWMA HM can be rewritten as:

$$\begin{cases} \frac{dS_t}{S_t} &= \mu_t dt + \sqrt{V_t} dW_t \\ dV_t &= \frac{1}{\tau_1} (\nu_t^2 - V_t) dt + \xi \nu_t \sqrt{V_t} dB_t \\ dm_t &= \frac{1}{\tau_2} \left(\frac{dS_t}{S_t} - m_t dt \right), \end{cases}$$

where ν_t is equal to (2.3).

Since β is strictly positive, there exists a threshold value of m below which the attraction value of the volatility follows an affine relation with m_t (see figure 5.1). This critical point is reached when $\alpha - \beta m_t$ is zero. Thus, this value is defined by

$$\bar{m} = \frac{\alpha}{\beta}.$$

Using \bar{m} , we can rewrite the attraction volatility as follows:

$$\nu_t = \underline{\nu} + \beta(\bar{m} - m_t)_+.$$

Equivalently, the attraction variance is equal to

$$\nu_t^2 = \underline{\nu}^2 + \mathbb{1}_{\{m_t < \bar{m}\}} (\beta^2 (\bar{m} - m_t)^2 + 2\underline{\nu}\beta(\bar{m} - m_t)).$$

Thus, when the past trend of the asset measured by an EWMA of its past returns is below the threshold value \bar{m} , all variations of this EWMA result in a proportional change in its attraction volatility:

$$\frac{d\nu_t}{dm_t} = \begin{cases} 0 & \text{for } m_t \geq \bar{m} \\ -\beta & \text{else.} \end{cases}$$

This result also allows us to deduce the occurrence of a reversal effect if the past market trend is negative. Let us assume that the market portfolio price dynamics follow the SDE system associated with the thresholded version of the EWMA HM, and the trend parameter of the price is given by

the following affine relationship:

$$\mu_t = r + \lambda\sqrt{V_t},$$

where r is the free-risk rate. To determine the type of relationship linking the respective dynamics of m and μ , we consider the following simplified case:

$$\lim_{\tau_1, \xi \rightarrow 0} V_t = \nu_t^2.$$

Under this hypothesis, we have

$$\frac{d\mu_t}{dm_t} = \begin{cases} 0 & \text{for } m_t \geq \bar{m} \\ -\beta\lambda & \text{else.} \end{cases}$$

Consequently, in this simplified framework, if $m_t < \bar{m}$, the joint dynamics of m and μ are negatively correlated even if the two Brownian motion $\{B\}_{t \in T}$ and $\{W\}_{t \in T}$ are independent (i.e. $\rho = 0$). More generally, this relation remains true if τ_1 is strictly positive and $\rho \in [-1 : 0]$. In financial terms,

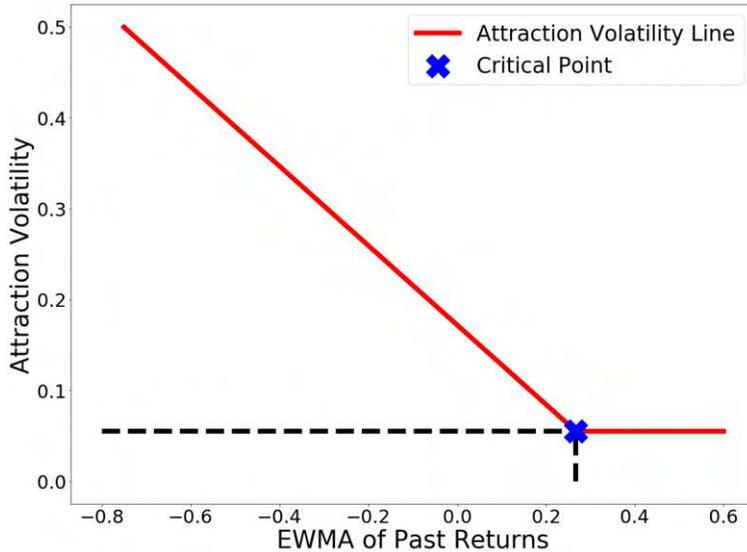


Figure 2.1: The relationship between the attraction volatility process and the trend parameter m_t in the thresholded EWMA Heston model.

which causes an asymmetry depending on the type of trend. Thus, positive trends such as $m_t > \bar{m}$ do not produce this type of reversal effect which results from a movement in the attraction volatility.

this relationship can be viewed as a reversal effect, in the sense that a negative past trend (such as $m_t < \bar{m}$) increases the actual price trend. However, note that unlike the assumption in behaviorist explanations ([84]), this model reversal effect is not a market anomaly. Indeed, the increase in the returns trend of a given asset following a strong fall in its price is due not to the correction of a market overreaction but to a pure market premium effect. In addition, this reversal effect emerges only if $m_t < \bar{m}$

2.2.3 Comparison with volatility models based on quadratic Hawkes processes

The EWMA HM has some links to the quadratic volatility models introduced by Blanc *et al.* ([39]) and developed by Dandapani *et al.* ([80]). To highlight the similarities between these models, we focus on the case of the thresholded version of the EWMA HM where $\mu, \xi, \tau_1 = 0$. In this specific case, we have:

$$\begin{cases} \frac{dS_t}{S_t} &= \sqrt{V_t} dW_t \\ V_t &= \underline{\nu}^2 + \mathbb{1}_{\{m_t < \bar{m}\}} (\beta^2 (\bar{m} - m_t)^2 + 2\underline{\nu}\beta(\bar{m} - m_t)) \\ m_t &= \frac{1}{\tau} \int_0^t e^{\frac{1}{\tau}(u-t)} \frac{dS_u}{S_u}. \end{cases}$$

This system of equations can be compared with the system of equations satisfied by different quadratic volatility models. First, let us focus on the pure quadratic case of the volatility model introduced by Blanc *et al.* ([39]), where the kernel function takes the following exponential form $k(t) = \sqrt{\frac{2}{\tau}} e^{-\tau t}$. For simplicity, we describe this model as the pure quadratic volatility model (PQVM). Notice that both the PQVM and the EWMA HM satisfy the following equations:

$$\begin{cases} \frac{dS_t}{S_t} &= \sqrt{V_t} dW_t \\ V_t &= \underline{\nu}^2 + b_t (b_t (\bar{m} - m_t)^2 + 2\sigma(\bar{m} - m_t)) \\ m_t &= \frac{1}{\tau} \int_0^t e^{\frac{1}{\tau}(u-t)} \frac{dS_u}{S_u}. \end{cases}$$

with respectively

	Pure Quadratic Volatility Model	EWMA Heston Model
b_t	$\sqrt{\frac{2}{\tau}}$	$\mathbb{1}_{\{m_t < \bar{m}\}} \beta$
\bar{m}	0	$\frac{\alpha}{\beta}$
σ	0	$\underline{\nu}$

such as $\tau, \beta, \underline{\nu} \in \mathbb{R}_+, \gamma \in [0 : 1[$.

Using this formulation, we can see that these two models present several similarities. The most important one is that, in both models, the variance is a quadratic function of an EWMA of past returns. Note however that, in contrast to the PQVM, in the EWMA HM the linear term in m_t is non-zero. This difference stems from the linear sensitivity to m_t of the volatility (if $m_t < \bar{m}$) in the EWMA HM, while in the PQVM, it is the variance that is linearly sensitive to m_t . Another important specificity of the model which constitutes one of the main differences with the various quadratic volatility models discussed in the literature ([39], [80], [106]) is the switching behavior caused by the indicator function $\mathbb{1}_{\{m_t < \bar{m}\}}$. This is confirmed if we consider the quadratic rough HM (QRHM). In contrast to the PQVM, the QRHM encodes the empirical asymmetry of the feedback

effect. However, this asymmetry is not captured in exactly the same way as in the EWMA HM. To highlight this point, note that both the EWMA HM considered and the QRHM described in section 4 in [106] respect the following equations:

$$\begin{cases} \frac{dS_t}{S_t} &= \sqrt{v_t}dW_t \\ V_t &= \bar{\theta} + b_t(b_t(\bar{m} - m_t)^2 + 2\sigma(\bar{m} - m_t)) \\ m_t &= m_0 - \int_0^t f(t-s)m_s ds + \int_0^t f(t-s)\eta \frac{dS_u}{S_u}, \end{cases}$$

with respectively:

	Quadratic Rough Heston Model	EWMA Heston Model
b_t	a	$\mathbb{1}_{\{m_t < \bar{m}\}}\beta$
σ	0	p
η	ν	1
$f(u)$	$u^{\alpha-1} \frac{\lambda}{\Gamma(\alpha)}$	τ

such as $\nu, a, \lambda, \tau, p \in \mathbb{R}_+, \alpha \in [0.5 : 1]$.

Again, one of the main differences between the two models concerns the parameter b_t , and the related switching behavior of the EWMA HM considered. Whereas in the QRHM this parameter is a positive constant, in the EWMA HM it is the product of a constant and an indicator function. This difference implies an alternative way to capture the empirical asymmetry between positive and negative trends in volatility. In the QRHM, this asymmetry is covered only by the parameter \bar{m} . In concrete terms, this means that even though positive and negative trends have asymmetric impacts on the volatility, a strong positive trend such as $m_t > \bar{m}$ increases volatility in the QRHM. However, since $m_t > \bar{m}$, the trend measured by m has no impact on the volatility in the EWMA HM considered. In addition, while in the QRHM the variance is an affine function of the squared difference between m_t and \bar{m} , the variance depends also on the absolute value of this difference. At the same time, unlike in the PQVM, the trend parameter m in the QRHM is not an EWMA of past returns. In this case the kernel used to model the dependency between past price trends and spot volatility, is a rough one. This change has several important implications. First, the model is no longer Markovian in terms of its variables (S, V) . Second, an important consequence of this choice is that the "memory" of m decays as a power law in the QRHM, while in the EWMA HM the memory decays exponentially. Note however that in both cases, m is highly sensitive to recent returns. Third, the reasons for the emergence of irregular behavior of the volatility over short time scales are different between the QRHM and the general EWMA HM. In the QRHM, the roughness of the volatility paths stems from the characteristics of its power law kernel whereas in the EWMA HM, this erratic volatility process behavior emerges under certain conditions due to antagonist effects produced on the one hand by the reversal effect of the variance process, and on the other hand by the volatility of this variance.

2.2.4 Model discretization scheme

As referred to in the introduction, a major strength of the EWMA HM compared to rough volatility models is its ease of use for Monte-Carlo simulations. The simplest way to run discrete-time simulations of the thresholded version of the EWMA HM is to use a Euler-type discretization scheme ([151]), analogous to a full truncated Euler scheme for the standard HM ([142]). We thus propose the following simple discretization scheme⁴:

$$\begin{aligned} S_{t+1} &= S_t + S_t(\mu_t \Delta t + \sqrt{(V_t)_+} Z_{t+1}) \\ m_{t+1} &= m_t + \frac{1}{\tau_2} \left(\frac{S_{t+1} - S_t}{S_t} - m_t \Delta t \right) \\ \nu_{t+1} &= \underline{\nu} + \max(0, \alpha - \beta m_{t+1}) \\ V_{t+1} &= V_t + \frac{1}{\tau_1} (\nu_{t+1}^2 - (V_t)_+) \Delta t + \xi \nu_{t+1} \sqrt{(V_t)_+} (\rho Z_{t+1} + \sqrt{1 - \rho^2} X_{t+1}), \end{aligned}$$

where X_t and Z_t are i.i.d. random variables associated with the Gaussian distribution $\mathcal{N}(0, \Delta t)$.

It should be noted that this scheme requires a sufficiently small time step Δt to ensure its stability. First, it is clear that $\Delta t/\tau_1$ and $\Delta t/\tau_2$ must be strictly less than 1. Moreover, the value of Δt determines the probability of obtaining a negative value for V without the adjustment applied by $(\cdot)_+$. Indeed:

$$V_{t+1} \sim \mathcal{N} \left(V_t + \frac{1}{\tau_1} (\nu_{t+1}^2 - V_t) \Delta t, \xi \nu_{t+1} \sqrt{V_t \Delta t} \right).$$

Consequently:

$$\mathbb{P}(V_{t+1} \leq 0) = \Phi \left(\frac{\frac{1}{\tau_1} (V_t - \nu_{t+1}^2) \Delta t - V_t}{\xi \nu_{t+1} \sqrt{V_t \Delta t}} \right),$$

where Φ is the cumulative distribution function of the standard normal distribution. Also, since $\nu, \epsilon > 0$ and as soon as $\Delta t/\tau_1 < 1$:

$$\lim_{\Delta t \rightarrow 0_+} \Phi(-\infty) = 0.$$

Regarding the trend component of $\{S\}_{t \in T}$, for the following simulations we assume $\mu_t = \lambda \sqrt{V_t}$. This implies replacing μ_t with $\lambda \sqrt{(V_t)_+}$ in the discretization scheme. In addition, to simulate future potential volatility trajectories for an asset from a given date t , requires m_t to be estimated. The

⁴The convergence of this discretization scheme is still to be demonstrated.

natural estimator of this parameter is given by the following discrete EWMA:

$$m_t = \frac{1}{\tau_2} \sum_{i=1}^n e^{-\frac{i\Delta t}{\tau_2}} \frac{S_{t-i\Delta t} - S_{t-(i+1)\Delta t}}{S_{t-(i+1)\Delta t}}.$$

To fix the value of n , we can write that:

$$\frac{1}{\tau_2} \int_{t-n\Delta t}^t e^{\frac{1}{\tau_2}(u-t)} du = 1 - e^{-\frac{\tau n \Delta t}{\tau_2}}.$$

Consequently, for a given proportion γ of the total weighting of the EWMA (in a continuous time framework), we have:

$$\begin{aligned} \gamma &= 1 - e^{-\frac{\tau n \Delta t}{\tau_2}} \\ n &= -\tau_2 \frac{\log(1 - \gamma)}{\Delta t}. \end{aligned}$$

The above equation allows us to fix n , given $\tau_2, \Delta t$ and γ .

2.3 Consistency of the model with market data

To assess the effectiveness of the model for capturing empirical financial phenomena, we use as market data the price and volatility dynamics of the S&P500 for the period July 23, 2001, to July 23, 2021. To estimate the volatility, we use the square root of the realized variance computed from 5-min samples provided by the Oxford-Man Institute of Quantitative Finance⁵. We compare these market data with a set of synthetic data comprised of 1000 simulations for each of 20 years, generated from the Euler scheme of the thresholded version of the EWMA HM, using a time step equal to $\frac{1}{25200}$ expressed in years. To fit the model, we use the ad-hoc estimation procedure presented in appendix 2.B. This results in the following parameters:

λ	ρ	τ_1	ξ	τ_2	$\underline{\nu}$	α	β
0.5575	-0.465	0.0013	42.95	0.276	0.0595	0.1033	0.419

Table 2.1: The parameters obtained from the calibration procedure.

These parameter values need some explanation. First, let us consider the value taken by λ which is setting the long-term drift. To obtain an order of magnitude for the expected value of the drift (i.e. $\mathbb{E}[\mu_t]$), we can use the product of the mean of the realized volatility of the S&P500 by this parameter λ as a proxy. The mean of the realized volatility is 13.1%, so the value obtained is 7.3% ($0.5575 \times 0.131 \approx 7.3\%$). This rough estimation of $\mathbb{E}[\mu_t]$ is consistent with the annualized empirical

⁵The data are available at: <https://realized.oxford-man.ox.ac.uk/data>.

daily mean of the returns which is equal to 7.3% (the S&P500 empirical mean of daily returns is equal to 0.028%: thus $(1 + 0.028\%)^{252} \approx 7.3\%$). Also interesting is that the value of τ_1 is extremely low compared to the range of values it takes in the standard HM fitted on market data ([160]). Since the unit of τ_1 is in years, τ_1 equal to 0.0013 means that the average duration of the deviation of the variance from this attraction value is of the order of 0.5 days ($0.0013 \times 365 \approx 0.5$). In other words, the reversion of the variance process toward the attraction variance is a short term phenomenon. This should be understood in the context of the very high value taken by ξ , implying that over very short time scales the variance process is dominated by randomness. The combination of these two effects - high randomness of V_t at the time scale dt and short time reversion toward ν_t^2 - results in erratic volatility behavior over the short term, and allows the roughness of empirical volatility paths to be mimicked. Consequently, while in the standard HM the volatility serial correlation is highly dependent on τ_1 , here, it depends almost exclusively on the attraction variance ν_t^2 . As a result and given the value of τ_2 , the autocorrelation of the volatility depends on a medium-term price trend of the order of 3.3 months ($0.276 \times 12 \approx 3.3$), and not directly on past endogenous volatility movements. Thus, there is a decoupling between the short and the longer-term behavior of the volatility, a property whose importance was emphasized notably by Bannengsen *et al.* ([29]).

2.3.1 The relationship between EWMA of returns and the volatility

If the EWMA HM is consistent, the market data should exhibit an affine relationship between the EWMA of past returns and realized volatility when the EWMA of past returns is significantly negative. However, if the EWMA of past returns is significantly positive, these two variables should be fairly independent. These relationships are confirmed in figure 2.2. Another interesting fact that emerges in relation to the optimal slope coefficient β (optimal in the least square sense) of the regression of the form $\sqrt{V_t} = \nu + (\alpha - \beta m_t)_+ + \epsilon_t$ is that for different past EWMA of the returns (with different values of τ), the optimal slope coefficient is very close to⁶: $\sqrt{\frac{2\tau}{\pi}}$. $\sqrt{2\tau}$ occurs because it corresponds to the inverse of the asymptotic standard deviation of an EWMA with parameter τ of a Brownian motion (see appendix 2.C). The rationale for $\sqrt{\pi}$ is not so obvious. As it stands, this could either be a simple value with no particular significance or suggest a deeper relationship, possibly related to $\frac{1}{\Gamma(0.5)}$, with $\Gamma(\cdot)$ the gamma function.

⁶For other examples of this empirical relationship, see appendix 2.D.

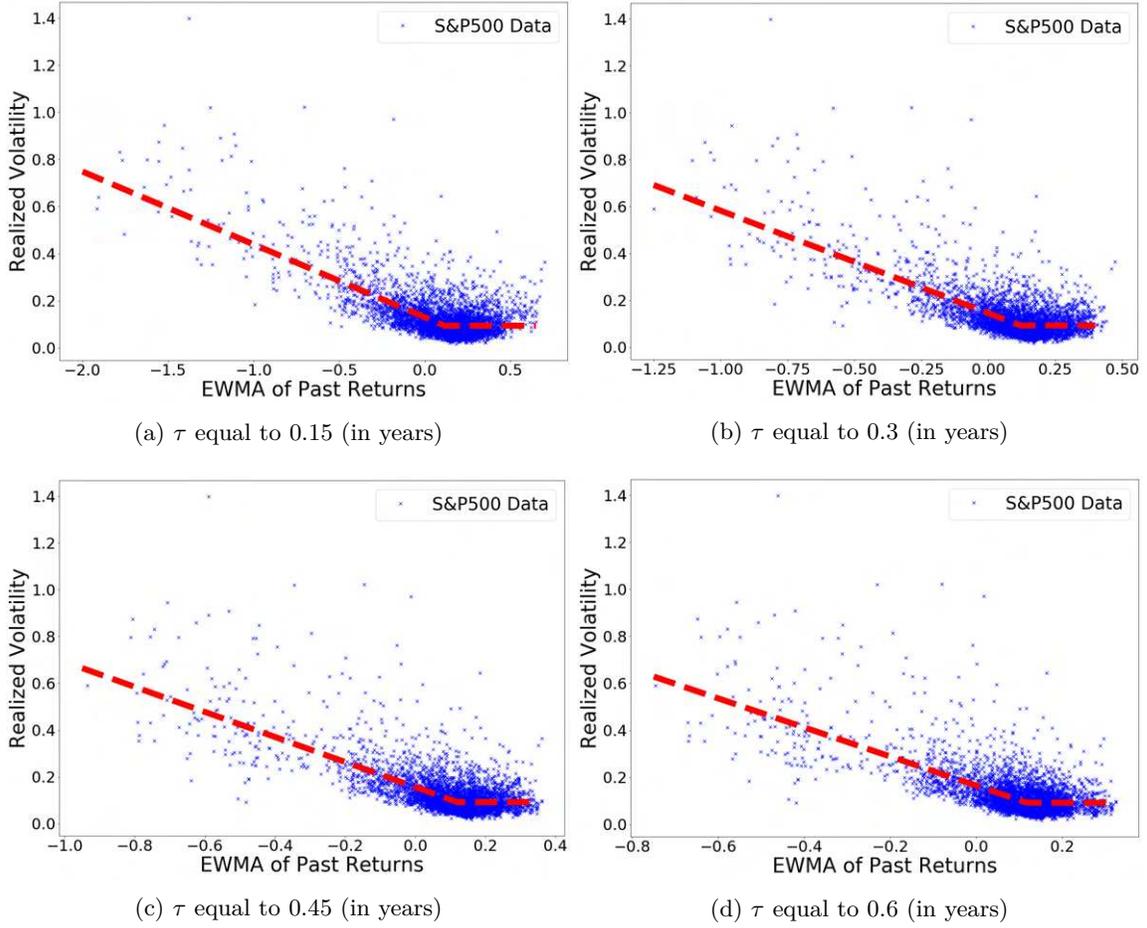


Figure 2.2: The empirical relationship between the realized volatility of the S&P500 and the EW-MAs of its past returns for different values of τ expressed in years. The red lines are regressions of the form $\sqrt{V_t} = \underline{\nu} + (\alpha - \beta m_t)_+$ with a slope coefficient equal to $\sqrt{\frac{2\tau}{\pi}}$. These regressions are respectively associated with the following R -squared: 0.558, 0.547, 0.523, 0.498.

2.3.2 The dynamics of price and volatility

Let us begin our comparison of the synthetic data produced by the EWMA HM and the S&P500 empirical data, by focusing on the joint price and volatility dynamics. A general observation that emerges from this comparison (see figures 2.3 and 2.4) is that the simulated price and volatility trajectory features resemble those of the empirical data.

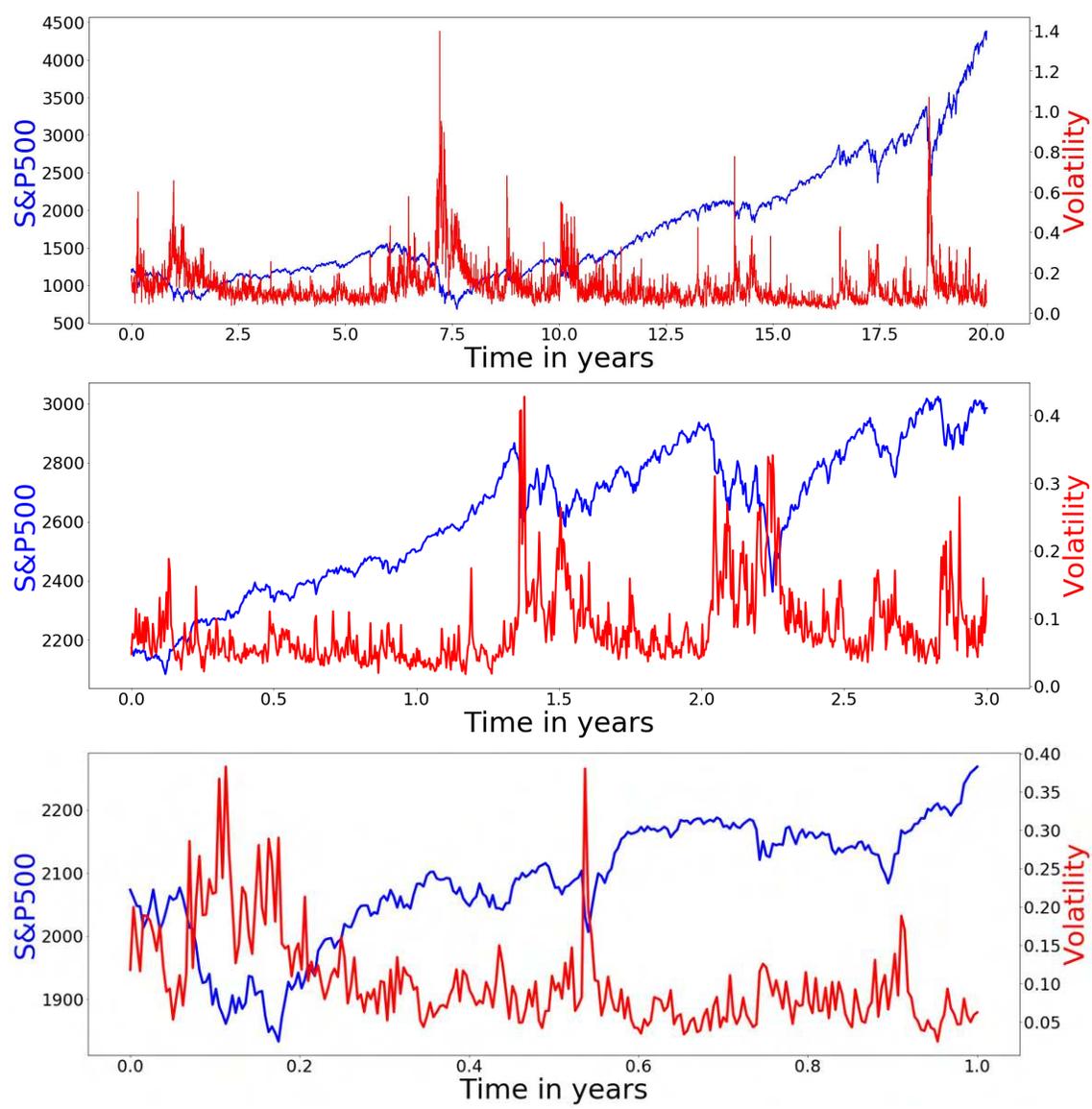


Figure 2.3: Empirical joint evolution of the S&P500 and its annualized daily realized volatility over 20, 3 and 1 years.

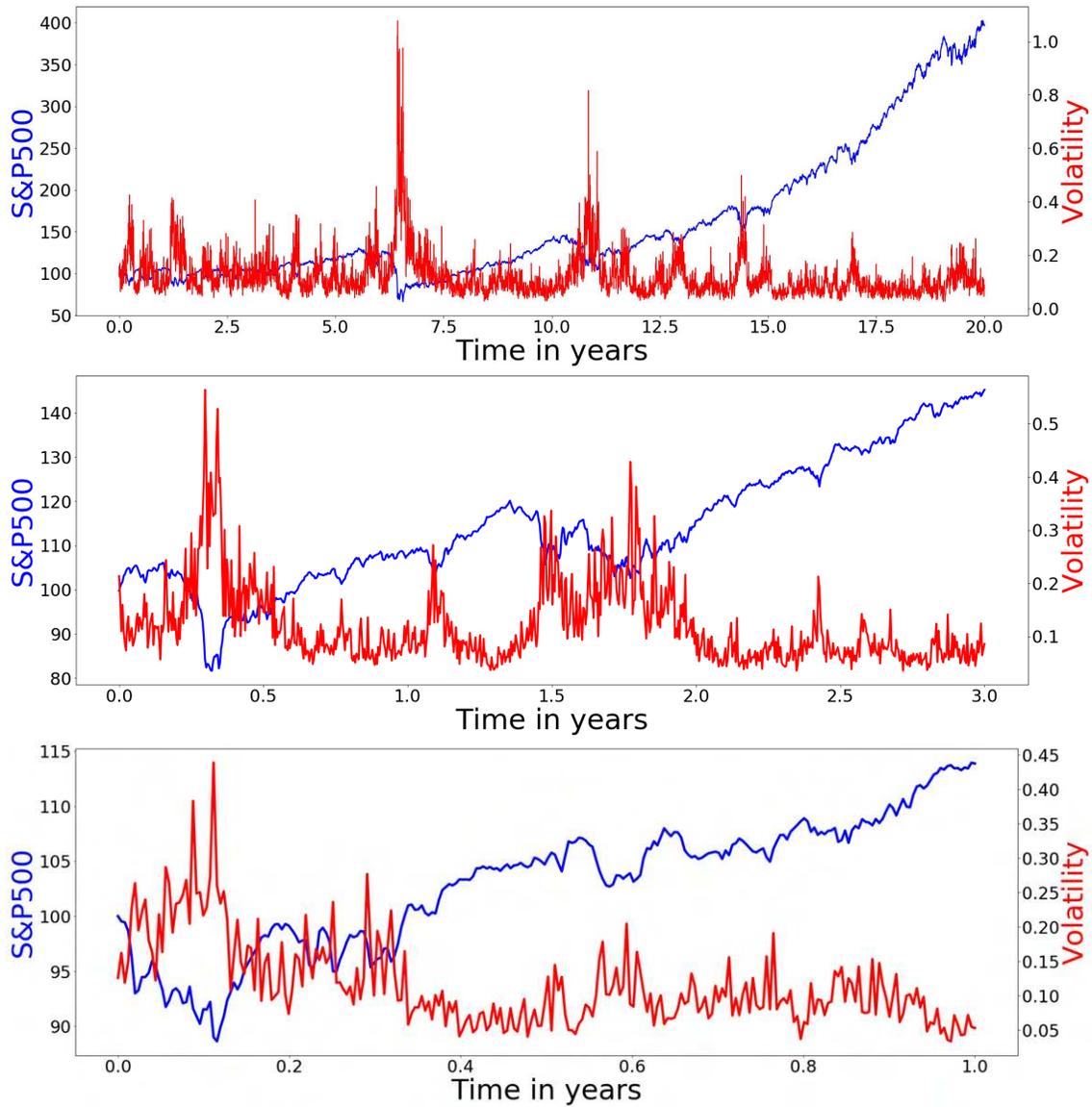


Figure 2.4: Joint evolution of the S&P500 and its associated volatility over 20, 3 and 1 years simulated from the fitted EWMA Heston model.

In particular, in both cases, volatility spikes immediately after large negative returns of the index and then decreases gradually. Also, in each case the long positive price trends are periods where volatility is relatively stable and low. If we consider the volatility dynamics more specifically, we see that the model is able also to capture the empirical coexistence of long periods of low volatility with other periods when volatility is very high. Moreover, the irregular behavior of volatility over short time scales is, from a visual standpoint, accurately captured by the model.

These general observations require completion by a more in-depth comparison between the features

characterizing the volatility dynamics produced by the model and the features of the realized volatility. We next focus on two features of the volatility: its autocorrelation function (ACF) and its roughness.

2.3.2.1 The autocorrelation function of the volatility process

The serial correlation of the volatility is an important feature of its dynamics. In figure 2.5 we observe that the ACF of the simulated volatility paths is generally consistent with the market data.

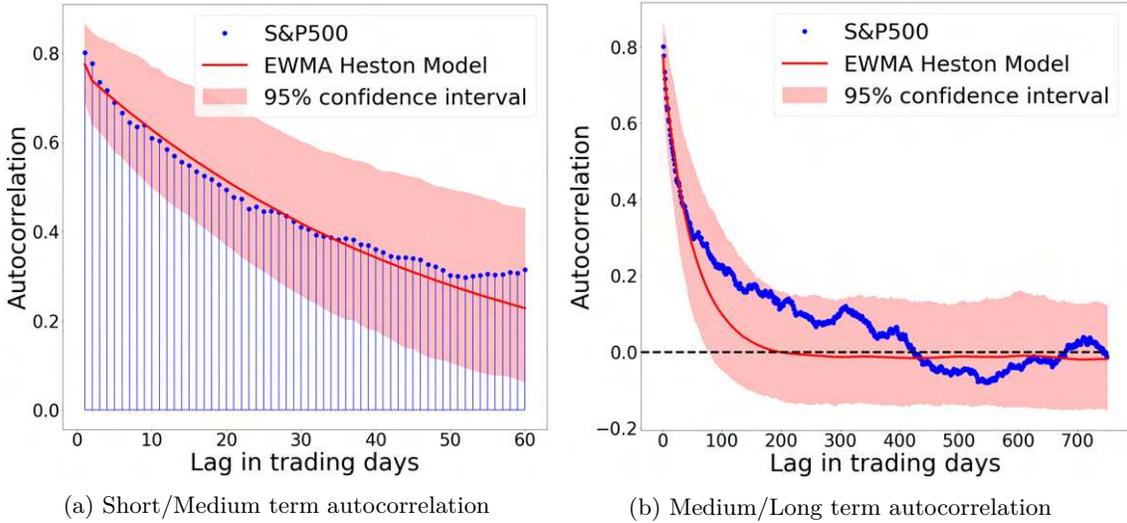


Figure 2.5: Autocorrelation of the volatility in the short and longer-term.

However, the short-medium term (between 1 and 60 trading days) and the longer-term must be analyzed separately. In the case of small lags (less than 60 trading days), the autocorrelations associated with the volatility paths generated by the model are very close to their empirical counterparts. Over this time horizon, not only are all empirical autocorrelations within the model's estimated confidence intervals⁷, they are also close to their expected values. In the case of longer lags, the situation changes slightly. The ACF exhibits faster decay in the model compared to the empirical data. However, this difference between the empirical and the model data should be tempered by the fact that the empirical autocorrelations are within the 95% confidence interval of the model. Consequently, it is difficult to be definitive about the capacity of the model to account accurately for the ACF in the real volatility process. However, the academic literature ([83], [70]) tends towards the hypothesis that the model ACF decays too rapidly compared to the market data. This fast decay is due to the fact that the ACF of the volatility is determined primarily by an exponential kernel through the trend process m_t . A possible modification of the thresholded version of the EWMA HM model allowing a volatility process with a more flexible memory structure would involve using a linear combination of several EWMA's of past returns rather than a unique EWMA.

⁷This confidence interval is estimated using the 2.5% and 97.5% simulation quantiles.

In these cases, m_t is replaced by \bar{m}_t defined as follows:

$$\bar{m}_t = \sum_{k=1}^d w_k \int_0^t e^{-\tau_k(t-u)} \frac{dS_u}{S_u} \quad (2.4)$$

In this framework, the model still belongs to the family of EWMA HM introduced previously (2.1), but no longer corresponds to the specific case of 2.2.

2.3.2.2 The question of volatility roughness

Although visually the EWMA HM would seem to capture the empirical roughness of the volatility paths, this requires deeper investigation. First, note that the EWMA HM is not a rough volatility model. Indeed, the EWMA HM is not based on a fractional process similar to a fractional Brownian motion. Rather, it is a diffusion model whose two sources of randomness are standard Brownian motions which by definition both have a Hurst exponent equal to 0.5. However, Rogers ([178]) and more recently Cont and Das ([72]) show that Brownian diffusion models may mimic some of the features of rough models of interest for volatility modeling. Cont and Das emphasize also that the empirical roughness of the volatility paths might be caused not by the volatility process but by a microstructure noise. Here, we leave aside consideration of the origins of the empirical rough behavior of realized volatility and adopt a phenomenological approach aimed at evaluating how well the EWMA HM mimics the empirical roughness of the volatility paths. To do this, we measure the roughness of both the empirical and model volatility time series using two different methods.

The first method is the most popular in the academic literature. It consists of estimating a roughness index by running an ordinary least square (OLS) regression of the form:

$$\log(\gamma(k\Delta, 2)) = c + a \cdot \log(k\Delta) + \epsilon_k, \quad k = 1, \dots, m,$$

where Δ is the time step between observation of the volatility process, m is a bandwidth parameter, and $\gamma(k\Delta, 2)$ is the empirical second order variogram defined as follows:

$$\gamma(k\Delta, 2) = \frac{1}{N-k} \sum_{i=1}^{N-k} |\log(\sigma_{(i+k)\Delta}) - \log(\sigma_{\Delta i})|^2,$$

where N is the number of observations in the sample considered. Due to the nature of the data used, sampling in Δ is conducted daily, and not on a logarithmic scale. In order to remove the bias induced by this sampling, we weight each observation k by

$$w_k = \log(k+1) - \log(k).$$

The slope coefficient \hat{a} obtained from the regression is used to estimate a roughness index, denoted \hat{H} , equal to $0.5\hat{a}$. In the case of a rough process, \hat{H} can be considered a Hurst exponent estimator ([105]). However, in the present case, \hat{H} is treated only as a metric to compare the behaviors of

the real and the EWMA HM volatility processes.

The second method used to measure the roughness of time series was proposed by Bennedsen *et al.* ([29]). Unlike the first method described, it has the advantage of being able to capture a possible non-affine relationship between $\gamma(k\Delta, 2)$ and $\log(k\Delta)$. The method consists of a non-linear least squares (NLLS) regression of the form:

$$\gamma(k\Delta, 2) = c + b(k\Delta)^{2a+1} + \epsilon_t, \quad k = 1, \dots, m.$$

The index \hat{H} is equal to $\hat{a} + 0.5$.

Table 2.3 reports the roughness index \hat{H} of the S&P500 and simulated volatilities obtained using OLS and NLLS methods with different bandwidth parameters.

Bandwidth parameter	10	20	60	125	250	500	750
\hat{H} of S&P500 data (OLS regression)	0.137	0.136	0.137	0.135	0.13	0.126	0.124
\hat{H} of model data (OLS regression)	0.112	0.121	0.142	0.15	0.149	0.14	0.134
\hat{H} of S&P500 data (NLLS regression)	0.082	0.078	0.116	0.072	0.046	0.091	0.07
\hat{H} of model data (NLLS regression)	0.054	0.126	0.283	0.173	0.042	-0.071	-0.118

Table 2.2: Roughness index \hat{H} of both S&P500 and simulated volatilities computed from OLS and NLLS methods with different bandwidth parameters.

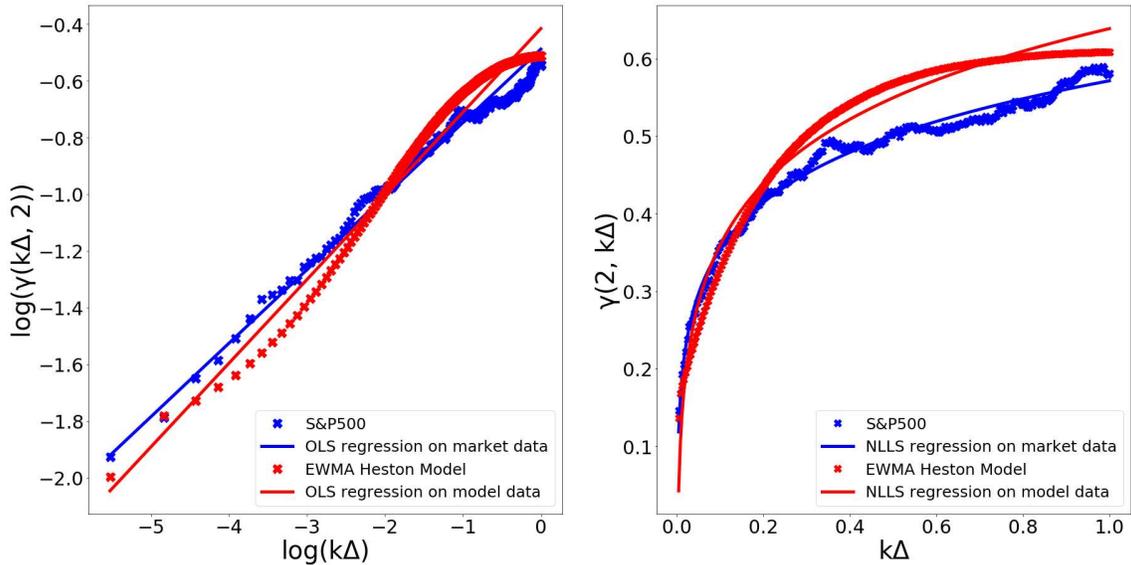


Figure 2.6: The relationships between $\log(\gamma(k\Delta, 2))$ and $\log(k\Delta)$ (left plot) and between $\gamma(k\Delta, 2)$ and $k\Delta$ (right plot): S&P500 data vs EWMA Heston Model.

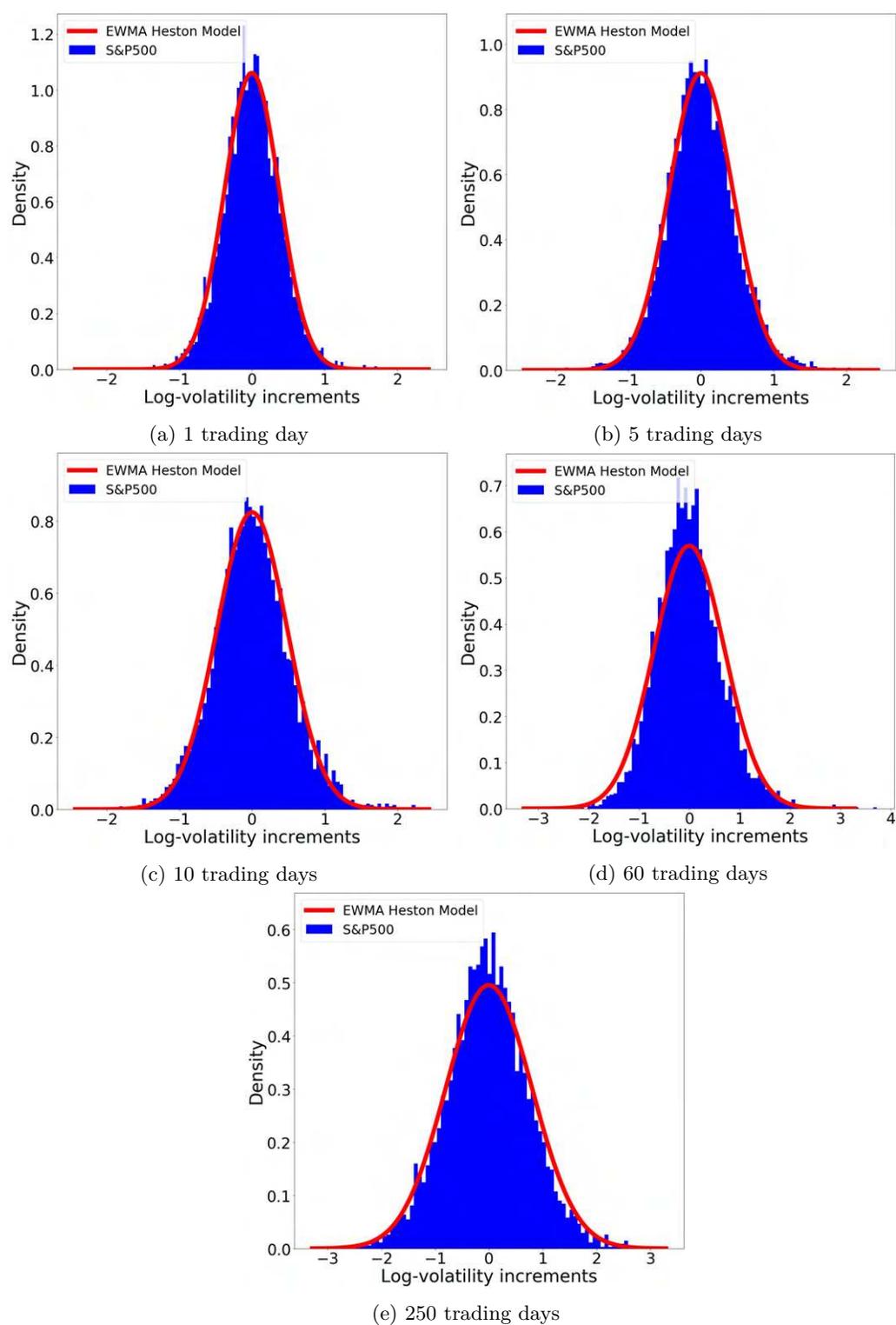


Figure 2.7: Log-volatility increments distribution for different time horizons: S&P500 data vs EWMA Heston Model.

First, when considering the \hat{H} exponents obtained from both estimation methods, while the realized volatility \hat{H} depends only slightly on the bandwidth parameters used, this parameter has a more significant influence on the value of \hat{H} for the volatility generated by the model. The reason for this difference becomes clearer when looking at figure 2.6. In terms of the empirical data, the relationship between $\log(\gamma(k\Delta, 2))$ and $\log(k\Delta)$ is captured very well by an OLS model, regardless of the time scale considered. However, the relationship between these two variables is s-shaped in the case of the volatility process generated by the model: it is convex up to about 50 trading days and then turns concave. Therefore, the model fails to capture the empirical relationship between the log of the expected increments in log-volatility and the log of the time scale. As a result, it does not reproduce the monofractal behavior observed in empirical volatility series. This is particularly important, as this empirical relationship is a defining characteristic of volatility behavior, irrespective of the type of asset considered [105]. The significance of this model limitation requires some qualification insofar as over reasonable time scales the distribution of the log-volatility increments generated by the model remain close to the empirical increments, as shown in figure 6.27. This is reflected in the fact that the distributions are close to Gaussian for both the empirical and simulated data. Furthermore, the low (but significant) divergence between the variances in the model log-volatility increments and their empirical counterparts over different time scales explains a relatively similar order of magnitude of the OLS estimators of the Hurst coefficient. In contrast, the more important divergence of the exponent \hat{H} obtained using NLLS is explained by the fact that this regression captures the nonlinearity of the relationship linking $\gamma(k\Delta, 2)$ and $\log(k\Delta)$. In addition, it should be noted that \hat{H} takes negative values, which emphasize that in the present context it is not relevant to consider \hat{H} as a Hurst exponent estimator ([146]). Nevertheless, in the same way as a volatility process with a long-range autocorrelation, this model limitation could be fixed by replacing m_t with a linear combination of multiple EWMA of past returns (equation 2.4).

2.3.3 Volatility and log-returns distributions for different time horizons

In what follows we focus on the ability of the model to reproduce the empirical log-returns and average volatility distributions for different time horizons. First, it should be pointed out that assessing the quality of the model on these criteria using standard statistical tools is not straightforward ([63]). These probability distributions do not originate from i.i.d. samples but from realizations of a path-dependent process. The practical implications of this are that more often than not, the results of a Kolmogorov-Smirnov test on log-returns or volatility distributions from two different simulations of the fitted model (with the same parameters) reject the null hypothesis at a 99% confidence level. Consequently, rather than proceeding with this type of statistical test based on the i.i.d. data assumption, it is better to check whether the different moments of the distributions from the market data are in the same range as those emerging from the fitted EWMA HM simulations. To do this, we use the already mentioned synthetic set of 1000 simulations for each 20 years. These simulations allow us to compute the mean as well as the 1st and the 9th deciles of the first four moments of the log-returns distributions and the average volatility distributions for different time horizons. Tables 2.3 and 2.4 report these metrics and the corresponding empirical moments.

Variable	Time horizon						
	1	5	20	60	125	250	750
Mean of market data	0.0003	0.0013	0.0051	0.0154	0.0321	0.0641	0.1923
Avg. mean of sim.	0.0002	0.001	0.0038	0.0114	0.0237	0.0474	0.1423
1 st decile of mean of sim.	-3e-6	-1e-5	-5e-5	-0.0002	-0.0003	-0.0007	-0.002
9 th decile of mean of sim.	0.0004	0.0018	0.007	0.0211	0.0439	0.0878	0.2635
Std. dev. of market data	0.0113	0.0245	0.0338	0.0789	0.1178	0.1673	0.2341
Avg. of st.deviation	0.0088	0.0233	0.047	0.0791	0.1098	0.1479	0.2304
1 st decile of std. deviation of sim.	0.0074	0.0193	0.0373	0.0581	0.0751	0.0931	0.1287
9 th decile of std. deviation of sim.	0.0105	0.028	0.0581	0.103	0.1502	0.2132	0.3631
Skewness of market data	-0.39	-0.97	-1.7	-1.49	-1.71	-1.51	-1.2
Avg. skewness of sim.	-1.12	-1.19	-1.55	-1.77	-1.64	-1.25	-0.55
1 st decile of skewness of sim.	-1.77	-1.93	-2.62	-2.9	-2.72	-2.24	-1.24
9 th decile of skewness of sim.	-0.63	-0.61	-0.78	-0.92	-0.8	-0.48	0.08
Kurtosis of market data	9.32	7.77	8.42	5.25	5.57	3.3	0.72
Avg. kurtosis of sim.	13.23	10.6	9.04	7.03	4.81	2.3	-0.09
1 st decile of kurtosis of sim.	5.08	3.51	2.51	1.69	0.81	-0.1	-0.97
9 th decile of kurtosis of sim.	24.96	19.96	17.68	14.49	10.88	6.13	0.86

Table 2.3: The first four moments of log-returns distributions for different time horizons in trading days: S&P500 data vs synthetic data generated from the fitted EWMA HM.

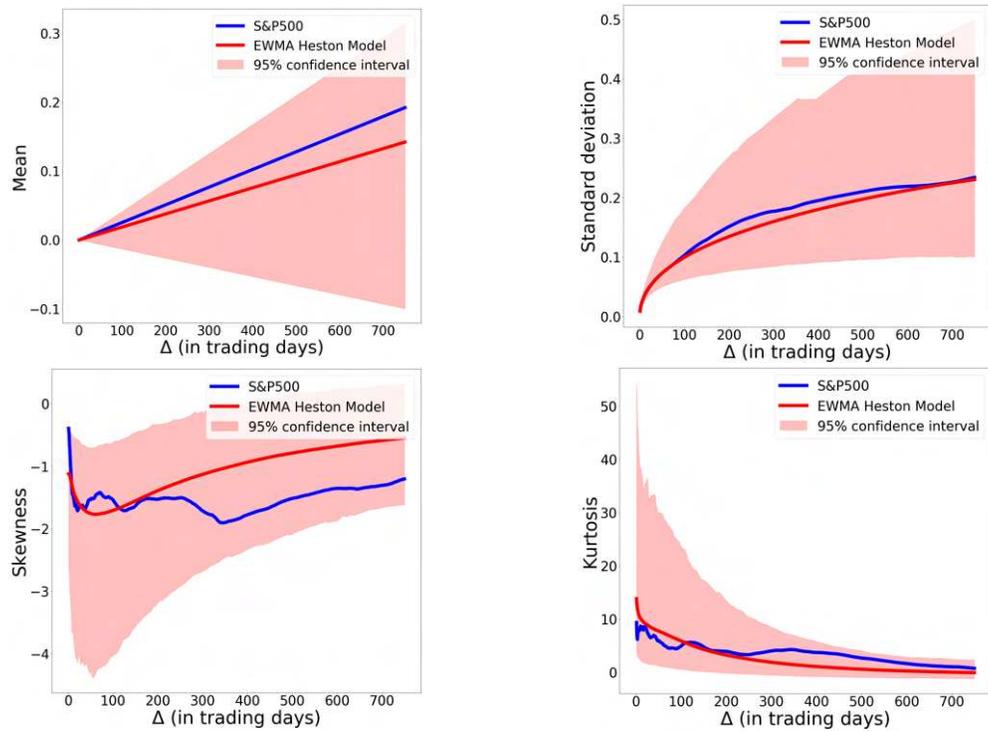


Figure 2.8: Evolution of the four first moments of the log-returns distributions in the function of the considered time-horizon: S&P500 data vs synthetic data generated from the fitted EWMA HM.

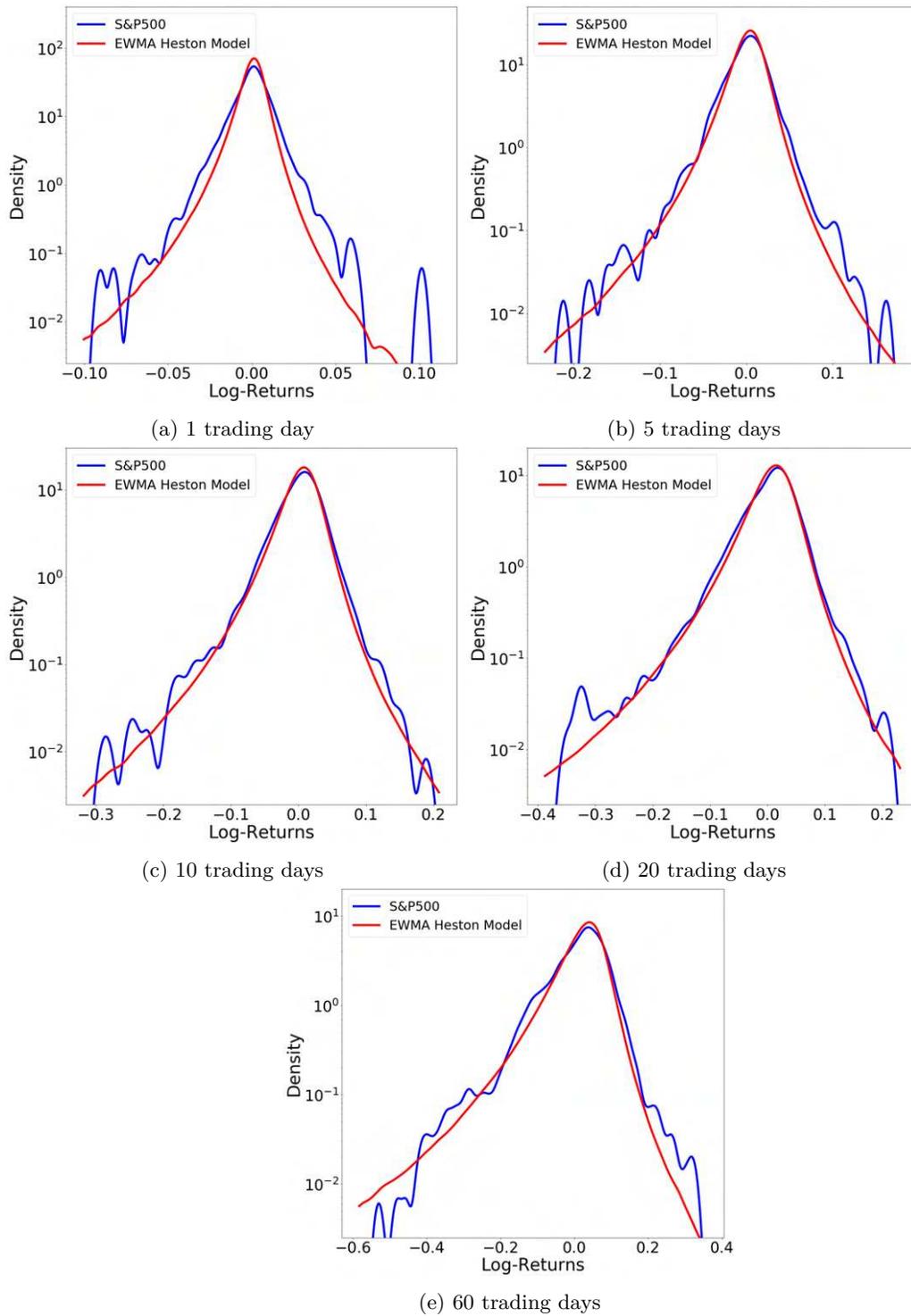


Figure 2.9: Log-returns distributions for different time intervals with logarithmic y-axis: S&P500 data vs EWMA Heston Model.

Variable \ Time horizon	1	5	20	60	125	250	750
Mean of market data	0.131	0.131	0.131	0.131	0.131	0.131	0.131
Avg. mean of sim.	0.129	0.129	0.129	0.129	0.129	0.129	0.129
1 st decile of mean of sim.	0.115	0.115	0.115	0.115	0.115	0.115	0.115
9 th decile of mean of sim.	0.145	0.145	0.145	0.145	0.145	0.145	0.145
Std. dev. of market data	0.0999	0.0903	0.0829	0.0732	0.0651	0.057	0.0403
Avg. of std.dev. of sim.	0.0887	0.0795	0.0741	0.0657	0.0559	0.0439	0.0253
1 st decile of std. dev. of sim.	0.0669	0.0581	0.0528	0.0449	0.0366	0.0272	0.0139
9 th decile of std. dev. of sim.	0.1163	0.1065	0.1009	0.0909	0.0796	0.0651	0.0402
Skewness of market data	3.42	3.18	2.88	2.41	2.05	1.55	0.93
Avg. skewness of sim.	2.27	2.4	2.23	1.98	1.67	1.26	0.54
1 st decile of skewness of sim.	1.83	1.45	1.34	1.12	0.85	0.48	-0.12
9 th decile of skewness of sim.	3.79	3.53	3.34	3.08	2.72	2.21	1.24
Kurtosis of market data	19.53	15.47	11.82	7.97	5.33	2.5	-0.37
Avg. kurtosis of sim.	12.95	9.64	7.84	5.75	3.8	1.8	-0.28
1 st decile of kurtosis of sim.	5.1	3	1.94	1.01	0.14	-0.57	-1.18
9 th decile of kurtosis of sim.	23.4	18.67	16.2	12.29	8.78	5.46	0.64

Table 2.4: The first four moments of average annualized volatility distributions for different time horizons: market data vs synthetic data generated from the fitted EWMA HM.

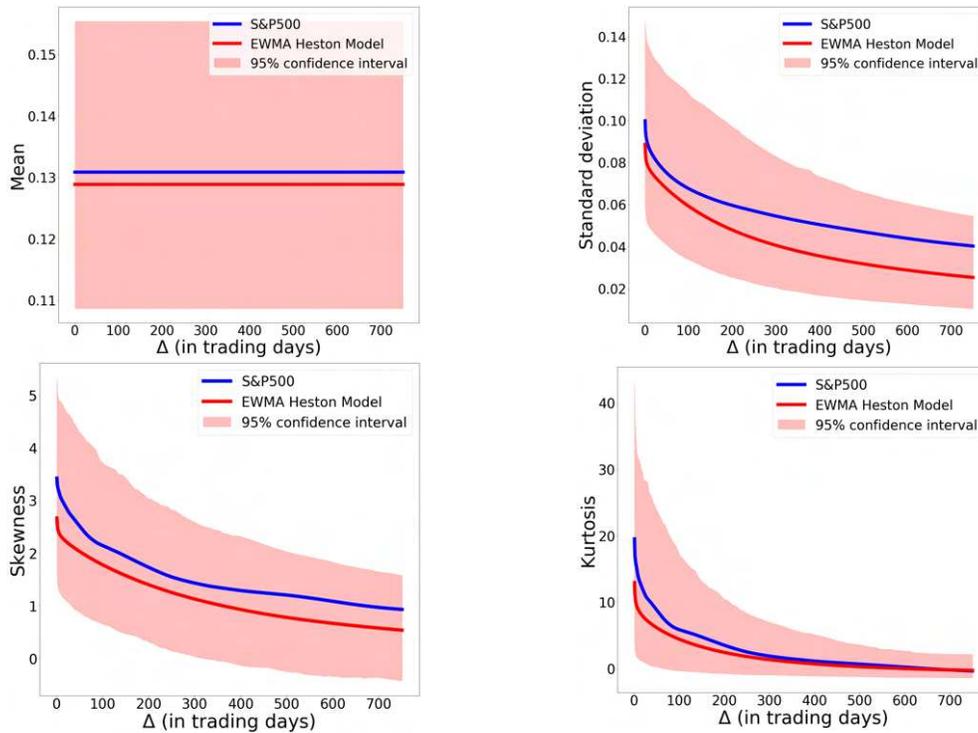


Figure 2.10: Evolution of the first four moments of annualized volatility distributions in the function of the considered time-horizon: S&P500 data vs synthetic data generated from the EWMA HM.

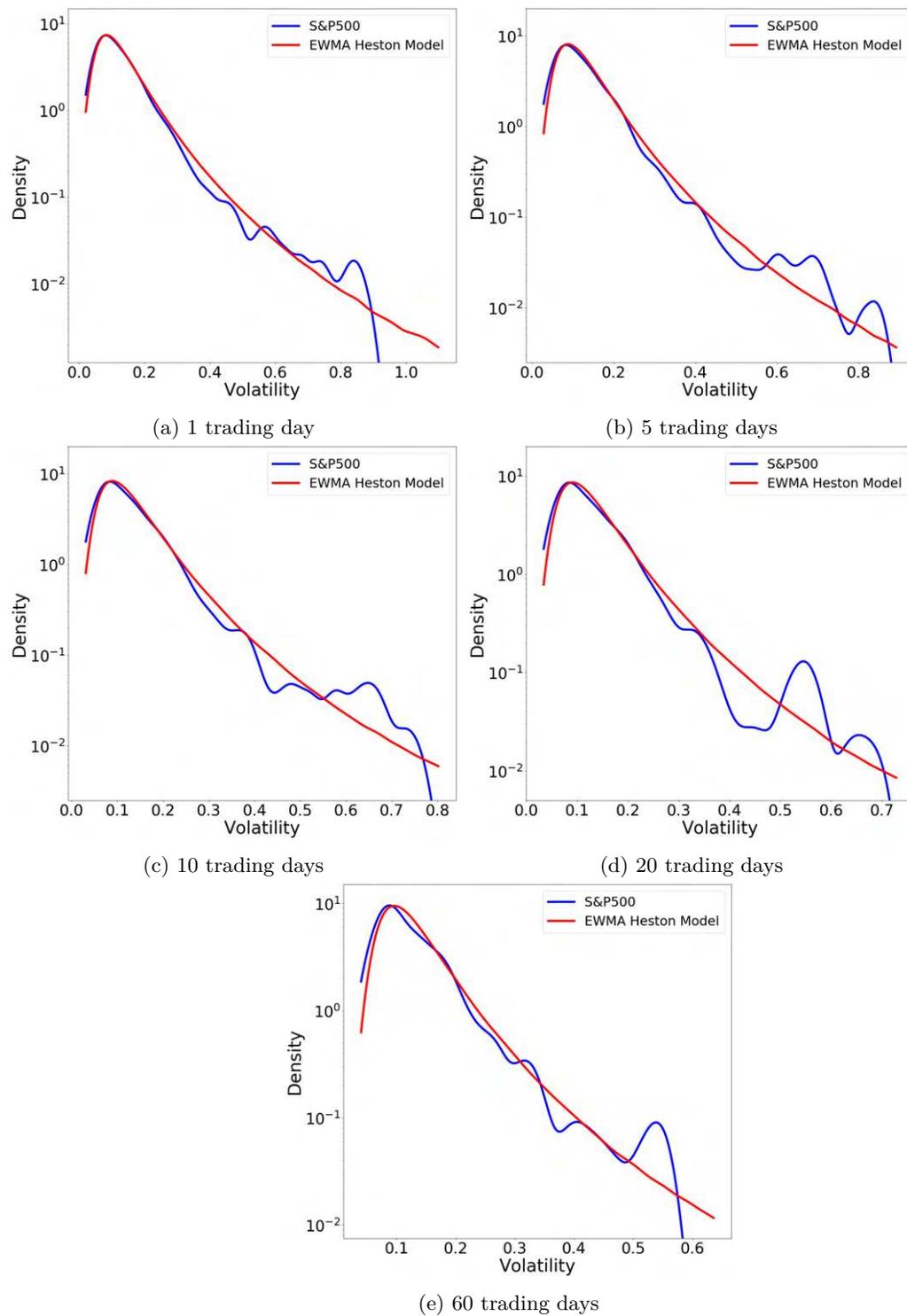


Figure 2.11: Annualized average volatility distributions for different time intervals with logarithmic y-axis: S&P500 data vs EWMA Heston Model.

2.3.3.1 The log-returns distributions

The log-returns distributions generated by the model are mostly in line with those based on the empirical data. For example, with the exception of the 1 trading day time horizon, the four first moments of the empirical log-returns distributions are between the corresponding 1st and the 9th deciles of the model for all the other time horizons considered⁸. In addition, the relationships between the moments of the log-returns distribution and the time-horizon are generally well reproduced by the model. In particular, as per the market data, in the short-medium terms, the log-returns distribution becomes more and more negatively skewed as the time horizon lengthens. This increasing asymmetry emerges clearly in figure for both the empirical and model distributions. However, the relationship is not monotonic and beyond a certain time-horizon threshold (approximately 65 trading days for the synthetic data), the skewness of the log-returns distribution decreases and in the long run tends to zero. In the case of the fourth moment of the log-returns distributions, the negative convex relationship between the value of the kurtosis and the time-horizon produced by the model is coherent also with the market data. Another characteristic of the empirical log-returns distributions is their heavy tails. Again, the tails of the log-returns generated by the model are mostly consistent with the empirical data, as shown in figure 2.9. This means that the model is able to capture the possibility of large price swings, a sign of the "wild randomness" of financial markets ([147]). However, despite these model qualities, it should be noted that unlike the log-returns distributions associated with longer time horizons, the model daily log-returns distributions exhibit significant divergence from the empirical distribution (see figure 2.3.3.1). This divergence is demonstrated in the too low standard deviation and a too negative skewness associated with the model distribution compared to the counterpart empirical data. This raises the question of the possibility of a normal source of randomness to accurately capture price fluctuations over short time scales.

2.3.3.2 The volatility distributions

We now consider the average volatility distributions for different time-horizons (see table 2.3 and figure 2.11), and observe a consistency between the model and the market data similar to that observed for the log-returns distributions. Thus, for all the time horizons considered, the first four moments of the average volatility distributions computed using the market data are between the corresponding 1st and the 9th deciles of the model and are close to their expected values. In addition, in the case of both the empirical data and the data generated by the model, the standard deviation, skewness, and kurtosis of the average volatility distribution adopt a decreasing convex relationship with the time horizon considered. More generally, the shapes of the model and empirical distributions are very similar particularly in the case of the daily annualized volatility distribution generated by the model which almost exactly replicates the empirical distribution associated with the S&P500 data. Although the divergence between the empirical and the model average volatility distributions increases slightly for longer time horizons, the goodness of fit remains very good.

⁸The 1st and 9th deciles can be considered proxies for the lower and upper bounds of the 80% confidence intervals of the moments of the model.

2.3.4 The time-reversal asymmetry

Finally, we examine the ability of the model to reproduce the time-reversal asymmetry (TRA) of financial time series highlighted Zumbach [203]). To assess this aspect of the model, we use the methodology in [39], and consider the following time asymmetry ratio (TAR) for different lags:

$$\zeta(l) = \frac{\sum_{l'=1}^l C(l') - C(-l')}{2 \sum_{l'=1}^l \max(|C(l')| - |C(-l')|)}$$

where C is the cross-correlation function of the realized volatility and absolute returns

$$C(l) = \frac{\sum_{t=1}^n (\sigma_t^R - \bar{\sigma}^R) (|r_{t-l}| - |\bar{r}|)}{\sqrt{\sum_{t=1}^n (|r_t| - |\bar{r}|)^2} \sqrt{\sum_{t=1}^n (\sigma_t^R - \bar{\sigma}^R)^2}},$$

where $|r_t|$ and σ_t^R are respectively the absolute return and the realized volatility at time t , $\bar{\sigma}^R$ is the mean of the realized volatility, and $|\bar{r}|$ is the mean of the absolute returns.

Table 2.5 reports the TAR for both the empirical and synthetic data for different lags:

Lag (in trad. days)	1	2	5	10	20	60
TAR S&P500 data	0.072	0.047	0.032	0.019	0.011	-0.007
Average TAR of sim. data	0.074	0.04	0.013	0.004	-0.000	-0.002

Table 2.5: The TAR for different lags: S&P500 data vs EWMA HM.

The numbers show that the EWMA HM provides a good representation of the empirical TRA.

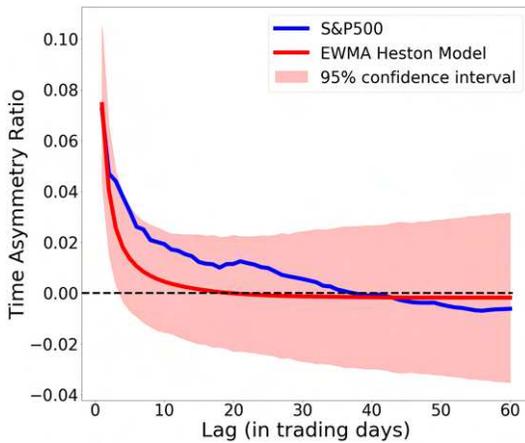


Figure 2.12: TAR: S&P500 data vs EWMA HM.

Indeed, as in the case of the S&P500 data, the TAR of the data simulated by the model follows a convex decreasing trend as a function of the lag size. Moreover, the order of magnitude of the TRA generated by the EWMA HM matches the empirical data, and the empirical TAR is within the estimated 95% confidence interval of the model for all considered lags. More broadly, the positive value of the TRA for small lags of these synthetic data confirms that the EWMA HM is able to capture the following aspect of the Zumbach effect: past absolute returns used to forecast future volatilities are more effective than use of past volatilities to forecast future absolute returns. This must be emphasized because it contradicts the idea that "models that

use Brownian SDEs are TRS by construction and cannot reproduce this asymmetry" ([39] p.17). Note also that TAR becomes slightly negative after about 20 trading days for the synthetic data and after about 35 trading days for the empirical data. In the case of our empirical financial time series, we cannot say whether or not this is a structural effect. However, for the data generated by the EWMA HM, the number of lags before the sign of TAR changes depends on the set of parameters used. For instance, for higher values of τ_2 , TAR remains positive for longer lags. Similarly, when using a combination of multiple exponentials to approximate a power-law kernel of $t^{-0.379}$, the decay of the TAR slows down and more closely resembles the behavior observed in market data.

2.4 Conclusion

In this chapter, we proposed a EWMA HM, a Markovian model which is very tractable for Monte-Carlo simulations and is generally consistent with market data. The model is based on a specific empirical relationship between the EWMA of returns and the realized volatility, which structure makes it naturally adaptable to accounting for a "strong Zumbach effect". We have shown that this model presents some similarities with stochastic volatility models based on quadratic Hawkes processes, models originally conceived to capture this effect ([39], [80], [106]). Besides these theoretical aspects, we show that based on simulations using a Euler discretization scheme, many aspects of the model are consistent with empirical stock market data. First, it generates realistic evolutions of an asset price combined with its associated volatility, reproducing the empirical irregular behavior of volatility paths occurring over short time scales. We show also that both log-returns and the model volatility distributions are consistent with the empirical results. In particular, the model is able to capture the respective deformations of these distributions in the function of the time horizon. Finally, the simulations show that the model also reproduces the empirical time-reversal asymmetry of financial time series.

However, despite all these strong qualities, the model has some shortcomings. First, in the model serial correlation of the volatility process seems to decay much too quickly compared to the empirical data. Relatedly, the type of relationship linking the expected absolute value of the log-volatility increments and the considered time scale differs between the model generated data and the market data. These two limitations both stem from the exponential memory of the volatility process in the model compared to an empirical volatility process based on a power-law memory ([105]). However, these shortcomings could be overcome by using a combination of several EWMA of returns in order to mimic a long-memory property. This potential extension could be conducted using the more general EWMA HM framework and should be the focus of future work.

Another limitation is that we focus only on use of the EWMA HM as a time series generator

⁹This choice of parameterization derives from the estimated values of \hat{H} obtained from market data in section 2.3.2.2, which range between 0.12 and 0.14. Thus, by taking $H = 0.13$, we have $t^{H-0.5} = t^{0.13-0.5} = t^{-0.37}$.

for Monte Carlo simulations. Its use for pricing issues needs investigation along with deeper study of the model's mathematical properties. We also identified a statistical regularity related to the slope coefficient between the EWMA's of past returns and realized volatility, a statistical regularity which might indicate the microstructural foundation of the EWMA HM. This is also worthy of further investigation and substantive consideration. Thus, this chapter offers some directions for further research and a new volatility modeling agenda.

2.5 Acknowledgments

I would like to thank Jean-Paul Laurent and Thierry Roncalli for useful comments and fruitful discussions.

Appendix 2.A Convergence of the attraction volatility under given specification

Let us consider the following differential equation

$$d\nu_t = \left(\frac{\psi}{\nu_t - \underline{\nu}} - \nu_t + \underline{\nu} + \alpha \right) \frac{dt}{\tau_2}$$

such as $\psi, \tau_2, \underline{\nu} \in \mathbb{R}_+$, $\alpha \in \mathbb{R}$, and $\nu > \underline{\nu}$. If this differential equation converges as $\lim_{t \rightarrow +\infty} \nu_t = \nu^* < +\infty$, thus $\frac{d\nu^*}{dt} = 0$. Let us start by noticing the following equalities:

$$\begin{aligned} 0 &= \frac{d\nu^*}{dt} \\ 0 &= \frac{\psi}{\nu_t - \underline{\nu}} - \nu_t + \underline{\nu} + \alpha \\ 0 &= \nu_t^2 - (2\underline{\nu} + \alpha)\nu_t + \underline{\nu} + \alpha\underline{\nu} - \psi \end{aligned}$$

Therefore:

$$\nu^* = \frac{2\underline{\nu} + \alpha \pm \sqrt{\alpha^2 + 4\psi}}{2}$$

Moreover, because $\nu_0 > \underline{\nu}$, the following inequalities must be respected:

$$2\underline{\nu} < 2\underline{\nu} + \alpha \pm \sqrt{\alpha^2 + 4\psi} \Leftrightarrow -\alpha < \pm \sqrt{\alpha^2 + 4\psi}$$

Since $\psi, \tau \in \mathbb{R}_+$, therefore $(\pm) = +$. It follows:

$$\nu^* = \frac{2\underline{\nu} + \alpha + \sqrt{\alpha^2 + 4\psi}}{2}$$

Moreover, $\frac{d\nu^*}{dt}$ is a continuous function positive on $] \underline{\nu} : \nu^*]$ and negative on $[\nu^* : +\infty [$. Consequently:

$$\lim_{t \rightarrow +\infty} \nu_t = \frac{2\underline{\nu} + \alpha + \sqrt{\alpha^2 + 4\psi}}{2} = \nu_\infty.$$

It follows that if β is 0, the EWMA HM degenerates into the following standard HM:

$$\begin{cases} \frac{dS_t}{S_t} = \mu_t dt + \sqrt{V_t} dW_t \\ dV_t = \frac{1}{\tau_1} \left((\nu^*)^2 - V_t \right) dt + \nu^* \sqrt{V_t} dB_t. \end{cases}$$

Appendix 2.B The model calibration procedure

To obtain numerical results and enable comparison with empirical data, we need to fit the model from \mathbf{H} , the daily empirical time-series of the S&P500 and its realized volatility from July 23, 2001, to July 23 2021. This involves identifying the following parameters

$$\theta_{\mathcal{G}} = (\lambda, \rho, \tau_1, \xi, \tau_2, \bar{\nu}, \alpha, \beta).$$

Due to the relative complexity of the model, obtaining these parameters using standard statistical methods such as maximum likelihood estimation is tricky. For this reason, to calibrate the model we use an ad-hoc procedure involving a neural network (NN), with NN denoted \mathcal{P} . This NN, which is used to provide an estimator of the vector of the parameters, takes as its input a vector encoding the index path and its volatility. In other words, the market data matrix \mathbf{H} passes through an encoder denoted \mathcal{E} which transforms it into an input vector \mathbf{E} , and the elements of this vector then constitute the input layer of \mathcal{P} which passes through the network to provide an estimate of $\theta_{\mathcal{G}}$. In summary:

$$\mathcal{P}\left(\underbrace{\mathcal{E}(\mathbf{H})}_{\mathbf{E}}; \theta_{\mathcal{P}}\right) = \hat{\theta}_{\mathcal{G}}$$

with $\theta_{\mathcal{P}}$ the parameters of \mathcal{P} , and $\hat{\theta}_{\mathcal{G}}$ the estimator of $\theta_{\mathcal{G}}$. More specifically, the encoder \mathcal{E} extracts from the time-series the following features:

1. The 4 first moments of the log-returns for the following lags in trading days: 1, 5, 20, 60, 125, 250.
2. The 4 first moments of the average volatility over the following time intervals in trading days: 1, 5, 20, 60, 125, 250.
3. The serial corellation for the following lags expressed in trading days: 1, 2, 3, 4, 5, 10, 20, 60, 125, 250.

Thus, the dimension of \mathbf{E} is thus equal to 1×30 . Obviously, the NN responsible for providing the model parameters derived from the encoded market data requires to be calibrated from a learning set. For this purpose, we randomly generated 20 000 parameter vectors such that each parameter was obtained using a uniform distribution within the bounds defined in table 2.6:

	λ	ρ	τ_1	ξ	τ_2	$\bar{\nu}$	α	β
Lower bound	0.2	-1	0.0008	$0.05\nu\sqrt{\frac{2}{\tau_1}}$	0.15	0.01	0.01	$0.9\sqrt{\frac{2\tau_1}{\pi}}$
Upper bound	0.5	0	2	$\nu\sqrt{\frac{2}{\tau_1}}$	1	0.1	0.15	$1.1\sqrt{\frac{2\tau_1}{\pi}}$

Table 2.6: The respective bounds of the uniform distributions used to generate the training set of parameter vectors.

With each of the 20 000 vectors of parameters obtained, we generate a sample path of the index and its volatility from \mathcal{G} . We thus pass through the encoder \mathcal{E} each of the 20 000 synthetic time-series, in order to obtain the following set $\{E^{(i)}\}_{1 \leq i \leq 20000}$. This step provides we have our learning set $\left\{ \left(E^{(i)}, \theta_{\mathcal{G}}^{(i)} \right) \right\}_{1 \leq i \leq 20000}$ necessary to fit \mathcal{P} . The loss function used for this purpose is:

$$\mathcal{L}(\theta_{\mathcal{G}}^{(i)}, \tilde{\theta}_{\mathcal{G}}^{(i)}) = \left\| w \odot \left(\theta_{\mathcal{G}}^{(i)} - \tilde{\theta}_{\mathcal{G}}^{(i)} \right) \right\|_2^2,$$

where w is a weighting vector, such that the value of the k^{th} coordinate is equal to the inverse of the standard deviation of $\left\{ \left(\theta_{\mathcal{G}}^{(i)} \right)_k \right\}_{1 \leq i \leq 20000}$. Given that, \mathcal{P} is fitted by solving the following optimization problem:

$$\theta_{\mathcal{P}}^* = \arg \min_{\theta_{\mathcal{P}} \in \Theta_{\mathcal{P}}} \sum_{i=1}^{20000} \left\| w \odot \left(\theta_{\mathcal{G}}^{(i)} - \mathcal{P}(E_i; \theta_{\mathcal{P}}) \right) \right\|_2^2$$

The NN being fitted, we calibrate finally $\theta_{\mathcal{G}}$ from market data:

$$\mathcal{P}(\mathcal{E}(\mathbf{H}); \theta_{\mathcal{P}}^*) = \hat{\theta}_{\mathcal{G}}$$

We obtain the following parameters:

λ	ρ	τ_1	ξ	τ_2	ν	α	β
0.5575	-0.465	0.0013	42.95	0.276	0.0595	0.1033	0.419

Table 2.7: The parameters obtained from the calibration procedure.

Appendix 2.C Standard deviation of an EWMA of a Brownian motion

The asymptotic variance of the EWMA of a Brownian motion is given by the following equalities:

$$\text{Var} \left(\frac{1}{\tau} \int_{-\infty}^t e^{\frac{1}{\tau}(u-t)} dW_u \right) = \frac{1}{\tau^2} \int_{-\infty}^t e^{\frac{2}{\tau}(u-t)} du = \frac{1}{\tau^2} \left[\frac{\tau}{2} e^{\frac{2}{\tau}(u-t)} \right]_{-\infty}^t = \frac{1}{2\tau}$$

It follows that the asymptotic standard deviation of the EWMA of a Brownian motion is equal to $\sqrt{\frac{1}{2\tau}}$.

Appendix 2.D Complementary results

The type of relationship between the EWMA of past returns and the volatility described by the EWMA HM is not unique to the S&P500 and can apply also to Nasdaq and EuroStoxx50 data¹⁰. Also, the optimal slope coefficient β is very close to $\sqrt{\frac{2\tau}{\pi}}$ as in the S&P500. The presence of this specific value coefficient in different market data sets would seem to suggest a fundamental reason underlying this statistical regularity.

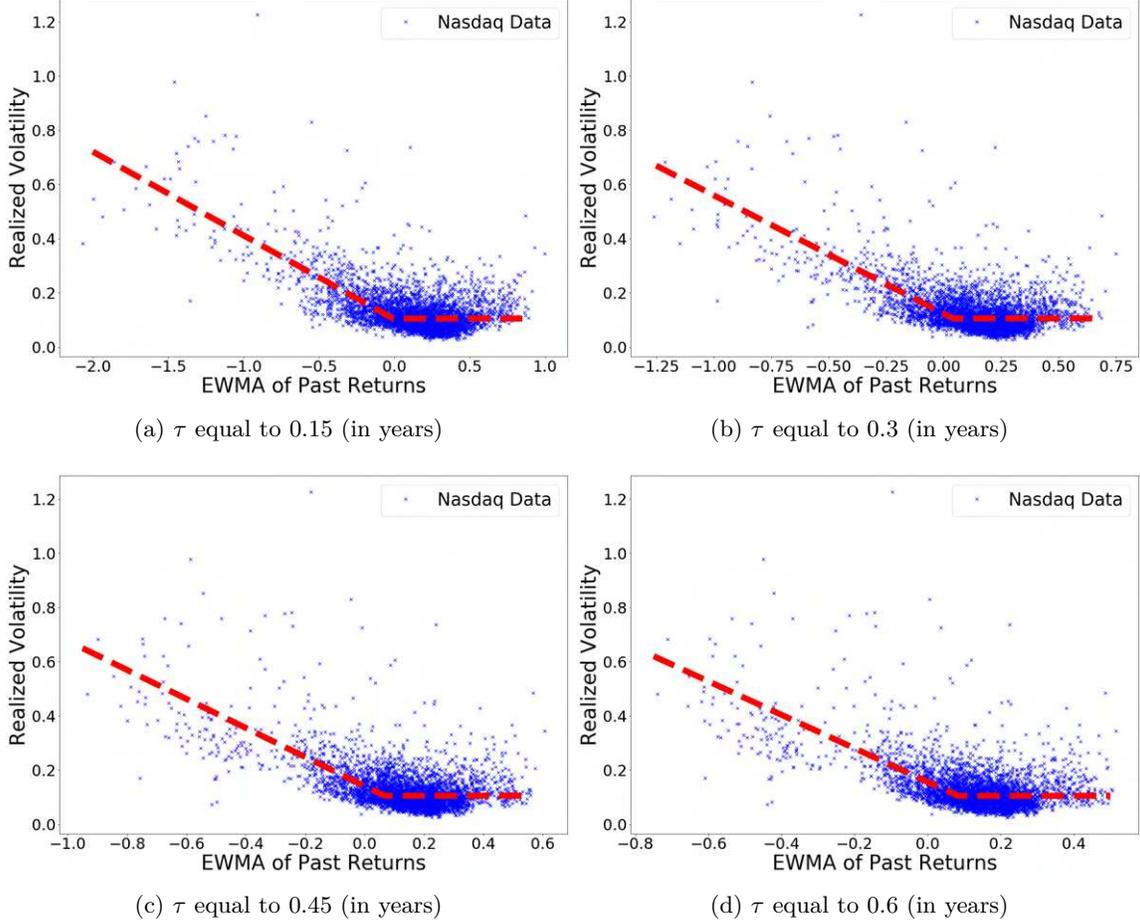


Figure 2.13: The empirical relationship between the realized volatility of the Nasdaq and the EWMA of its past returns for different values of τ expressed in years. The red lines are regressions of the form $\sqrt{V_t} = \underline{\nu} + (\alpha - \beta m_t)_+$ with a slope coefficient equal to $\sqrt{\frac{2\tau}{\pi}}$. These regressions are respectively associated with the following R -squared: 0.476, 0.434, 0.405, 0.384.

¹⁰As in the S&P500 example, the data sets cover the period July 23 2001 to July 23 2021, and as an estimator we use volatility, and the square root of the realized variance computed from 5-min samples provided by the Oxford-Man Institute of Quantitative Finance.

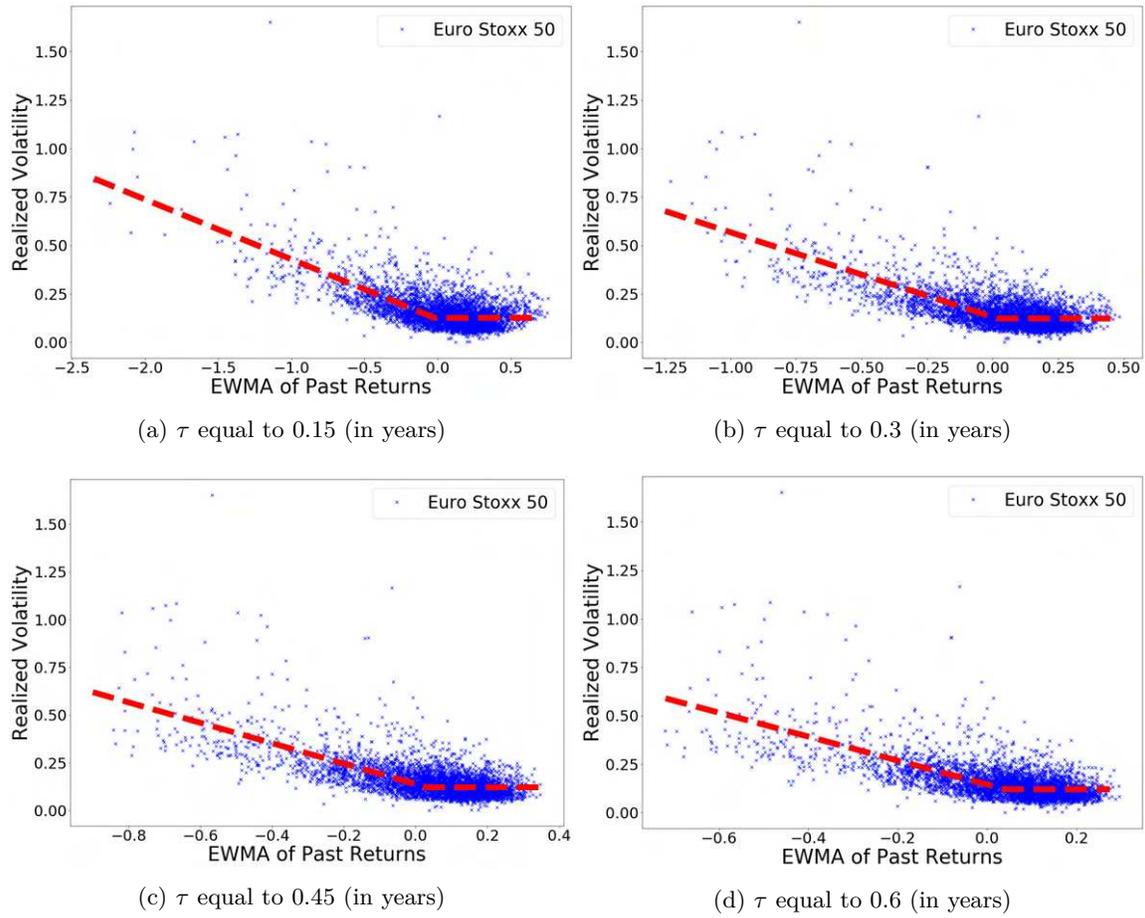


Figure 2.14: The empirical relationship between the realized volatility of the EuroStoxx50 and the EWMA of its past returns for different values of τ expressed in years. The red lines are regressions of the form $\sqrt{V_t} = \underline{\nu} + (\alpha - \beta m_t)_+$ with a slope coefficient equal to $\sqrt{\frac{2\tau}{\pi}}$. These regressions are respectively associated with the following R -squared: 0.476, 0.466, 0.445, 0.427.

CHAPTER 3

Rough Path-Dependent Volatility Models

Abstract

This chapter introduces a family of volatility models, referred to as rough path-dependent volatility (RPDV) models, which capture two key empirical features of volatility: its rough behavior and its path dependence. After presenting its form and connections to existing models in the literature, we provide a Markovian multi-factor approximation for this family, building on the work of Abi Jaber ([1], [2]). A comparative study of various RPDV model specifications is then conducted using realized volatility data from five stock indices, in order to provide insights into the question: what are the most effective ways to model rough path-dependent volatility within the proposed framework? We then discuss the implications of the obtained results in terms of modeling choices, particularly regarding the necessity of incorporating an exogenous source of randomness and how to account for it within the chosen framework. Additionally, the systematic outperformance of models incorporating a historical volatility factor observed in this study raises the question of whether this result indicates the need to include such a factor due to structural positive volatility feedback, or if it could be attributed to the exogenous component of volatility. This question is investigated through numerical experiments based on a RPDV model that does not incorporate a historical volatility factor but includes an exogenous source of randomness. Through these experiments, we demonstrate that, on the one hand, this type of modeling can replicate most of the properties that characterize empirical volatility dynamics. On the other hand, we show that the apparent positive volatility feedback observed in market data can be explained, at least partially, by the effect of the exogenous component of volatility.

3.1 Introduction

Continuous-time volatility modeling is a central issue in quantitative finance that has given rise to prolific academic literature in recent years. Among these new approaches, two particularly important classes of models emerged: the rough volatility (RV) and the path-dependent volatility (PDV) models.

While not strictly a rough volatility model, the Multifractal Random Walk (MRW), introduced by

Bacry *et al.* ([14], [162]), can be seen as a precursor to this literature due to its contributions and the connections it shares with the rough volatility approach ([198]). However, it was only several years later that the seminal article *Volatility is Rough* ([104]) was published, presenting empirical evidence of volatility roughness across a wide range of financial markets. From this point onward, various rough volatility models have emerged, including the rough Bergomi (RB) model ([23]), the rough-Heston (RH) model ([89], [90]), the super-Heston rough (SHR) model ([80]), and the quadratic rough Heston (QRH) model ([105]). These have shown strong practical interest to deal with pricing and hedging issues, but also for volatility forecasting ([104], [89], [90], [181], [182]).

In parallel, in recent years, there has been a resurgence of interest in PDV models ([116], [39], [106]) motivated by the fact that empirical volatility clearly "depends on the path followed by the asset price in the recent past" ([116], p.1). This fact already shown by Zumbach works ([203], [204], [205]) has been strongly reaffirmed by the recent article by Guyon and Lekeufack ([117]).

Given the significance of these two approaches, the primary objective of this chapter is to introduce a general class of volatility models that reconciles their respective advantages, drawing upon existing models in the literature. For simplicity, models in this class will be referred to as rough path-dependent volatility (RPDV) models¹. The variety of these models, and consequently the different approaches to capturing both the rough behavior and path-dependence of volatility, is vast and diverse. Therefore, the second objective of this chapter is to provide insights into the question: what are the most effective ways to model rough path-dependent volatility within the proposed framework?

In line with these two objectives, the chapter is organized as follows. In section 3.2, we introduce a general form of PDV model that includes several volatility models popular in the academic literature. On that basis, we define the family of RPDV models. Then, using the results of Abi Jaber ([1],[2]), we expose a Markovian multi-factor approximation of these models. Section 3.3 involves a comparative study of various RPDV model specifications based on the historical volatility of 5 stock indices. In this section, we analyze the respective performance of these different specifications in explaining realized volatility and discuss the implications of these results in terms of modeling choices. Two key points are discussed in particular. First, the necessity of incorporating an exogenous source of randomness and how to account for it within the chosen framework. Second, the necessity (or not) of including a historical volatility factor to explicitly account for positive volatility feedback. This second point is further examined in section 3.4 through numerical experiments based on an RPDV model that excludes a historical volatility factor but includes an exogenous source of randomness. These experiments demonstrate that, on the one hand, this type of modeling can replicate most of the properties that characterize empirical volatility dynamics. On the other hand, they show that the apparent positive volatility feedback observed in market data can be explained, at least partially, by the effect of an exogenous volatility component.

¹This nomenclature does not imply that the family of models introduced here exhaustively encompasses all path-dependent rough volatility models.

3.2 Rough path-dependent volatility models

3.2.1 A general form of path-dependent volatility model

Let us introduce a general form of path-dependent volatility model defined by the following equation system:

$$\begin{cases} \frac{dP_t}{P_t} &= \mu_t dt + \sigma_t dB_t \\ (\sigma_t)^p &= \beta_0 + \beta_1^{(+)} (m_{1,t} - \bar{m}_1)_+^{a_1} + \beta_1^{(-)} (\bar{m}_1 - m_{1,t})_+^{a_1} + \beta_2 (m_{2,t})^{p/a_2} \\ m_{1,t} &= \int_{I_t} \mathcal{K}_1(t-u) \frac{dP_u}{P_u} + \kappa_1 \int_{I_t} \mathcal{K}_1(t-u) (\theta_{1,u} - m_{1,u}) du, \\ m_{2,t} &= \int_{I_t} \mathcal{K}_2(t-u) (\sigma_u)^{a_2} du + \kappa_2 \int_{I_t} \mathcal{K}_2(t-u) (\theta_{2,u} - m_{2,u}) du \end{cases} \quad (3.1)$$

with B a Brownian motion, I_t the time interval of integration with an upper bound less than or equal to t , where $p, a_1, a_2 \in \{1, 2\}$, and such as θ_j are \mathcal{F}_t -measurable processes, and \mathcal{K}_j are kernel functions continuous decreasing on \mathbb{R}_+ .

In this model, the path dependency of price dynamics is captured through processes m_1 and m_2 . The process m_1 corresponds to a past price trend variable and can be understood as a sort of weighted moving average of past returns. It allows to capture the dependence of the volatility process to the historical price path. The variable m_2 is a historical volatility factor that capture the asset price activity regardless of its trend. It can be viewed as a moving average of the volatility process either $p = 1$, or of the variance process σ^2 if $p = 2$. The impact of these variables on the volatility process depends on the specification of the second equation of 4.1. In this, the value of the parameter p plays a central role. In case $p = 1$, the multilinear function expressed by the right-hand term of the second equation of 4.1 defines the volatility, while if $p = 2$ this function defines the variance process.

Furthermore, if θ_1 and θ_2 are deterministic functions, the volatility process is a function of a single source of randomness B : the Brownian motion driving the asset price. The model is then purely path-dependent. Conversely, the functions θ_j can also include an exogenous source of randomness by taking, for example, the following form:

$$\theta_{j,t} dt = \bar{\theta}_{j,t} dt + \sigma_t dW_t, \quad (3.2)$$

where $\bar{\theta}_j$ is a deterministic process and W is a Brownian motion independent of B . In this case, the model is partially path-dependent and corresponds to a stochastic volatility (SV) model in the terminology of Guyon and Lekeufack ([117]).

Thanks to its highly flexible form, the model includes a certain number of existing volatility models. Table 3.1 reports some of them and their respective specifications.

	Model 1	Model 2	Model 3	Model 4	Model 5
p	2	2	2	1	1
a_1	1	2	2	1	1
a_2	–	2	–	–	2
$\mathcal{K}_1(\tau)$	$\frac{\tau^{\alpha-1}}{\Gamma(\alpha)}$	$\gamma_1 e^{-\gamma_1 \tau}$	$\frac{\tau^{\alpha-1}}{\Gamma(\alpha)}$	$\gamma e^{-\gamma \tau}$	$(\delta_1 + \tau)^{-\alpha_1}$
$\theta_1(u)$	–	–	–	$m_{1,u}$	$m_{1,u}$
$\beta_1^{(+)}$	β_1	β_1	β_1	0	β_1
$\beta_1^{(-)}$	$-\beta_1$	β_1	β_1	$-\beta_1$	$-\beta_1$
\bar{m}_1	0	0	\bar{m}_1	0	0
$\mathcal{K}_2(\tau)$	–	$\gamma_2 e^{-\gamma_2 \tau}$	–	–	$(\delta_2 + \tau)^{-\alpha_2}$
$\theta_2(u)$	–	$m_{2,u}$	–	–	$m_{2,u}$
β_2	–	–	–	–	β_2

Table 3.1: Some model specifications that correspond to existing models in SV literature.

Model 1 relates to the rough Heston model ([89], [90]), in which the variance process is given by:

$$(\sigma_t)^2 = \beta_0 + \beta_1 m_{1,t}$$

$$m_{1,t} = \int_0^t \frac{(t-u)^{\alpha-1}}{\Gamma(\alpha)} \frac{dP_u}{P_u} + \int_0^t \frac{(t-u)^{\alpha-1}}{\Gamma(\alpha)} \kappa_1 (\theta_{1,u} - m_{1,u}) du,$$

with $\alpha \in [0.5, 1]$ and

$$\theta_{1,t} dt = (\theta_t - \beta_0) dt + \sigma_t \nu dW_t,$$

where W is an independent Brownian motion from B , and θ is an explicit \mathcal{F}_0 -measurable process ([89]). Here thus, when $\nu = 0$, the variance process is linearly dependent on a moving average of past returns with power law weighting.

Model 2 proposed by Blanc *et al.* ([39]) corresponds to a limiting model with a microstructural foundation that uses quadratic Hawkes processes. The variance is given by:

$$\begin{aligned}(\sigma_t)^2 &= \beta_0 + \beta_1(m_{1,t})^2 + \beta_2 m_{2,t}, \\ m_{1,t} &= \gamma_1 \int_0^t e^{\gamma_1(u-t)} \frac{dP_u}{P_u}, \\ m_{2,t} &= \gamma_2 \int_0^t e^{\gamma_2(u-t)} (\sigma_u)^2 du.\end{aligned}$$

Here, m_1 and m_2 are respectively an exponentially weighted moving average (EWMA) of past returns and an EWMA of past variance. Note that other similar models based on quadratic Hawkes processes with other kernel type has also been studied in [80].

Model 3 corresponds to a quadratic rough Heston model (QRHM) introduced by Gatheral *et al.* ([105]) defined as

$$\begin{aligned}(\sigma_t)^2 &= \beta_0 + \beta_1(\bar{m}_1 - m_{1,t})^2, \\ m_{1,t} &= \int_0^t \frac{(t-u)^{\alpha-1}}{\Gamma(\alpha)} \frac{dP_u}{P_u} + \int_0^t \frac{(t-u)^{\alpha-1}}{\Gamma(\alpha)} (\theta_1(u) - m_{1,u}) du,\end{aligned}$$

with $\bar{m}_1 > 0$. On the contrary to model 1 and 2, it encodes an asymmetry of the feedback effect of the price trend on the volatility. Indeed, for the same absolute value of m_1 , volatility is higher when m_1 is negative than when it is positive.

Model 4 is a specific case of the EWMA HM introduced in chapter 2, in which

$$\begin{aligned}\sigma_t &= \beta_0 + \beta_1(\bar{m}_1 - m_{1,t})_+, \\ m_{1,t} &= \gamma_1 \int_0^t e^{(u-t)\gamma_1} \frac{dP_u}{P_u},\end{aligned}$$

with $\bar{m}_1 > 0$. Like the QRHM, this model encodes an asymmetry of the feedback effect of the price trend on the volatility but not in exactly the same way. Indeed, in this model, when m_1 falls below \bar{m}_1 , the feedback effect disappears. Another difference with models 1,2 and 3, is that $p = 1$ which means that the volatility process is linearly sensitive to the asset trend process m_1 when $m_{1,t} < \bar{m}_1$.

Model 5 is the PDV model introduced by Guyon and Lekeufack ([117]), defined by

$$\begin{aligned}\sigma_t &= \beta_0 + \beta_1 m_{1,t} + \beta_2 \sqrt{m_{2,t}}, \\ m_{1,t} &= \int_{-\infty}^t (\delta_1 + t - u)^{-\alpha_1} \frac{dP_u}{P_u}, \\ m_{2,t} &= \int_{-\infty}^t (\delta_2 + t - u)^{-\alpha_2} (\sigma_u)^2 du.\end{aligned}$$

As in model 2, the volatility dynamics is not only driven by the past price dynamics but also by the variance price path through m_2 . However, contrary to model 4, the volatility is linearly sensitive to m_1 . In addition, it is worth to note that in the case $\delta_1 = \delta_2 = 0$, the model is rough.

These different examples show thus the diversity of approaches to model the path-dependence of volatility which is included in the PDV model (4.1). Moreover, it is on this basis that the RPDV model family will be defined in the following section.

3.2.2 Rough path-dependent volatility models

We now define a *rough path-dependent volatility* as a specification of 4.1 for which \mathcal{K}_1 and \mathcal{K}_2 are power law kernels of the form

$$\mathcal{K}_j(\tau) = \tau^{-\alpha_j} \tag{3.3}$$

where α_j is a strictly positive constant.

It may be noted that given this definition, models 1 and 3, and model 5 in the case $\delta_1 = \delta_2 = 0$ (see table 3.1) belong to the family of the RPDV models. These particular cases share with other RPDV models to be structurally adapted to capture two main features of empirical volatility: its rough behavior and its path dependency.

First, the ability of the model to capture the roughness of volatility is made possible by the power law kernels associated with processes m_1 and m_2 . This characteristic is important to the extent that empirical data tends to show that the rough behavior of volatility is verified for all classes of financial assets, at least at the daily scale ([104]).

Second, the path-dependency of the RPDV models is clear given that dynamics of the volatility process are driven (either fully or partially) by past price dynamics through the processes $m_{1,t}$ and $m_{2,t}$. This path-dependence allows capturing important empirical phenomena occurring in the financial market like the Zumbach effect ([105]), but more generally, the fact that the volatility dynamics are mostly explainable by the past price path ([117]).

In addition to their interest considered separately, the relevance of combining these two notions in a parametric model is also justified by the results in [182], showing that machine learning methods for volatility forecasting tend to confirm that volatility is both rough and path-dependent.

Besides these properties, the model allows the decoupling short- and long-term behavior of volatility, a property whose importance was emphasized notably in [29]. For instance, α_1 could take a low value and α_2 a high one. In this example, the long-term behavior of the volatility would be driven by a medium/long-term past price trend while its short-term behavior would be dominated by recent price activity regardless of its trend.

At a more technical level, it should be noted that the existence and uniqueness of the solution to the equation defining the volatility process are not guaranteed for all possible specifications. However, these conditions have been demonstrated for a certain number of cases, among which, for example, the modeling of volatility by a Rough Volterra square-root process ([3]). For other specifications of 4.1, these existence and uniqueness conditions have been demonstrated in the context where \mathcal{K}_j are smooth kernels, but remain an open question for the case where these kernels take the form 3.3. This is the case, for instance, for a wide range of Volterra processes. For these types of specifications, the lack of an existence and uniqueness proof is not necessarily a problem, particularly if we limit its use as a simulation tool, since power law kernels can be approximated by smooth kernels ([1], [2], [3]) as we will see in more detail in the following section (section 3.2.3). In this context, the theoretical inversion, which involves considering the power-law kernel as an approximation of the actual smooth kernel, can be performed. The rough model then becomes only an ideal-type and a way for parsimoniously approximating the actual model using a kernel with only one parameter α .

3.2.3 Markovian multi-factor approximation of the RPDV model

The RPDV models are non-Markovian and non-semimartingale, which makes efficient simulation challenging. Accordingly, the purpose of this section is to propose a Markovian multi-factor approximation of RPDV models to address this issue.

3.2.3.1 Rough kernel approximation

The key idea is to approximate the kernel $\mathcal{K}(\tau) = \tau^{-\alpha}$ by a sum of n exponential kernels defined by

$$\hat{\mathcal{K}}(\tau) = \sum_{i=1}^n w_i e^{-\gamma_i \tau}, \quad (3.4)$$

where $w_i, \gamma_i > 0 \forall i \in \{1, \dots, n\}$, in order to approximate variables m_1 and m_2 as sums of solutions of stochastic differential equations.

Different methods exist for determining the parameters $(w_i, \gamma_i)_{1 \leq i \leq n}$ in the academic literature

([40], [1], [2]). Here we adopt the approach proposed by Abi Jaber ([1], [2]). This takes as its starting point that the fractional kernel $\mathcal{K}(\tau)$ may be write as the Laplace transform of a positive measure μ^2

$$\mathcal{K}(\tau) = \int_0^\infty e^{-\gamma\tau} \mu(d\gamma), \quad \mu(d\gamma) = \frac{\gamma^{\alpha-1}}{\Gamma(\alpha)}.$$

By approximating μ by a finite sum of Dirac measures $\sum_{i=1}^n w_i \delta_{\gamma_i}$, we obtain a kernel \mathcal{K}_E of the form 3.4. Then, using a geometric partition of $(w_i, \gamma_i)_{1 \leq i \leq n}$,

$$w_i = \int_{\eta_{i-1}}^{\eta_i} \mu(dx), \quad \gamma_i = \frac{1}{w_i} \int_{\eta_{i-1}}^{\eta_i} x \mu(dx),$$

with $\mu(dx) = \frac{x^{\alpha-1}}{\Gamma(\alpha)} dx$ and $\eta_i = x^{i-\frac{n}{2}}$ we obtain:

$$w_i = \frac{(1-x^{-\alpha})x^{\alpha(i-\frac{n}{2})}}{\alpha\Gamma(\alpha)}, \quad \gamma_i = \frac{\alpha(x^{1+\alpha}-1)x^{i-1-\frac{n}{2}}}{(1+\alpha)(x^\alpha-1)}, \quad (3.5)$$

with $x > 1$ whose value can be determined by solving

$$x^* = \arg \min_{x>1} \|\hat{\mathcal{K}} - \mathcal{K}\|_{L^2(t_-, t_+)}.$$

With this specification method of $(w_i, \gamma_i)_{1 \leq i \leq n}$, $\hat{\mathcal{K}}$ converges toward \mathcal{K} in the L^2 sense when $n \rightarrow \infty$ (see proof in [2]).

Note that in [1] or [181] t_- is zero. However, it may be consistent to choose a strictly positive value for this parameter, particularly when n is little. Indeed, because of the characteristic of rough type kernel, the very recent past (t close to zero) may represent a too high share of $\|\hat{\mathcal{K}} - \mathcal{K}\|_{L^2(0, t_+)}$ in comparison of its practical importance. For instance, in a context of medium-long term volatility simulation, the divergence between $\mathcal{K}(\tau)$ and $\hat{\mathcal{K}}(\tau)$ for a very small value of τ (ex.: τ is equal to one second) does not matter much, particularly if the discretization timestep is proportionally large (of the order of an hour or more). In such a case, it is coherent to set $t_- > 0$, because it allows choosing a value of x which minimizes the difference between $\mathcal{K}(\tau)$ and $\hat{\mathcal{K}}(\tau)$ for the range of values of τ which really impacts the quality of the estimation in the context considered.

²The considered kernel being $\mathcal{K}(\tau) = \frac{\tau^{\alpha-1}}{\Gamma(\alpha)}$ in the original article, the measure μ differs according to the kernel specificity.

3.2.3.2 Markovian approximation of the RPDV models

Using results exposed in section 3.2.3.1, the RPDV models may be approximated through the following multi-factor model Markovian in variables:

$$\left\{ \begin{array}{l} \frac{dP_t}{P_t} = \mu_t dt + \sigma_t dB_t, \\ \sigma_t = \left(\beta_0 + \beta_1^{(+)} (m_{1,t} - \bar{m}_1)_+^{a_1} + \beta_1^{(-)} (\bar{m}_1 - m_{1,t})_+^{a_1} + \beta_2 (m_{2,t})^{p/a_2} \right)^{1/p}, \\ dM_{1,t} = \mathbf{1}_n \cdot \left(\frac{dP_t}{P_t} + \kappa_1 (\theta_1(t) - m_{1,t}) dt \right) - \Lambda_1 \odot M_{1,t} dt, \\ dM_{2,t} = \mathbf{1}_n \cdot \left((\sigma_t)^{a_2} + \kappa_2 (\theta_2(t) - m_{2,t}) \right) dt - \Lambda_2 \odot M_{2,t} dt, \\ m_{1,t} = \langle W_1, M_{1,t} \rangle, \\ m_{2,t} = \langle W_2, M_{2,t} \rangle, \end{array} \right.$$

where $\mathbf{1}_n$ is a $n \times 1$ vector of ones, W_j the vector of weights $(w_{j,i})_{1 \leq i \leq n}$ and Λ_j the vector of discount coefficients $(\gamma_{j,i})_{1 \leq i \leq n}$, with $(w_{j,i}, \gamma_{j,i})_{1 \leq i \leq n}$ given by equations 3.5, such as

$$W_j = \begin{bmatrix} w_{j,1} \\ \dots \\ w_{j,n} \end{bmatrix}, \quad \Lambda_j = \begin{bmatrix} \gamma_{j,1} \\ \dots \\ \gamma_{j,n} \end{bmatrix}.$$

As shown in appendix 3.A, we have

$$m_{1,t} = \int_0^t \underbrace{\sum_{i=1}^n w_{1,i} e^{-\gamma_{1,i}(t-u)}}_{\hat{\kappa}_1(t-u)} \left(\frac{dP_u}{P_u} + \kappa_1 (\theta_{1,u} - m_{1,u}) du \right),$$

and

$$m_{2,t} = \int_0^t \underbrace{\sum_{i=1}^n w_{2,i} e^{-\gamma_{2,i}(t-u)}}_{\hat{\kappa}_2(t-u)} \left((\sigma_t)^{a_2} + \kappa_2 (\theta_{2,u} - m_{2,u}) \right) du.$$

This results in a Markovian model in variables that approximates the RPDV models through smooth kernels.

3.3 Comparison of model specifications

The objective of this section is to evaluate the explanatory power of various fully path-dependent RPDV models regarding historical volatility and to derive insights on a consistent approach to jointly modeling the rough behavior and path dependence of volatility. To achieve this, we adopt a comparative approach inspired by Guyon and Lekeufack ([117]), in which we evaluate the performance of discrete-time versions of RPDV models used as regression models to explain spot realized volatility. The dataset used in this study consists of historical data from five stock market indices: the S&P500, Nasdaq, FTSE, DAX, and Euro Stoxx 50, covering the period from January 2000 to June 2021.

3.3.1 The tested RPDV models

3.3.1.1 The building blocks of models

In order to enhance the clarity of understanding of the different tested RPDV model specifications, we begin by breaking down their components into variables that serve as the elementary building blocks of the tested models. Additionally, we provide a discrete approximation of these variables, enabling their computation from empirical time series. These elements are reported in table 3.2.

	Expression	Approximated expression
$\mathbf{R}_{1,t}$	$\int_{-\infty}^{t-\epsilon} (t-u)^{-\alpha_1} \frac{dP_u}{P_u}$	$\sum_{i=1}^p (t-0.5(t_{i-1}+t_i))^{-\alpha_1} \frac{P_{t_i}-P_{t_{i-1}}}{P_{t_{i-1}}}$
$\mathbf{\Theta}_{1,t}$	$\int_{-\infty}^{t-\epsilon} (t-u)^{-\alpha_1} \theta_u du$	$\sum_{i=1}^p (t-0.5(t_{i-1}+t_i))^{-\alpha_1} \theta_{t_{i-1}} (t_i-t_{i-1})$
$\mathbf{S}_{1,t}$	$\int_{-\infty}^{t-\epsilon} (t-u)^{-\alpha_1} \sigma_u du$	$\sum_{i=1}^p (t-0.5(t_{i-1}+t_i))^{-\alpha_1} \sigma_{t_{i-1}} (t_i-t_{i-1})$
$\mathbf{S}_{2,t}$	$\int_{-\infty}^{t-\epsilon} (t-u)^{-\alpha_2} \sigma_u du$	$\sum_{i=1}^p (t-0.5(t_{i-1}+t_i))^{-\alpha_2} \sigma_{t_{i-1}} (t_i-t_{i-1})$
$\mathbf{V}_{2,t}$	$\int_{-\infty}^{t-\epsilon} (t-u)^{-\alpha_2} (\sigma_u)^2 du$	$\sum_{i=1}^p (t-0.5(t_{i-1}+t_i))^{-\alpha_2} (\sigma_{t_{i-1}})^2 (t_i-t_{i-1})$

Table 3.2: The building blocks of models and their numeral quadratures.

Several important remarks should be made here. First, the elementary building blocks defined in table 3.2 are integrals over $]-\infty : t-\epsilon]$. Here, ϵ is a small positive constant close to zero, representing a latency in the transmission of the dynamics of the explanatory variables to the volatility level. However, the reason for choosing an upper bound of $t-\epsilon$ instead of t is primarily technical. In the case where $\epsilon=0$, the values of α are constrained to be less than 0.5 to avoid inducing divergence in the volatility process. By choosing a strictly positive value for ϵ , this constraint can be relaxed. In this regard, it can be seen as an alternative to the shifted power-law.

Another important point to explain concerns the choice of approximating the integrals that constitute the elementary building blocks of the tested models. In practice, these integrals are truncated to a length of 1260 trading days (≈ 5 years) and then estimated using a Riemann method. This type of approximation is employed, for instance, by Gatheral to approximate the integral defining the RFSV model³. If all the different building blocks of the models are approximated using Riemann sums, the quadrature method used to approximate $R_{1,t}$ differs from the method used for the other variables. Specifically, the midpoint method is employed for $R_{1,t}$, while the left rectangle method is used for the remaining variables. If the choice of the left rectangle method for approximating Θ_t , $R_{1,t}$, $S_{1,t}$, $S_{2,t}$, and $V_{2,t}$ may seem peculiar at first glance, it is, in fact, driven by the objective of avoiding any tautological or quasi-tautological formulations. Thus, contrary to the midpoint method, the chosen numeral quadrature ensures that the estimation of σ_t does not depend on itself or a proxy of itself. This difference in the approximation method also implies that only the building block R_{1,t_i} requires information known after $t - 1$: the price at time t , P_t . On the other hand, all other discretized variables depend solely on elements known at time $t - 1$.

3.3.1.2 The tested models

Given the previously mentioned flexibility of the RPDV model, conducting a comprehensive study of all possible specifications is impractical. Instead, our focus lies in examining the main RPDV models discussed in academic literature, along with some other specifications. Therefore, we consider the model specifications reported in table 3.3.

	Model	Free paramaters
M.1	$\sqrt{b_0 + b_1 R_{1,t}}$	$b_0, \alpha_1 \in \mathbb{R}_+, b_1 \in \mathbb{R}$
M.2	$b_0 + b_1 R_{1,t}$	$b_0, \alpha_1 \in \mathbb{R}_+, b_1 \in \mathbb{R}$
M.3	$\sqrt{b_0 + b_1 (R_{1,t} - \bar{R}_1)^2}$	$b_0, \alpha_1 \in \mathbb{R}_+, b_1, \bar{R}_1 \in \mathbb{R}$
M.4	$b_0 + b_1 R_{1,t} - \bar{R}_1 $	$b_0, \alpha_1 \in \mathbb{R}_+, b_1, \bar{R}_1 \in \mathbb{R}$
M.5	$b_0 + b_1 R_{1,t} + b_2 \sqrt{V_{2,t}}$	$b_0, \alpha_1, \alpha_2 \in \mathbb{R}_+, b_1, b_2 \in \mathbb{R}$
M.6	$b_0 + b_1 R_{1,t} + b_2 S_{2,t}$	$b_0, \alpha_1, \alpha_2 \in \mathbb{R}_+, b_1, b_2 \in \mathbb{R}$
M.7	$b_0 + b_1 R_{1,t} + b_1 S_{1,t} + b_3 \Theta_{1,t}$	$b_0, \alpha_1 \in \mathbb{R}_+, b_1, b_2, b_3 \in \mathbb{R}$

Table 3.3: The RPDV models tested.

³For more details, refer to the section "Implement variance forecast in Python" of the article *Rough Volatility with Python*, accessible at https://tpq.io/p/rough_volatility_with_python.html.

M.1, M.3, and M.5 are existing models in the volatility literature. M.1 can be viewed as a version of the RH model ([89], [90]) in which the asset and the variance process share the same Brownian motion. M.3 corresponds to the QRH model ([106]), and M.5 to a version of the PDV model proposed by Guyon and Lekeufack. Models M.2 and M.4 can be viewed as the counterparts of M.1 and M.3, respectively, for $p = 1$. Similarly, M.6 is the counterpart of M.5 for $a_2 = 1$. This allows for a comparison of two different approaches to capturing the autoregressivity of the volatility process. Regarding model M.7, several variations will be tested based on the specification of θ . The first version considered, denoted as M.7.1, corresponds to the case where $\theta = 0$ (or, equivalently, $b_3 = 0$). This first version of M.7 is equivalent to model M.6 under the constraint $\alpha_1 = \alpha_2$, thus enabling us to assess the impact of this constraint on the model’s explanatory power. We also consider two other specifications of M.7, for which θ is a path-dependent process. The idea is to evaluate whether the addition of a second layer of path-dependence enhances the explanatory power of M.7. More precisely, for M.7.2, we consider:

$$\theta_t = \sqrt{\int_{-\infty}^t e^{\gamma(u-t)} (\sigma_u)^2 du}.$$

In this case, the term Θ_1 represents, similar to the term S_1 but with a different structure, a historical volatility factor. For M.7.3, we consider:

$$\theta_t = \left| \int_{-\infty}^t e^{\gamma(u-t)} \frac{dP_u}{P_u} \right|.$$

In this specification, Θ_1 is a power-law average of the absolute value of an exponentially weighted moving average of returns. This term thus allows for capturing a potential impact of past price trends on volatility, regardless of the trend’s direction.

Models are trained using data from 2000 to the end of 2014, minimizing the mean squared error. The test set, on the other hand, covers the period from 2015 to 2021. This data segmentation allows for evaluating the generalization capability of the different specifications by assessing their performance not only on the training data (in-sample) but also on the out-of-sample data from the test set.

3.3.2 Comparative regression results

3.3.2.1 Model performance evaluation

The results of the regressions obtained for the models defined in section 3.3.1.2 are presented in table 3.4. A first general observation is that models incorporating a historical volatility process (represented by the building blocks S_i and V_i) outperform significantly the models solely based on the price trend process R_1 . This finding, which had already been highlighted by Guyon and Lekeufack, is reflected in the much better performance of models M.5, M.6, and M.7 compared to models M.1, M.2, M.3, and M.4, both in-sample and out-of-sample.

		S&P500		Nasdaq		FTSE		DAX		Stoxx 50	
		Train	Test	Train	Test	Train	Test	Train	Test	Train	Test
M.1	r^2	0.379	0.136	0.439	0.053	0.346	-0.15	0.406	-0.47	0.348	0.145
	RMSE	0.092	0.08	0.068	0.08	0.087	0.094	0.082	0.088	0.09	0.083
M.2	r^2	0.596	0.442	0.613	0.141	0.467	0.105	0.547	0.009	0.494	0.365
	RMSE	0.074	0.065	0.057	0.076	0.078	0.083	0.071	0.073	0.079	0.072
M.3	r^2	0.6	0.543	0.638	0.377	0.5	0.22	0.56	0.122	0.484	0.417
	RMSE	0.074	0.058	0.055	0.065	0.076	0.077	0.07	0.068	0.08	0.069
M.4	r^2	0.605	0.517	0.617	0.245	0.483	0.181	0.555	0.088	0.494	0.369
	RMSE	0.073	0.06	0.057	0.071	0.077	0.079	0.071	0.07	0.079	0.071
M.5	r^2	0.765	0.736	0.789	0.731	0.69	0.545	0.777	0.67	0.7	0.614
	RMSE	0.056	0.044	0.042	0.042	0.06	0.059	0.05	0.042	0.061	0.056
M.6	r^2	0.766	0.733	0.791	0.724	0.691	0.56	0.775	0.669	0.7	0.621
	RMSE	0.056	0.045	0.042	0.043	0.06	0.058	0.05	0.042	0.061	0.055
M.7.1	r^2	0.757	0.728	0.786	0.729	0.681	0.541	0.77	0.663	0.692	0.608
	RMSE	0.057	0.045	0.042	0.043	0.061	0.059	0.051	0.042	0.062	0.056
M.7.2	r^2	0.767	0.739	0.79	0.734	0.696	0.556	0.774	0.675	0.704	0.62
	RMSE	0.056	0.044	0.042	0.042	0.059	0.058	0.05	0.042	0.061	0.055
M.7.3	r^2	0.765	0.737	0.79	0.736	0.689	0.552	0.776	0.671	0.697	0.623
	RMSE	0.057	0.044	0.042	0.042	0.06	0.058	0.05	0.042	0.061	0.055

Table 3.4: r^2 and RMSE associated with various RPDV regression models for different stock indices.

If we delve deeper into the results, the performance of models M.1, M.2, M.3, and M.4 itself varies significantly. It is clear that the explanatory power of M.1, especially out-sample, is extremely weak and inferior to M.2, M.3, and M.4. The comparison is particularly interesting with M.2, as it corresponds to the counterpart of M.1 with $p = 1$ ($p = 2$ for M.1). Hence, the respective results obtained by these two models suggests that a modeling approach where the instantaneous volatility is linearly sensitive to a price trend process (here R_1) is more preferable than a modeling approach where the instantaneous variance is linearly sensitive to such a process. About the relative underperformance of M.2 out-of-sample compared to M.3 and M.4, this one can be explained by the specificity of the test period considered. Indeed, this period includes the Covid-19 crisis of March 2020, which led to a significant drop in stock indices followed by a strong rebound. While this rebound resulted in a sharp increase in the variable R_1 and thus a decrease in the volatility

estimated by the model (due to $b_1 < 0$), the realized volatilities remained higher than their median level. Besides these results, it is worth to note that for these 4 models (M.1-M.4) and all indices, the optimal value of α_1 obtained from the regression falls within the range of 0.4 to 0.5, aligning with the standard assumption in rough volatility literature where $\alpha_1 \in [0, 0.5[$ and $\epsilon = 0$.

On the contrary, for models incorporating a historical volatility factor (M.5-M.7), the selected value of the exponent α_1 is greater than 0.5: between 0.65 and 0.8 for models M.5 and M.6, and in the vicinity of 1.3 for models M.7. This higher value of α_1 results in a more significant weighting of the recent past in the trend process R_1 in practice. As for the exponent α that determines the kernel of the historical volatility factor for models M5-M.7 (α_2 in the case of M5 and M.6, and α_1 for model M.7), the selected value is between 1.25 and 1.35. In the median case $\alpha = 1.3$, this implies that the periods $t - 1$, $t - 2$, and $t - 3$ contribute approximately 46%, 11%, and 6% of the total weighting, respectively, considering a period of 2520 trading days. Furthermore, the cumulative contribution of the last 20 days amounts to 82% of the total weighting. Furthermore, for all these models (M.5-M.7), volatility is primarily explained by historical volatility factors, and the contribution of the R_1 factor in the explained volatility is relatively minor. Consequently, spot volatility becomes predominantly explained by an autoregressive relationship that heavily weights the recent past.

It is also interesting to note that the performances of models M.5-M.7 are fairly homogeneous. Thus, the similar performances of models M.5 and M.6 suggest that the definition of the historical volatility factor either as a moving average of past volatility ($p = 1$) or as the square root of a moving average of past variance ($p = 2$) has little impact on the explanatory power of the model. Similarly, M.7.1 only slightly underperforms M.5 and M.6, which suggests that differentiating the exponent α for the historical volatility factor and the price trend factor is relatively marginal, at least within this static regression framework. In a similar manner, the addition of a factor θ (models M.7.2 and M.7.3) provides a non-zero but small improvement in performance.

3.3.2.2 Residuals and the missing randomness question

To delve deeper into the analysis of the regression results reported in section 3.3.2.1, it is interesting to examine the residuals of these regressions. Firstly, it should be noted that the residuals of the regressions exhibit significant disparities in structure depending on the model under consideration. A notable manifestation of these disparities concerns the autocorrelation function of these residuals. Here again, a clear separation appears between the models incorporating a historical volatility factor and the other models. For the former, the autocorrelation is weak, and the Durbin-Watson statistic is very close to 2, indicating insignificant autocorrelation. On the contrary, the residuals of the models not incorporating a historical volatility factor exhibit long and significant autocorrelation, which can explain the relatively poor out-of-sample performance of these models ([111]). The figure 3.1 illustrating the respective autocorrelation function (ACF) of the residuals for models M.2 and M.7.1 on the training data of the S&P500 provides an illustration of this disparity.

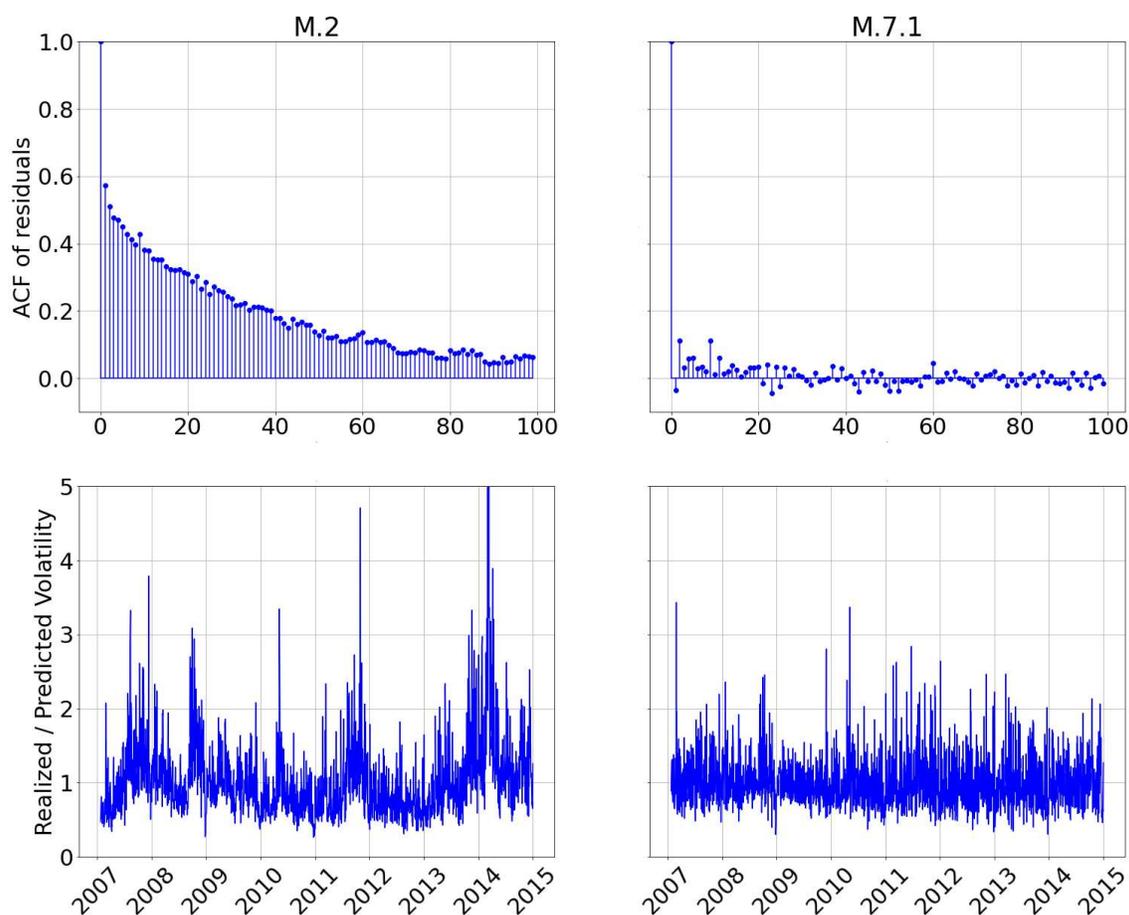


Figure 3.1: Comparison of residuals from models M.2 and M.7.1 in-the-sample. The ACF graphs show that the autocorrelation of M.2 model residuals is strong and persistent, while that of the M.7.1 model is lowly significant at all lags. This difference is clearly noticeable in the graphs displaying the residual ratio of realized volatility on predicted volatility for these two models.

In addition to its usefulness in explaining the performance of different models, this distinction between models M.1-M.4 and M.5-M.7 seems to be tailored to different assumptions regarding the integration of exogenous sources of randomness. Firstly, it is worth noting that the incorporation of such an exogenous randomness component appears necessary for all the considered models. In other words, volatility appears to be only partially path-dependent as long as we assume that the price follows a purely diffusive process with a single stochastic term $\sigma_t dB_t$ ⁴. Indeed, standard deviations of log-returns of realized volatilities are significantly higher than the standard deviations of log-returns of volatilities simulated from discrete models using the parameters obtained through regressions. As an example, for the case of the S&P500, the empirical standard deviation of daily log-returns of volatility is 0.347. However, this standard deviation is approximately 0.2 and

⁴Outside of this framework, assuming, for instance, that the price process includes a jump component, the conclusion is less straightforward, as volatility-adjusted returns (or log-returns) can significantly deviate from a normal distribution ([155]).

0.11 for the simulations carried out using models M.2 and M.7.1, respectively (see figure 3.2a)⁵. One hypothesis could be that this differential stems from the variance of the noise in the realized volatility estimator. However, the results presented in Appendix 3.D indicate that the magnitude of this differential is too significant to be attributed solely to the estimator’s error, particularly in models that incorporate a historical volatility factor.

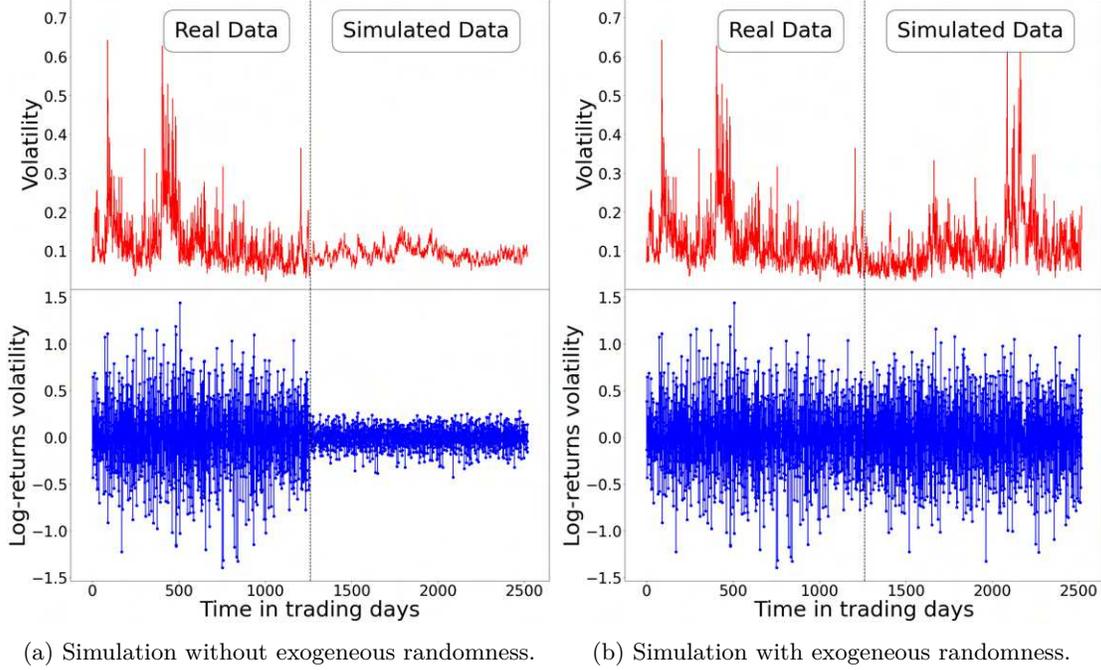


Figure 3.2: Comparison of simulated data from model M.7.1 calibrated on historical data (S&P500), without incorporating an exogenous source of randomness (figure 3.2a), and with the incorporation of an exogenous source of randomness (figure 3.2b). In the first case, the model produces volatility trajectories with characteristics that are very different from those of the historical volatility. In the second case, the model generates realistic volatility trajectories that capture most of the features characterizing historical volatility dynamics.

For these latter ones (M.5-M.7), a modeling approach like the one proposed by Guyon and Lekeufack ([117]) seems particularly coherent. This involves defining the volatility expression as the product (in the sense of multiplication) of a purely path-dependent model with a stochastic process, where the random component is independent of the asset’s price Brownian motion. The volatility then takes the form:

$$\sigma_t = \sigma(R_{1,t}, \dots)Y_t, \quad (3.6)$$

where $\sigma(\cdot)$ is the purely-path dependent model and Y is a stochastic process responsible for capturing the exogenous component of volatility. Guyon and Lekeufack suggest modeling Y using either an OU process or an exponential-OU process. This second option appears particularly suitable when

⁵The simulations are performed using the ARPDV model introduced in the appendix 3.C.

considering the properties of the empirical ratios of realized volatility to predicted volatility for models M.5-M.7. Indeed, these ratios approximately follow a log-normal distribution $\mathcal{LN}(-0.5s^2, s^2)$, where s corresponds to the daily standard deviation of the logarithm of this quantity. Moreover, the asymptotic distribution of an exponential-OU process is a log-normal distribution. These factors, combined with the fact that the autocorrelation of the residuals in models M.5-M.7 is very low, even at a horizon of 1 day, suggest an average duration of the deviation of the process Y_t from 1 of less than 1 trading day. Consequently, the mean reversion parameter must be extremely high, implying that exogenous random frictions constitute an intraday phenomenon. In the case of discrete modeling with a time step equal to or greater than 1 trading day, one can directly use a log-normal distribution $\mathcal{LN}(-0.5s^2, s^2)$, where the realizations are independent from one trading day to the next. Figure 3.2b illustrates the impact of integrating an exogenous randomness component into the M.7.1 model, employed as a simulation device. While the purely path-dependent version of the model generated volatility dynamics considerably divergent from their empirical counterparts, this modified version, wherein $Y_t = 1 \forall t$ is replaced with $Y_t \sim \mathcal{LN}(-0.5s^2, s^2)$, produces much more realistic trajectories. Specifically, the standard deviation of daily log-returns of both historical and simulated volatility trajectories becomes approximately equal. More generally, this modification enables the model M.7.1 to accurately replicate the main characteristics of empirical volatility behavior.

If we now turn our attention to models M.1-M.4, a question must be posed right away: given the poor performance of these models in explaining spot volatility compared to models M.5-M.7, are they truly relevant for modeling volatility, even when incorporating an exogenous source of randomness? For their part, Guyon and Lekeufack interpret these results as evidence "of the importance of including (an) historical volatility factor". While this observation is coherent, it is, however, not the only possible rational explanation. Indeed, the poor performance of models M.1-M.4 could be attributed, at least partially, to the neglect of exogenous random frictions. However, it does not seem coherent to integrate these in the same manner as for models M.5-M.7. Indeed, for models M.1-M.4, the residuals of the regressions exhibit long memory, and the behavior of the realized volatility to predicted volatility ratio is too erratic to be modeled by OU or exponential OU processes (see figure 3.1). Therefore, for these models, it appears more coherent to incorporate exogenous randomness through θ , using a modeling approach of the form 3.2. In other words, using the terminology of Guyon and Lekeufack ([117]), it makes more sense to favor a "stochastic volatility" type modeling. Under this approach, the volatility dynamics partly depend on a process that is not observable from the asset price data. A phenomenon then may arise, that can be designated as *spurious volatility feedback*, denoting an apparent structural dependence of volatility on its past trajectory, generated by a temporal dependence of volatility on an unobserved third variable. In this case, volatility is not inherently dependent on its past values; however, the fact that both current and past volatilities depend on the same variable creates the illusion of intrinsic autoregressiveness. This phenomenon could potentially explain the superior performance of models M.5-M.7, even though volatility is not structurally dependent on a historical volatility factor. Section 3.4 illustrates this through numerical experiments.

3.4 Numerical experiments

In this section, we conduct numerical experiments to examine the behavior of the volatility process generated by a simple RPDV model that excludes a historical volatility factor while incorporating an exogenous source of randomness. The objective is to evaluate whether this type of modeling can capture the principal empirical features of volatility dynamics while simultaneously displaying apparent volatility feedback, thereby testing the plausibility of the spurious volatility feedback hypothesis introduced in section 3.3.2.2.

3.4.1 Experimental setup

3.4.1.1 The model used

In this section, we consider the following simple RPDV model, from which we will conduct numerical experiments:

$$\begin{cases} \frac{dP_t}{P_t} = \sigma_t(\lambda dt + dB_t) \\ \sigma_t = \beta \int_{-\infty}^t (t-u)^{-\alpha} ((\bar{\sigma} - \sigma_u)du + \sigma_u dZ_u), \end{cases} \quad (3.7)$$

with $\langle dB_t, dZ_t \rangle = \rho$. This model does not include a historical volatility factor and is only a partially path-dependent RPDV since $\rho \notin (-1, 1)$. Indeed, the parameters used for the simulations are as follows:

α	ρ	λ	$\bar{\sigma}$	β
0.425	$-\sqrt{0.5}$	0.4	0.15	0.35

Table 3.5: The parameters used for simulations.

A few remarks on these parameters and their implications are worth making. First, the value of α is set to the value of α_1 obtained in Section 3.3 for model M.2 calibrated on the S&P500 data (rounded to the nearest 0.005). The value of the correlation coefficient $\rho = -\sqrt{0.5}$ is chosen such that there is equal weighting between the Brownian motion B associated with the SDE of the price and the Brownian motion Z modeling the exogenous random shocks that also drive the volatility⁶. Regarding β , this is set using the estimated value of b_1 for Model M.2 and the S&P500 index, such that $\beta \approx \hat{b}_1 \rho$. The value of $\bar{\sigma}$ has been determined through trial and error so that the average value of the volatility process is approximately equal to the average value of the realized volatility of the S&P 500. Furthermore, the values taken by $\bar{\sigma}$ and λ imply that when volatility equals its attraction value ($\sigma_t = \bar{\sigma}$), the drift of the asset price is equal to 6% ($0.15 \times 0.4 = 0.06$).

⁶Indeed, $dZ_t = \rho dB_t + \sqrt{1-\rho^2}dW_t$ where $\langle dB_t, dW_t \rangle = 0$. Thus, since $\rho = -\sqrt{0.5}$, we have $dW_t = -\sqrt{0.5}dB_t + \sqrt{0.5}dW_t$.

3.4.1.2 Simulation methodology

The simulations are conducted using an Euler scheme for the Markovian approximation of model 3.7, a discretization method commonly employed in the literature ([181]). This scheme is presented in the appendix 3.F. The chosen time step Δt is equal to $\frac{1}{252 \times 78}$ expressed in years. This choice is motivated by the fact that our proxy for daily realized volatility corresponds to the square root of the sum of squares of 78 5-minute returns. In order to obtain a similar proxy, the same methodology for calculating daily volatility is used for synthetic data. As for the state variables of the model, those are all initialized to $\frac{\bar{\sigma}}{\sum_{i=1}^n w_i}$, and a 20-year diffusion of the process is performed to minimize this initialization bias. The state variables obtained after this phase are then extracted to carry out the simulation of the volatility process (and the asset price process) that are considered. This initialization phase is repeated for each simulation, constituting the synthetic dataset, ensuring that the simulations begin with diverse state variables distributed in alignment with the model's structure. Following this procedure, 1000 time-series spanning 20 years each are generated. These time series constitute the synthetic dataset analyzed in the following section.

3.4.2 Results of numerical experiments

3.4.2.1 Comparison with market data

The purpose of this section is to compare certain key properties of the simulated volatility trajectories generated from model 3.7 (section 3.4.1.1) with the properties of actual realized volatility data. Since the parameter choice is tailored to the S&P500 data, the comparison of synthetic data is specifically made here with the real data of this index.

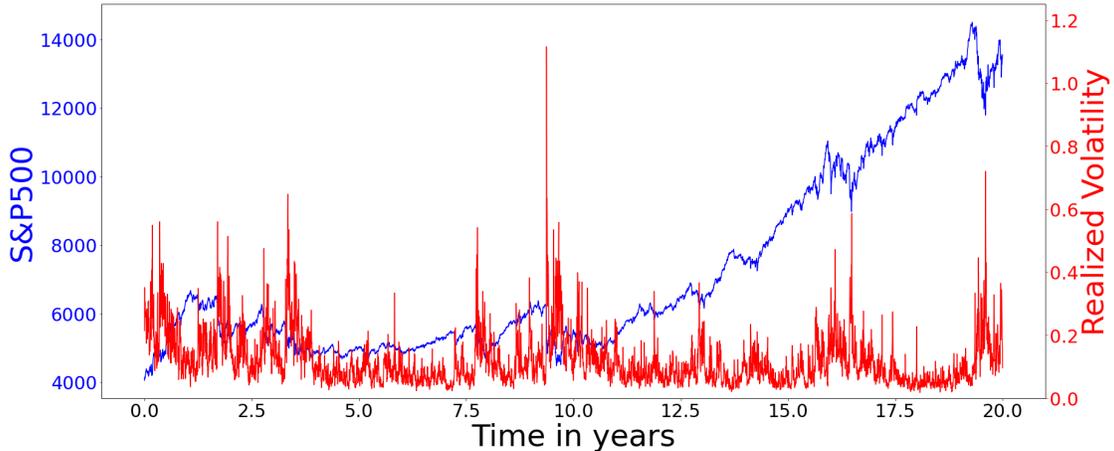


Figure 3.3: Example of joint dynamics between the S&P500 and its volatility generated by the model 3.7.

If we begin by examining the unconditional distribution of volatility, the synthetic data are generally highly consistent with the empirical data. For instance, as shown in figure 3.4, the distribution of daily volatility closely resembles the empirical daily volatility of the S&P500. In a more objective

assessment, table 3.6 demonstrates the consistency of the distributions of volatilities generated by the model at various time scales with their empirical counterparts by comparing their respective moments. Thus, for all considered time horizons, the first four moments of the empirical distributions of the S&P500 are within the first and the ninth decile of the model distributions.

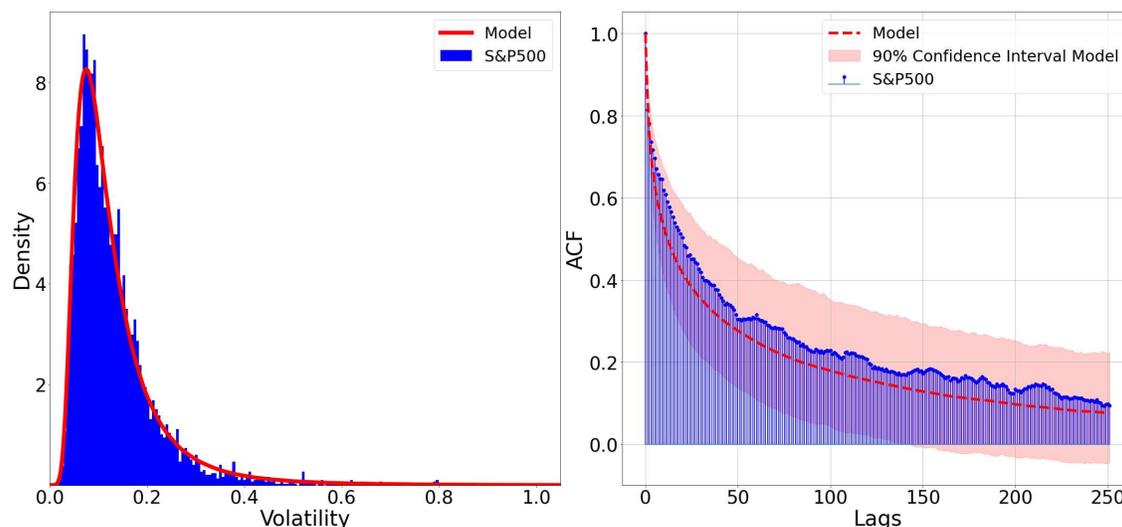


Figure 3.4: The left plot compares the distribution of daily volatility generated by the model 3.7 with the empirical distribution of daily volatility of the S&P500. The right plot compares their respective autocorrelation functions.

Time horizon	1	5	21	63	126	252
Mean of market data	0.134	0.134	0.134	0.134	0.134	0.134
Avg. mean of sim.	0.135	0.135	0.135	0.135	0.135	0.135
1 st decile of mean of sim.	0.105	0.105	0.105	0.105	0.105	0.105
9 th decile of mean of sim.	0.168	0.168	0.168	0.168	0.168	0.168
Std. dev. of market data	0.098	0.088	0.081	0.071	0.064	0.056
Avg. std.deviation of sim.	0.097	0.088	0.078	0.068	0.06	0.051
1 st decile of std. deviation of sim.	0.063	0.056	0.047	0.039	0.033	0.027
9 th decile of std. deviation of sim.	0.137	0.128	0.114	0.099	0.092	0.081
Skewness of market data	3.357	3.05	2.756	2.268	1.935	1.414
Avg. skewness of sim.	2.982	2.663	2.224	1.781	1.458	1.115
1 st decile of skewness of sim.	2.075	1.753	1.347	0.955	0.681	0.39
9 th decile of skewness of sim.	4.299	3.876	3.291	2.787	2.381	1.988
Kurtosis of market data	19.598	14.881	11.437	7.573	5.249	2.35
Avg. kurtosis of sim.	17.594	13.184	8.383	4.817	2.886	1.343
1 st decile of kurtosis of sim.	6.839	4.553	2.227	0.602	-0.154	-0.681
9 th decile of kurtosis of sim.	34.748	26.037	16.32	10.713	6.845	4.134

Table 3.6: The first four moments of the volatility distributions for different time horizons in trading days: S&P500 data vs synthetic data generated from the model 3.7.

In addition, for both the empirical data and the data generated by the model, the standard deviation, skewness, and kurtosis of the volatility distribution follow a decreasing convex relationship with the time horizon considered. Similarly, the autocorrelation of the volatility process generated by the model is fairly close to the empirical autocorrelation of S&P500 volatility (see figure 3.4). Hence, all empirical autocorrelations fall within the model's estimated 90% confidence intervals, and, the model effectively reproduces the long memory of volatility observed in financial time series, characterized by a slower decay of autocorrelation compared to exponential decay. More broadly, most of the phenomena characterizing the joint dynamics of price and volatility are well reproduced by the model. As shown in figure 3.3, the combined phenomenon of volatility spikes leading to sharp price drops akin to near-jumps is captured by the model, just as the presence of relatively long periods of volatility coinciding with positive price trends. Furthermore, certain phenomena that are unattainable from a purely path-dependent model without incorporating a historical volatility factor are replicated here. Among these, for instance, are the "rallies" following a price drop that rapidly propels the asset price to a level significantly higher than

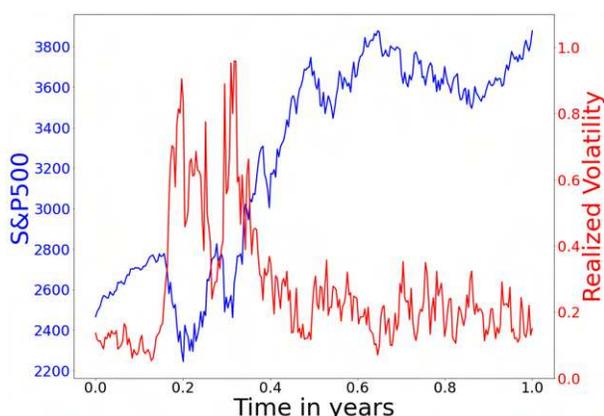


Figure 3.5: Example of a rally phenomenon following an increase in volatility, generated by the model 3.7.

its previous value. This is illustrated in figure 3.5. The phenomenon is as follows: the decrease in the Brownian motion B initially leads to an increase in the volatility level and a decrease in the asset price. This dual movement is even stronger if W also experiences a significant increase. When the dynamics of B becomes positive, the asset price starts to increase again. The sensitivity of the price dynamics to B is directly related to volatility. As a result, there are two polar cases. In the first case, W follows a downward trajectory:

the rise in W combined with the decrease in B leads to a rapid drop in volatility, resulting in low exposure to the upward dynamics of B . Consequently, the rebound in the asset price is very moderate. In the second case, which corresponds to the scenario illustrated in figure 3.5, W follows an upward trajectory: the rise in B , which exerts downward pressure on the volatility level, is slowed down by the opposing effect of the increase in W . The volatility remains at a high level, allowing the asset price to be highly exposed to the upward dynamics of B , leading to a very strong rebound in the price.

This example illustrates another crucial point implied by this modeling framework: the rate of volatility decrease following a spike is highly dependent on the dynamics of the exogenous volatility component. Thus, the phenomenon of slow volatility decay after such a spike is contingent

within the considered modeling framework. However, empirical volatility data suggests that this phenomenon is more systematic, making the time reversal asymmetry of volatility "clear to the naked eye" ([117]). While many properties are well captured by the model 3.7, this last point seems to advocate for a modeling approach that incorporates a historical volatility factor. This question is further explored in the following section.

3.4.2.2 Regressions on synthetic data from different RPDV models

In this section, we perform the same regressions as in section 3.3, but this time using the entire set of synthetic time series generated from the model 4.1. The aim is to determine whether the properties of the synthetic data produce regression results similar to those obtained from market data, and thus to assess the plausibility of the spurious volatility feedback' hypothesis set out in the section.

First and foremost, the results presented in table 3.7 exhibit a consistent pattern with the outcomes derived from market data outlined in table 3.4. Thus, on one hand, the performance ranking of the models remains preserved, and on the other hand, the magnitude of the metrics for these models aligns with the results obtained from empirical data. Another noteworthy element, which might seem surprising, the average performance gap between models M.1-M.4 and models M.5-M.7 is slightly more pronounced for synthetic data than for empirical data. Hence, the outperformance of models incorporating a historical volatility factor compared to models without such a factor is even more evident here, even though the synthetic data considered are generated from a model that does not include such a factor. This element thus seems to lend credence to the hypothesis of spurious volatility feedback, or at the very least, does not allow us to reject this hypothesis. In other words, the outperformance of models incorporating a historical volatility factor does not necessarily imply a structural volatility feedback effect, but could also be a result of exogenous frictions that are not directly measurable. In this hypothesis, the inclusion of a historical volatility factor helps account for the impact of these frictions on volatility, but it does not reveal a causal relationship between historical volatility and spot volatility. However, this hypothesis, in its pure form, meaning a complete absence of positive volatility feedback effects, appears challenging to align with the excessive symmetry between volatility increase and decrease movements generated by the model, as discussed in section 3.4.2.1.

A less stringent assumption, which may better reconcile the various phenomena under consideration, is that positive volatility feedback does exist; however, its magnitude appears greater than its actual scale due to the influence of exogenous randomness. In this hypothesis, the historical volatility factor obtained through regressions performed on empirical data (as those presented in section 3.3 or in [117]) encompasses two heterogeneous components: a genuine positive volatility feedback effect and a component capturing the impact on spot volatility of past exogenous random frictions. This hypothesis thus justifies a modeling approach that, on one hand, includes a historical volatility factor, and on the other hand, incorporates the exogenous randomness component in the

same manner as its endogenous counterpart through the process θ_1 (equations 4.1). Therefore, unlike a modeling approach of the form 3.6, which aims to explain volatility as much as possible through endogenous factors, the exogenous source of randomness of volatility is, in this type of modeling, no longer treated as a residual but as a component of the same nature as the source of randomness associated with the price dynamics. If these two modeling forms constitute competing approaches, they are not necessarily incompatible and could represent two valid ways of describing volatility dynamics from two different angles. Nevertheless, this perspective implies the existence of a mathematical equality relationship between these two modeling forms, allowing for a dialectical resolution of this apparent contradiction.

		Median		Q1		Q3		D1		D9	
		Train	Test	Train	Test	Train	Test	Train	Test	Train	Test
M.1	r^2	0.35	0.105	0.291	-0.15	0.419	0.245	0.229	-0.52	0.479	0.351
	RMSE	0.063	0.067	0.052	0.056	0.076	0.087	0.045	0.05	0.096	0.116
M.2	r^2	0.406	0.174	0.328	-0.1	0.488	0.323	0.233	-0.57	0.576	0.437
	RMSE	0.059	0.065	0.049	0.053	0.073	0.084	0.041	0.046	0.091	0.114
M.3	r^2	0.469	0.281	0.419	0.11	0.539	0.378	0.374	-0.19	0.599	0.481
	RMSE	0.056	0.059	0.047	0.049	0.068	0.078	0.04	0.043	0.085	0.108
M.4	r^2	0.455	0.286	0.407	0.093	0.531	0.368	0.357	-0.21	0.596	0.448
	RMSE	0.056	0.06	0.047	0.05	0.068	0.079	0.04	0.044	0.085	0.109
M.5	r^2	0.82	0.78	0.795	0.748	0.849	0.814	0.776	0.728	0.877	0.85
	RMSE	0.032	0.032	0.028	0.028	0.039	0.04	0.024	0.024	0.047	0.049
M.6	r^2	0.821	0.782	0.796	0.749	0.85	0.816	0.778	0.728	0.877	0.85
	RMSE	0.032	0.032	0.028	0.028	0.039	0.04	0.024	0.024	0.047	0.049
M.7.1	r^2	0.821	0.78	0.796	0.749	0.85	0.816	0.777	0.728	0.876	0.85
	RMSE	0.032	0.032	0.028	0.028	0.039	0.04	0.024	0.024	0.047	0.048
M.7.2	r^2	0.821	0.78	0.796	0.749	0.852	0.816	0.778	0.729	0.877	0.85
	RMSE	0.032	0.032	0.028	0.028	0.039	0.04	0.024	0.024	0.047	0.049
M.7.3	r^2	0.822	0.78	0.797	0.749	0.851	0.816	0.778	0.728	0.877	0.851
	RMSE	0.032	0.032	0.028	0.028	0.039	0.04	0.024	0.024	0.047	0.049

Table 3.7: r^2 and RMSE associated with various RPDV regression models applied to synthetic data generated from model 3.7.

In a more pragmatic perspective of financial engineering, both approaches have their own weaknesses and strengths. For models incorporating a historical volatility factor in the form of 3.6, they hold the significant advantage of making the purely path-dependent component dependent solely on observable or quasi-observable variables - returns and past volatility - and of a single source of endogenous randomness. This property offers undeniable advantages in terms of tractability, ease of computation, and simulation, as listed in section 2 of the article by Guyon and Lekeufack, when the model is purely path-dependent (Y is a constant). However, this comparative advantage diminishes when incorporating an exogenous source of randomness through Y , which appears to be necessary based on empirical data (see section 3.3.2.2). On the other hand, the stochastic volatility modeling approach allows for the incorporation of an exogenous component of random frictions without the need for an additional process Y . However, this more natural way of integrating random frictions within the volatility process can lead to greater complexity in terms of calibration. Consequently, in a financial engineering context, the specific problem at hand may lead to a preference for either one of these two approaches.

3.5 Conclusion

The contribution of this chapter has been twofold. First, it introduced a class of models enabling the joint modeling of both the rough behavior and the path dependency of volatility. Secondly, it provided insights into the question: what are the appropriate ways to model these two empirical characteristics in a manner consistent with market data?

To begin with, section 3.2 initiated by outlining a general form of path-dependent model, which encompasses several significant volatility models found in the current academic literature. Based on this foundation, we then defined what we have referred to as the family of RPDV models. In this modeling framework, volatility is a function of two stochastic processes, which can be seen respectively as a type of price trend process and a type of volatility trend process. Being neither Markovian nor semimartingale since their properties actually define them as a rough models, RPDV models are not very tractable for simulations. To overcome this issue, we therefore introduced a Markovian multi-factor approximation of RPDV models using Abi Jaber's findings ([1], [2]).

The section 3.3 assessed various forms of RPDV models based on their respective ability to explain the realized volatility of five stock indices. The results obtained highlighted several significant findings. Firstly, as previously demonstrated by Guyon and Lekeufack ([117]), purely path-dependent models incorporating a historical volatility factor significantly outperform purely path-dependent models solely based on a price trend process. In addition, it has been shown that these two categories of models differ from each other in the structure of their respective residuals. On one hand, the models incorporating a historical volatility factor exhibit little to no significant autocorrelation

in their residuals. On the other hand, the models that do not incorporate such factors show a very strong and persistent autocorrelation in their residuals. This result has been paralleled with the fact that all the models calibrated on market data - with or without a historical volatility factor - implied a too low randomness of volatility in comparison to empirical volatility randomness. It has thus been demonstrated that the conjunction of these results suggests the need to differentiate how an exogenous randomness source is integrated, depending on whether the model includes a historical volatility factor. In the first case, it is coherent to model volatility as the product of a purely path-dependent model (integrating a historical volatility factor) with a stochastic process whose random component is independent of the asset's price Brownian motion, process that mean-reverts very fast. This corresponds to the approach of Guyon and Lekeufack ([117]). For models without a historical volatility factor, considering the autocorrelation structure of their associated residuals, it is more consistent to integrate the exogenous randomness source as a component of the same nature as the endogenous randomness associated with price dynamics. This aligns with the standard approach of stochastic volatility models. Moreover, this way of integrating the exogenous component of volatility has led to the formulation of the "spurious volatility feedback" hypothesis, which posits that the apparent structural dependence of volatility on its past is actually driven by its temporal dependence on an unobservable source of randomness, rather than a true structural feedback mechanism in volatility.

The credibility of this hypothesis has been investigated in section 3.4 through numerical experiments based on a fairly simple RPDV model that do not incorporate a historical volatility factor but includes an exogenous source of randomness. First of all, based on these experiments, it was shown that most of the phenomena that characterize the empirical dynamics of volatility can be reproduced from this type of model. Furthermore, it has been demonstrated that by replicating the same comparison carried out in section 3.3, this time using synthetic data generated from the considered model, the obtained results were closely aligned with those obtained using empirical data. In particular, similar to the assessments conducted on market data, regression models incorporating a historical volatility factor greatly outperform those that do not include such a factor, even though the data considered in this case are generated by a model that lacks a historical volatility process. These results thus seem to lend credence to the hypothesis of spurious volatility feedback, or at least, do not allow its rejection. However, it has been pointed out that these findings should be tempered by the observation that the volatility increase and decrease movements generated by the model are slightly too symmetrical compared to empirical data. The connection of these two findings has thus led to favoring the hypothesis that there is indeed a positive feedback effect of volatility to some extent, but the magnitude of this effect may be overestimated by the estimation method used due to the presence of exogenous randomness. This hypothesis thus justifies a volatility modeling approach that, on one hand, includes a historical volatility factor, and on the other hand, incorporates the exogenous randomness component in the same manner as standard stochastic volatility models. However, this does not delegitimize the primarily path-dependent approach as adopted in [117], which treats the exogenous part of volatility as a residual. Indeed, it has been emphasized that, on one hand, these two ways of modeling volatility are not necessarily incompatible and, on the other

hand, they each present distinct advantages that position them as complementary approaches in a financial engineering perspective.

3.6 Acknowledgments

I would like to thank my thesis supervisor Jean-Paul Laurent for his support and useful advice in preparing this chapter, as well as Julien Guyon for its feedback.

Appendix 3.A Stochastic differential equations for the Markovian approximation of the RPDV model

We want to solve the following SDE:

$$dm_{1,t}^{(i)} = \frac{dP_t}{P_t} + \left(\kappa_1 \theta_{1,t} - \kappa_1 m_{1,t} - \gamma_i m_{1,t}^{(i)} \right) dt,$$

To consider the dynamics of $m_{1,t}^{(i)}$, we set $g\left(m_{1,t}^{(i)}, t\right) = e^{\gamma_i t} m_{1,t}^{(i)}$ and apply the Itô lemma:

$$dg\left(m_{1,t}^{(i)}, t\right) = e^{\gamma_i t} \left(\frac{dP_t}{P_t} + \kappa_1 (\theta_{1,t} - m_{1,t}) dt \right).$$

Consequently, if $m_{1,0}^{(i)} = 0$

$$m_{1,t}^{(i)} = \int_0^t e^{-\gamma_i(t-u)} \left(\frac{dP_u}{P_u} + \kappa_1 (\theta_{1,u} - m_{1,u}) du \right)$$

Thus:

$$\sum_{i=1}^n w_{1,i} m_{1,t}^{(i)} = \int_0^t \underbrace{\sum_{i=1}^n w_{1,i} e^{-\gamma_{1,i}(t-u)}}_{\hat{\kappa}_1(t-u)} \left(\frac{dP_u}{P_u} + \kappa_1 (\theta_{1,u} - m_{1,u}) du \right)$$

Analogously, with

$$dm_{2,t}^{(i)} = \left((\sigma_t)^{a_2} + \kappa_2 \theta_{2,t} - \kappa_2 m_{2,t} - \gamma_i m_{2,t}^{(i)} \right) dt,$$

by applying same steps, we obtain:

$$\sum_{i=1}^n w_{2,i} m_{2,t}^{(i)} = \int_0^t \sum_{i=1}^n w_{2,i} e^{-\gamma_{2,i}(t-u)} \left((\sigma_u)^{a_2} + \kappa_2 (\theta_{2,u} - m_{2,u}) \right) du.$$

Appendix 3.B Numerical quadratures

We are aiming to approximate the different integrals that make up the various elementary building blocks of the tested RPDV models. The first integral we are seeking to approximate is as follows:

$$\int_{-\infty}^{t-\epsilon} (t-u)^{-\alpha} \frac{dP_u}{P_u}. \quad (3.8)$$

First, using the midpoint method, we have the following approximation:

$$\int_{t_{i-1}}^{t_i} (t-u)^{-\alpha} \frac{dP_u}{P_u} \approx (t-0.5(t_{i-1}+t_i))^{-\alpha} \frac{P_{t_i}-P_{t_{i-1}}}{P_{t_{i-1}}}.$$

Therefore, because ϵ is assumed to be negligible ($\epsilon \approx 0$), we can approximate the integral 3.8 as follows:

$$\sum_{t_i \leq t} (t-0.5(t_{i-1}+t_i))^{-\alpha} \frac{P_{t_i}-P_{t_{i-1}}}{P_{t_{i-1}}}.$$

The other integrals that we are seeking to approximate are of the form:

$$\int_{-\infty}^{t-\epsilon} (t-u)^{-\alpha} f(u) du. \tag{3.9}$$

Using the midpoint rectangle rule, we have the following approximation:

$$\int_{t_{i-1}}^{t_i} (t-u)^{-\alpha} f(u) du \approx (t-0.5(t_{i-1}+t_i))^{-\alpha} f(t_{i-1})(t_i-t_{i-1}).$$

Consequently, we can approximate the integral 3.9 as follows:

$$\sum_{t_i \leq t} (t-0.5(t_{i-1}+t_i))^{-\alpha} f(t_{i-1})(t_i-t_{i-1}).$$

Appendix 3.C The ARPDV model as a simulation device

The purpose of this appendix is to introduce a discrete version of the model M.6 (and consequently of M.2 and M.7 when $\theta = 0$) formulated as an extended autoregressive model, enabling the direct utilization of parameters obtained from the regression method discussed in section 3.3.1. For simplification purposes, we assume here that $\mu_t = 0$, which is equivalent to assuming that the asset price is a martingale. Furthermore, in order to incorporate the exogenous component of volatility, we assume here that volatility follows a relation of the form 3.6, with Y being an exponential OU process (refer to Appendix 3.E). Based on this premise, we propose the following discrete model,

which we refer to as the *Autoregressive Rough Path-Dependent Volatility* (ARPDV) model:

$$r_t = \sigma_{t-1} \sqrt{\Delta t} X_t,$$

$$P_t = P_{t-1}(1 + r_t)$$

$$\sigma_t = \left(b_0 + b_1 \sum_{i=1}^l w_{1,i} r_{t-i} + b_2 \sum_{i=1}^l w_{2,i} \sigma_{t-i} \right) \exp \left(-0.5s^2 + sZ_t \right), \quad w_{j,i} = \left(t - (i + 0.5)\Delta t \right)^{-\alpha_j},$$

where X_t and Z_t are i.i.d realizations from a standard normal distribution. In the case of $s = 0$, the model is purely path-dependent. As soon as $s \neq 0$, the path-dependent component is multiplied by a log-normal distribution with an expected value of 1.

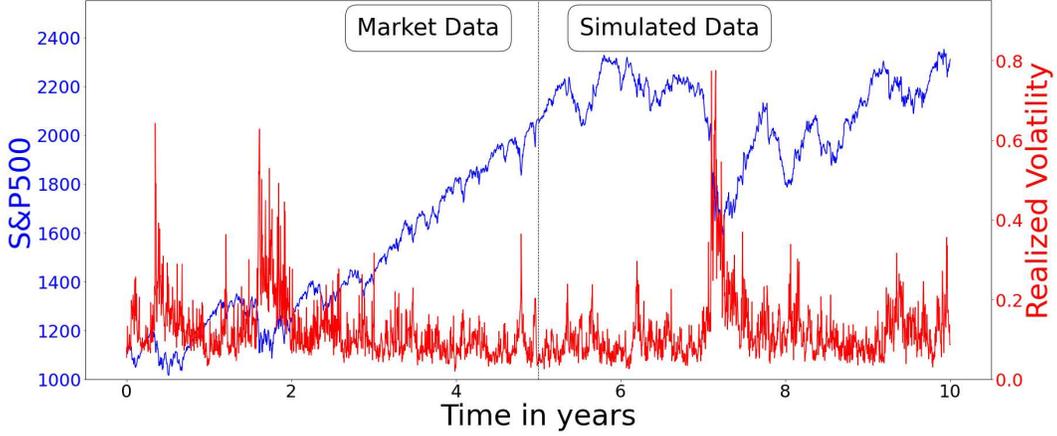


Figure 3.6: Example of joint dynamics between the S&P500 and its volatility. The first 5 years consist of real data, while the subsequent 5 years are generated from 3-parameter ARPDV model with $\alpha = 1.315$, $b_1 = -3.42$, $s = 0.316$.

Remark Interestingly, when we impose the constraint $\alpha_1 = \alpha_2$, the optimal parameters (in the least squares sense) for the various indices are such that $b_0 \approx 0$ and $b_2 \sum_{i=1}^l w_{2,i} \approx 1$. Given that, it makes sense to use the following 3-parameter ARPDV model:

$$r_t = \sigma_{t-1} \sqrt{\Delta t} X_t,$$

$$P_t = P_{t-1}(1 + r_t)$$

$$\sigma_t = \left(b_1 \sum_{i=1}^l w_i r_{t-i} + \sum_{i=1}^l w_i \sigma_{t-i} \right) \exp \left(-0.5s^2 + sZ_t \right), \quad w_i = \frac{\left(t - (i + 0.5)\Delta t \right)^{-\alpha}}{\sum_{k=1}^l \left(t - (k + 0.5)\Delta t \right)^{-\alpha}}.$$

Thus, as shown in figure 3.6, even though very parsimonious, this model enables the generation of realistic joint dynamics of price and volatility.

Appendix 3.D Estimation of the minimum value of the standard deviation of the log-returns of volatility

The estimator of realized volatility used is calculated as follows:

$$\hat{\sigma}_t = \sqrt{\frac{1}{n\Delta t} \sum_{i=1}^n r_{t,i}^2}.$$

Let $r_{t,i} \sim \mathcal{N}(0, \sigma_t^2 \Delta t)$. Under this assumption, the estimated log-returns of volatility can be decomposed as follows:

$$\underbrace{\log\left(\frac{\hat{\sigma}_{t+1}}{\hat{\sigma}_t}\right)}_{\hat{R}_\sigma} \stackrel{d}{=} \log\left(\frac{\sigma_{t+1} \sqrt{\frac{1}{n} \sum_{i=1}^n X_i^2}}{\sigma_t \sqrt{\frac{1}{n} \sum_{i=1}^n Z_i^2}}\right) = \underbrace{\log\left(\frac{\sigma_{t+1}}{\sigma_t}\right)}_{R_\sigma} + \underbrace{\frac{1}{2} \log\left(\frac{\sum_{i=1}^n X_i^2}{\sum_{i=1}^n Z_i^2}\right)}_{\mathcal{E}},$$

where $\{X_i\}_{1 \leq i \leq n} \cup \{Z_i\}_{1 \leq i \leq n}$ is a i.i.d. sample such as $\forall i : X_i, Z_i \sim \mathcal{N}(0, 1)$. Consequently, denoting by ρ the correlation between R_σ and \mathcal{E} , we have:

$$\begin{aligned} \text{Var}(\hat{R}_\sigma) &= \text{Var}(R_\sigma) + \text{Var}(\mathcal{E}) + 2\text{Std}(R_\sigma)\text{Std}(\mathcal{E})\rho \\ 0 &= \text{Var}(\mathcal{E}) - \text{Var}(\hat{R}_\sigma) + 2\rho\text{Std}(\mathcal{E})\text{Std}(R_\sigma) + \text{Std}(R_\sigma)^2. \end{aligned}$$

Because $\text{Std}(R_\sigma) > 0$:

$$\text{Std}(R_\sigma) = \sqrt{\text{Var}(\mathcal{E})(\rho^2 - 1) + \text{Var}(\hat{R}_\sigma) - \rho\text{Std}(\mathcal{E})}.$$

In addition, since $\text{Var}(\hat{R}_\sigma) - \text{Var}(\mathcal{E}) > 0$

$$\frac{\partial \text{Std}(R_\sigma)}{\partial \rho} = \text{Std}(\mathcal{E}) \left(\frac{\text{Std}(\mathcal{E})\rho}{\sqrt{\text{Var}(\mathcal{E})\rho^2 + \text{Var}(\hat{R}_\sigma) - \text{Var}(\mathcal{E})}} - 1 \right) < 0.$$

Consequently, $\text{Std}(R_\sigma)$ increases with ρ over the interval $[-1 : 1]$. It follows that the minimum value of $\text{Std}(R_\sigma)$ is:

$$\min \text{Std}(R_\sigma) = \text{Std}(\hat{R}_\sigma) - \text{Std}(\mathcal{E}).$$

In the case considered in section 3.3.2.2, $\text{Std}(\hat{R}_\sigma) \approx 0.347$. Moreover, with $n = 78$ and using the relation $\mathcal{E} = 0.5 \cdot \log(F)$, where F follows a Fisher distribution $\mathcal{F}(78, 78)$, we obtain $\text{Std}(\mathcal{E}) \approx 0.114$. Consequently, the standard deviation of the daily log-returns of volatility is at a minimum around 0.23 ($0.347 - 0.114 = 0.233$).

Appendix 3.E Exponential-OU process

We know that if Y follows

$$dY_t = \eta(\bar{Y} - Y_t)dt + v dZ_t,$$

the conditional distribution of Y_t given \mathcal{F}_0 is ([153]):

$$Y_t \sim \mathcal{N}\left(\bar{Y} + (Y_0 - \bar{Y})e^{-\eta t}, \frac{v^2}{2\eta}(1 - e^{-2\eta t})\right).$$

Therefore, if $\bar{Y} = -\frac{v^2}{4\eta}$, by setting $s^2 = \frac{v^2}{2\eta}$ we have:

$$\lim_{t\eta \rightarrow +\infty} \exp(Y_t) \sim \mathcal{LN}(-0.5s^2, s^2).$$

Appendix 3.F The model used for numerical experiments

By utilizing the approach detailed in section 3.2.3, the model 4.1 can be approximated by the following Markovian model:

$$\begin{cases} \frac{dP_t}{P_t} &= \sigma_t(\lambda dt + dB_t), \\ dM_t &= \mathbf{1}_n \cdot (\sigma_t dZ_t + (\bar{\sigma} - \sigma_t)dt) - \Lambda \odot M_t dt, \\ \sigma_t &= \beta \cdot \langle \mathbf{W}, M_t \rangle, \end{cases}$$

where \mathbf{W} is the vector of weights and Λ is the vector of discount coefficients, whose values are determined using the method outlined in section 3.2.3 for a given pair (α, n) . For the numerical experiments conducted in section 3.4, we utilize a 10-factor approximation ($n = 10$). In addition, to perform simulations from this model, we use the following explicit Euler discretization scheme:

$$\begin{cases} P_{t+\Delta t} &= P_t \left(1 + \sigma_t(\lambda \Delta t + \sqrt{\Delta t} X_{t+\Delta t})\right), \\ M_{t+\Delta t} &= M_t \odot (\mathbf{1}_n - \Lambda \cdot \Delta t) + \mathbf{1}_n \cdot \left(\sigma_t \sqrt{\Delta t} (\rho X_{t+\Delta t} + \sqrt{1 - \rho^2} W_{t+\Delta t}) + (\bar{\sigma} - \sigma_t) \Delta t\right), \\ \sigma_{t+\Delta t} &= \beta \cdot \langle \mathbf{W}, M_{t+\Delta t} \rangle, \end{cases}$$

where X_t and W_t are i.i.d realizations from a standard normal distribution. It is worth to note that in order to ensure the stability of the scheme, all coordinates of Λ must lower than $\frac{1}{\Delta t}$. This condition is verified for the specification considered in section 3.2.3.

CHAPTER 4

Price and Volatility Dynamics Under Path-Dependent Models

Abstract

This chapter explores the implications of various specifications of a path-dependent volatility (PDV) model¹ that takes the form of a stochastic convolutive equation, under the assumption that price drift follows a quadratic relationship with volatility. Initially, the chapter examines the impact of assumptions about the risk premium on volatility dynamics. It thus demonstrates how, within the considered framework, the components of the risk premium influence the speed of mean reversion of the volatility process toward its long-term level, as well as the level of attraction to which volatility is drawn. Several refinements of the model are then considered, which underlie different assumptions about the mechanisms of price and volatility formation. The first type considered assumes that the level of attraction of volatility is itself a path-dependent process, leading to an entanglement of path-dependence. The second type of refinement aims to articulate the dynamics of price and volatility with the relationship between price and fundamental value. Two hypotheses in this regard are then explored. In the first, the level of attraction of volatility acts as an adjustment operator between price and fundamental value, producing an effect on the level of volatility and consequently influencing the drift to bring the price closer to the fundamental value. In the second hypothesis, the relationship is reversed: the risk premium serves as the adjustment variable, resulting in a modification of the mean reversion strength of volatility based on the differential between price and fundamental value.

4.1 Introduction

Section 3.4 of the previous chapter has shown that a partially path-dependent volatility model, in the form of a stochastic convolutive equation with a single kernel, can replicate most of the characteristics of price and volatility dynamics. Moreover, their relative simplicity in terms of structure makes them interesting analytical tools for understanding the relationships that link the joint dynamics of price, drift, and volatility.

¹It is understood here that PDV models include both fully and partially path-dependent models.

Building on these insights, this chapter explores the implications of various possible specifications of the following PDV model², which belongs to the general class (4.1) introduced in section 3.2 of chapter 3:

$$\begin{cases} \frac{dP_t}{P_t} &= \mu_t dt + \sigma_t dB_t \\ \sigma_t &= \beta_0 + \beta_1 m_{1,t}, \\ m_{1,t} &= \int_{I_t} \mathcal{K}(t-u) \left(\frac{dP_u}{P_u} + \kappa_1(\theta_{1,u} - \sigma_u) du \right), \\ \theta_{1,t} dt &= \left(\theta_t - \frac{r_t}{\kappa_1} \right) dt + \nu \sigma_t dW_t, \end{cases} \quad (4.1)$$

where B and W are independent Brownian motions, θ is a \mathcal{F}_t -measurable process, \mathcal{K} is a positive decreasing convolution kernel on \mathbb{R}_+ , $\alpha > 0, \beta_0 \geq 0, \beta_1 < 0, \kappa_1, \nu \in \mathbb{R}$, and with r the free risk rate. For the sake of clarity, we rewrite the expression of the volatility as:

$$\sigma_t = \beta_0 + \int_{I_t} \mathcal{K}(t-u) \left((\kappa(\theta_u - \sigma_u) - \beta_1 r_u) du + \left(\bar{\nu} \sigma_u dW_u + \beta_1 \frac{dP_u}{P_u} \right) \right), \quad (4.2)$$

with $\kappa = \kappa_1 \beta_1$ and $\bar{\nu} = \nu \beta_1 \kappa_1$. Equivalently, by introducing a Brownian motion Z such that $dZ_t = \rho dW_t + \sqrt{1 - \rho^2} dB_t$, and setting $\beta_1 = -\beta \rho$ and $\nu = -\beta \sqrt{1 - \rho^2}$, we can also express the volatility process as follows:

$$\sigma_t = \beta_0 + \int_{I_t} \mathcal{K}(t-u) \left((\kappa(\theta_u - \sigma_u) + \beta(r_u - \mu_u)) du - \beta \sigma_t dZ_u \right).$$

As clearly shown in this last equation, the modeling framework considered implies an influence of the drift μ on the volatility dynamics due to its path-dependent property. In this chapter, we will examine its influence in the context where the risk premium of the asset ([34]) is an affine function of the volatility, such as:

$$\frac{\mu_t - r_t}{\sigma_t} = \lambda_1 + \lambda_2 \sigma_t, \quad (4.3)$$

where λ_1 and λ_2 are assumed to be positive or zero³. This leads to the asset price drift following the quadratic relationship:

$$\mu_t = r_t + \lambda_1 \sigma_t + \lambda_2 (\sigma_t)^2, \quad (4.4)$$

where λ_1 and λ_2 can be interpreted as the volatility and variance risk premia, respectively. This type of asset price drift, defined by a risk-free rate component and a risk premium component,

²Here, the notion of PDV model includes both partially and fully path-dependent models.

³The assumption of the constancy of λ_1 and λ_2 is lifted in section 4.4.1.

aligns with a modeling approach similar to that of the intertemporal capital asset pricing model proposed by Merton ([154]), under the assumption that the volatility of the considered asset is linked by a linear relationship with the instantaneous covariance of its returns with the returns of the market portfolio (covariance, which is assumed to be positive). Regarding the quadratic form 4.4 similar to the one proposed in [118], this allows for the modeling of a potential convex relationship between the level of volatility and the drift.

Using the expression for the drift 4.4, the volatility can be rewritten as the following Brownian semistationary process ([20]):

$$\sigma_t = \beta_0 + \int_{I_t} \mathcal{K}(t-u) f_u(\sigma_u) du - \beta \int_{I_t} \mathcal{K}(t-u) \sigma_u dZ_u, \quad (4.5)$$

where

$$f_t(\sigma_t) = \kappa\theta_t - (\kappa + \beta\lambda_1)\sigma_t - \beta\lambda_2(\sigma_t)^2. \quad (4.6)$$

In addition, to bound our analysis, we only consider the cases where the following constraints are respected $\forall t \in \mathbb{R}_+$:

1. $\forall \sigma_t \in [0 : \bar{\sigma}_t[: f_t(\sigma_t) \geq 0$,
2. $f_t(\bar{\sigma}_t) = 0$,
3. $\forall \sigma_t \in]\bar{\sigma}_t, +\infty[: f_t(\sigma_t) \leq 0$, and f_t is concave and decreasing.

To describe the behavior of the volatility dynamics for different model specifications, we will refer to $\bar{\sigma}_t$ as the *attraction volatility* whenever the equation $f_t(\bar{\sigma}_t) = 0$ admits a unique solution for each $t \in \mathbb{R}_+$. Moreover, the existence of an attraction volatility for f_t ensures that a mean-reverting force prevents the volatility process from diverging, even in the case where \mathcal{K} is a power law kernel.

Having established the general modeling framework under consideration, a few preliminary remarks deserve to be made.

First, in the case where $\bar{\nu} = 0$ and θ is a deterministic function (equation 4.2), the model is purely path-dependent. Conversely, if $\bar{\nu}$ is nonzero, the model is only partially path-dependent and includes an exogenous source of randomness. Moreover, the proposed model is a stochastic Volterra function when $I_t = [0 : t]$. While this type of process is commonly used in the volatility literature, the application of these Volterra processes in this context is largely confined to modeling variance through a Volterra square-root or Volterra Heston process [1, 2, 4, 5]. However, the type of convolutive equation considered here for the volatility (in the case where $I_t = [0, t]$) has already been the subject of literature that has led to a number of important technical results ([165], [174],[167], [75], [5]). In particular, when f is deterministic and affine (i.e., θ is constant and $\lambda_2 = 0$), σ_t

belongs to the class of affine Volterra processes for which existence and uniqueness results for the corresponding stochastic convolution equations are available ([5]).

The remainder of this chapter will set aside these technical aspects to focus on the impact of different specifications of f (equation 4.6). Several objectives are pursued. First, we aim to define the conditions for the consistency of the specifications of f while clarifying the influence of the different parameters on volatility dynamics. In this way, we also seek to make explicit the economic hypotheses underlying the model specifications. Additionally, this chapter aims to examine competing assumptions regarding the mechanisms of price and volatility formation, employing model 4.5 as an analytical framework.

To achieve these objectives, the chapter is organized according to the hypotheses formulated on θ . Section 4.2 analyzes the impact of the model parameters on the level of attraction volatility as well as on the forces driving volatility towards this value, under the assumption that θ is a constant. Section 4.3 then considers specifications in which θ itself is a path-dependent process to discuss the impact of this entanglement of path dependence. Finally, section 4.4 presents θ as a long-run coherence operator between price and value. Two hypotheses are discussed: the first assumes that the mechanisms driving the price's return to the vicinity of "fundamental value" in the long run occur through variations in the level of attraction volatility, while the second implies that this mean-reverting behavior is caused by adjustments of the risk premium.

4.2 The function θ as a constant

The θ function plays a central role in the dynamics of volatility. A natural specification is to consider it as a constant.

The first option is to assume that θ is strictly positive. In such a case, for $f(0) > 0$, κ must also be positive. Under this specification, because of the term $\kappa(\theta - \sigma_t)$, θ could seem to correspond to the long-term volatility level towards which the volatility process tends to revert like in the Heston ([121]) or Ornstein-Uhlenbeck (OU) ([186]) models. However, except in the risk-neutral measure hypothesis (i.e. when λ_1 and λ_2 are zero), the volatility level towards which the volatility tends to revert is not equal to θ . Indeed, in addition to the difference between θ and σ_t , the price drift μ_t also plays a mean-reverting role by inducing a risk premium effect of $-\beta(\lambda_1\sigma_t + \lambda_2(\sigma_t)^2)$, which is a negative, continuously decreasing function of the volatility on \mathbb{R}_+ .

The second option is to assume that θ and κ are negative. The underlying hypothesis is the existence of positive volatility feedback through $-\kappa\sigma_t$. At first sight, under this specification and if \mathcal{K} is a power law kernel, the model may appear to be divergent due to the presence of positive

feedback. However, due to the risk premium effect mentioned above, this is not the case under appropriate specifications of the drift process μ_t , which are detailed in the rest of this section.

Risk premium assumptions thus play a central role in the dynamics of volatility, regardless of the value of θ . Let us consider the cases $\lambda_2 = 0$ and $\lambda_2 > 0$ in turn.

4.2.1 The case $\lambda_2 = 0$

If we consider the case $\lambda_2 = 0$, f is defined by⁴:

$$f(\sigma_t) = \kappa\theta - (\kappa + \beta\lambda_1)\sigma_t. \quad (4.7)$$

So that σ_t not to explode, $-\kappa \leq \beta\lambda_1$. In addition, κ and θ must share the same sign in order that $f(0) \geq 0$. If κ is positive, the condition $-\kappa \leq \beta\lambda_1$ is clearly satisfied. However, this inequality is more interesting for $\kappa < 0$. In this case, it means concretely that the risk premium effect that puts downward pressure on volatility must be greater than the positive volatility feedback. As long as this condition is met, regardless the κ value, f decrease linearly with the volatility and admit unique constant attraction volatility given by:

$$\bar{\sigma} = \frac{\kappa\theta}{\kappa + \beta\lambda_1}. \quad (4.8)$$

The attraction volatility results thus from two effects: a volatility feedback effect and a risk premium effect. If $\kappa > 0$ both ones share the same direction. Conversely, if $\kappa < 0$, these two effects are antagonistic. In both cases, f can be rewritten as follows:

$$f(\sigma_t) = (\kappa + \beta\lambda_1)(\bar{\sigma} - \sigma_t), \quad (4.9)$$

and therefore,

$$\sigma_t = \beta_0 + \bar{\kappa} \int_{I_t} \mathcal{K}(t-u)(\bar{\sigma} - \sigma_u)du - \beta \int_{I_t} \mathcal{K}(t-u)\sigma_u dZ_u, \quad (4.10)$$

with $\bar{\kappa} = \kappa + \beta\lambda_1$. Equation 4.10 highlights thus the mean-reverting component that pushes the volatility process towards the attraction level $\bar{\sigma}$. Thus, the greater the risk premium λ_1 , the lower the attraction volatility and the higher the mean-reverting force $\bar{\kappa}$.

The $\theta = 0$ case is specific since it does not admit an attraction volatility. The volatility is then equal to

$$\sigma_t = \beta_0 - \int_{I_t} \mathcal{K}(t-u)\sigma_u(\beta dZ_u - \bar{\kappa}du).$$

It is interesting to note that if $\bar{\kappa} = 0$ ($\kappa = -\beta\lambda_1$), the volatility becomes an affine function of a

⁴We denote f without a time index when its time dependency only depends on σ_t .

power-law moving average of the past returns adjusted to its deterministic component. In economic terms, it may be interpreted as the fact that the risk premiums and the instantaneous volatility are known by the market and, therefore, only stochastic fluctuations (modeled by Brownian dynamics) have an impact on the dynamics of volatility.

4.2.2 The case $\lambda_2 > 0$

Now consider the behavior of f under the assumption $\lambda_2 > 0$. Given this configuration, f admits as unique attraction volatility

$$\bar{\sigma} = \frac{\kappa + \beta\lambda_1 - \sqrt{(\kappa + \beta\lambda_1)^2 + 4\beta\lambda_2\kappa\theta}}{-2\beta\lambda_2}, \quad (4.11)$$

and f may be rewritten as (see appendix 4.A):

$$f(\sigma_t) = (\kappa + \beta\lambda_1)(\bar{\sigma} - \sigma_t) + \beta\lambda_2(\bar{\sigma}^2 - (\sigma_t)^2). \quad (4.12)$$

The existence of a non-zero variance risk premium, therefore, led to the emergence of a second mean-reverting term - $\beta\lambda_2(\bar{\sigma}^2 - (\sigma_t)^2)$ - which depends linearly on the differential between the attraction variance $\bar{\sigma}^2$ and the variance process $(\sigma_t)^2$. However, two cases must be treated separately: $-\kappa \leq \beta\lambda_1$ and $-\kappa > \beta\lambda_1$.

In the first case ($-\kappa \leq \beta\lambda_1$), f is continuously decreasing in volatility (on \mathbb{R}_+), but contrary to the case defined by equation 4.7, this relationship is strictly concave. Consequently, the force of mean reversion is stronger for $\sigma_t = \bar{\sigma} + \Delta\sigma$ than for $\sigma_t = \bar{\sigma} - \Delta\sigma$ for all $\Delta\sigma > 0$. On a practical level, this direct consequence of the quadratic form of the drift process reduces the probability of extremely high volatility levels compared to the pure volatility premium hypothesis (i.e., $\lambda_2 = 0$).

Now, if we consider the case where $-\kappa > \beta\lambda_1$, the behaviour of f change dramatically as shown in table 4.1. In this specification, f is no more homogeneous but concave downward because of the antagonistic effects of the positive volatility feedback and the risk premium effect (see figure 4.2). Indeed, the volatility feedback component $-\kappa\lambda_1$ is linearly increasing in volatility while the risk premium effect $-\beta(\lambda_1\sigma_t + \lambda_2(\sigma_t)^2)$ is strictly concave decreasing of volatility. Therefore, when the volatility is between 0 and the attraction volatility (equation 4.11), the volatility feedback effect dominates and f is positive. Conversely, when σ_t is greater than the attraction volatility, the risk premium effect dominates and f is positive. This specification is particularly interesting in the case where $\mathcal{K}(\cdot)$ is a rough kernel, as it allows for a strict positive volatility feedback (strict in the sense that $\kappa + \beta\lambda_1 > 0$) while preventing the divergence of the process via the variance premium.

σ_s	0	$-\frac{\kappa + \beta\lambda_1}{2\beta\lambda_2}$	$\frac{\kappa + \beta\lambda_1 - \sqrt{(\kappa + \beta\lambda_1)^2 + 4\beta\lambda_2\kappa\theta}}{-2\beta\lambda_2}$	$+\infty$
$f'(\sigma_s)$	+	0	-	-
$f(\sigma_s)$	0	$\kappa\theta + \frac{(\kappa + \beta\lambda_1)^2}{4\beta\lambda_2}$	0	$-\infty$

Figure 4.1: The table of variation of $f(\sigma_t)$ when $\lambda_2 > 0$ and $-\kappa > \beta\lambda_1$.

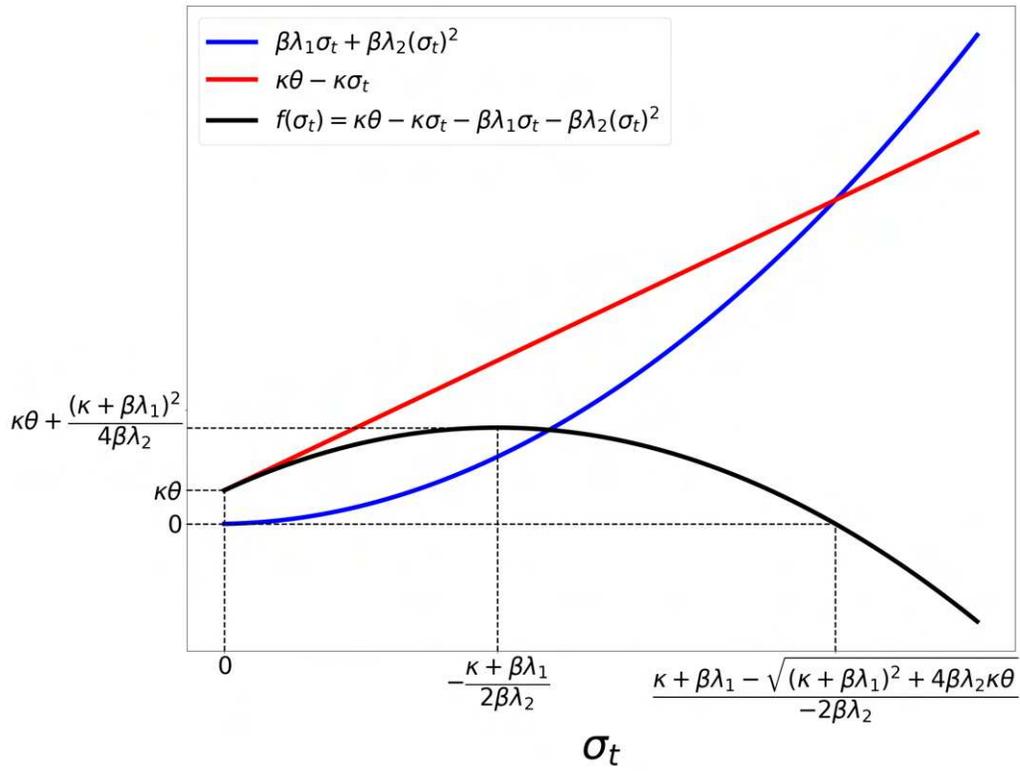


Figure 4.2: Plot of $f(\sigma_t)$ and its 2 components $\kappa\theta - \kappa\sigma_t$ and $\beta\lambda_1\sigma_t + \beta\lambda_2(\sigma_t)^2$, when $-\kappa > \beta\lambda_1$.

The study of these different specifications of 4.2 emphasizes the importance of assumptions relating to the drift process on the mechanisms that push back the volatility towards an attraction level. More precisely, it has been shown through equations 4.8 and 4.11 that the level of attraction volatility is a decreasing function of risk premia⁵. In other words, all things being equal, the higher the risk premia are, the lower the mean volatility. Conversely, the mean-reversion speed is an increasing function of risk premia (equations 4.9 and 4.12). Consequently, the higher the risk premia, the lower the dispersion of the volatility distribution. Broadly speaking, it has shown that even when we limit to the case θ as constant, the model covers various competing assumptions about volatility dynamics.

In sections 4.3 and 4.4, we will consider hypotheses in which θ is time-dependent.

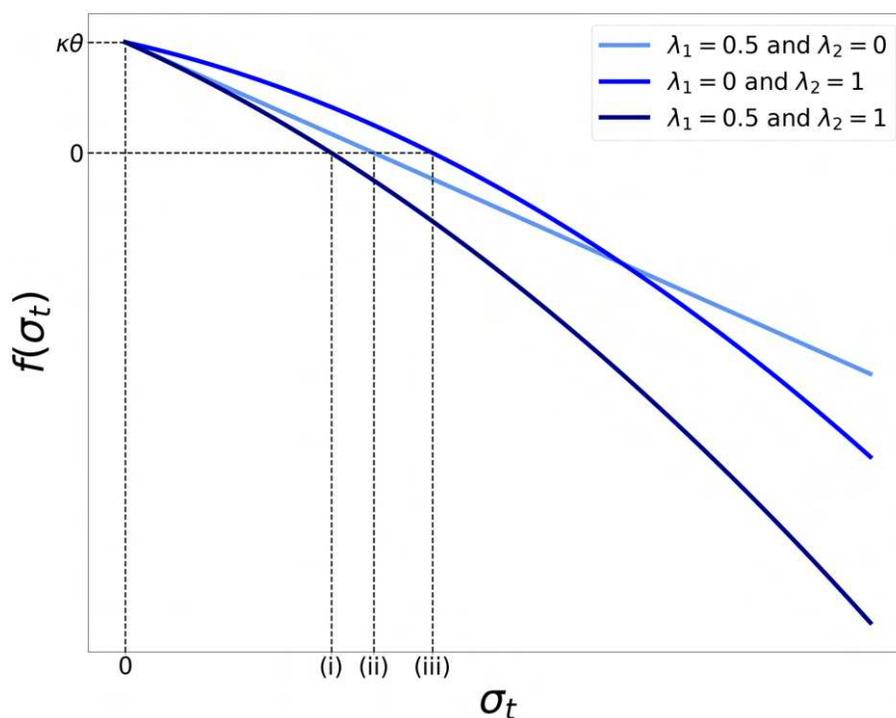


Figure 4.3: Plot of $f(\sigma_t)$ when κ and θ are strictly positive, with different hypotheses about the risk premium, and the associated zeros of $f(\sigma_t)$: (i) = $\frac{\sqrt{(\kappa + \beta\lambda_1)^2 + 4\kappa\theta\beta\lambda_2} - \kappa - \beta\lambda_1}{2\beta\lambda_2}$, (ii) = $\frac{\kappa\theta}{\kappa + \beta\lambda_1}$, and (iii) = $\frac{\sqrt{\kappa^2 + 4\kappa\theta\beta\lambda_2} - \kappa}{2\beta\lambda_2}$.

⁵When we refer to "risk premia" in this chapter, we are referring to λ_1 and λ_2 , which determine the total risk premium λ_t (equation 4.3).

4.3 The function θ as a path-dependent process

A possible alternative to understanding θ as a constant is to specify it as a path-dependent process. The idea is to take inspiration from the structure of the super-Heston raw volatility model (SHRV) introduced in [80] in the specification of θ . This SHRV model is defined by:

$$\begin{aligned}(\sigma_t)^2 &= (\sigma_0)^2 + \beta \int_0^t \frac{(t-u)^{\alpha-1}}{\Gamma(\alpha)} \sigma_u dZ_u + \kappa \int_0^t \frac{(t-u)^{\alpha-1}}{\Gamma(\alpha)} (\theta_u - (\sigma_u)^2) du, \\ \theta_t &= \bar{\theta}_t + (X_t)^2, \\ X_t &= \int_0^t K(t-u) \frac{dP_u}{P_u},\end{aligned}$$

with $\bar{\theta}$ a deterministic function, K a kernel function, and Z a Brownian motion independent on B^6 . Therefore, the process X and thus θ are path-dependent processes allowing to capture of the strong Zumbach effect. Concretely, θ which plays the role of attraction variance (the level towards which the process of variance tends), is a function of the square of a moving average of past returns (i.e. $(X_t)^2$). Due to the independence between W and B , the path dependency is only captured by this variable θ . Regarding this point, the SHR model and the path-dependent volatility model under consideration differ. Indeed, for a model of the form 4.2, the path dependency of θ adds to the path dependency depending on

$$\int_{I_t} \mathcal{K}(t-u) \frac{dP_u}{P_u}. \quad (4.13)$$

There is then a mechanism of entanglement of path dependencies. The question then arises about the interest to introduce this new layer of complexity. One of the main potential motivations is better accounting of the time-reversal asymmetry that characterizes financial time series ([203], [204], [205]). This consideration appears relevant in light of the results presented in section 3.4 of chapter 3, which indicate that the upward and downward movements in volatility generated by models that do not a historical volatility factor are "slightly too symmetrical compared to empirical data". The aim is to model the fact that a sharp increase in volatility is typically triggered by a price drop through a fall in 4.13. However, even when 4.13 returns to its level prior to the volatility spike, volatility tends to remain elevated due to a latency phenomenon. More broadly, this approach allows the attraction volatility to be modeled as a path-dependent process, thereby explaining phenomena that are often perceived as exogenous structural breaks in the volatility formation mechanism. If $\kappa > 0$, a possible specification of θ in line with this objective is

$$\theta_t = \bar{\theta} + |X_t|^{\frac{1}{\alpha}}, \quad (4.14)$$

$$X_t = \int_{I_t} K(t-u) \left(\frac{dP_u}{P_u} \right)^{\alpha}, \quad (4.15)$$

⁶Note also that the SHRV model is a RPDV model as defined in chapter 3.

with $a \in \{1, 2\}$, $\bar{\theta} > 0$ and K a kernel whose integral on \mathbb{R}_+ is convergent (for instance an exponential kernel).

First, note that since θ is necessarily positive and given results obtained in section 4.2, the existence and unicity of the attraction volatility is guaranteed at each time t . Moreover, this attraction volatility is given by equation 4.8 if $\lambda_2 = 0$ and by equation 4.11 if $\lambda_2 > 0$ (we just replace θ by $\theta(t)$ defined by 4.14)⁷. Whatever the assumption about risk premia, by the relationship between θ and the attraction volatility, the expression 4.14 makes the level towards which volatility tends a function of a process dependent on the historical price path: X . The nature of this process depends on the value taken by the parameter a .

If $a = 1$, X is a moving average of returns like the term 4.13, and may be apprehended as a past price trend. Consequently, the volatility process becomes linearly dependent on a moving average of the absolute value of a moving average of past returns through the term⁸:

$$\kappa \int_{I_t} \mathcal{K}(t-u) \underbrace{\left| \int_{-\infty}^u K(u-s) \frac{dP_s}{P_s} \right|}_{X_u} du. \quad (4.16)$$

This entanglement of moving averages has several implications. First, it allows the introduction of inertia in the impact of past price trends on the volatility level coherent with the empirical time reversal asymmetry mentioned above. Second, it implies that the longer a price trend the higher 4.16. Another important element, on the contrary to the term which is linearly sensitive to a past price trend, θ is linearly sensitive to the absolute value of a past price trend. Therefore, a positive price trend such as $X_t > 0$ increases θ_t and by extension the attraction volatility. Consequently, a strong positive price trend produces two antagonistic effects on the volatility level: on the one hand it decreases the volatility component 4.13, on the other hand, it increases the component 4.16. Conversely, in a negative price trend circumstance, both component puts upwards pressure on the volatility process. This asymmetry in the impact of the price trend on the volatility process is consistent with the empirical relationship highlighted in chapter 2, which links the EWMA of past returns to the volatility levels observed in stock markets.

If $a = 2$, X corresponds to a weighted moving average of the realized variance. Therefore, X is no longer a price trend variable but a variable representing asset price activity, similar to m_2 in the

⁷For instance, if $\lambda_2 = 0$, the attraction volatility is

$$\bar{\sigma}_t = \frac{\kappa \left(\bar{\theta} + |X_t|^{\frac{1}{a}} \right)}{\kappa + \beta \lambda_1}.$$

⁸The volatility may be written as

$$\sigma_t = \beta_0 + \int_{I_t} \mathcal{K}(t-u) \left(\kappa (\bar{\theta} + |X_u|) - (\kappa + \beta \lambda_1) \sigma_u - \beta \lambda_2 (\sigma_u)^2 \right) du + \bar{\beta} \int_{I_t} \mathcal{K}(t-u) \sigma_u dZ_u.$$

general PDV model 4.1 introduced in chapter 3. The volatility process is thus linearly sensitive to

$$\kappa \int_0^t \mathcal{K}(t-u) \sqrt{\underbrace{\int_0^u K(u-s)(\sigma_s)^2 ds}_{X_u}} du. \quad (4.17)$$

Through this term, in a similar way to the case $a = 1$, the volatility process is subject to a force of inertia that depends on the duration of the volatility trend due to the fact that X is a moving average. Consequently, a short volatility spike does not increase the attraction volatility in the same proportion as a longer period of high volatility. Furthermore, this specification is particularly suited to capture time-reversal asymmetry. Indeed, even after a strong rebound following a negative trend causing a period of high volatility, the volatility process tends to remain higher than normal for some time due to the force of inertia induced by the term 4.17.

Thus, if the case $a = 1$ and $a = 2$ imply different hypotheses about the determinant of the attraction volatility, both allow make the RPDV model 4.2 strongly time-reversal asymmetry.

4.4 The function θ as a long-run coherence operator between price and value

A third way to apprehend θ is to link it to the long-term consistency between "fundamental" value and price. The idea is thus to model the hypothesis formulated by Black in 1986 ([36]) that, while prices may significantly diverge from their "fundamental value" in the short to medium term, there are "forces tending to cause prices to return to value" in the long term. This hypothesis aligns with the empirical work of Campbell and Shiller ([57], [58]), which shows that fundamental indicators - such as dividends per share or the price-to-earnings ratio in the stock market - tend to revert to a long-term level, thereby preventing a complete disconnection between price and fundamentals. θ will thus take the role of coherence operator between price and value, allowing to reconcile this empirical evidence while preserving the hypothesis on price drift as a quadratic function of volatility.

For this purpose, we assume, first, the existence and the uniqueness of attraction volatility denoted $\bar{\sigma}_t$, and second that the following conditions are met:

$$r_t + \lambda_1 \bar{\sigma}_t + \lambda_2 (\bar{\sigma}_t)^2 \begin{cases} \leq q_t & \text{for } V_t \leq P_t, \\ = q_t & \text{for } P_t = V_t, \\ \geq q_t & \text{for } V_t \geq P_t, \end{cases} \quad (4.18)$$

where V is the value, q its growth rate such as $q_t \geq r_t$ ⁹. Concretely, these conditions mean that the drift tends to revert toward: (1) a level superior or equal to the growth rate of the value when the price is inferior to the value, (2) the growth rate of the value when the price is equal to the value, (3) a level inferior or equal to the growth rate of the value when the price is superior to the value, and these adjustments pass by a variation of the level towards which the volatility tends.

A hypothesis consistent with 4.18 is to assume that

$$\lambda_1 \bar{\sigma}_t + \lambda_2 (\bar{\sigma}_t)^2 = F(V_t, P_t)(q_t - r_t), \quad (4.19)$$

where F is a function that respects $\forall (V, P) \in \mathbb{R}_+^2$, $F(V, P) > 0$ and such as

$$F(V, P) \begin{cases} \leq 1 & \text{for } V_t < P_t, \\ = 1 & \text{for } P_t = V_t, \\ \geq 1 & \text{for } V_t > P_t. \end{cases} \quad (4.20)$$

This supplementary condition ensures a positive risk price, an important property since market players are supposed to be risk-averse. Obviously, its definition remaining quite general, F admits a multiplicity of specifications.

The simplest case satisfying conditions 4.20 is $F(V, P) = 1$, $\forall (V, P) \in \mathbb{R}_+^2$. Within this framework, the attraction volatility does not depend on V_t and P_t . This property is interesting from a practical since it dispenses with determining the value V : only its growth rate q is required to specify both θ and the attraction volatility. In consequence, the differential between price and value has no impact on the drift of the asset price.

In all other specifications of F , the term $F(V, P)$ is a path-dependent component of the volatility process. Indeed, it implies necessarily that the right-hand term in equation 4.19 whom the volatility process depends is greater in case $V > P$ than in case $V < P$. The particularity of this path-dependence is that it is not purely endogenous, but linked with the joint past path of price and value. It implies the existence of a mechanism for an increase in the drift when $V_t > P_t$ and a decrease in it when $V_t < P_t$. Obviously, the introduction of such a mechanism must be done in accordance with the conditions previously defined, in particular the existence of attraction volatility (by definition strictly positive). For instance, a simple specification that respects conditions 4.20 is to define F as the ratio value-on-price, i.e.

$$F(V_t, P_t) = \frac{V_t}{P_t}. \quad (4.21)$$

⁹If the value follows an ordinary stochastic differential equation, the value dynamics are thus given by:

$$\frac{dV_t}{V_t} = q_t dt.$$

The impact of this specification on the dynamics of the volatility process differs significantly depending on the drift components ensuring that the equality 4.19 holds.

We will consider below two competing hypotheses. In the first one, risk premia are constant and equality 4.19 is guaranteed through a variation of the attraction volatility. Conversely, in the second hypothesis, the attraction volatility is constant and equation 4.19 is realized at each instant by risk premia adjustments.

4.4.1 The attraction volatility as an adjustment variable

Let us begin with the case where risk premia are constants, and the attraction volatility constitutes the adjustment variable allowing equation 4.19 to be respected. Again, the cases $\lambda_2 = 0$ and $\lambda_2 > 0$ needs to be considered separately.

If $\lambda_2 = 0$, as shown in section 4.2.1, it exists attraction volatility only if $-\kappa < \beta\lambda_1$. Under this condition, the attraction volatility is defined by:

$$\bar{\sigma}_t = \frac{F(V_t, P_t)(q_t - r_t)}{\lambda_1}, \quad (4.22)$$

and $\theta(t)$ (using equation 4.8), by

$$\theta_t = \frac{F(V_t, P_t)(q_t - r_t)(\kappa + \beta\lambda_1)}{\kappa\lambda_1}. \quad (4.23)$$

In this case, there is so a linear relationship with $F(V_t, P_t)(q_t - r_t)$ for both θ_t and the attraction volatility. Therefore, for a given pair (V, P) , the attraction volatility is an increasing linear function of the differential between the growth rate of value and the risk-free rate. Thus, for a constant free-risk rate of the value, a variation Δq of the growth rate of the value results in a variation $\Delta\bar{\sigma}$ of the attraction value. Symmetrically, for a constant growth rate of the value, a variation Δr of the risk-free rate results in a variation $-\Delta\bar{\sigma}$ of the attraction value. Of course, the independence between q and r is not required.

In case $\lambda_2 > 0$, the attraction volatility is defined by:

$$\bar{\sigma}_t = \frac{-\lambda_1 + \sqrt{(\lambda_1)^2 + 4F(V_t, P_t)(q_t - r_t)\lambda_2}}{2\lambda_2}, \quad (4.24)$$

and $\theta(t)$ is given (using equation 4.11) by¹⁰:

$$\theta_t = \frac{\beta F(V_t, P_t)(q_t - r_t)\lambda_2 + 0.5\kappa \left(\sqrt{(\lambda_1)^2 + 4F(V_t, P_t)(q_t - r_t)\lambda_2} - \lambda_1 \right)}{\kappa\lambda_2}. \quad (4.25)$$

¹⁰See appendix 4.B.

In addition, to ensure that $f_u(0) > 0$, the following inequality must be respected

$$F(V_t, P_t)(q_t - r_t) > \frac{\kappa}{\beta\lambda_2} \left(\lambda_1 + \frac{\kappa}{\beta} \right). \quad (4.26)$$

The hypothesis then makes then the attraction volatility a strictly concave increasing function of $F(V_t, P_t)(q_t - r_t)$. It follows that for a given pair (V, P) , a movement of the differential $q - r$ has more impact on the attraction volatility when $q - r$ is low. In the same way, a variation of $F(V_t, P_t)$ has more impact on the attraction volatility when volatility is low. Moreover, it should be emphasized that the condition 4.26 is a strong constraint when F is different from $F(V, P) = 1 \forall (V, P) \in \mathbb{R}_+^2$. For instance, the hypothesis of F as the value-on-price ratio is not compatible with $\lambda_2 > 0$ due to this condition, because in such case, the inequality 4.26 has a non-zero probability to be violated.

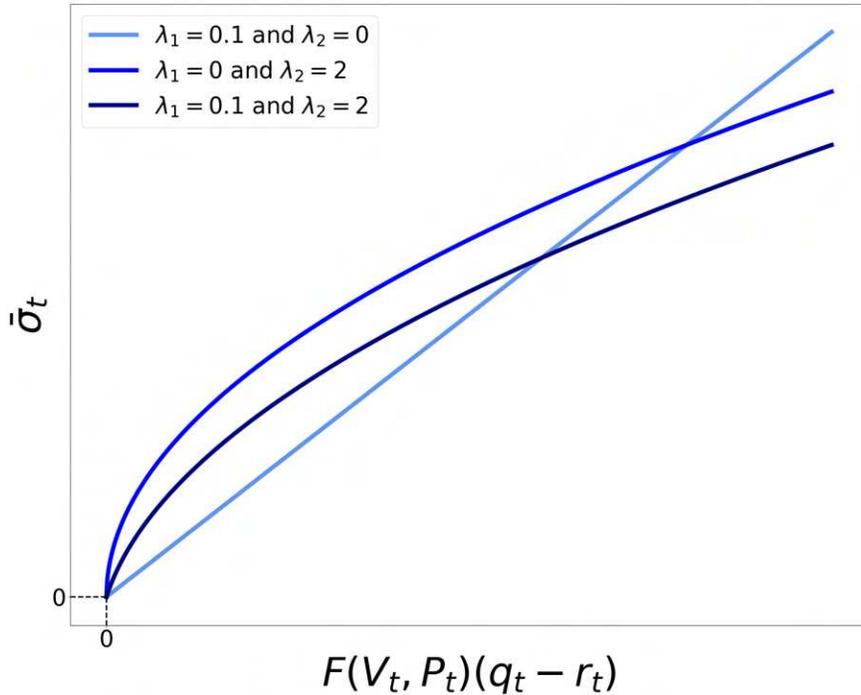


Figure 4.4: The relationship between the attraction volatility $\bar{\sigma}_t$ and $F(V_t, P_t)(q_t - r_t)$ under different hypothesis about risk premia values.

In both considered assumptions relating to risk premia, all things being equal, for all specifications of F different of $F(V, P) = 1 \forall (V, P) \in \mathbb{R}_+^2$ the attraction volatility is strictly higher when $V_t > P_t$ than in the situation $V_t < P_t$. The price adjustment to value in the long run pass then through variations of the attraction volatility that impact the volatility and thus, due to the positive relation between μ_t and σ_t , the asset price drift. For instance, if F corresponds to the value-to-price ratio (equation 4.21), the higher the price relative to the value, the lower the attraction volatility is. In this specification of F , a significant drop in price that is not due to a change in the fundamentals generates an increase in the attraction value as long as the value-to-price ratio has not returned to

its original level. This increase is higher in absolute terms the more that the value-to-price ratio is high. For example, under the hypothesis $\lambda_2 = 0$ and $\lambda_1 > 0$, an instant 20% price drop increases attraction volatility by 25%, while a 50% price drop increases it by 100%. Furthermore, as in the hypotheses considered in section 4.3, this mechanism joined to the negative relationship between past price trends and the volatility process makes the model structurally adapted to capture a strong reversal asymmetry.

4.4.2 The risk premium as an adjustment variable

An alternative hypothesis is that the attraction volatility is constant, with the risk premium and its component variables serving as the adjustment variables that ensure the satisfaction of equation 4.18. This reversal has strong theoretical implications. Indeed, under this hypothesis, the implied risk aversion of the market becomes endogenous. The price of the risk is no more the result of exogenous market participant preferences but a consequence on the one hand of the difference between the growth rate of the value and the risk-free rate, on the other hand of the long-run level of volatility. In other words, the price of the risk becomes a variable socially constituted by the interaction between agents that treat information related to fundamentals, and thus cannot be deduced by the aggregation of individual risk aversions.

To understand the practical implications of this reversal, let us start by considering the value taken by risk premia in this hypothesis and its implication on the volatility dynamics. To this end, we now assume that the risk premium takes the following form:

$$\lambda_t = \frac{\mu_t - r_t}{\sigma_t} = \psi_t \underbrace{(\lambda_1 + \lambda_2 \sigma_t)}_{\bar{\lambda}_t},$$

where $\bar{\lambda}_t$ is the component of the risk premium independent of V_t (and its dynamics) and ψ_t is the adjustment operator for the risk premium. Equation 4.19 may then be rewritten as:

$$\psi_t(\lambda_1 + \lambda_2 \bar{\sigma})\bar{\sigma} = F(V_t, P_t)(q_t - r_t).$$

It follows that the adjustment operator for the risk premium is given by:

$$\psi_t = \frac{F(V_t, P_t)(q_t - r_t)}{(\lambda_1 + \lambda_2 \bar{\sigma})\bar{\sigma}}.$$

Thus, in this framework, the price drift and the risk premium are defined by:

$$\mu_t = r_t + \underbrace{\frac{F(V_t, P_t)(q_t - r_t)}{(\lambda_1 + \lambda_2 \bar{\sigma})\bar{\sigma}}}_{\lambda_t} (\lambda_1 + \lambda_2 \sigma_t) \sigma_t.$$

Consequently, for a given volatility level, the risk premium is linearly dependent on the differential between the growth rate of value and the risk-free rate multiplied by $F(V_t, P_t)$. From an economic

perspective, this means that an increase (resp. decrease) of the differential between the growth rate of value and the risk-free rate increase (resp. decrease) the implied risk aversion. Similarly, if F is a strictly increasing function of the differential between price and value, an increase (resp. decrease) in the value/price ratio also increases (resp. decreases) the implicit risk aversion.

Besides, if θ were constant, an increase in the risk premium would result in a reduction of the attraction volatility (equations 4.8 and 4.11). However, as mentioned above, we here assume constant attraction volatility. Accordingly, using the results obtained in appendix 4.A, θ is defined as follows:

$$\theta_t = \frac{\bar{\sigma}}{\kappa} (\kappa + \beta\lambda_{1,t} + \bar{\sigma}\beta\lambda_{2,t}),$$

where $\lambda_{1,t} = \psi_t\lambda_1$ and $\lambda_{2,t} = \psi_t\lambda_2$. This definition ensures that the attraction volatility remains invariant by neutralizing the effects of variations in risk premia on the attraction volatility level. However, this invariance does not imply neutrality of the risk premium λ_t level on the volatility dynamics. Indeed, as already remarked in section 4.2, the function f may be written as

$$f_t(\sigma_t) = (\kappa + \beta\lambda_{1,t})(\bar{\sigma} - \sigma_t) + \beta\lambda_{2,t}(\bar{\sigma}^2 - (\sigma_t)^2),$$

and therefore, the reverting force that pushes the volatility process towards its attraction level is an increasing function of risk premia. Due to this relationship, all things being equal, the higher the risk premium λ (for a given value of w) the lower the dispersion of the volatility distribution. In this manner, an increase in the risk premium, which can be originated by an increase of the differential between the growth rate of value and the risk-free rate or an increase of the value-on-price ratio (or both), tends to reduce the randomness of the volatility.

4.5 Conclusion

This chapter explored the implications of various specifications of a PDV model that takes the form of a stochastic convolutive equation, under the assumption that the price drift follows a quadratic function of the volatility process.

First, it was shown that, in the considered framework, the combination of the path-dependence property of volatility with the assumption about the drift leads to an impact of the risk premium on the level to which the volatility tends to revert (the attraction volatility) and the strength of its mean reversion. More precisely, the attraction volatility is a decreasing function of the volatility and variance risk premia, while the speed of mean reversion increases with these premia. Moreover, the structure of the risk premium - that is, in the considered framework, the respective values of the volatility and variance risk premia - influences the type of mean reversion. In particular, the

presence of a strictly positive variance premium results in a stronger mean-reversion force toward the attraction volatility when the difference between the attraction volatility and the current volatility is negative, compared to when it is positive, for the same absolute differential. In contrast, if the variance premium is zero, the mean-reversion force is symmetric.

This chapter also explored competing hypotheses regarding the mechanisms of asset price and volatility formation by considering more specific forms of the general model.

Thus, after presenting the simpler specifications in which both the attraction volatility and the mean-reverting force are constant, we introduced a specification in which the attraction volatility level itself is a path-dependent process, leading to an entanglement of path dependencies. This specification reinforces the time-reversal asymmetry property of the process, which is an important empirical characteristic of financial data ([203], [204], [205], [39]). In this way, it addresses the shortcomings of models of the type considered in section 3.4 of chapter 3, in which the upward and downward movements in volatility are "slightly too symmetrical compared to empirical data".

The last type of specification considered in this chapter aimed to propose a formalization that reconciles the highly endogenous nature of price and volatility dynamics implied by the path-dependence property of volatility, with the hypothesis that, in the long term, the price tends to return to the vicinity of its fundamental value. Two competing hypotheses were then considered.

In the first, the adjustment between price and fundamental value occurs through a variation in the level of attraction volatility. In this configuration, the attraction volatility increases with both the differential between the growth rate of value and the risk-free rate, and the differential between price and value. Thus, changes in attraction volatility lead to changes in the volatility level which, in turn, through a risk premium effect (with the structure $\lambda_t = \lambda_1 + \lambda_2\sigma_t$ remaining unchanged), results in an adjustment of the drift level.

The second specification reverses the causal relationship, assuming that it is the risk premium that acts as the adjustment variable, preventing a total disconnection between price and fundamental value. In this specification, both the volatility premium and the variance premium increase with the differential between the growth rate of value and the risk-free rate, as well as the differential between price and value. Consequently, while the attraction volatility remains constant, changes in the structure of the risk premium create an increasing relationship between the mean-reversion force and these differentials.

Appendix 4.A Highlighting the mean-reverting component of the volatility process

Using expression of the attraction volatility 4.11, we have:

$$\begin{aligned}\theta &= \frac{(2\beta\lambda_2\bar{\sigma} + \kappa + \beta\lambda_1)^2 - (\kappa + \beta\lambda_1)^2}{4\beta\lambda_2\kappa} \\ \theta &= \frac{\beta\lambda_2\bar{\sigma}^2}{\kappa} + \frac{\bar{\sigma}(\kappa + \beta\lambda_1)}{\kappa} \\ \theta &= \frac{\bar{\sigma}}{\kappa}(\kappa + \beta\lambda_1 + \beta\lambda_2\bar{\sigma}).\end{aligned}$$

Replacing θ in equation 4.6, we then obtain

$$f(\sigma_t) = (\kappa + \beta\lambda_1)(\bar{\sigma} - \sigma_t) + \beta\lambda_2(\bar{\sigma}^2 - (\sigma_t)^2).$$

Appendix 4.B Zero of the function f in polynomial drift hypothesis

The unique positive solution to¹¹ $y_t + \lambda_1\bar{\sigma} + \lambda_2\bar{\sigma}^2 = 0$ is

$$\bar{\sigma}_t = \frac{-\lambda_1 + \sqrt{(\lambda_1)^2 + 4y_t\lambda_2}}{2\lambda_2}.$$

In addition, from equation 4.11 in section 4.2, we now that

$$\bar{\sigma}_t = \frac{\sqrt{(\kappa + \beta\lambda_1)^2 + 4\kappa\beta\lambda_2\theta_t} - \kappa - \beta\lambda_1}{2\beta\lambda_2}.$$

¹¹In the considered context set out in section 4.4,

$$y_t = F(V_t, P_t)(q_t - r_t).$$

By combining the two previous equalities, we have:

$$\begin{aligned}
\frac{\sqrt{(\kappa + \beta\lambda_1)^2 + 4\kappa\beta\lambda_2\theta_t} - \kappa - \beta\lambda_1}{2\beta\lambda_2} &= \frac{-\lambda_1 + \sqrt{(\lambda_1)^2 + 4y_t\lambda_2}}{2\lambda_2} \\
\sqrt{(\kappa + \beta\lambda_1)^2 + 4\kappa\beta\lambda_2\theta(t)} - \kappa - \beta\lambda_1 &= -\beta\lambda_1 + \beta\sqrt{(\lambda_1)^2 + 4y_t\lambda_2} \\
(\kappa + \beta\lambda_1)^2 + 4\kappa\beta\lambda_2\theta_t &= \left(\kappa + \beta\sqrt{(\lambda_1)^2 + 4y_t\lambda_2}\right)^2 \\
(\kappa + \beta\lambda_1)^2 + 4\kappa\beta\lambda_2\theta_t &= \kappa^2 + \beta^2 \left((\lambda_1)^2 + 4y_t\lambda_2\right) + 2\kappa\beta\sqrt{(\lambda_1)^2 + 4y_t\lambda_2} \\
\theta_t &= \frac{\beta y_t \lambda_2 + 0.5\kappa \left(\sqrt{(\lambda_1)^2 + 4y_t\lambda_2} - \lambda_1\right)}{\kappa\lambda_2}.
\end{aligned}$$

It is clear that if κ is positive, θ_t is also positive $\forall y_t > 0$ and therefore $\theta_t\kappa > 0$ which ensures that $f(\sigma_t) > 0$. If κ is negative, to ensure that $\theta(t)\kappa \geq 0$:

$$\begin{aligned}
\beta y_t \lambda_2 + 0.5\kappa \left(\sqrt{(\lambda_1)^2 + 4(q_t - r_t)\lambda_2} - \lambda_1\right) &> 0 \\
\left(\frac{\beta\lambda_2}{0.5\kappa} y_t - \lambda_1\right)^2 - (\lambda_1)^2 - 4\lambda_2 y_t &> 0 \\
\frac{4(\beta\lambda_2)^2}{\kappa^2} (y_t)^2 - 4\lambda_2 \left(\frac{\beta\lambda_1}{\kappa} + 1\right) y_t &> 0 \\
4\lambda_2 y_t \left(\frac{\beta^2\lambda_2}{\kappa^2} y_t - \left(\frac{\beta\lambda_1}{\kappa} + 1\right)\right) &> 0
\end{aligned}$$

Consequently,

$$y_t > \frac{\kappa}{\beta\lambda_2} \left(\lambda_1 + \frac{\kappa}{\beta}\right).$$

CHAPTER 5

Deep Estimation for Volatility Forecasting

Abstract

The use of deep neural networks (DNNs) for the calibration of volatility models applied to pricing and hedging issues has led to abundant academic literature. In contrast, few works utilize these tools for model estimation with a focus on volatility forecasting. Building on this observation, this chapter introduces an innovative deep estimation method using historical data, specifically designed for volatility forecasting. To illustrate the approach, we focus on estimating a rough path-dependent volatility (RPDV) model from the class introduced in chapter 3, which is particularly well-suited for forecasting but challenging to estimate with standard techniques. After formalizing the estimation problem within the framework of Bayesian decision theory, the chapter details the methodology for constructing the estimator function. Finally, a comprehensive evaluation of the estimation approach is conducted using both synthetic and market data to assess its performance.

5.1 Introduction

The use of machine learning methods in finance has experienced a significant boom in recent years. Among the most important use cases is the deep calibration of volatility models which consists in using artificial neural networks (NNs) to determine a set of parameters for a certain model in order to best meet given pricing or hedging objectives. If this topic resulted in abundant literature ([24], [21], [125], [181]), research on the use of NNs for model estimation in the context of volatility forecasting is much scarcer. However, this subject also represents an important issue within quantitative finance.

The potential interest in deep estimation within this framework is manifold. Firstly, the last decade has seen the emergence of numerous new volatility models that could serve as powerful tools for volatility forecasting. However, due to their complexity, some of these models are practically difficult

to estimate using standard approaches such as maximum likelihood-like methods. In addition, as decision theory demonstrates, the most probable set of parameters is not necessarily the optimal choice from a consequentialist perspective ([168], [30], [32]). In fact, the optimality of a parameter set depends not only on the parameter distribution but also on the model's intended use for estimation ([120]). For instance, the optimality criterion for the parameters of a model used in an options pricing perspective is typically defined by a fitting criterion, either of an option price map as in [125], or of an implied volatility surface as in the article by Rosenbaum and Zhang ([181]). Analogously, the optimality criterion used for parameter estimation in the context of volatility forecasting must reflect this objective to be fully consistent. However, the associated optimization problem can be very difficult, if not impossible, to solve analytically.

Accordingly, the present chapter aims to introduce a deep estimation method for volatility models based on historical data, grounded in the theoretical framework of Bayesian decision theory (BDT). The core principle of this method involves estimating a volatility model through the interaction of two neural networks (NNs): the first NN associates a historical dataset with a vector of parameters and state variables of the considered model, while the second NN uses this vector to estimate different moments of the associated volatility. This second NN compensates for the lack of analytical formulas for the volatility moments of the model, playing a role similar to the NN pricing map approximator proposed by Horvath *et al.* ([125]). The volatility model under consideration for estimation is a version of the rough path-dependent volatility (RPDV) model, which belongs to the class of models introduced in chapter 3. It is a good candidate for the proposed approach since, on the one hand, it provides a framework for capturing the main empirical features that characterize volatility dynamics, making it a potentially suitable model for volatility forecasting, and on the other hand, it is very complex to estimate using traditional approaches. Obviously, although this chapter focuses on this model, the general principle of the proposed estimation method can be applied to other volatility models as well.

The chapter is organized as follows. Section 5.2 provides the definition of the RPDV model intended for estimation, along with an explanation of its role in forecasting. This leads to the formalization of the estimation problem as an optimization issue within the BDT framework. In section 5.3, a method for constructing an estimator function is presented, aiming to address this problem by utilizing two deep neural networks within a collaborative game framework. Lastly, in section 5.4, a comprehensive evaluation of the resulting estimator function is conducted from various perspectives, using both synthetic and market data.

5.2 Exposition of the estimation problem

5.2.1 The model to be estimated

5.2.1.1 The considered rough path-dependent volatility model

In the present chapter, we aim to estimate the following RPDV model:

$$\left\{ \begin{array}{l} \frac{dP_t}{P_t} = (\lambda_1 \sigma_t + \lambda_2 (\sigma_t)^2) dt + \sigma_t dB_t, \\ \sigma_t = \beta_0 + \beta_1 R_{1,t} + \beta_2 \sqrt{R_{2,t}}, \\ R_{1,t} = \int_{-\infty}^{t-\epsilon} (t-u)^{-\alpha_1} \left(\frac{dP_u}{P_u} - \kappa_1 \cdot R_{1,u} du \right), \\ R_{2,t} = \int_{-\infty}^{t-\epsilon} (t-u)^{-\alpha_2} (\sigma_u^2 - \kappa_2 \cdot R_{2,u}) du \end{array} \right. . \quad (5.1)$$

Here, the asset price thus depends on λ_1 and λ_2 , which are positive risk premia, B is a Brownian motion that constitutes the unique source of randomness, and σ is the volatility process. This volatility process is a multilinear function with β_0 a positive constant, $\beta_1 \leq 0$ a sensitivity parameter to R_1 that can be viewed as an asset price trend variable, and $\beta_2 \geq 0$ a sensitivity parameter to R_2 that can be viewed as a variable measuring recent market price activity regardless of the sign of the trend. ϵ is a positive parameter close to zero that encodes a latency of the impact of price dynamics on the volatility process. Technically, this parameter allows for values of α_j greater than 0.5 to be given without causing divergence issues. Furthermore, the memory of the processes R_1 and R_2 depends on the respective positive parameters α_1, κ_1 and α_2, κ_2 .

This model has several remarkable properties that make it highly suitable for volatility forecasting issues. First, it is structurally adapted to jointly capture two important empirical features, which are the rough behavior and the path-dependence of the volatility process. The rough volatility dynamics are determined by the rough kernels $K_1(\tau) = \tau^{-\alpha_1}$ and $K_2(\tau) = \tau^{-\alpha_2}$, while the model incorporates path-dependency through the processes R_1 and R_2 . Additionally, this version of the RPDV model shares a similar structure with the PDV model introduced by Guyon and Lekeufack ([117]), which has demonstrated strong predictive capabilities in volatility forecasting.

5.2.1.2 The Markovian approximation of the model and its discretization scheme

Like other rough volatility models, the RPDV model is non-Markovian, which makes it difficult to simulate efficiently. However, as shown in chapter 3, we can approximate model 5.1 by the following Markovian model (see appendix 5.A.1), which will be referred to as the M-RPDV model.

This model substitutes rough kernels $t^{-\alpha_j}$ with kernels of the form $\tilde{K}_j(\tau) = \sum_{i=1}^n w_{j,i} \gamma_{j,i} e^{-\gamma_{j,i} \tau}$:

$$\left\{ \begin{array}{l} \frac{dP_t}{P_t} = (\lambda_1 \sigma_t + \lambda_2 (\sigma_t)^2) dt + \sigma_t dB_t, \\ \sigma_t = \beta_0 + \beta_1 R_{1,t} + \beta_2 \sqrt{R_{2,t}}, \\ dR_{1,t} = \Gamma_1 \cdot \left(\frac{dP_t}{P_t} - \kappa_1 R_{1,t} dt \right) - \Gamma_1 \odot R_{1,t} dt, \\ dR_{2,t} = \Gamma_2 \cdot (\sigma_t^2 - \kappa_2 R_{2,t}) dt - \Gamma_2 \odot R_{2,t} dt, \\ R_{1,t} = W_1^\top R_{1,t}, \\ R_{2,t} = W_2^\top R_{2,t}, \end{array} \right. \quad (5.2)$$

where W_j the vector of weights $(w_{j,i})_{1 \leq i \leq n}$ and Γ_j the vector of discount coefficients $(\gamma_{j,i})_{1 \leq i \leq n}$, such that

$$W_j = \begin{bmatrix} w_{j,1} \\ \dots \\ w_{j,n} \end{bmatrix}, \quad \Gamma_j = \begin{bmatrix} \gamma_{j,1} \\ \dots \\ \gamma_{j,n} \end{bmatrix}.$$

The method used to obtain these vectors is presented in appendix 5.A.2. It should be noted that model 5.2 depends on the parameter vector

$$\phi = (\lambda_1, \lambda_2, \beta_0, \beta_1, \beta_2, \alpha_1, \alpha_2, \kappa_1, \kappa_2),$$

and that all relevant information at time T for the volatility dynamics is aggregated into the following vector of state variables:

$$R_T = \left(R_{1,T}^{(1)}, \dots, R_{1,T}^{(n)}, R_{2,T}^{(1)}, \dots, R_{2,T}^{(n)} \right).$$

Consequently, the estimation procedure will consist of estimating the $2n + 9$ vector

$$\theta_T = \left(\lambda_1, \lambda_2, \beta_0, \beta_1, \beta_2, \alpha_1, \alpha_2, \kappa_1, \kappa_2, R_{1,T}^{(1)}, \dots, R_{1,T}^{(n)}, R_{2,T}^{(1)}, \dots, R_{2,T}^{(n)} \right), \quad (5.3)$$

with T a given period. This vector is therefore composed of 9 parameters and $2n$ state variables. To perform simulations from the model required by the estimation procedure, we use the following

explicit Euler discretization scheme:

$$\left\{ \begin{array}{l} P_{t+\Delta t} = P_t \left(1 + (\lambda_1 \sigma_t + \lambda_2 (\sigma_t)^2) \Delta t + \sigma_t (B_{t+\Delta t} - B_t) \right), \\ R_{1,t+\Delta t} = R_{1,t} \odot (\mathbf{1}_n - \Gamma_1 \cdot \Delta t) + \Gamma_1 \cdot \left(\frac{P_{t+\Delta t} - P_t}{P_t} - \kappa_1 R_{1,t} \Delta t \right), \\ R_{2,t+\Delta t} = R_{2,t} \odot (\mathbf{1}_n - \Gamma_2 \cdot \Delta t) + \Gamma_2 \cdot (\sigma_t^2 - R_{2,t}) \Delta t, \\ R_{1,t+\Delta t} = W_1^\top R_{1,t+\Delta t}, \\ R_{2,t+\Delta t} = W_1^\top R_{2,t+\Delta t} \\ \sigma_{t+\Delta t} = \beta_0 + \beta_1 R_{1,t+\Delta t} + \beta_2 \sqrt{R_{2,t}}, \end{array} \right. \quad (5.4)$$

with Δt being the time step of simulations and $(B_{t+\Delta t} - B_t) \sim \mathcal{N}(0, \Delta t)$. It is important to note that in order to ensure the stability of the scheme, all coordinates of Γ_j must be lower than $\frac{1}{\Delta t}$. If one wishes to eliminate this condition, an alternative is to opt for an implicit-explicit scheme analogous to the scheme proposed by Rosenbaum and Zhang ([181]) for the quadratic rough Heston model. The time step Δt used in this chapter is $\frac{1}{19656}$ ¹ year, and the larger discount factor is equal to 10000 (expressed in years). Therefore, because $\forall i, j, \gamma_{j,i} \Delta t < 1$, this stability issue does not arise.

5.2.2 The Bayesian estimation problem to solve: a forecasting objective-based estimation problem

5.2.2.1 The forecasting issue

The estimation method presented in this chapter for model 5.4 is specifically designed to address a particular forecasting problem. More precisely, we place in a context in which we have a data matrix D of the form

$$D = \begin{pmatrix} P_{t_1} & \tilde{\sigma}_{t_1} \\ \dots & \dots \\ P_{t_N} & \tilde{\sigma}_{t_N} \end{pmatrix}, \quad (5.5)$$

where $t_1 < \dots < t_N = T$, P represents the price of a financial asset, and $\tilde{\sigma}$ is a proxy of its realized volatility defined as the square root of the sum of squares of a sample of 78 observations of logarithmic returns over the considered period². From this $N \times 2$ data matrix, we want to get

¹The reason for choosing this discretization time step is that a trading day is approximately equal to $\frac{1}{252}$ of a year, and the realized volatility estimator used in this chapter is calculated using 78 price observations per trading day: $\frac{1}{252} \times \frac{1}{78} = \frac{1}{19656}$.

²Regarding the simulated data, this proxy is calculated from 78 log-returns evenly distributed over a period of $\frac{1}{252}$ year. For the real data used in section 5.4.2, $\tilde{\sigma}$ is calculated from a sample of 78 5-minute log-returns.

an estimator as accurate as possible of the following set of conditional moment vectors:

$$\Omega_M = \left\{ \left(\mathbb{E}[\sigma_{T+\delta_k} | D], \text{Std}[\sigma_{T+\delta_k} | D] \right) \right\}_{k=1}^P, \quad (5.6)$$

where $\delta_{k=1 \leq k \leq P}$ represents different time horizons. In this chapter, we will consider the horizons of 1, 5, 21, 42, and 63 trading days, which are defined here as $\frac{1}{252}$ year.

This objective, based on the first two moments of volatility, makes sense in practice for any goal of predicting the volatility distribution over the considered horizons, such as implementing volatility arbitrage strategies. Indeed, on the one hand, log-volatility increments closely follow a Gaussian distribution ([104]), and empirical volatility distributions closely resemble log-normal distributions ([162], [193]); on the other hand, the conditional volatility distributions generated by the RPDV model also exhibit similar log-normal behavior. Consequently, the mean and standard deviation of volatility almost fully capture the information about its distribution, both in an empirical context and within the RPDV model framework.

The RPDV model will therefore serve as a tool to estimate these moments. Consequently, contrary to standard statistical approaches like maximum likelihood estimation, the estimation procedure will not consist of determining the most likely vector θ_T , but rather the vector $\hat{\theta}_T$ that serves this forecasting goal the best.

5.2.2.2 The Bayesian estimation problem

To propose an appropriate estimation method for the RPDV model that aligns with the forecasting objective defined in section 6.2.2.1, we adopt the theoretical framework of Bayesian decision theory ([30], [32]). As a result, we assume that the dynamics of (P, σ) follow a model given by 5.4, and θ_t is considered as a random vector with a prior distribution π (i.e., $\theta_t \sim \pi$), $\forall t$. Under these assumptions and following the principles of BDT, an estimator $\hat{\theta}_T$ of θ_T is optimal given D and a loss function L if it minimizes the expected posterior loss defined as follows:

$$\mathbb{E}_{\pi_D} \left[L(\theta_T, \hat{\theta}_T) \right] = \int_{\mathbb{R}^{2n+9}} L(\theta_T, \hat{\theta}_T) d\pi_D(\theta_T), \quad (5.7)$$

where π_D represents the posterior distribution for θ_T given D^3 . Regarding the loss function L , its purpose is to capture the objective of the estimation, which is to obtain an estimator of the conditional moments in Ω_M . This function L is defined as follows:

$$L(\theta_T, \hat{\theta}_T) = \sum_{k=1}^P c_k \cdot C \left(M(\theta_T, \delta_k), M(\hat{\theta}_T, \delta_k) \right), \quad (5.8)$$

³In practice, updating π with the information contained in D (i.e., determining π_D) is not a trivial task. The estimation procedure presented in section 5.3 does not require directly computing this posterior measure.

where $\{c_k\}_{k=1}^p$ are positive weights, C is another loss function, and M is a function defined as:

$$M(\theta_T, \delta_k) = \left(\mathbb{E}[\sigma_{T+\delta_k} | \theta_T], \text{Std}[\sigma_{T+\delta_k} | \theta_T] \right), \quad \forall \boldsymbol{\pi}(\theta_T) \neq 0, \delta_k \in \mathbb{R}_+, \quad (5.9)$$

where $\mathbb{E}[\sigma_{T+\delta_k} | \theta_T]$ and $\text{Std}[\sigma_{T+\delta_k} | \theta_T]$ represent the conditional mean and standard deviation of volatility at horizon δ_k given θ_T . In other words, the cost associated with an estimator $\hat{\theta}_T$ given the true θ -vector θ_T is a function of the prediction error in the mean and standard deviation of volatility for time horizons $T + \delta_1, \dots, T + \delta_p$ induced by this choice of θ -estimator. This cost is influenced by the form of C , which will be specified in section 5.2.3. Irrespective of the specific form of C and within the previously established framework, the Bayes estimator of θ_T under the posterior measure $\boldsymbol{\pi}_D$ is a solution to the following optimization program:

$$\arg \min_{\hat{\theta}_T \in \mathbb{R}^{2n+9}} \sum_{k=1}^p c_k \cdot \mathbb{E}_{\boldsymbol{\pi}_D} \left[C \left(M(\theta_T, \delta_k), M(\hat{\theta}_T, \delta_k) \right) \right]. \quad (5.10)$$

The objective of the estimation method introduced in this chapter, which will be presented in section 5.3, is to find an approximate solution to this optimization problem.

5.2.3 The loss function: a sum of proxy divergence measures

As mentioned in section 5.2.2, the choice of the loss function is crucial as it implicitly encodes preferences regarding estimation errors. The mean squared error (MSE) is commonly used as a loss function in forecasting problems due to its simplicity. However, although the MSE has certain advantages, it may not be the most suitable loss function for the forecasting objective. In this case, using the MSE would give excessive weight to situations where the expected volatility and volatility of volatility are high compared to cases where these quantities are low. Therefore, we will employ an ad-hoc loss function that can be interpreted as a sum of proxy divergence measures.

To do so, we use the fact that we can approximate the distribution of $\sigma_{T+\delta}$ by a log-normal distribution - as mentioned in section 6.2.2.1 - with parameters $m(\theta_{T+\delta})$ and $s(\theta_{T+\delta})$. Leveraging the analytical expressions for the expectation and variance of the log-normal distribution, we can express them as follows:

$$\mathbb{E}[\sigma_{T+\delta} | \theta_T] \approx e^{m(\theta_{T+\delta}) + s(\theta_{T+\delta})^2 / 2}, \quad \text{Var}[\sigma_{T+\delta} | \theta_T] \approx \left(e^{s(\theta_{T+\delta})^2} - 1 \right) e^{2m(\theta_{T+\delta}) + s(\theta_{T+\delta})^2}.$$

Equivalently, we can write (see details in appendix 5.B):

$$\begin{aligned} \tilde{m}(\theta_{T+\delta_k}) &= \log \left(\mathbb{E}[\sigma_{T+\delta} | \theta_T] \right) - 0.5 \log \left(\frac{\text{Var}[\sigma_{T+\delta} | \theta_T]}{\mathbb{E}[\sigma_{T+\delta} | \theta_T]^2} + 1 \right), \\ \tilde{s}(\theta_{T+\delta_k})^2 &= \log \left(\frac{\text{Var}[\sigma_{T+\delta} | \theta_T]}{\mathbb{E}[\sigma_{T+\delta} | \theta_T]^2} + 1 \right), \end{aligned} \quad (5.11)$$

where $\tilde{m}(\theta_{T+\delta_k})$ and $\tilde{s}(\theta_{T+\delta_k})^2$ are approximations of $m(\theta_{T+\delta})$ and $s(\theta_{T+\delta})$, respectively. Furthermore, the divergence between two log-normal distributions, $\mathcal{LN}(m_1, (s_1)^2)$ and $\mathcal{LN}(m_2, (s_2)^2)$, can be expressed as an analytical function of m_1 , s_1 , m_2 , and s_2 . Specifically, for the case of Kullback-Leibler (KL) divergence, we have ([107]):

$$\mathcal{D}_{KL}(\mathbb{P}_1, \mathbb{P}_2) = \frac{(m_1 - m_2)^2 + (s_1)^2 - (s_2)^2}{2(s_2)^2} + \log\left(\frac{s_2}{s_1}\right),$$

where $\mathbb{P}_1 = \mathcal{LN}(m_1, (s_1)^2)$ and $\mathbb{P}_2 = \mathcal{LN}(m_2, (s_2)^2)$.

Considering these elements, we specify the loss function as follows⁴:

$$L(\theta_T, \hat{\theta}_T) = \sum_{k=1}^p \underbrace{\left(\frac{(\tilde{m}(\hat{\theta}_{T+\delta_k}) - \tilde{m}(\theta_{T+\delta_k}))^2 + \tilde{s}(\hat{\theta}_{T+\delta_k})^2 - \tilde{s}(\theta_{T+\delta_k})^2}{2\tilde{s}(\theta_{T+\delta_k})^2} + \log\left(\frac{\tilde{s}(\theta_{T+\delta_k})}{\tilde{s}(\hat{\theta}_{T+\delta_k})}\right) \right)}_{\mathcal{D}_{KL}(\hat{\mathbb{P}}_i, \mathbb{P}_i)},$$

with $\mathbb{P}_i = \mathcal{LN}(\tilde{m}(\theta_{T+\delta_k}), \tilde{s}(\theta_{T+\delta_k})^2)$ and $\hat{\mathbb{P}}_i = \mathcal{LN}(\tilde{m}(\hat{\theta}_{T+\delta_k}), \tilde{s}(\hat{\theta}_{T+\delta_k})^2)$.

Therefore, the loss function L can be understood as the summation of estimated KL divergences between the predicted volatility distribution and the true volatility distribution at various time horizons. It quantifies the discrepancy between these distributions. It is worth noting that although the log-normal distribution is an approximation of the distribution of $\sigma_{T+\delta}$ given θ_T , the KL divergence $\mathcal{D}_{KL}(\hat{\mathbb{P}}_i, \mathbb{P}_i)$ achieves its minimum value (which is 0) when $\mathbb{E}[\sigma_{T+\delta}|\theta_T] = \mathbb{E}[\sigma_{T+\delta}|\hat{\theta}_T]$ and $\text{Var}[\sigma_{T+\delta}|\theta_T] = \text{Var}[\sigma_{T+\delta}|\hat{\theta}_T]$.

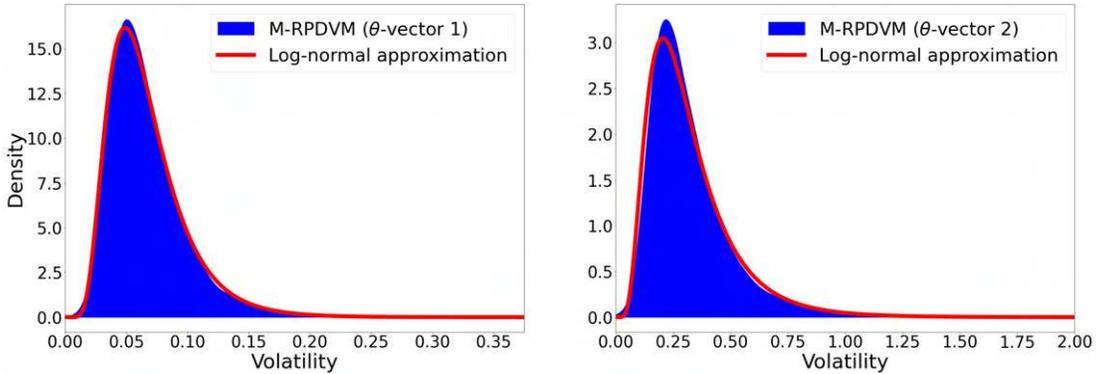


Figure 5.1: Example of two volatility distributions at a one-week horizon generated from the M-RPDV model using two different θ -vectors and their respective log-normal approximations. The KL-divergence between the two associated distributions is equal to 4.66, while the KL-divergence between the log-normal approximations of these distributions is equal to 3.92.

⁴Note that L is a loss function of the form 5.8.

5.3 Construction of the Bayesian estimator function

Section 5.2.2 outlined the estimation problem we seek to solve, defined by the optimization program 5.10. However, two main obstacles need to be overcome: firstly, an analytical formula for the function M is not available, and secondly, updating the prior distribution $\boldsymbol{\pi}$ to obtain the posterior distribution $\boldsymbol{\pi}_D$ is a highly complex task. To address the first issue, we will adopt a strategy similar to that of Horvath *et al.* ([125]), which compensates for the lack of an analytical formula for the option price by using "a neural network that maps parameters of a stochastic model to pricing functions" in their calibration process. In a comparable fashion, we introduce in section 5.3.1 an estimator for the function M in the form of an NN that maps (θ_T, δ) to the conditional mean and standard deviation $\mathbb{E}[\sigma_{T+\delta}|\theta_T]$ and $\text{Std}[\sigma_{T+\delta}|\theta_T]$. We will use this proxy of M to calibrate a second NN that will play the role of a Bayesian estimator function. Therefore, the objective of this second NN, the architecture of which will be detailed in section 5.3.2, is to provide, for any D , a proxy for the Bayesian estimator of θ_T , i.e., an approximate solution to the optimization program 5.10. Section 5.3.3 will detail an estimation procedure followed to achieve this situation, a method that circumvents the challenge of estimating the posterior distribution $\boldsymbol{\pi}_D$ directly by indirectly addressing the original problem, solving a related problem under the prior measure $\boldsymbol{\pi}$.

5.3.1 The neural network as a proxy for function M

The initial stage in constructing the estimator function involves developing an estimator for the function M , denoted as \mathcal{M} . This estimator maps (θ, δ) to the conditional moments $\mathbb{E}[\sigma_{T+\delta}|\theta_T]$ and $\text{Std}[\sigma_{T+\delta}|\theta_T]$. The objective is for \mathcal{M} to approximate M for all $\theta : \boldsymbol{\pi}(\theta_T) \neq 0$ and $\delta \in \delta_1, \dots, \delta_p$, as expressed by the approximation

$$\mathcal{M}(\theta_T, \delta) \approx M(\theta_T, \delta) \quad \text{and} \quad \nabla \mathcal{M}(\theta_T, \delta) \approx \nabla M(\theta_T, \delta). \quad (5.12)$$

The objective is for \mathcal{M} not only to be a good approximation of the function M , but also for the gradient of \mathcal{M} to be approximately equal to the gradient of M for all $\theta : \boldsymbol{\pi}(\theta_T) \neq 0$ and $\delta \in \delta_1, \dots, \delta_p$. This property ($\nabla \mathcal{M}(\theta_T, \delta) \approx \nabla M(\theta_T, \delta)$) is crucial in the role that \mathcal{M} will play in learning the estimator of the θ -vector. To achieve this, \mathcal{M} will be implemented as a neural network (NN) with a specialized and tailored architecture designed for this task. In this section, we will provide a detailed description of the adopted network architecture.

5.3.1.1 General structure of the network

As mentioned earlier, the architecture of \mathcal{M} is specifically adapted to ensure that the gradient passed to the estimator neural network contains informative information about the implications of the chosen θ -vector in moment predictions. This choice is motivated by the observation that a more standard network structure often results in a fitted NN that is primarily sensitive to state vectors R_1 and R_2 , thus missing the main purpose of \mathcal{M} . In order to address this issue and satisfy the requirements described in equations 5.12, \mathcal{M} adopts an architecture schematized in figure 5.2.

Firstly, the neural network (NN) consists of 2 distinct input layers: the first layer receives the θ -vectors and thus has $9 + 2n$ input neurons, while the second layer has a dimension of 1 to specify the temporal horizon δ for which the conditional moments are to be computed. These 2 input layers feed several subregions of the network, which can be segmented into 2 main components. The first component is responsible for estimating $\mathbb{E}[\sigma_{T+\delta}|\theta_T]$, while the second component is tasked with estimating $\text{Std}[\sigma_{T+\delta}|\theta_T]$.

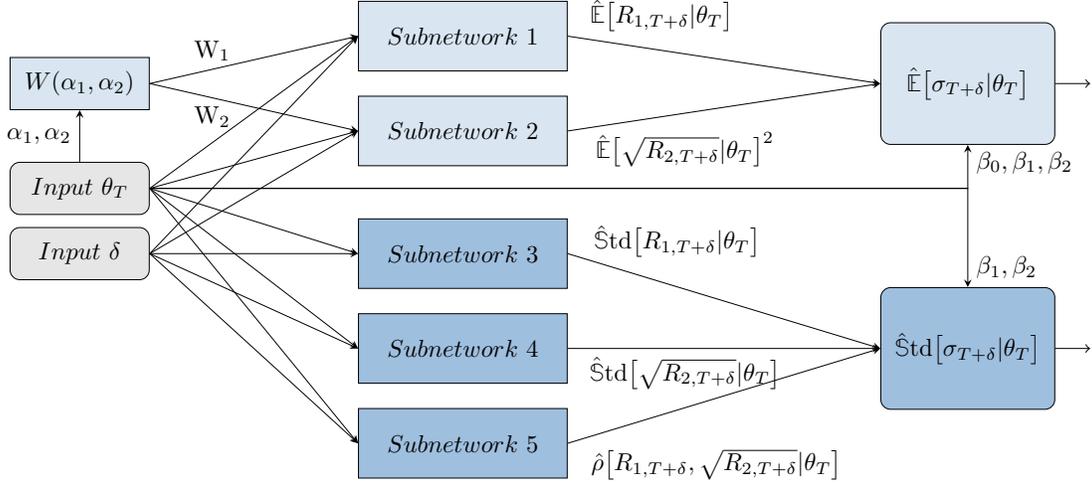


Figure 5.2: The architecture of the neural network \mathcal{M} .

5.3.1.2 The part of the network responsible for estimating $\mathbb{E}[\sigma_{T+\delta}|\theta_T]$

The part of the network responsible for predicting $\mathbb{E}[\sigma_{T+\delta}|\theta_T]$ utilizes the fact that this expectation can be expressed as follows:

$$\mathbb{E}[\sigma_{T+\delta}|\theta_T] = \beta_0 + \beta_1 \mathbb{E}[R_{1,T+\delta}|\theta_T] + \beta_2 \mathbb{E}[\sqrt{R_{2,T+\delta}}|\theta_T] \quad (5.13)$$

The approach is to estimate $\mathbb{E}[R_{1,T+\delta}|\theta_T]$ and $\mathbb{E}[\sqrt{R_{2,T+\delta}}|\theta_T]$ (more precisely, $\mathbb{E}[\sqrt{R_{2,T+\delta}}|\theta_T]^2$) separately using two parallel subnetworks: subnetworks 1 and 2 as depicted in figure 5.2.

Equation 5.13 is then used to calculate an estimation of $\mathbb{E}[\sigma_{T+\delta}|\theta_T]$. These subnetworks each consist of 6 hidden layers, with the first 5 layers containing 100 ReLU neurons each, and the final layer consisting of 20 linear neurons. The last hidden layer connects to an output layer with a single neuron (each subnetwork has its own output neuron/layer), which receives input from the two input layers associated with θ_T and δ , as well as from a function that computes the vectors W_1 and W_2 from θ_T .

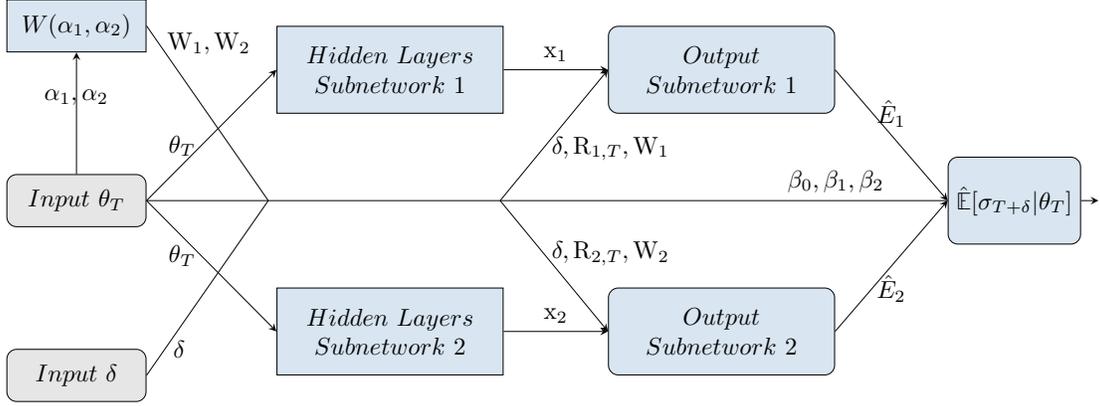


Figure 5.3: The part of the network \mathcal{M} responsible for estimating $\mathbb{E}[\sigma_{T+\delta}|\theta_T]$.

The activation function associated with this output neuron is of the form:

$$\hat{E}_j(\delta, R_{j,T}, W_j, x_j) = \underbrace{\sum_{i=1}^n w_{j,i} e^{-\gamma_i \delta} R_{j,i,T}}_{(1)} + \underbrace{\sum_{k=1}^{20} (1 - e^{-g_k \delta}) x_{j,k}}_{(2)}, \quad (5.14)$$

where x_j is the output vector of the last hidden layer associated with subnetwork j . The exponential weights $(g_k)_{1 \leq k \leq 20}$ are defined as follows:

$$g_k = \exp \left(\log \left(\frac{1}{1000} \right) + \frac{\log(100) - \log \left(\frac{1}{1000} \right)}{20 - 1} (k - 1) \right)^{-1}. \quad (5.15)$$

The specific form of this activation function is particularly suited to its objective in view of the analytical expressions of $\mathbb{E}[R_{1,T+\delta}|\theta_T]$ and $\mathbb{E}[\sqrt{R_{2,T+\delta}}|\theta_T]$. First of all, when $\delta = 0$,

$$\hat{E}_1(0, R_{1,t}, W_1, x_1) = R_{1,t} \quad \text{and} \quad \hat{E}_2(0, R_{2,t}, W_2, x_2) = R_{2,t}.$$

Thus, vectors x_j have no impact on the calculation since we use the analytical formulas for $R_{1,t}$ and $R_{2,t}$ which are θ -measurable. On the other hand, when $\delta = \infty$, the term (1) in equation 5.13 becomes zero, and we have

$$\hat{E}_1(\infty, R_{1,t}, W_1, x_1) = \sum_{k=1}^{20} x_{1,k} \quad \text{and} \quad \hat{E}_2(\infty, R_{2,t}, W_2, x_2) = \sum_{k=1}^{20} x_{2,k},$$

which can be respectively interpreted as the estimated asymptotic value of the expectation of R_1 and the squared estimated asymptotic value of the expectation of $\sqrt{R_2}$. Besides these polar cases, component (2) aims to estimate clearly identified variables. For the network responsible for

estimating $\mathbb{E}[R_{1,T+\delta}|\theta_T]$, the objective is that for all δ :

$$\begin{aligned} \sum_{k=1}^{20} (1 - e^{-g_k\delta}) x_{1,k} &\approx \mathbb{E}[R_{1,T+\delta}|\theta_T] - \sum_{i=1}^n \gamma_{1,i} w_{1,i} e^{-\gamma_i\delta} R_{1,i,T} \\ &= \mathbb{E} \left[\int_T^{T+\delta} \sum_{i=1}^n \gamma_{1,i} w_{1,i} e^{-\gamma_{1,i}(T+\delta-u)} \left(\frac{dP_u}{P_u} - \kappa_1 R_{1,i,u} du \right) \middle| \theta_T \right] \\ &= \mathbb{E} \left[\int_T^{T+\delta} \hat{K}_1(T+\delta-u) \left(\frac{dP_u}{P_u} - \kappa_1 R_{1,u} du \right) \middle| \theta_T \right]. \end{aligned}$$

For the network responsible for estimating $\mathbb{E}[\sqrt{R_{2,T+\delta}}|\theta_T]$, the objective is that for all δ^5 :

$$\begin{aligned} \sum_{k=1}^{20} (1 - e^{-g_k\delta}) x_{2,k} &\approx \mathbb{E}[R_{2,T+\delta}|\theta] - \text{Var}[\sqrt{R_{2,T+\delta}}|\theta] - \sum_{i=1}^n \gamma_{2,i} w_{2,i} e^{-\gamma_i\delta} R_{2,i,T} \\ &= \mathbb{E} \left[\int_T^{T+\delta} \sum_{i=1}^n \gamma_{2,i} w_{2,i} e^{-\gamma_{2,i}(T+\delta-u)} \left(\sigma_u^2 du - \kappa_2 R_{2,i,u} du \right) \middle| \theta_T \right] - \text{Var}[\sqrt{R_{2,T+\delta}}|\theta_T]. \\ &= \mathbb{E} \left[\int_T^{T+\delta} \hat{K}_2(T+\delta-u) \left(\sigma_u^2 du - \kappa_2 R_{2,i,u} du \right) \middle| \theta_T \right] - \text{Var}[\sqrt{R_{2,T+\delta}}|\theta_T]. \end{aligned}$$

These output layers of the two sub-networks thus produce an estimation of $\mathbb{E}[R_{1,T+\delta}|\theta_T]$ and $\mathbb{E}[\sqrt{R_{2,T+\delta}}|\theta_T]$, respectively, which feed into a global output neuron of the network \mathcal{M} : the neuron whose output is the estimator $\hat{\mathbb{E}}[\sigma_{T+\delta}|\theta_T]$. This neuron, also fed by the input layer associated with θ_T , then computes the estimator of $\mathbb{E}[\sigma_{T+\delta}|\theta_T]$ using the analytical formula 5.13:

$$\hat{\mathbb{E}}[\sigma_{t+\delta t}|\theta] = \beta_0 + \beta_1 \hat{E}_1 + \beta_2 \sqrt{(\hat{E}_2)_+}, \quad (5.16)$$

with \hat{E}_1 the output of the sub-network 1 and \hat{E}_2 the output of the sub-network 2.

5.3.1.3 The part of the network responsible for estimating $\text{Std}[\sigma_{T+\delta}|\theta_T]$

The part of the network responsible for predicting $\text{Std}[R_{1,T+\delta}|\theta_T]$, $\text{Std}[\sigma_{T+\delta}|\theta_T]$ uses the fact that the variance is equal to (see appendix 5.D.3):

$$\begin{aligned} \text{Var}[\sigma_{T+\delta}|\theta_T] &= (\beta_1 \text{Std}[R_{1,T+\delta}|\theta_T])^2 + (\beta_2 \text{Std}[\sqrt{R_{2,T+\delta}}|\theta_T])^2 \\ &\quad + 2\beta_1\beta_2 \text{Std}[R_{1,T+\delta}|\theta_T] \text{Std}[\sqrt{R_{2,T+\delta}}|\theta_T] \rho[R_{1,T+\delta}, \sqrt{R_{2,T+\delta}}|\theta_T], \end{aligned} \quad (5.17)$$

where $\rho[R_{1,T+\delta}, \sqrt{R_{2,T+\delta}}|\theta_T]$ is the correlation between $R_{1,T+\delta}$ and $\sqrt{R_{2,T+\delta}}$ given θ_T . Similar to the branch of \mathcal{M} responsible for estimating $\mathbb{E}[\sigma_{T+\delta}|\theta_T]$, the objective is to estimate $\text{Std}[R_{1,T+\delta}|\theta_T]$,

⁵It arises from the relationship:

$$\text{Var}[\sqrt{R_{2,T+\delta}}|\theta_T] = \mathbb{E}[R_{2,T+\delta}|\theta_T] - \mathbb{E}[\sqrt{R_{2,T+\delta}}|\theta_T]^2 \Leftrightarrow \mathbb{E}[\sqrt{R_{2,T+\delta}}|\theta_T]^2 = \mathbb{E}[R_{2,T+\delta}|\theta_T] - \text{Var}[\sqrt{R_{2,T+\delta}}|\theta_T].$$

$\text{Std}[\sqrt{R_{2,T+\delta}}|\theta_T]$, and $\rho[R_{1,T+\delta}, \sqrt{R_{2,T+\delta}}|\theta_T]$ using 3 parallel sub-networks assigned to each of these components. These 3 sub-networks are each composed of 6 hidden layers, with the first layer being fed by the input layer θ_T . The first 5 layers of each sub-network consist of 100 ReLU neurons. The last hidden layer is composed of linear neurons, with 3 neurons for sub-networks 3 and 4, which are responsible for estimating $\text{Std}[R_{1,T+\delta}|\theta_T]$ and $\text{Std}[\sqrt{R_{2,T+\delta}}|\theta_T]$, respectively, and 21 neurons for the 5th sub-network responsible for estimating the correlation $\rho[R_{1,T+\delta}, \sqrt{R_{2,T+\delta}}|\theta_T]$. The last hidden layer of each sub-network feeds an output neuron, which also receives input from the δ layer.

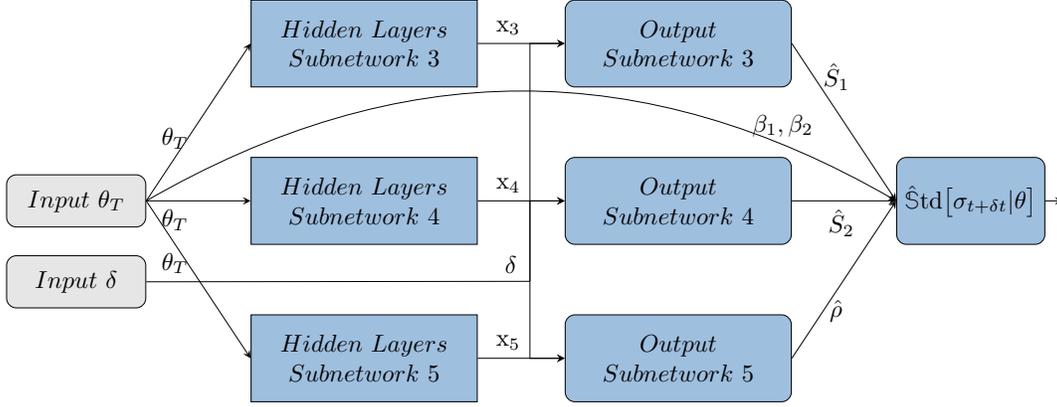


Figure 5.4: The part of the network \mathcal{M} responsible for estimating $\text{Std}[\sigma_{t+\delta}|\theta_T]$.

For sub-networks 3 and 4, the output neuron is associated with an activation function of the following form:

$$\hat{S}_j(\delta, x_j) = \exp \left(x_{j,1} + \max(x_{j,2}, \epsilon) \cdot \log \left(\frac{\delta}{1 + (x_{j,3})_+ \cdot \delta} \right) \right)$$

where ϵ is a positive constant close to zero. The choice of this activation function is motivated by the fact that, due to the properties of the RPDV model, the logarithms of the respective standard deviations of R_1 and $\sqrt{R_2}$ for the considered time horizons approximately follow a relationship of the form:

$$\hat{\text{Std}}[(R_{j,T+\delta})^{1/j}|\theta_T] \approx \exp(a + b \cdot \log(\delta)),$$

where b is positive. In the chosen activation function, $\log(\delta)$ is replaced with $\log(\delta/(1 + x_{j,3} \cdot \delta))$ to potentially capture concavity of the relationship for certain θ -vectors. Additionally, it is also interesting to note that this function ensures that the estimated standard deviations are zero when $\delta = 0$, consistent with the fact that $R_{1,T}$ and $R_{2,T}$ are θ -measurable.

For subnetwork 5 responsible for estimating the correlation between $R_{1,T+\delta}$ and $\sqrt{R_{2,T+\delta}}$ given θ_T , the output neuron is associated with the following activation function:

$$\hat{\rho}(\delta, y) = \min \left(2; x_{5,1} + \sum_{k=2}^{21} (1 - e^{-g_k \delta}) x_{5,k} \right) - 1.$$

This activation function ensures that the output range is limited to the interval $[-1, 1]$, which is suitable for estimating a correlation. Furthermore, the instantaneous correlation and the asymptotic correlation between R_1 and $\sqrt{R_2}$ conditioned on θ_T are given by:

$$\hat{\rho}(0, x_5) = \min(2; x_{5,1}) - 1 \quad \text{and} \quad \hat{\rho}(\infty, x_5) = \min \left(2; x_{5,1} + \sum_{k=2}^{21} x_{5,k} \right) - 1.$$

Each output layer of subnetworks 3, 4, and 5 feeds the output neuron of \mathcal{M} responsible for estimating $\text{Std}[\sigma_{T+\delta} | \theta_T]$. This output neuron, also fed by the input layer θ_T , is associated with the following activation function using the analytical expression 5.17:

$$\hat{\text{Std}}[\sigma_{T+\delta} | \theta_T] = \sqrt{(\beta_1 \hat{S}_1)^2 + (\beta_2 \hat{S}_2)^2 + 2\beta_1 \beta_2 \hat{S}_1 \hat{S}_2 \hat{\rho}},$$

with \hat{S}_1, \hat{S}_2 and $\hat{\rho}$ being the respective outputs of sub-networks 1, 2, and 3.

5.3.2 The estimator function

As mentioned in the introduction of section 5.3, the purpose of the function \mathcal{M} defined in section 5.3.1 is to assist in the training of a second neural network, the estimator function Θ , which is responsible for estimating the Bayes estimator of θ_T from a data matrix D under the posterior measure π_D . The objective is to construct an estimator function Θ that satisfies the following criterion:

$$\mathbb{E}_{\pi_D} [L(\theta_T, \Theta(D))] \approx \min_{\hat{\theta}_T} \mathbb{E}_{\pi_D} [L(\theta_T, \hat{\theta}_T)], \quad \forall D : \pi(D) \neq 0. \quad (5.18)$$

Therefore, the architecture of Θ should be designed to extract all relevant information contained in D in order to achieve the stated objective. To this end, Θ take the following general form:

$$\Theta(D) = \mathcal{NN}(\mathcal{E}(D)),$$

where \mathcal{NN} is a neural network and \mathcal{E} is a time-series encoder generally defined by:

$$\mathcal{E}(D) = \begin{pmatrix} z_1 \\ \dots \\ z_{n_{\mathcal{E}}} \end{pmatrix}, \quad (5.19)$$

with $n_{\mathcal{E}}$ the number of features extract by \mathcal{E} . The estimator function first encodes with \mathcal{E} the raw data matrix D into a feature vector that is used as input for a neural network \mathcal{NN} to predict the θ -vector. The sections 5.3.2.1 and 5.3.2.2 provide a detailed description of the structure of these two components that form Θ .

5.3.2.1 Dual encoder structure: combining non-trainable and trainable methods

The role of the encoder \mathcal{E} is to extract informative features from D for estimating θ_T . Given the variety of methods available for encoding time series, \mathcal{E} can take different forms. These methods, found in the academic literature, include transforming time series into pattern variables ([131], [194]), imaging time series ([196], [21]), or using signature methods ([158]). Depending on the chosen method, the encoding can be predetermined, meaning that features are extracted using a fixed method determined in advance, or it can be learned during the training process, allowing the encoder to adapt to the specific data characteristics. In this work, the encoder \mathcal{E} combines both approaches by incorporating a non-trainable component and a trainable component. The aim is to leverage prior knowledge of the process by extracting informative metrics using the non-trainable component, while complementing them with the trainable component of the encoder.

5.3.2.1.1 The non-trainable encoder

The non-trainable component of the encoder \mathcal{E} , denoted as \mathcal{E}_1 , is a pre-determined method that extracts informative metrics from D for determining the θ -vector. Specifically, \mathcal{E}_1 computes the following features from D :

- The serial correlation of the log realized volatility for the following lag times expressed in trading days: 1, 2, 3, 4, 5, 10, 20, 60, 125, 252.
- The mean of the absolute value of realized log-volatility increments over the following time intervals in trading days: 1, 2, 3, 4, 5, 10, 20, 60, 125, 252.
- The first four moments of the distribution of returns and realized volatility for the time horizons of 1, 5, 21, 63, 252 expressed in trading days.
- The 20 percentiles of the distribution of returns and of the realized volatility for the time horizons of 1, 5, 21, 63, 252 expressed in trading days.
- The linear regression coefficient between the volatility increments and the returns for the following lag times expressed in trading days: 1, 2, 3, 4, 5, 10, 20, 60, 125, 252.
- The standardized exponential moving averages of returns, realized volatility, and realized variance, which are defined respectively as

$$m_{1,j} = \frac{\sum_{i=1}^N r_{t_i} e^{(t_i-t)g_j}}{\sum_{i=1}^N e^{(t_i-t)g_j}}, \quad m_{2,j} = \frac{\sum_{i=1}^N \tilde{\sigma}_{t_i} e^{(t_i-t)g_j}}{\sum_{i=1}^N e^{(t_i-t)g_j}}, \quad m_{3,j} = \frac{\sum_{i=1}^N \tilde{\sigma}_{t_i}^2 e^{(t_i-t)g_j}}{\sum_{i=1}^N e^{(t_i-t)g_j}},$$

where $g_j \in (g_k)_{1 \leq k \leq 20}$ and their values are defined in equation 5.15 (section 5.3.1.2).

The metrics computed by \mathcal{E}_1 are diverse, allowing for a multifaceted approach to the data in D . These metrics, along with those from the trainable component of the encoder \mathcal{E} , will be used as inputs to the \mathcal{NN} network.

5.3.2.1.2 The trainable encoder

The trainable component of \mathcal{E} , denoted as \mathcal{E}_2 , aims to complement the metrics calculated by \mathcal{E}_1 , adopting a more agnostic approach. It consists of a convolutional neural network (CNN) with a structure similar to that of a multi-scale CNN (MCNN) proposed by Cui *et al.* ([79]). The input layer of \mathcal{E}_2 takes the raw data matrix D as input and feeds it into four branches. The first layer of each branch is associated with a function defined as:

$$\mathcal{A}(D, l) = \begin{pmatrix} \frac{P_{t_1+l}}{P_{t_1}} & \tilde{\sigma}_{[t_1:t_1+l]} & r_{[t_1:t_1+l]} & \tilde{\sigma}_{t_1+l} - \tilde{\sigma}_{t_1} \\ \dots & \dots & \dots & \dots \\ \frac{P_{t_N}}{P_{t_1}} & \tilde{\sigma}_{[t_1:t_1+l]} & r_{[t_{N-l}:t_N]} & \tilde{\sigma}_{t_N} - \tilde{\sigma}_{t_{N-l}} \end{pmatrix}, \quad (5.20)$$

where

$$\tilde{\sigma}_{[t_i:t_{i+l}]} = \sqrt{\frac{1}{l} \sum_{k=1}^l \tilde{\sigma}_{t_i+k}^2} \quad \text{and} \quad r_{[t_i:t_{i+l}]} = \frac{P_{t_{i+l}} - P_{t_i}}{P_{t_i}}.$$

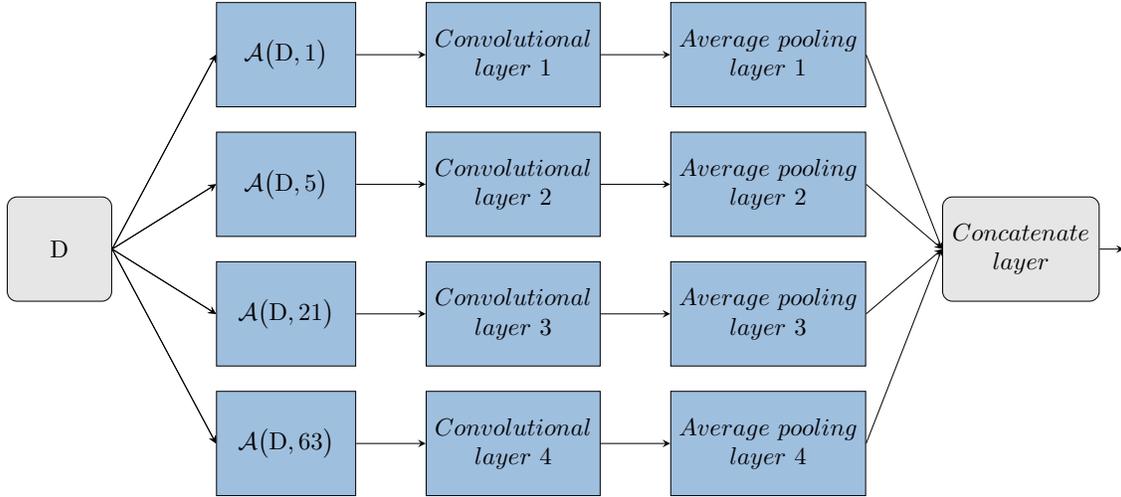


Figure 5.5: The architecture of the encoder \mathcal{E}_2 .

This function transforms the original $N \times 2$ matrix into a $(N - l) \times 4$ matrix. The first column represents the normalized price with respect to the date t_1 . The second column corresponds to the integrated volatility over a time window of l trading days, while the third column denotes the asset return over the same time window. The fourth column captures the variation of volatility over the

l -day period. For branches 1, 2, 3, and 4, the time window parameter l is fixed at 1, 5, 21, and 63, respectively. This choice allows for the augmentation of the original matrix D with two additional columns and captures important information at different time scales using multiple l values. Each augmented matrix is processed by a convolutional layer followed by an average pooling layer for each of the four branches. Each convolutional layer consists of 50 filters of size 5×4 , with a stride of 1. The pooling layers perform global average pooling for each filtered time series. The outputs of the four pooling layers associated with the branches of \mathcal{E}_2 are finally flattened and concatenated with the output of \mathcal{E}_1 . This combined output is then used as input to \mathcal{NN} .

5.3.2.2 The network \mathcal{NN} : from encoded data to θ -vector

The \mathcal{NN} network is responsible for predicting the Bayesian estimator of the θ -vector using the feature vector provided by the encoder \mathcal{E} . To accomplish this, \mathcal{NN} proceeds sequentially by first estimating the 9 parameters of the model and then, in a second step, estimating the state variables. These two operations are performed by two separate multilayer perceptrons (MLPs): \mathcal{NN}_1 , which is responsible for estimating $\hat{\phi}$, and \mathcal{NN}_2 , which is responsible for estimating \hat{R}_T .

To begin with, the input layer of \mathcal{NN}_1 is fed by the output of \mathcal{E} , which results in $n_{\mathcal{E}}$ input neurons. This is followed by 6 ReLU layers, each with 100 neurons. The last ReLU layer feeds into the output layer, which consists of 9 neurons that correspond to the parameters to be predicted. The output neuron responsible for estimating β_1 is associated with an inverted ReLU activation function, while the other output neurons are associated with a standard ReLU. The output layer of \mathcal{NN}_1 is then fed into the network \mathcal{NN}_2 , as well as into the output layer of \mathcal{NN} (and therefore the output layer of Θ), where it is concatenated with the output of the \mathcal{NN}_2 network.

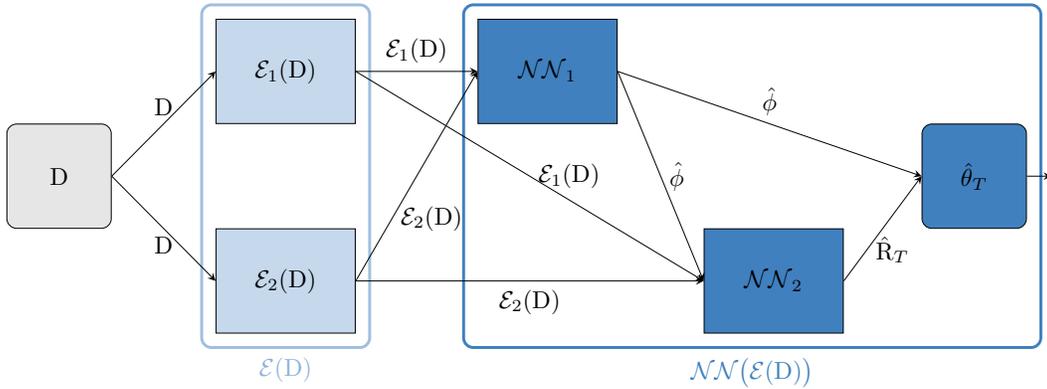


Figure 5.6: The architecture of the estimator function Θ .

The MLP \mathcal{NN}_2 receives as input both the output of \mathcal{E} and the output of \mathcal{NN}_1 , which results in $n_{\mathcal{E}} + 9$ input neurons. Like the input layer of \mathcal{NN}_1 , the input layer of \mathcal{NN}_2 is followed by 6 ReLU layers, each with 100 neurons. Finally, the output layer of \mathcal{NN}_1 consists of $2n$ linear neurons,

representing the $2n$ state variables. This output layer feeds into the output layer of \mathcal{NN} , where it is concatenated with the output of the \mathcal{NN}_2 network.

5.3.3 The estimation procedure

This section presents an estimation approach for the RPDV model, utilizing the functions \mathcal{M} and Θ introduced in sections 5.3.1 and 5.3.2, respectively. This method can be divided into three phases: the generation of training data, the training of \mathcal{M} , and the training of Θ . These three steps are subsequently described in this section. At the end of this procedure, Θ can be used as an estimator function for the θ -vectors.

5.3.3.1 Generation of initial data

To calibrate the functions \mathcal{M} and Θ , it is necessary to generate training sets. This process involves defining the method for generating the θ -vectors, which is equivalent to establishing the prior measure π . We will then outline the different steps involved in constructing the training sets for \mathcal{M} and Θ .

5.3.3.1.1 Defining the prior measure π

The estimation procedure requires defining the prior measure π . The specification of this measure incorporates prior knowledge about the parameters, even if this knowledge is vague. By constraining the parameter space, it is likely to increase the estimation quality without overly restricting the range of possible values. The idea is therefore to propose a generation procedure for parameter θ -vectors that exploits prior knowledge about price and volatility dynamics, without excessively constraining the parameter space.

In this context, certain coordinates of the random vector θ_T are assumed to be independent random variables, while others exhibit a correlation structure. Specifically, the parameters $\beta_0, \alpha_1, \alpha_2, \kappa_1, \kappa_2$ are distributed independently, as follows:

$$(\beta_0, \alpha_1, \alpha_2, \kappa_1, \kappa_2) \sim \mathcal{U}(0, 0.25) \times \mathcal{U}(0, 1) \times \mathcal{U}(0, 1) \times \mathcal{U}(0, 5) \times \mathcal{U}(0, 5) \times \mathcal{U}(0, 0.15).$$

Regarding the risk premia, their generation is slightly more complex. Firstly, the value of the price drift $\lambda_1\sigma_t + \lambda_2(\sigma_t)^2$ is generated when the volatility is equal to 15% ($\sigma_t = 0.15$) as follows:

$$\bar{\mu} \sim \mathcal{U}(0, 0.1).$$

In other words, under the measure π , the price drift ranges from 0 to 10% when the volatility level is 15%. Then, there is a $\frac{1}{3}$ probability that $\lambda_1 = \frac{\bar{\mu}}{0.15}$ and $\lambda_2 = 0$, a $\frac{1}{3}$ probability that $\lambda_1 = 0$ and $\lambda_2 = \frac{\bar{\mu}}{0.15^2}$, and a $\frac{1}{3}$ probability that

$$\lambda_1 \sim \mathcal{U}(0, 1) \cdot \frac{\bar{\mu}}{0.15}, \quad \lambda_2 = \frac{\bar{\mu} - 0.15\lambda_1}{0.15^2}.$$

Therefore, there is an equal probability of having a pure volatility premium, a pure variance premium, or a mixture of both. The last two remaining parameters are β_1 and β_2 . In both cases, it makes sense to consider the specificity of the kernel associated with the variables for which β_1 and β_2 determine the volatility sensitivity. With respect to β_1 , it is generated as follows:

$$\beta_1 \sim \mathcal{U}(-1.5, 0) \cdot \left(\sum_{i=1}^n \sum_{k=1}^n \frac{w_{1,i} w_{1,k} \gamma_{1,i} \gamma_{1,k}}{\gamma_{1,i} + \gamma_{1,k}} \right)^{-0.5}.$$

The term that weights $\mathcal{U}(-1.5, 0)$ is the inverse of the asymptotic standard deviation of the BSS process associated with the kernel \hat{K}_1 (see appendix 5.D.2). This weighting allows for the generation of reasonable values of β_1 given α_1 . Similarly, the parameter β_2 is generated as follows:

$$\beta_2 \sim \frac{\mathcal{U}(0, 1)}{\sum_{i=1}^n w_{2,i}}.$$

The term that weights $\mathcal{U}(0, 1)$ corresponds to the inverse of the integral over \mathbb{R}_+ of the kernel \hat{K}_2 (see appendix 5.D.1). This weighting prevents volatility from exploding regardless of the value of κ_2 .

Next, we define how the state variables $\{(R_1)_j\}_{j=1}^n$ and $\{(R_2)_j\}_{j=1}^n$ are generated. The approach consists of three steps. The first step is the initialization of the state variables. The state variables $\{(R_1)_j\}_{j=1}^n$ are simply initialized to zero, and the state variables $\{(R_2)_j\}_{j=1}^n$ are initialized as follows:

$$(R_{2,0})_i \sim (\beta_0)^2 \cdot \mathcal{U}(0.9, 1.1).$$

From these initial values and the associated parameter vector, a simulation of the volatility dynamics is performed over a period of 5 years. Finally, the values of the state variables at the end of this simulation are retained. However, the θ -vector is retained if and only if the initial value of the volatility is positive and lower than 300%, i.e., if $0 < \beta_0 + \beta_1 R_{1,T} + \beta_2 \sqrt{R_{2,T}} < 3$.

5.3.3.1.2 Constructing training sets

The objective is to generate data samples

$$\left\{ \mathbf{D}, \theta_T, \left\{ \bar{E}_{1,T+\delta_k}, \bar{E}_{2,T+\delta_k}, \bar{S}_{1,T+\delta_k}, \bar{S}_{2,T+\delta_k}, \bar{\rho}_{T+\delta_k} \right\}_{k=1}^p \right\},$$

where \mathbf{D} is a data matrix of the form 5.5 (section 6.2.2.1) generated from the generator 5.4 (section 5.2.1.2), θ_T is the value taken by the θ -vector at the end of the simulation of \mathbf{D} , and where $\bar{E}_{k,T+\delta_j}$, $\bar{S}_{1,T+\delta_k}$ and $\bar{\rho}_{T+\delta_k}$ are unbiased estimators of $\mathbb{E}[R_{k,T+\delta}|\theta_T]$, $\text{Std}[R_{k,T+\delta}|\theta_T]$ and $\rho[R_{1,T+\delta}, \sqrt{R_{2,T+\delta}}|\theta_T]$, respectively. To do this, we use the following algorithm.

Algorithm 1 Procedure for generating training sets.

Require: π, n_1, n_2

1. Generate an i.i.d. sample $\Omega_{\theta_0} = \left\{ \theta_{t_0}^{(i)} \right\}_{1 \leq i \leq n_1}$ from the distribution π .
 2. Generate from Ω_{θ_0} and model 5.4 the pairs $\left\{ D^{(i)}, \theta_T^{(i)} \right\}_{1 \leq i \leq n_1}$.
 3. Generate n_2 time series over the periods $T + \delta_1, \dots, T + \delta_p$ using model 5.4 for each $\theta \in \left\{ \theta_T^{(i)} \right\}_{1 \leq i \leq n_1}$, and extract from each series the sets $\Omega_R = \left\{ \left\{ R_{1,T+\delta_k}^{(i,k)}, R_{2,T+\delta_k}^{(i,j)} \right\}_{k=1}^p \right\}_{\substack{1 \leq i \leq n_1 \\ 1 \leq j \leq n_2}}$.
 4. Compute set $\left\{ \left\{ \bar{E}_{1,T+\delta_k}^{(i)}, \bar{E}_{2,T+\delta_k}^{(i)}, \bar{S}_{1,T+\delta_k}^{(i)}, \bar{S}_{2,T+\delta_k}^{(i)}, \bar{\rho}_{T+\delta_k}^{(i)} \right\}_{k=1}^p \right\}_{1 \leq i \leq n_1}$ from sample set Ω_R .
- return** $\left\{ D^{(i)}, \theta_T^{(i)}, \left\{ \bar{E}_{1,T+\delta_k}^{(i)}, \bar{E}_{2,T+\delta_k}^{(i)}, \bar{S}_{1,T+\delta_k}^{(i)}, \bar{S}_{2,T+\delta_k}^{(i)}, \bar{\rho}_{T+\delta_k}^{(i)} \right\}_{1 \leq k \leq p} \right\}_{1 \leq i \leq n_1}$
-

This data generation procedure aligns with the adopted Bayesian approach. The set of i.i.d. matrices $\{D^{(i)}\}_{1 \leq i \leq n_1}$ is generated from initial vectors $\{\theta_{t_0}^{(i)}\}_{1 \leq i \leq n_1}$ sampled from the prior distribution π , which incorporates vague knowledge about the model parameters. In this chapter, we specify $n_1 = 200\,000$ and $n_2 = 200$. The large value chosen for n_1 ensures good coverage of the parameter and state variable space. As for n_2 , its value allows for empirical moment estimators with reasonable variance on average.

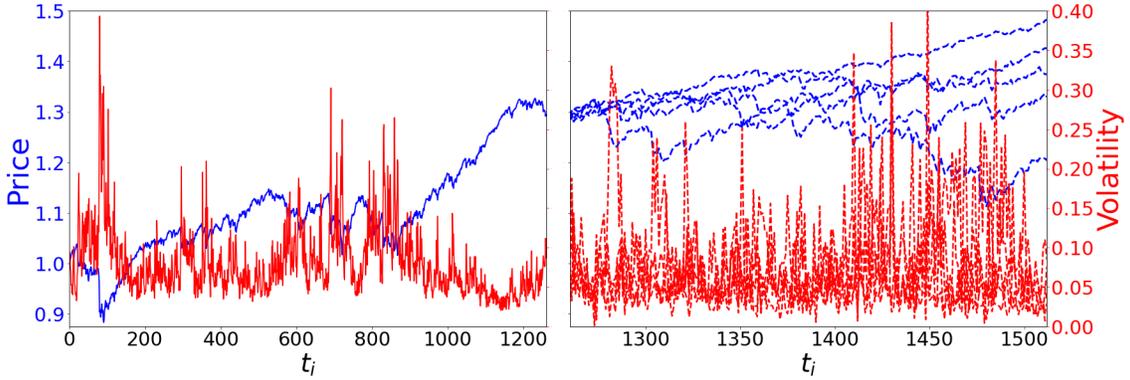


Figure 5.7: The left figure plots a historical trajectory of realized joint price and volatility of an asset over 1260 trading days contained in a matrix $D^{(i)}$, generated from the initial θ -vector value $\theta_0^{(i)}$. The right figure represents 5 continuations of this trajectory over 252 trading days generated from $\theta_{1260}^{(i)}$.

5.3.3.2 The training of \mathcal{M}

The second phase of the general procedure involves training \mathcal{M} using the datasets generated in section 5.3.3.1. The method to be proposed for this purpose is derived from the following proposition proved in appendix 5.C.1.

Proposition 1 *Let be $\theta_T^{(1)}, \dots, \theta_T^{(n_1)}$ a sequence of i.i.d. random variable following π , and $\{\bar{M}_{T+\delta_k}^{(1)}\}_{1 \leq j \leq p}, \dots, \{\bar{M}_{T+\delta_k}^{(n_1)}\}_{1 \leq j \leq p}$ a sequence of sets such as $\forall i, k, \bar{M}_{T+\delta_k}^{(i)}$ is an unbiased estimator of $M(\theta_T^{(i)}, \delta_k)$ calculated from a sample of size n_2 . If it exists \mathcal{M}^* , such as $\mathcal{M}^*(\theta_T, \delta_k) = M(\theta_T, \delta_k), \forall \theta_T : \pi(\theta_T) \neq 0$ and $\delta_k \in \{\delta_1, \dots, \delta_p\}$, thus $\forall \hat{\mathcal{M}}^*$ solution to*

$$\arg \min_{\mathcal{M}} \frac{1}{n_1} \sum_{i=1}^{n_1} \sum_{k=1}^p \left\| \mathcal{M}(\theta_T^{(i)}, \delta_k) - \bar{M}_{T+\delta_k}^{(i)} \right\|_2^2,$$

$\hat{\mathcal{M}}^*(\theta, \delta_k) = M(\theta_T, \delta_k), \forall \theta_T : \pi(\theta_T) \neq 0$ and $\delta_k \in \{\delta_1, \dots, \delta_p\}$.

Therefore, this proposition implies a way to make \mathcal{M} a convergent estimator of M , under suitable conditions of existence, for the time horizons $\delta_1, \dots, \delta_p$ and the θ -vectors associated with a non-zero probability under the measure π . It simply involves minimizing the mean squared difference between the estimators returned by \mathcal{M} and the unbiased estimator \bar{M} for each pair $(\delta^{(i)}, \theta_T^{(i)})$ in the training set. This minimization constitutes the second step of algorithm 2, which we propose for training \mathcal{M} .

Algorithm 2 Training procedure for \mathcal{M}

Require: $\left\{ \theta_T^{(i)}, \left\{ \bar{E}_{1,T+\delta_j}^{(i)}, \bar{E}_{2,T+\delta_j}^{(i)}, \bar{S}_{1,T+\delta_j}^{(i)}, \bar{S}_{2,T+\delta_j}^{(i)}, \bar{\rho}_{T+\delta_j}^{(i)} \right\}_{1 \leq k \leq p} \right\}_{1 \leq i \leq n_1}$

1. Optimize

$$\arg \min_{\mathcal{M}} \frac{1}{n_1} \sum_{i=1}^{n_1} \sum_{k=1}^p \left(\hat{\rho}_{T+\delta_k}^{(i)} - \bar{\rho}_{T+\delta_k}^{(i)} \right)^2 + \sum_{j=1}^2 \left(\left(\hat{E}_{j,T+\delta_k}^{(i)} - \bar{E}_{j,T+\delta_k}^{(i)} \right)^2 + \left(\hat{S}_{j,T+\delta_k}^{(i)} - \bar{S}_{j,T+\delta_k}^{(i)} \right)^2 \right).$$

2. Optimize

$$\arg \min_{\mathcal{M}} \frac{1}{n_1} \sum_{i=1}^{n_1} \sum_{k=1}^p \left\| \mathcal{M}(\theta_T^{(i)}, \delta_k) - \bar{M}_{T+\delta_k}^{(i)} \right\|_2^2,$$

starting from the training of \mathcal{M} obtained at the end of step 1.

return $\hat{\mathcal{M}}^*$

The first step of the procedure serves only to prepare for the second (and final) step of the training of \mathcal{M} . It involves independently training the five subnetworks that compose \mathcal{M} (see figure 5.2).

The objective by the end of step 1 is as follows:

$$\begin{aligned}\hat{E}_{j,T+\delta_k}^{(i)} &\approx \mathbb{E} \left[(R_{j,T+\delta_k})^{1/p} | \theta_T^{(i)} \right]^p, & \hat{S}_{j,T+\delta_k}^{(i)} &\approx \text{Std} \left[R_{j,T+\delta_k} | \theta_T^{(i)} \right] \quad \text{and} \\ \hat{\rho}_{T+\delta_k}^{(i)} &\approx \rho \left[R_{1,T+\delta_k}, \sqrt{R_{2,T+\delta_k}} | \theta_T^{(i)} \right],\end{aligned}$$

$\forall \{i, k\}$. In this first phase, the training of \mathcal{M} is conducted to align with the specific role of each of the 5 subnetworks within it. This first phase, which frames the outputs of the subnetworks in \mathcal{M} , is followed by step 2, which is aimed at achieving the objective 5.12 stated in the introduction of section 5.3.1. As previously mentioned, according to proposition 5.3.3.2, under suitable conditions of existence, as n_1 and n_2 approach infinity, solving the optimization problem associated with step 2 involves finding $\hat{\mathcal{M}}^*$ such that:

$$\hat{\mathcal{M}}^*(\theta, \delta_k) = M(\theta_T, \delta_k), \forall \theta_T : \pi(\theta_T) \neq 0 \quad \text{and} \quad \delta_k \in \{\delta_1, \dots, \delta_p\}.$$

The assumption of the existence of \mathcal{M}^* is related to the flexibility of the network \mathcal{M} in approximating the function M . Moreover, the reasonableness of this assumption is guaranteed by the structure of the subnetworks that constitute \mathcal{M} , which can approximate any continuous function due to the universal approximation theorem ([124]).

5.3.3.3 The training of Θ

The third and final step of the estimation process involves training Θ to achieve the desired situation described in 5.18 at the beginning of section 5.3.2. The consistency of the method to be proposed in this section for this purpose stems from the following proposition demonstrated in appendix 5.C.2.

Proposition 2 *Let $\theta_{t_0}^{(1,1)}, \dots, \theta_{t_0}^{(n_1,1)}$ be a sequence of i.i.d. random variables following π , $D^{(1)}, \dots, D^{(n_1)}$ a set of time-series such that $D^{(i)}$ is generated from the M -RPDV associated with the θ -vector $\theta_{t_0}^{(i)}$, and $\theta_T^{(1,2)}, \dots, \theta_T^{(n_1,2)}$ the set of values taken by θ at time t_N for each time series $D^{(i)}$. If there exists Θ^* such that for all $D : \mathbb{P}_\pi(D) \neq 0^6$, $\Theta^*(D)$ is a Bayes estimator of θ_T under the posterior measure π , then for any $\hat{\Theta}^*$ solution to the optimization problem*

$$\arg \min_{\Theta} \lim_{n_1 \rightarrow +\infty} \frac{1}{n_1} \sum_{i=1}^{n_1} L \left(\theta_T^{(1,i)}, \Theta \left(D^{(i)} \right) \right),$$

$\hat{\Theta}^*(D)$ is a Bayes estimator of θ_T under the posterior measure $\pi_D, \forall D : \mathbb{P}_\pi(D) \neq 0$.

This proposition, therefore, has significant implications as it provides a way to calibrate Θ , such that asymptotically and under a suitable existence condition, $\Theta(D)$ becomes a Bayesian estimator of θ_T under the posterior measure π_D , for all $D : \mathbb{P}_\pi(D) \neq 0$. This result is even more remarkable considering that the estimation method does not require explicit calculation of the posterior measures (i.e., the measure π updated by a matrix D) at any point. Indeed, it simply involves

⁶ \mathbb{P}_π denotes the distribution of D induced by π .

following steps 1 and 2 of algorithm 5.4.1.1) to generate a set of pairs $\left\{D^{(i)}, \theta_T^{(i)}\right\}_{1 \leq i \leq n_1}$, and then minimizing the average losses measured by L between the θ -vectors predicted by Θ (i.e. $D^{(i)}$) and the true θ -vectors (i.e. $\theta_T^{(i)}$). However, since L depends on the function M , which is not known, we use the following proxy that replaces M with \mathcal{M} :

$$\hat{L}\left(\theta_T^{(i)}, \Theta(D^{(i)})\right) = \sum_{k=1}^p C\left(\mathcal{M}\left(\theta_T^{(i)}, \delta_k\right), \mathcal{M}\left(\Theta\left(D^{(i)}\right), \delta_k\right)\right).$$

The idea is that after training \mathcal{M} (algorithm 2), $\mathcal{M}(\theta_T, \delta) \approx M(\theta_T, \delta)$ and $\nabla \mathcal{M}(\theta_T, \delta) \approx \nabla M(\theta_T, \delta)$, and therefore:

$$\hat{L}\left(\theta_T^{(i)}, \Theta(D^{(i)})\right) \approx L\left(\theta_T^{(i)}, \Theta(D^{(i)})\right) \quad \text{and} \quad \nabla \hat{L}\left(\theta_T^{(i)}, \Theta(D^{(i)})\right) \approx \nabla L\left(\theta_T^{(i)}, \Theta(D^{(i)})\right).$$

Consequently, the quality of the approximation of the function M by \mathcal{M} is crucial in the training of Θ . Regarding the assumption of the existence of Θ^* on which proposition 5.3.3.3 relies, its reasonableness depends on both the encoder's ability to extract all relevant information contained in the time series D and the plasticity of the networks \mathcal{NN}_1 and \mathcal{NN}_2 .

Based on these elements, the training of Θ is carried out using the following algorithm.

Algorithm 3 Learning procedure for the estimator function Θ

Require: $\left\{D^{(i)}, \theta_T^{(i)}\right\}_{1 \leq i \leq n_1}$

1. Optimize

$$\arg \min_{\mathcal{E}_2, \mathcal{NN}_1} \frac{1}{n_1} \sum_{i=1}^{n_1} \left\| \mathcal{NN}_1\left(\mathcal{E}\left(D^{(i)}\right)\right) - \phi^{(i)} \right\|_2^2.$$

2. Optimize

$$\arg \min_{\mathcal{NN}_2} \frac{1}{n_1} \sum_{i=1}^{n_1} \left(\mathcal{NN}_2\left(\mathcal{E}\left(D^{(i)}\right), \phi^{(i)}\right) - R_T^{(i)} \right)^2.$$

3. Optimize

$$\arg \min_{\mathcal{NN}_2} \frac{1}{n_1} \sum_{i=1}^{n_1} \left(\mathcal{M}\left(\left(\phi^{(i)}, \mathcal{NN}_2\left(D^{(i)}\right)\right), 0\right) - \mathcal{M}\left(\theta^{(i)}, 0\right) \right)^2.$$

4. Optimize

$$\arg \min_{\Theta} \frac{1}{n_1} \sum_{i=1}^{n_1} \sum_{k=1}^p C\left(\mathcal{M}\left(\theta_T^{(i)}, \delta_k\right), \mathcal{M}\left(\Theta\left(D^{(i)}\right), \delta_k\right)\right).$$

return $\hat{\Theta}^*$

The first three learning steps for sub-regions of the network that constitutes Θ in practice serve only to prepare for the fourth and final step, which is directly derived from proposition 5.3.3.3. In the first phase, only the components of the network responsible for predicting the parameter vector ϕ , i.e., the trainable encoder \mathcal{E}_2 and the neural network \mathcal{NN}_1 , are trained. The aim is to guide the learning of Θ initially by minimizing the sum of squared differences between predicted and actual parameter vectors, thereby obtaining parameters consistent with the prior measure π . Next, the neural network \mathcal{NN}_2 responsible for predicting the state variable vector R is trained. The calibration of \mathcal{NN}_2 uses the encoder \mathcal{E} fitted during phase 1 and actual parameter vectors, rather than those estimated by \mathcal{NN}_1 . This approach allows for a more focused learning of the relationship between data and the vector of state variables to be predicted, without introducing any bias caused by estimation errors in \mathcal{NN}_1 . The first three steps of the algorithm described previously are not important in themselves, but serve only to prepare for the final learning step of Θ . This final step aims to achieve the objective defined in the introduction of section 5.3.2, namely that Θ returns an estimator of θ_T that is close (in terms of expected loss measured by L) to its Bayesian estimator under the posterior measure. The consistency of the optimization program solved in this step with this objective is established by the following proposition proved in the appendix 5.C.2.

5.4 Assessment of the estimation method

The purpose of this section is to evaluate the estimator function defined in section 5.3. For this purpose, we perform various tests, starting with synthetic data and then moving on to market data.

5.4.1 Evaluation of estimation method using synthetic data

In this section, the objective is to evaluate the estimation method presented in section 5.3 using synthetic data. We start by assessing the accuracy of the moment estimator \mathcal{M} in approximating the function M . Next, we evaluate the effectiveness of the estimator function Θ in providing consistent estimates of the θ -vector that align with our forecasting objectives.

5.4.1.1 Test dataset and evaluation metrics

The test dataset is generated using algorithm , as introduced in section 5.3.3.1.2, with parameters $n_1 = 10000$ and $n_2 = 1000$. Hence, we have the following elements that will be used to construct the test datasets:

$$\left\{ D^{(i)}, \theta_T^{(i)}, \left\{ \bar{M}_{T+\delta_k}^{(i)} \right\}_{1 \leq k \leq p} \right\}_{1 \leq i \leq 10000}.$$

The choice of $n_2 = 1000$ in Algorithm 1 allows us to consider $\bar{M}_{T+\delta_k}^{(i)}$ as reasonably accurate estimators of $M(\theta_T^{(i)}, \delta)$. This consistency enables us to use them as targeted values for comparison

with the predicted values generated by $\mathcal{M}(\theta_T^{(i)}, \delta)$ and $\mathcal{M}(\mathbf{D}^{(i)}, \delta)$.

The evaluation of the estimated conditional moments \hat{y}_i with the targeted conditional moment values y_i will be conducted using the following metrics:

- The root mean squared error (RMSE)

$$\text{RMSE} = \sqrt{\frac{\sum_{i=1}^{n_1} (y_i - \hat{y}_i)^2}{N}}.$$

- The mean absolute error (MAE)

$$\text{MAE} = \frac{\sum_{i=1}^{n_1} |y_i - \hat{y}_i|}{N}.$$

- The mean absolute percentage error (MAPE)

$$\text{MAPE} = \frac{1}{n_1} \frac{\sum_{i=1}^{n_1} |y_i - \hat{y}_i|}{y_i}$$

- The coefficient of determination

$$r^2 = 1 - \frac{\sum_{i=1}^{n_1} (y_i - \hat{y}_i)^2}{\sum_{i=1}^{n_1} \left(y_i - \frac{1}{n_1} \sum_{i=1}^{n_1} y_i \right)^2}.$$

The use of these different metrics allows for evaluating, from different angles, the moment estimators. RMSE is a classic metric that measures the difference between predicted and actual values, while taking into account the variance of errors. MAE, on the other hand, measures the average of absolute errors, providing an indication of the overall accuracy of the estimation. MAPE has the advantage of comparing the performance of the estimation, taking into account the heterogeneity of the magnitudes of the moments. Finally, the coefficient of determination r^2 measures the overall adequacy of the model by providing an indication of the proportion of variance explained by the model.

5.4.1.2 Evaluation of the ability of \mathcal{M} to approximate M

We aim to evaluate how closely \mathcal{M} approximates M . To do so, we compare the estimators

$$\mathcal{M}(\theta_T, \delta)_1 = \hat{\mathbb{E}}[\sigma_{T+\delta} | \theta_T] \quad \text{and} \quad \mathcal{M}(\theta_T, \delta)_2 = \hat{\text{Std}}[\sigma_{T+\delta} | \theta_T],$$

with the corresponding empirical estimators calculated from a sample of volatility trajectories using θ_T . This comparison is performed using the test dataset

$$\left\{ \left\{ \theta_T^{(i)}, \bar{M}_{T+\delta_k}^{(i)} \right\}_{1 \leq i \leq 10000} \right\}_{1 \leq k \leq 5},$$

which is extracted from the synthetic data defined in section 5.4.1.1. The results obtained are reported in tables 5.1 and 5.2.

	$\delta = 1$	$\delta = 5$	$\delta = 21$	$\delta = 42$	$\delta = 63$
RMSE	0.0055	0.0069	0.0095	0.0090	0.0100
MAE	0.0029	0.0035	0.0043	0.0045	0.0049
MAPE	0.0270	0.0325	0.0391	0.0421	0.0453
R-Squared	0.9958	0.9920	0.9832	0.9842	0.9800

Table 5.1: Evaluation metrics for the estimation of $\mathbb{E}[\sigma_{T+\delta}|\theta_T]$ by \mathcal{M} .

	$\delta = 1$	$\delta = 5$	$\delta = 21$	$\delta = 42$	$\delta = 63$
RMSE	0.0258	0.0279	0.0297	0.0310	0.0353
MAE	0.0107	0.0105	0.0122	0.0131	0.0141
MAPE	0.1081	0.0965	0.1046	0.1102	0.1143
R-Squared	0.9010	0.8887	0.8742	0.8675	0.8359

Table 5.2: Evaluation metrics for the estimation of $\text{Std}[\sigma_{T+\delta}|\theta_T]$ by \mathcal{M} .

The obtained results demonstrate that \mathcal{M} provides a reliable approximation of M across different time horizons. When estimating $\mathbb{E}[\sigma_{T+\delta}|\theta_T]$, the evaluation metrics indicate a significant agreement with the empirical estimators in terms of both absolute and relative deviation. For example, the MAE falls within the range of 0.0029 to 0.0049 for various time horizons δ , indicating that, on average, the estimated values of $\mathbb{E}[\sigma_{T+\delta}|\theta_T]$ provided by \mathcal{M} deviate from the empirical estimator by less than 0.005 units. Additionally, the MAPE metric reveals that the average absolute deviation between these two estimators ranges from 2.7% to 4.5% in relative terms, which is notably low. Furthermore, the consistently high R-squared values exceeding 98% confirm the excellent quality of approximation for the conditional expectation by \mathcal{M} .

The same observation applies to the estimation of the conditional standard deviations $\text{Std}[\sigma_{T+\delta}|\theta_T]$, albeit with some nuances. The evaluation metrics demonstrate a significant agreement between \mathcal{M} and the empirical estimators, indicating a reliable approximation of $\text{Std}[\sigma_{T+\delta}|\theta_T]$ for the different time horizons. However, it is worth noting that the deviation between these estimators is significantly larger compared to the estimation of $\mathbb{E}[\sigma_{T+\delta}|\theta_T]$. Specifically, we observe that the MAE ranges from 0.0107 to 0.0141 for different values of δ , suggesting that, on average, the estimated values provided by \mathcal{M} deviate from the empirical estimator by less than 0.015 units. In terms of the MAPE, the average absolute deviation between the estimators ranges from 9.65% to 11.43% in relative terms. Moreover, the R-squared values for $\text{Std}[\sigma_{T+\delta}|\theta_T]$ range from 83.59% to 90.10%, indicating a relatively high concordance between \mathcal{M} and the empirical estimators, but lower compared to the estimation of $\mathbb{E}[\sigma_{T+\delta}|\theta_T]$.

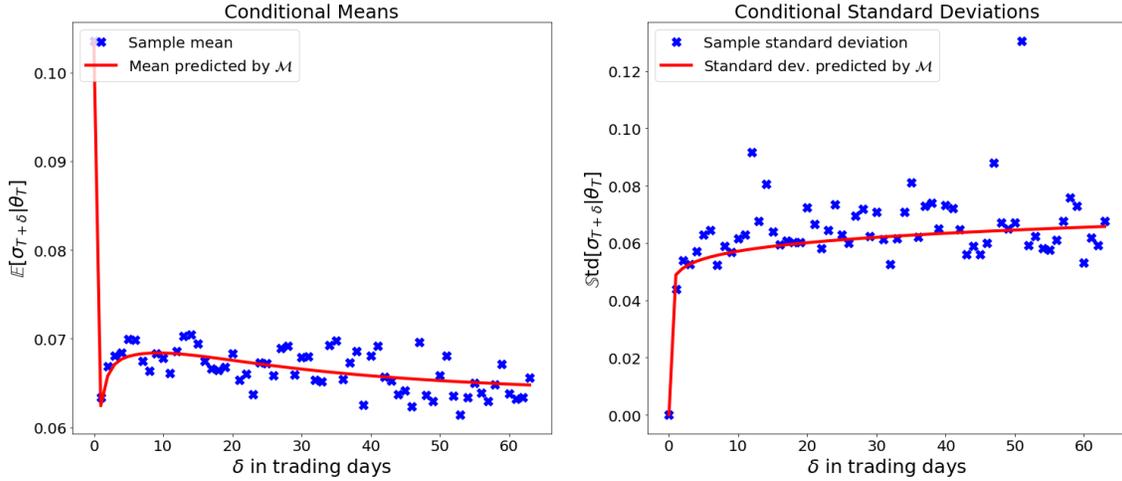


Figure 5.8: Example of estimated conditional means and standard deviations by \mathcal{M} compared to sample conditional estimators.

However, this discrepancy does not necessarily imply a lower quality of estimation for $\text{Std}[\sigma_{T+\delta t}|\theta_T]$. In fact, it can be mainly attributed to the higher variance of the empirical estimator used for estimating the standard deviations. This higher variance is illustrated by the example exhibited in figure 5.8, where the absolute percentage error for the 51 trading days horizon is greater than 50%. However, by replicating the experiment with 20 000 simulations, this absolute error decreases significantly to less than 2%. This suggests that the observed residuals are primarily due to the higher variance of the empirical estimators used rather than a misestimation of \mathcal{M} .

5.4.1.3 Assessment of the estimator function Θ

5.4.1.3.1 Evaluation based on conditional moments

To evaluate the performance of the estimator function Θ , we compare in this section the following estimators:

$$\mathcal{M}(\Theta(D^{(i)}), \delta)_1 = \hat{\mathbb{E}}[\sigma_{t+\delta}|D^{(i)}] \quad \text{and} \quad \mathcal{M}(\Theta(D^{(i)}), \delta)_2 = \hat{\text{Std}}[\sigma_{t+\delta}|D^{(i)}],$$

using two types of moment estimators: the empirical estimators already used in section 5.4.1.2, and the estimators computed from \mathcal{M} using the actual θ -vectors.

To begin the evaluation, we use the following test dataset, where the targeted values are the sample moments:

$$\left\{ \left\{ D^{(i)}, \bar{M}_{T+\delta_k}^{(i)} \right\}_{1 \leq i \leq 10000} \right\}_{1 \leq k \leq 5}.$$

The results obtained from this evaluation are presented in tables 5.3 and 5.4.

	$\delta = 1$	$\delta = 5$	$\delta = 21$	$\delta = 42$	$\delta = 63$
RMSE	0.0147	0.0129	0.0128	0.0122	0.0129
MAE	0.0083	0.0076	0.0074	0.0075	0.0077
MAPE	0.0904	0.0799	0.0753	0.0747	0.0761
R-Squared	0.9705	0.9725	0.9696	0.9705	0.9662

Table 5.3: Evaluation metrics comparing $\mathcal{M}(\Theta(\mathbf{D}), \delta)_1$ and $(\bar{M}_{T+\delta_k})_1$.

	$\delta = 1$	$\delta = 5$	$\delta = 21$	$\delta = 42$	$\delta = 63$
RMSE	0.0355	0.0368	0.0373	0.0400	0.0476
MAE	0.0156	0.0155	0.0165	0.0175	0.0185
MAPE	0.1857	0.1655	0.1610	0.1634	0.1655
R-Squared	0.8007	0.7944	0.7891	0.7688	0.7048

Table 5.4: Evaluation metrics comparing $\mathcal{M}(\Theta(\mathbf{D}), \delta)_2$ and $(\bar{M}_{T+\delta_k})_2$.

A first observation is that, consistently, the cost metrics are higher and the R-squared values are lower compared to the case studied in section 5.4.1.2, where the θ -vectors are known. However, the observed difference, although significant, remains relatively moderate, which suggests the quality of the estimation of the θ -vectors produced by Θ . However, to better interpret these results, it is important to note that this difference tends to decrease relative to the temporal horizon. Thus, while the MAPE between $\mathcal{M}(\theta_T, \delta)_1$ and $(\bar{M}_{T+\delta})_1$ increases from 2.7% for a 1-day trading horizon to 4.5% for a 3-month horizon (63 trading days), the MAPE between $\mathcal{M}(\Theta(\mathbf{D}^{(i)}), \delta)_1$ and $(\bar{M}_{T+\delta})_1$ decreases for the same periods from 9% to 7.6%. Similarly, the MAPE between $\mathcal{M}(\theta_T, \delta)_2$ and $(\bar{M}_{T+\delta})_2$ decreases from 10.8% for a 1-day trading horizon to 11.4% for a 3-month horizon, compared to a decrease from 18.6% to 16.5% for the same horizons when comparing $\mathcal{M}(\theta_T, \delta)_2$ and $(\bar{M}_{T+\delta})_2$. A first explanation for this phenomenon could be the decreasing significance of short-term information contained in the state variables associated with higher discount factors, as it has a diminishing impact on the conditional moments. However, these state variables are particularly challenging to estimate due to the daily observation frequency, which explains the reduction in the cost gaps between $\mathcal{M}(\Theta(\mathbf{D}), \delta)$ and $\mathcal{M}(\theta_T, \delta)$ as δ increases. This phenomenon may also be partly caused by the variance of the empirical estimator $\bar{M}_{T+\delta}$. To isolate the impact of estimating the θ -vectors using Θ , it is valuable to directly compare the estimator $\mathcal{M}(\Theta(\mathbf{D}), \delta)$ with the estimator $\mathcal{M}(\theta_T, \delta)$. To investigate this further, we employ the same evaluation procedure for Θ as discussed previously, but on the following test dataset:

$$\left\{ \left\{ \mathbf{D}^{(i)}, \mathcal{M}(\theta_T^{(i)}, \delta_k) \right\}_{1 \leq i \leq 10,000} \right\}_{1 \leq k \leq 5}.$$

The resulting outcomes from this evaluation are presented in the following tables.

	$\delta = 1$	$\delta = 5$	$\delta = 21$	$\delta = 42$	$\delta = 63$
RMSE	0.0128	0.0114	0.0113	0.0115	0.0118
MAE	0.0078	0.0072	0.0072	0.0074	0.0077
MAPE	0.0851	0.0763	0.0757	0.0797	0.0840
R-Squared	0.9744	0.9767	0.9750	0.9729	0.9711

Table 5.5: Evaluation metrics comparing $\mathcal{M}(\Theta(D), \delta)_1$ and $\mathcal{M}(\theta_T, \delta)_1$.

	$\delta = 1$	$\delta = 5$	$\delta = 21$	$\delta = 42$	$\delta = 63$
RMSE	0.0209	0.0214	0.0222	0.0229	0.0235
MAE	0.0108	0.0111	0.0116	0.0121	0.0125
MAPE	0.1603	0.1492	0.1431	0.1421	0.1423
R-Squared	0.9031	0.9034	0.9029	0.9014	0.8997

Table 5.6: Evaluation metrics comparing $\mathcal{M}(\Theta(D), \delta)_2$ and $\mathcal{M}(\theta_T, \delta)_2$.

In a consistent manner, the majority of cost metrics demonstrate lower values, and the R-squared value consistently shows higher values when using $\mathcal{M}(\theta_T, \delta)$ as targeted values instead of empirical moment estimators. Although the difference is relatively small for conditional expectation estimators, it becomes more significant for conditional standard deviation estimators. Specifically, the R-squared is 10 to 20 points lower when using $\mathcal{M}(\theta_T, \delta)_2$ as the targeted value compared to the sample standard deviation. This suggests that a significant portion of the discrepancy between the estimator $\mathcal{M}(\theta_T, \delta)_2$ and the empirical estimator of the conditional standard deviation can be attributed to the variance of the latter. In practical terms, this finding further strengthens the notion that the future volatility distributions associated with the θ -vectors estimated by Θ closely align with the actual future volatility distributions, not only in terms of the mean but also in terms of the standard deviation.

5.4.1.3.2 Evaluation based on conditional distributions using the Kolmogorov-Smirnov test

In addition to evaluating the estimator function through conditional moments, it is important to examine the consistency between the estimated θ -vectors and the true θ -vectors in terms of the associated conditional distributions. To address this aspect, we conduct a statistical experiment to assess the adequacy of the estimated θ -vectors by Θ .

The first step of this experiment involves estimating each θ -vector associated with each matrix $D^{(i)}$ using the Θ method. Subsequently, we generate 100 simulations for each estimated vector, considering the p time horizons of interest to us: 1, 5, 21, 42, and 63 trading days. For each combination of $(\theta_T^{(i)}, \hat{\theta}_T^{(i)})$ and for each time horizon, we employ the Kolmogorov-Smirnov (KS) test to calculate the p-value between the simulated volatility sample generated from the estimated θ -vector by Θ and the sample generated from the true θ -vector. The p-value indicates the likelihood of observing a discrepancy as large as or larger than the one observed, assuming both samples

are drawn from the same distribution. By computing the proportion of non-rejection of the null hypothesis of the KS test at different significance levels, we can evaluate the agreement between the estimated and true θ -vectors regarding the underlying conditional distributions. The results of this analysis are presented in the following table, providing valuable insights into the robustness and reliability of the estimation procedure conducted by Θ .

	$\delta = 1$	$\delta = 5$	$\delta = 21$	$\delta = 42$	$\delta = 63$
Proportion (%) at a significance level of 0.1%	93.0	94.8	96.4	97.1	97.2
Proportion (%) at a significance level of 1%	84.7	88.4	90.2	91.9	91.9
Proportion (%) at a significance level of 5%	76.1	80.9	83.5	84.9	85.2
Proportion (%) at a significance level of 10%	68.5	73.2	76.4	77.8	77.5

Table 5.7: Proportion of non-rejection of the null hypothesis of the KS test.

The results obtained demonstrate a high level of consistency between the conditional distributions generated from the estimated θ -vectors by Θ and those generated from the true θ -vectors. This consistency is evident through significant proportions of non-rejection observed across different significance levels for the various time horizons examined. Notably, even with a relatively high significance level of 10%, a substantial portion of the sample (ranging from 68.5% to 77.5% depending on the time horizon) does not reject the null hypothesis, indicating a strong agreement between the estimated and true θ -vectors concerning the associated conditional distributions. These findings emphasize the robustness and reliability of the estimation procedure performed by Θ in capturing the underlying future volatility distributions accurately.

Furthermore, it is interesting to note that the proportion of non-rejection of the null hypothesis increases with the time horizon. This phenomenon can be explained by two main factors already discussed in section 5.4.1.2. Firstly, as the time horizon increases, the data variability also increases, leading to conditional distributions with a larger standard deviation and, consequently, a greater acceptance of the null hypothesis. Secondly, as the time horizon grows, the significance of short-term information contained in the state variables associated with higher discount factors diminishes in its impact on the conditional distributions. However, these state variables are particularly challenging to estimate due to the daily observation frequency.

5.4.2 Evaluation of estimation procedure using market data

The objective of the estimation procedure presented in this chapter is to make the RPDV model a robust model for volatility prediction. The purpose of this section is therefore to evaluate the performance of the RPDV model on real data according to this objective.

5.4.2.1 Market data sets

To evaluate the performance of Θ on real data, the tests conducted in this section utilize historical data from 2000 to 2022 for five stock indices: S&P500, Nasdaq, FTSE, DAX, and Euro Stoxx 50.

These historical datasets consist of daily observations for each index, including its corresponding value and the realized volatility over the day, annualized. To create the input matrices D for the estimator Θ , a rolling window approach is employed. Specifically, a window of size 1260×2 is slid with a step of 1 trading day over the 22-year historical period. This process generates the matrices D of dimension 1260×2 that are used as inputs for the estimator Θ . Furthermore, the prediction horizons considered are 1, 5, 21, 42, and 63 trading days. Therefore, from a historical period of 5544 trading days, a total of 4222 ($5544 - 1260 + 1 - 63$) test pairs are obtained: $\left\{ D^{(i)}, \left\{ \tilde{\sigma}_{i+\delta_k} \right\}_{k=1}^5 \right\}_{i=1}^{4222}$.

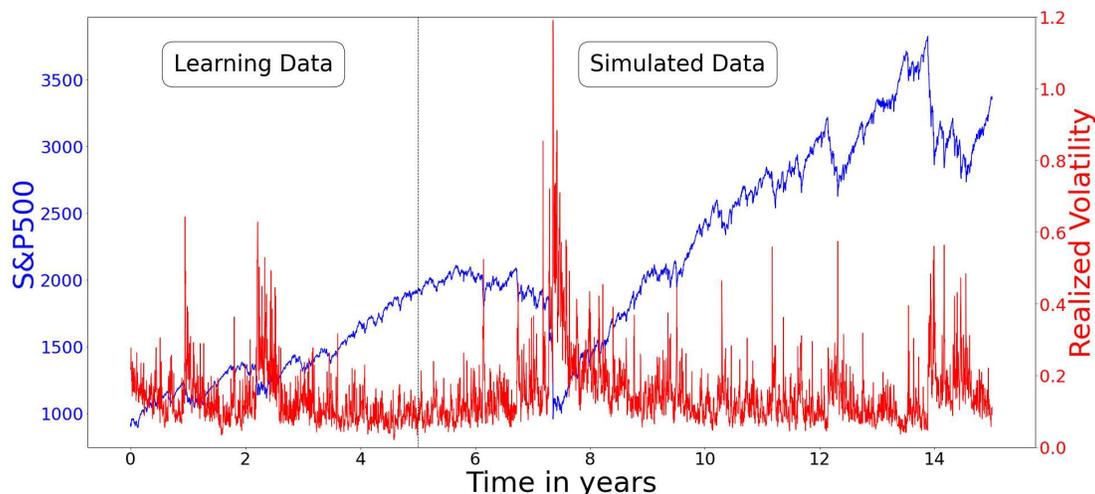


Figure 5.9: Example of joint evolution of the S&P500 and its realized volatility: the first 5 years are real data used to estimate θ_T from Θ , followed by 10 simulated years using this θ -vector.

5.4.2.2 Comparison of model forecasts with benchmark volatility models

The objective of this section is to assess the performance of the RPDV model on market data presented in section 5.4.2.1. For this purpose, we use as volatility forecaster, the estimator of $\mathbb{E}[\sigma_{T+\delta} | D]$:

$$\hat{\sigma}_{T+\delta} = \mathcal{M}(\Theta(D), \delta)_1,$$

which we then compare with the following benchmark models from academic literature ([104], [182]):

- The autoregressive (AR) models of order 5 and 21 (with trading days frequency), which take the following form for an order p model:

$$\sigma_t = a + \sum_{k=1}^p b_k \cdot \sigma_{t-k}.$$

- The heterogeneous autoregressive (HAR) models introduced by Corsi ([74]) HAR

$$\hat{\sigma}_t = a + b_1 \sigma_{t-1} + b_2 \sum_{k=1}^5 \sigma_{t-k} + b_3 \sum_{k=1}^{21} \sigma_{t-k}.$$

- The rough fractional stochastic volatility (RFSV) model introduced by Gatheral *et al.* ([105])

$$\hat{\sigma}_t = \exp \left(\frac{\cos(H\pi)}{\pi} \int_{-\infty}^{t-1} \frac{\sigma_s}{(t-s+1)(t-s)^{H+0.5}} ds + \frac{G(1.5-H)\nu^2}{2G(H+0.5)G(2-2H)} \right).$$

where $G(\cdot)$ denotes the gamma function. In practice, we truncate the integral to 1260 trading days and approximate it using a Riemann sum.

Each of these models is estimated using the data contained in the first column of the matrices D , which represents the historical realized volatility over the past 1260 trading days (approximately 5 years). The parameters of the AR and HAR models are estimated using the ordinary least squares method, while the FSVM is estimated using the method proposed by Gatheral. The accuracy of these different forecasters is evaluated by calculating the MSE between their respective forecasts and realized volatility, using the market data considered in section 5.4.2.1.

The results presented in table 5.8 demonstrate that, in most of the cases examined, the RPDV predictions outperform alternative models in terms of forecast accuracy. However, it is important to note that the predictive ability of RPDV heavily relies on the specific time horizon being considered. Specifically, for all 1-day horizon forecasts, the MSE-based performance of RPDV is inferior to that of the AR models or RFSV. Nevertheless, its relative performance significantly improves for the 5 trading days horizon, where it becomes the most accurate model in 3 out of 5 cases. Moreover, for longer time horizons such as 1, 2, and 3 months, RPDV consistently outperforms other volatility forecasters. In general, as the time horizon increases, RFSV forecasts tend to outperform those of other models.

The differential performance based on the considered time horizons could be attributed to two different reasons. The first one is the chosen model. This one is purely path-dependent, whereas empirical data suggests that volatility dynamics are also driven by exogenous frictions ([117], chapter 3). Additionally, the kernels associated with the variables R_1 and R_2 follow a power law, while shifted power law kernels better fit market data according to Guyon and Lekeufack ([117]). Therefore, using a model with more flexible kernels and incorporating an exogenous component of volatility could improve the performance of the approach. The second possible cause could be the estimation method itself. Indeed, unlike the other models considered, Θ is trained only using synthetic data with a constant observation time step of $\frac{1}{252}$ of a year, whereas the real data has an uneven observation frequency due to factors such as the presence of weekends. Thus, the empirical observation gap between two consecutive weekdays is $\frac{1}{365}$ of a year, and $\frac{3}{365}$ of a year between Friday and the following Monday. This bias can have a relatively strong impact on the estimation

of volatility at a short time horizon and tends to be smoothed out as the time horizon increases. Nevertheless, this is currently a limitation of the proposed estimation method that is worth highlighting. To address this limitation, one potential approach is to incorporate a combination of synthetic and real data in the training of Θ . By including real data with varying observation frequencies, the model can better adapt to the characteristics of empirical data. Additionally, introducing noise or biases in the training data can enhance the robustness of the θ -vector estimation and help reduce potential biases.

	AR(5)	AR(21)	HAR	RFSV	RPDV
SPX $\delta = 1$	0.0030	0.0031	0.0036	0.0031	0.0033
SPX $\delta = 5$	0.0052	0.0052	0.0052	0.0051	0.0050
SPX $\delta = 21$	0.0087	0.0083	0.0081	0.0080	0.0076
SPX $\delta = 42$	0.0106	0.0100	0.0098	0.0094	0.0087
SPX $\delta = 63$	0.0111	0.0103	0.0103	0.0099	0.0089
Nasdaq $\delta = 1$	0.0021	0.0022	0.0027	0.0022	0.0028
Nasdaq $\delta = 5$	0.0039	0.0040	0.0039	0.0038	0.0040
Nasdaq $\delta = 21$	0.0064	0.0063	0.0059	0.0057	0.0058
Nasdaq $\delta = 42$	0.0076	0.0072	0.0069	0.0066	0.0064
Nasdaq $\delta = 63$	0.0080	0.0075	0.0073	0.0069	0.0065
FTSE $\delta = 1$	0.0038	0.0039	0.0042	0.0038	0.0040
FTSE $\delta = 5$	0.0055	0.0053	0.0053	0.0052	0.0053
FTSE $\delta = 21$	0.0086	0.0077	0.0076	0.0073	0.0074
FTSE $\delta = 42$	0.0101	0.0090	0.0088	0.0083	0.0083
FTSE $\delta = 63$	0.0107	0.0094	0.0093	0.0088	0.0086
DAX $\delta = 1$	0.0026	0.0026	0.0031	0.0026	0.0029
DAX $\delta = 5$	0.0045	0.0044	0.0042	0.0042	0.0041
DAX $\delta = 21$	0.0075	0.0067	0.0064	0.0062	0.0060
DAX $\delta = 42$	0.0090	0.0082	0.0076	0.0073	0.0068
DAX $\delta = 63$	0.0096	0.0086	0.0081	0.0078	0.0072
Stox50 $\delta = 1$	0.0039	0.0040	0.0046	0.0039	0.0042
Stox50 $\delta = 5$	0.0062	0.0061	0.0061	0.0060	0.0058
Stox50 $\delta = 21$	0.0099	0.0091	0.0088	0.0084	0.0081
Stox50 $\delta = 42$	0.0110	0.0105	0.0101	0.0095	0.0090
Stox50 $\delta = 63$	0.0114	0.0106	0.0105	0.0099	0.0092

Table 5.8: MSE for the AR, HAR, RFSV and RPDV predictors.

5.4.2.3 Evaluation by density

The evaluation of standard deviations for conditional volatility distributions is not directly possible from historical data since, by definition, we only have a single realization for each date. Therefore, we proceed indirectly by using the approximation of volatility distributions with the log-normal distribution introduced in section 5.2.3. In this framework, our estimator for the volatility distribution

at horizon $T + \delta$ at time T is the log-normal distribution $\mathcal{LN}(\tilde{m}_{T+\delta}, (\tilde{s}_{T+\delta})^2)$, where:

$$\tilde{m}_{T+\delta} = \log(\mathcal{M}(\Theta(\mathbf{D}), \delta)_1) - 0.5(\tilde{s}_{T+\delta})^2 \quad \text{and} \quad (\tilde{s}_{T+\delta})^2 = \log\left(\left(\frac{\mathcal{M}(\Theta(\mathbf{D}), \delta)_2}{\mathcal{M}(\Theta(\mathbf{D}), \delta)_1}\right)^2 + 1\right).$$

Using the properties of the log-normal distribution, we define the estimator of the cumulative distribution function (CDF) of $\sigma_{T+\delta}$ at time T as follows:

$$\hat{F}_{T+\delta}(\sigma) = 0.5 + 0.5 \cdot \operatorname{erf}\left(\frac{\log(\sigma) - \tilde{m}_{T+\delta}}{\tilde{s}_{T+\delta}\sqrt{2}}\right). \quad (5.21)$$

We then proceed to calculate the proportion of observed realized volatility values that fall within different confidence intervals constructed based on this estimated CDF. Specifically, we compute the included proportion of the sample that falls within:

- the bilateral confidence interval $[\alpha/2 : 1 - \alpha/2]$:

$$p_{1-\alpha} = \frac{\sum_{k=1}^N \mathbb{1}\left\{\frac{\alpha}{2} \leq \hat{F}_{T_k+\delta}(\sigma_{T_k+\delta}) \leq 1 - \frac{\alpha}{2}\right\}}{N},$$

- the upper unilateral confidence interval $[\alpha : 1]$:

$$p_{1-\alpha}^{(+)} = \frac{\sum_{k=1}^N \mathbb{1}\left\{\alpha \leq \hat{F}_{T_k+\delta}(\sigma_{T_k+\delta})\right\}}{N},$$

- the lower unilateral confidence interval $[0 : 1 - \alpha]$:

$$p_{1-\alpha}^{(-)} = \frac{\sum_{k=1}^N \mathbb{1}\left\{\hat{F}_{T_k+\delta}(\sigma_{T_k+\delta}) \leq 1 - \alpha\right\}}{N}.$$

These calculations are performed for the following values of α : 0.05, 0.1, 0.25, 0.5. The idea is then to compare the theoretical proportion, which should be $1 - \alpha$, with the calculated proportions $p_{1-\alpha}$, $p_{1-\alpha}^{(+)}$, and $p_{1-\alpha}^{(-)}$. Indeed, the closer the calculated proportions $p_{1-\alpha}$, $p_{1-\alpha}^{(+)}$, and $p_{1-\alpha}^{(-)}$ are to $1 - \alpha$, the stronger the indication that $\hat{F}_{T_k+\delta}$ is a reliable estimator of the conditional distributions of volatility at horizon δ .

The figures reported in the table 5.9 demonstrate that our estimator of conditional volatility distributions generally provides a good approximation of the actual conditional distributions. The proportions of realized volatility included in the estimated confidence intervals are typically close to the theoretical proportions (i.e., $1 - \alpha$), indicating that the estimator effectively captures the characteristics of the conditional distributions. This, in turn, suggests that $\mathcal{M}(\Theta(\mathbf{D}), \delta)_2$ produces a good estimation of conditional standard deviations. However, it should be noted that, as already mentioned in section 5.4.2.2, the quality of the model estimations is quite sensitive to the consid-

ered time horizon. Thus, while the difference between the theoretical proportion and the observed proportion inside the confidence intervals is around 10 points in most cases for a 1-day horizon, this difference is almost always less than 5 points for horizons equal to or greater than 1 month.

	$p_{0.95}$	$p_{0.95}^{(+)}$	$p_{0.95}^{(-)}$	$p_{0.90}$	$p_{0.90}^{(+)}$	$p_{0.90}^{(-)}$	$p_{0.75}$	$p_{0.75}^{(+)}$	$p_{0.75}^{(-)}$	$p_{0.50}$	$p_{0.50}^{(+)}$	$p_{0.50}^{(-)}$
SPX $\delta = 1$	0.86	0.90	0.91	0.80	0.85	0.85	0.65	0.72	0.72	0.44	0.50	0.50
SPX $\delta = 5$	0.89	0.92	0.91	0.83	0.88	0.85	0.67	0.74	0.69	0.44	0.53	0.47
SPX $\delta = 21$	0.90	0.92	0.93	0.85	0.88	0.88	0.70	0.74	0.71	0.46	0.52	0.48
SPX $\delta = 42$	0.90	0.92	0.94	0.86	0.87	0.89	0.71	0.75	0.73	0.48	0.52	0.48
SPX $\delta = 63$	0.91	0.92	0.94	0.86	0.88	0.89	0.73	0.75	0.74	0.49	0.52	0.49
Nasdaq $\delta = 1$	0.87	0.90	0.91	0.81	0.85	0.86	0.65	0.70	0.73	0.43	0.47	0.53
Nasdaq $\delta = 5$	0.89	0.93	0.89	0.82	0.88	0.83	0.68	0.75	0.70	0.45	0.51	0.49
Nasdaq $\delta = 21$	0.89	0.93	0.92	0.84	0.88	0.86	0.69	0.75	0.72	0.46	0.50	0.50
Nasdaq $\delta = 42$	0.90	0.92	0.94	0.85	0.88	0.88	0.71	0.75	0.73	0.48	0.52	0.48
Nasdaq $\delta = 63$	0.90	0.92	0.94	0.86	0.88	0.88	0.72	0.75	0.74	0.49	0.51	0.49
FTSE $\delta = 1$	0.86	0.88	0.89	0.79	0.82	0.82	0.63	0.68	0.68	0.42	0.47	0.52
FTSE $\delta = 5$	0.88	0.89	0.92	0.82	0.83	0.87	0.65	0.71	0.73	0.43	0.49	0.51
FTSE $\delta = 21$	0.90	0.91	0.93	0.84	0.86	0.86	0.68	0.72	0.74	0.46	0.49	0.51
FTSE $\delta = 42$	0.90	0.92	0.94	0.85	0.86	0.88	0.69	0.72	0.75	0.47	0.50	0.50
FTSE $\delta = 63$	0.91	0.91	0.94	0.85	0.86	0.88	0.70	0.72	0.76	0.48	0.50	0.50
DAX $\delta = 1$	0.88	0.89	0.91	0.83	0.83	0.87	0.66	0.71	0.73	0.45	0.50	0.50
DAX $\delta = 5$	0.89	0.90	0.94	0.84	0.85	0.89	0.68	0.73	0.75	0.47	0.51	0.49
DAX $\delta = 21$	0.90	0.91	0.94	0.86	0.86	0.90	0.69	0.73	0.76	0.49	0.50	0.50
DAX $\delta = 42$	0.91	0.91	0.94	0.86	0.87	0.91	0.70	0.73	0.77	0.49	0.50	0.50
DAX $\delta = 63$	0.91	0.91	0.94	0.87	0.87	0.92	0.71	0.73	0.78	0.50	0.50	0.50
Stox $\delta = 1$	0.87	0.89	0.91	0.82	0.84	0.83	0.65	0.70	0.70	0.43	0.49	0.51
Stox $\delta = 5$	0.89	0.91	0.92	0.84	0.86	0.85	0.67	0.71	0.71	0.45	0.50	0.50
Stox $\delta = 21$	0.90	0.92	0.92	0.85	0.87	0.87	0.69	0.73	0.72	0.48	0.51	0.49
Stox $\delta = 42$	0.91	0.92	0.94	0.86	0.88	0.88	0.70	0.73	0.73	0.49	0.51	0.49
Stox $\delta = 63$	0.91	0.92	0.94	0.87	0.87	0.88	0.71	0.73	0.74	0.50	0.50	0.50

Table 5.9: Proportions of realized volatility samples included in estimated confidence intervals.

Another interesting point is that the narrower the confidence interval, the more accurate the model estimation, in the sense that the empirical proportions approach the theoretical proportions. Moreover, the empirical proportions are generally lower than the theoretical proportions, especially when considering wider confidence intervals. This phenomenon can be explained by several factors. The first, which is certainly the most important, is that observations outside the confidence intervals are simply the result of a poor model prediction for a part of the sample. This hypothesis is supported by the fact that these discrepancies are strongly correlated with the model's relative performance reported in table 5.8. Another factor explaining this discrepancy is an underestimation by the model of the conditional standard deviations. In fact, even if the model perfectly predicted the conditional means, such an underestimation of the standard deviations would lead to $p_{1-\alpha} < 1 - \alpha$. Finally, part of this difference could be also explained by the use of the log-normal approximation. Indeed, with constant mean and standard deviation, the kurtosis of the volatility distributions directly generated from RPDV model tends to be slightly higher than their log-normal approximations (it can be seen in figure 5.1).

5.5 Conclusion

The present chapter introduced an innovative deep estimation method for volatility models, specifically designed for volatility forecasting within the theoretical framework of Bayesian decision theory. To illustrate this method, the chapter focused on the estimation of an RPDV model.

The objective of the proposed approach was formally outlined in section 5.2. It involves constructing an estimator function for the considered model that, from a historical matrix of price and realized volatility data, returns optimal parameters and state variables according to Bayesian decision theory principles and based on the criterion defined in this section. This criterion, arising from a forecasting objective across different horizons, has been defined as a function of the first two conditional moments of volatility at various time horizons.

The estimation method itself has been exposed in section 5.3. It involves 2 NNs. The first one, denoted as Θ , serves as the estimator function. It takes a historical matrix of price and realized volatility data as input and returns a θ -vector containing the parameters and state variables defining a Markovian approximation of the RPDV model at a specific time instant. The second neural network, denoted as \mathcal{M} , estimates the mean and standard deviation of the volatility under the considered RPDV model for a given pair of θ -vector and time horizon. This NN thus addresses the absence of an analytical formula for these moments. The proposed estimation method first involves training this neural network \mathcal{M} , and then, in a second step, training Θ through interaction with \mathcal{M} . In this approach, \mathcal{M} is used to compute the cost of the θ -vectors predicted by Θ , thereby adjusting the parameters of the neural network Θ accordingly. Importantly, it has been demonstrated that under certain conditions, following the proposed estimation procedure, Θ behaves asymptotically as a Bayesian estimator aligned with the forecasting objective outlined in section 5.2. Consequently, the outputs of Θ offer estimations of the optimal θ -vectors tailored to the specified forecasting goal.

Section 5.4 has presented a comprehensive evaluation of the practical effectiveness of the estimator function Θ using both synthetic and market data. The evaluation on synthetic data demonstrated that the estimated θ -vectors by Θ yield volatility distribution estimates that closely align with the real distributions at different time horizons. These results highlighted the efficacy of the proposed estimation method within the analytical framework, where the estimation data are noise-free and generated from the model being estimated. The evaluation using market data, spanning 22 years of data from 5 stock indices, provided insights under less favorable conditions. The results showed a generally positive outcome, although with more mixed findings compared to the tests conducted on synthetic data. Notably, the model's performance as a volatility forecaster varied depending on the chosen time horizon. For the 1 trading day volatility forecast, the model exhibited lower

performance compared to benchmark forecasters such as AR and RSFV. However, for a 1-week horizon (5 trading days), the model's performance became comparable to, or even slightly better than, the benchmark models. Moreover, the model consistently outperformed other models for longer horizons of 1 month or more, including the HAR and RSFV models that are known for their effectiveness in volatility prediction over longer timeframes.

The reasons put forward to explain the differential performance based on the considered time horizon are of two kinds: the first is related to the chosen volatility model, the second to the estimation procedure itself. Regarding the first reason, a model allowing more flexible kernels and enabling the incorporation of an exogenous component of volatility could be better adapted to capture the empirical dynamics of volatility, thereby improving short-term forecasting. Regarding the estimation procedure itself, the fact that Θ is trained solely on synthetic data with a constant observation time step, while the observation frequency varies for empirical data, could introduce a bias in the prediction of state variables, which diminishes as the prediction time horizon decreases. To address this limitation, one potential approach could be to incorporate a combination of synthetic and real data in the training of Θ . Additionally, introducing noise or biases in the training data may also enhance the robustness of the θ -vector estimation and help reduce potential biases. These avenues for improvement could refine the estimation framework presented in this chapter, which already demonstrates promising results, particularly in utilizing RPDV as a volatility predictor for medium to long horizons.

5.6 Acknowledgments

I would like to express my gratitude to my thesis supervisor, Jean-Paul Laurent, for his invaluable support and insightful advice throughout the preparation of this chapter.

Appendix 5.A Approximation of the considered RPDV model

5.A.1 Stochastic differential equations for the Markovian approximation of the considered RPDV model

We aim to solve the following SDE:

$$dR_{1,i,t} = \gamma_i \left(\frac{dP_t}{P_t} - (\kappa_1 R_{1,t} + R_{1,i,t}) dt \right).$$

To consider the dynamics of $R_{1,i,t}$, we set $g(R_{1,i,t}, t) = e^{\gamma_i t} R_{1,i,t}$ and apply the Itô lemma:

$$de^{-\gamma_i t} R_{1,i,t} = \gamma_i e^{\gamma_i t} R_{1,i,t} dt + e^{\gamma_i t} dR_{1,i,t} = \gamma_i e^{\gamma_i t} \left(\frac{dP_t}{P_t} - \kappa_1 R_{1,t} dt \right).$$

Consequently:

$$R_{1,i,t} = R_{1,i,0} e^{-\gamma_i t} + \gamma_i \int_0^t e^{-\gamma_i(t-u)} \left(\frac{dP_u}{P_u} - \kappa_1 R_{1,u} du \right).$$

Thus:

$$\sum_{i=1}^n w_{1,i} R_{1,i,t} = R_{1,t} = \sum_{i=1}^n \gamma_i w_{1,i} e^{-\gamma_{1,i} t} R_{1,i,0} + \underbrace{\int_0^t \sum_{i=1}^n \gamma_i w_{1,i} e^{-\gamma_{1,i}(t-u)} \left(\frac{dP_u}{P_u} - \kappa_1 R_{1,u} du \right)}_{\hat{K}_1(t-u)}.$$

It follows that

$$\lim_{t \rightarrow +\infty} R_{1,t} = \int_0^t \hat{K}(t-u) \left(\frac{dP_u}{P_u} - \kappa_1 R_{1,u} du \right).$$

Analogously, with

$$dR_{2,i,t} = \left((\sigma_t)^2 - \kappa_2 R_{2,t} - \gamma_i R_{2,i,t} \right) dt,$$

by applying same steps, we obtain:

$$\sum_{i=1}^n w_{2,i} R_{2,i,t} = R_{2,t} = \sum_{i=1}^n \gamma_i w_{2,i} e^{-\gamma_{2,i} t} R_{2,i,0} + \int_0^t \sum_{i=1}^n w_{2,i} e^{-\gamma_{2,i}(t-u)} \left((\sigma_u)^2 - \kappa_2 R_{2,t} \right) du,$$

and therefore

$$\lim_{t \rightarrow +\infty} R_{2,t} = \int_0^t \hat{K}(t-u) \left((\sigma_u)^2 - \kappa_2 R_{2,t} \right) du.$$

5.A.2 Approximation of the power law kernel

Chapter 3 has shown that the vectors W_j and Λ_j can be determined using the work of Abi Jaber ([1], [2]), which is based on the expression of the kernel $K_j(\tau_j) = \tau^{-\alpha_j}$ as the Laplace transform of a positive measure⁷. However, this method has several drawbacks. The first is that the convergence between K_j and \hat{K}_j is relatively slow with respect to n , the number of exponential kernels that make up \hat{K}_j . Consequently, when n is small ($n \leq 10$), there exists kernels of the same form (and with the same n) that better approximate the power kernel in the L^2 sense. Furthermore, the discount coefficients obtained through this method depend on α . While this is not inherently a problem, it complexify the estimation problem in practice. For these reasons, an alternative approximation of K_j is used here, in which the discount coefficients are constant (they do not depend on α_j) and only the weight vector W_j vary.

In order to fix Λ_j , we start to remark that the inverse of the discount coefficients $\gamma_{j,i}$ corresponds to the duration of $(R_{j,t})_i$. Based on that, we start by defining the shortest and longest durations as $\tau_- = \gamma_{j,1}^{-1}$ and $\tau_+ = \gamma_{j,n}^{-1}$, respectively. In this case, we set $\tau_- = \frac{1}{10000}$ and $\tau_+ = 1000$ expressed in years. Subsequently, we perform a uniform logarithmic discretization between these two bounds to determine the values of the remaining $n - 2$ discounting coefficients $(\gamma_{j,i})_{2 \leq i \leq n-1}$, as follows:

$$\gamma_{j,i} = \exp \left(\log(\tau_-) + \frac{\log(\tau_+) - \log(\tau_-)}{n-1} (i-1) \right)^{-1}.$$

The idea is to have a set of exponential kernels with durations distributed in such a way as to be able to well approximate any power law kernel $K(\tau) = \tau^{-\alpha}$ with $\alpha \in]0 : 1[$. With the value of Λ_j fixed, we then solve the following least-square problem:

$$\arg \min_{W_j \geq \mathbf{0}_n} \|y_j - A_j W_j\|^2$$

with $\mathbf{0}_n$ a n -dimensional vector of zeros,

$$A_j = \begin{bmatrix} \gamma_{j,1} e^{-\gamma_{j,1} \tau_1} & \dots & \gamma_n e^{-\gamma_{j,n} \tau_1} \\ \vdots & \ddots & \vdots \\ \gamma_{j,1} e^{-\gamma_{j,1} N \tau_N} & \dots & \gamma_n e^{-\gamma_{j,n} N \tau_N} \end{bmatrix}, \quad y_j = \begin{bmatrix} \tau_1^{-\alpha} \\ \dots \\ \tau_N^{-\alpha} \end{bmatrix}.$$

In order to evaluate the quality of the approximation obtained by this method, we will compare it with the method proposed by Abi Jaber ([1]), using $n = 10$ in both cases. Accordingly, we compute the L^1 and L^2 norms of the difference over the time interval $[\frac{1}{10000} : 10]$ between power-law kernels and their associated approximations using each of these 2 methods. Table 5.10 reports the results obtained.

⁷Bayer and Breneis ([22]) consider an analogous approximation approach that will not be considered here.

α	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9	1
$\ \hat{K}_1 - K\ _{L^1(T)}$	2.443	1.475	0.873	0.586	0.518	0.489	0.563	0.844	1.604
$\ \hat{K}_2 - K\ _{L^1(T)}$	0.016	0.026	0.041	0.065	0.107	0.188	0.355	0.449	0.970
$\ \hat{K}_1 - K\ _{L^2(T)}$	0.618	0.22	0.105	0.218	1.391	9.51	65.5	436.14	2809.4
$\ \hat{K}_2 - K\ _{L^2(T)}$	0.0004	0.004	0.033	0.228	1.412	8.11	43.8	66.51	470.8

Table 5.10: Comparison of two power-law kernel approximation methods based on L^1 and L^2 norms evaluated over the time interval $[\frac{1}{10000} : 10]$. \hat{K}_1 is the approximation method introduced by Abi Jaber, while \hat{K}_2 uses the method proposed in this section.

Based on the metrics considered, the method proposed here generally provides a better approximation of the power-law kernel than Abi Jaber’s method for most of the considered α values. If we focus on the L^1 norm criterion, this method produces systematically a better approximation for all the considered α values. When we consider the L^2 norm criterion, Abi Jaber’s approximation outperforms the method introduced in the present section for the cases where α is equal to 0.5 and 0.6. However, even in these two cases, the difference in performance is very small.

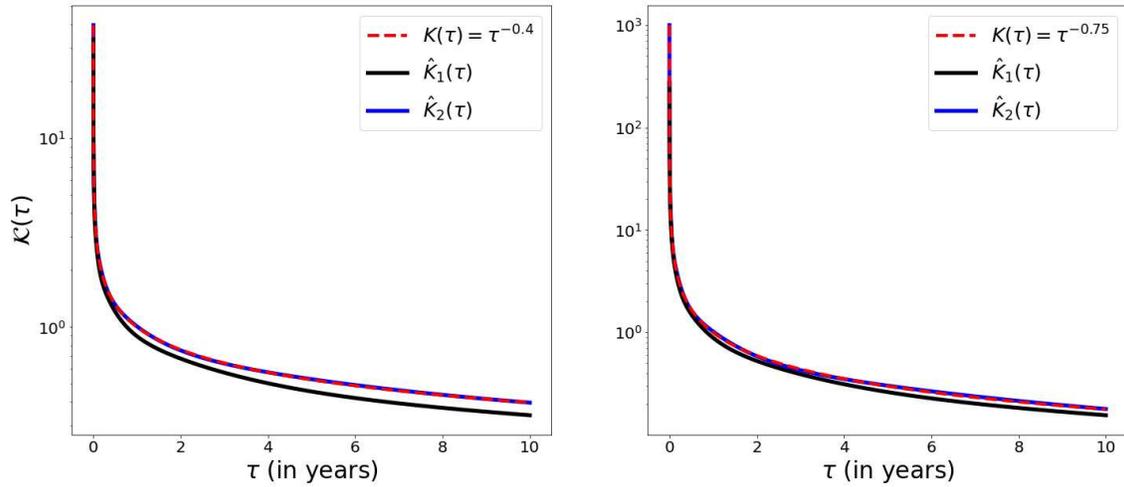


Figure 5.10: Examples of approximations of two power-law kernels using 2 different approximation methods. Kernel \hat{K}_1 is obtained using the Abi Jaber approximation, while kernel \hat{K}_2 is obtained using the approximation described in this section.

Appendix 5.B Parameters of the log-normal approximation to the conditional volatility distribution

We want to express m and s in terms of $\mathbb{E}[\sigma_{T+\delta}|\theta_T]$ and $\text{Var}[\sigma_{T+\delta}|\theta_T]$ given:

$$\mathbb{E}[\sigma_{T+\delta}|\theta_T] = e^{m+\frac{s^2}{2}}, \quad \text{Var}[\sigma_{T+\delta}|\theta_T] = (e^{s^2} - 1) e^{2m+s^2}.$$

It is clear that the first equation can be rewrite as

$$m = \log \left(\mathbb{E}[\sigma_{T+\delta}|\theta_T] \right) - \frac{s^2}{2}.$$

Injecting this result in the second equation, we obtain:

$$\begin{aligned} \text{Var}[\sigma_{T+\delta}|\theta_T] &= (e^{s^2} - 1) e^{2\log(\mathbb{E}[\sigma_{T+\delta}|\theta_T])} \\ s^2 &= \log \left(\frac{\text{Var}[\sigma_{T+\delta}|\theta_T]}{\mathbb{E}[\sigma_{T+\delta}|\theta_T]^2} + 1 \right) \end{aligned}$$

Therefore,

$$m = \log \left(\mathbb{E}[\sigma_{T+\delta}|\theta_T] \right) - 0.5 \log \left(\frac{\text{Var}[\sigma_{T+\delta}|\theta_T]}{\mathbb{E}[\sigma_{T+\delta}|\theta_T]^2} + 1 \right).$$

Appendix 5.C Proofs of convergence results

5.C.1 Convergence of the estimation procedure of the function \mathcal{M}

Proposition 1 *Let be $\theta_T^{(1)}, \dots, \theta_T^{(n_1)}$ a sequence of i.i.d. random variable following π , and $\left\{ \bar{M}_{T+\delta_k}^{(1)} \right\}_{1 \leq j \leq p}$, $\dots, \left\{ \bar{M}_{T+\delta_k}^{(n_1)} \right\}_{1 \leq j \leq p}$ a sequence of sets such as $\forall i, k, \bar{M}_{T+\delta_k}^{(i)}$ is an unbiased estimator of $M(\theta_T^{(i)}, \delta_k)$ calculated from a sample of size n_2 . If it exists \mathcal{M}^* , such as $\mathcal{M}^*(\theta_T, \delta_k) = M(\theta_T, \delta_k)$, $\forall \theta_T : \pi(\theta_T) \neq 0$ and $\delta_k \in \{\delta_1, \dots, \delta_p\}$, thus $\forall \hat{\mathcal{M}}^*$ solution to*

$$\arg \min_{\mathcal{M}} \frac{1}{n_1} \sum_{i=1}^{n_1} \sum_{k=1}^p \left\| \mathcal{M}(\theta_T^{(i)}, \delta_k) - \bar{M}_{T+\delta_k}^{(i)} \right\|_2^2,$$

$\hat{\mathcal{M}}^*(\theta, \delta_k) = M(\theta_T, \delta_k)$, $\forall \theta_T : \pi(\theta_T) \neq 0$ and $\delta_k \in \{\delta_1, \dots, \delta_p\}$.

Proof of proposition 1. The density being by definition positive or zero, we have the following inequality:

$$\begin{aligned} \int_{\mathbb{R}^{2n+9}} \left(\min_{\mathcal{M}} \sum_{k=1}^p \|\mathcal{M}(\theta_T, \delta_k) - M(\theta_T, \delta_k)\|_2^2 \right) d\boldsymbol{\pi}(\theta_T) \\ \leq \min_{\mathcal{M}} \int_{\mathbb{R}^{2n+9}} \sum_{k=1}^p \|\mathcal{M}(\theta_T, \delta_k) - M(\theta_T, \delta_k)\|_2^2 d\boldsymbol{\pi}(\theta_T). \end{aligned}$$

In addition, $\forall M \in \mathbb{R}^2$, M is the unique solution to

$$\arg \min_{\hat{M}} \|M - \hat{M}\|_2^2.$$

It follows that if it exists \mathcal{M}^* such as $\mathcal{M}^*(\theta_T, \delta_k) = M(\theta_T, \delta_k)$, $\forall \theta : \boldsymbol{\pi}(\theta_T) \neq 0$, and $\delta_k \in \{\delta_1, \dots, \delta_p\}$, thus $\forall \hat{\mathcal{M}}^*$ solution to

$$\arg \min_{\mathcal{M}} \int_{\mathbb{R}^{2n+9}} \sum_{k=1}^p \|\mathcal{M}(\theta_T, \delta_k) - M(\theta_T, \delta_k)\|_2^2 d\boldsymbol{\pi}(\theta),$$

$\hat{\mathcal{M}}^*(\theta_T, \delta_k) = M(\theta_T, \delta_k)$, $\forall \theta : \boldsymbol{\pi}(\theta) \neq 0$. Moreover, because $\bar{M}_{T+\delta_k}^{(i)}$ is an unbiased estimator of $M(\theta_T^{(i)}, \delta_k)$ calculated from a sample of size n_2 ⁸ :

$$\lim_{n_2 \rightarrow +\infty} \bar{M}_{T+\delta_k}^{(i)} = M(\theta_T^{(i)}, \delta_k).$$

Similarly, by the law the of large numbers

$$\lim_{\substack{n_1 \rightarrow +\infty \\ n_2 \rightarrow +\infty}} \frac{1}{n_1} \sum_{i=1}^{n_1} \sum_{k=1}^p \left\| \mathcal{M}(\theta_T^{(i)}, \delta_k) - \bar{M}_{T+\delta_k}^{(i)} \right\|_2^2 = \int_{\mathbb{R}^{2n+9}} \sum_{k=1}^p \|\mathcal{M}(\theta_T, \delta_k) - M(\theta_T, \delta_k)\|_2^2 d\boldsymbol{\pi}(\theta_T).$$

Therefore, using previous results, under the existence condition of \mathcal{M}^* , $\forall \hat{\mathcal{M}}^*$ solution to

$$\arg \min_{\mathcal{M}} \lim_{\substack{n_1 \rightarrow +\infty \\ n_2 \rightarrow +\infty}} \frac{1}{n_1} \sum_{i=1}^{n_1} \sum_{k=1}^p \left(\mathcal{M}(\theta_T^{(i)}, \delta_k) - \bar{M}_{T+\delta_k}^{(i)} \right)^2,$$

$\hat{\mathcal{M}}^*(\theta_T, \delta_k) = M(\theta_T, \delta_k)$, $\forall \theta_T : \boldsymbol{\pi}(\theta_T) \neq 0$ and $\delta_k \in \{\delta_1, \dots, \delta_p\}$. QED.

5.C.2 Convergence of the estimation procedure of the estimator function Θ

Proposition 2 Let $\theta_{t_0}^{(1,1)}, \dots, \theta_{t_0}^{(n_1,1)}$ be a sequence of i.i.d. random variables following $\boldsymbol{\pi}$, $D^{(1)}, \dots, D^{(n_1)}$ a set of time-series such that $D^{(i)}$ is generated from the M-RPDV associated with the θ -vector $\theta_{t_0}^{(i)}$,

⁸We assume here that, $\forall \theta_T : \boldsymbol{\pi}(\theta_T) \neq 0$ and $[0 : \delta_p]$, $\sigma_{T+\delta_k}$ has finite variance.

and $\theta_T^{(1,2)}, \dots, \theta_T^{(n_1,2)}$ the set of values taken by θ at time t_N for each time series $D^{(i)}$. If there exists Θ^* such that for all $D : \mathbb{P}_\pi(D) \neq 0$, $\Theta^*(D)$ is a Bayes estimator of θ_T under the posterior measure π , then for any $\hat{\Theta}^*$ solution to the optimization problem

$$\arg \min_{\Theta} \lim_{n_1 \rightarrow +\infty} \frac{1}{n_1} \sum_{i=1}^{n_1} L\left(\theta_T^{(i)}, \Theta\left(D^{(i)}\right)\right),$$

$\hat{\Theta}^*(D)$ is a Bayes estimator of θ_T under the posterior measure $\pi_D, \forall D : \mathbb{P}_\pi(D) \neq 0$.

Proof of proposition 2. The expectation of the cost under the prior measure π is defined by:

$$\mathbb{E}_\pi \left[L(\boldsymbol{\theta}_T, \Theta(\mathbf{D})) \right] = \int_{\mathbb{R}_+^{N \times 2}} \mathbb{E}_{\pi_D} \left[L(\boldsymbol{\theta}_T, \Theta(\mathbf{D})) \right] d\mathbb{P}_\pi(\mathbf{D}).$$

Using this expression, and given density being by definition positive or zero, we have the following inequality:

$$\int_{\mathbb{R}_+^{N \times 2}} \left(\min_{\Theta} \mathbb{E}_{\pi_D} \left[L(\boldsymbol{\theta}_T, \Theta(\mathbf{D})) \right] \right) d\mathbb{P}_\pi(\mathbf{D}) \leq \min_{\Theta} \mathbb{E}_\pi \left[L(\boldsymbol{\theta}_T, \Theta(\mathbf{D})) \right].$$

It follows that if it exists Θ^* such as $\forall D : \mathbb{P}_\pi(D) \neq 0$, $\Theta^*(D)$ is a Bayes estimator of θ_T under the posterior measure π_D , if $\hat{\Theta}$ is solution to

$$\min_{\Theta} \mathbb{E}_\pi \left[L(\boldsymbol{\theta}_T, \Theta(\mathbf{D})) \right],$$

$\hat{\Theta}(D)$ is a Bayes estimator of θ under the posterior measure $\pi_D, \forall D : \mathbb{P}_\pi(D) \neq 0$.

In addition, if $D^{(1)}, \dots, D^{(n)}$ is a set of i.i.d. of time-series such as $D^{(i)} \sim \mathbb{P}_\pi$, by the law of large numbers

$$\lim_{n_1 \rightarrow +\infty} \frac{1}{n_1} \sum_{i=1}^{n_1} L\left(\theta_T^{(i)}, \Theta\left(D^{(i)}\right)\right) = \mathbb{E}_\pi \left[L(\boldsymbol{\theta}_T, \Theta(\mathbf{D})) \right].$$

Combining the above propositions, if there exists Θ^* such that for all $D : \mathbb{P}_\pi(D) \neq 0$, $\Theta^*(D)$ is a Bayes estimator of θ_T under the posterior measure π , then for any $\hat{\Theta}^*$ solution to the optimization problem

$$\arg \min_{\Theta} \lim_{n_1 \rightarrow +\infty} \frac{1}{n_1} \sum_{i=1}^{n_1} L\left(\theta_T^{(i)}, \Theta\left(D^{(i)}\right)\right),$$

$\hat{\Theta}^*(D)$ is a Bayes estimator of θ_T under the posterior measure $\pi_D, \forall D : \mathbb{P}_\pi(D) \neq 0$. QED.

Appendix 5.D Annex results

5.D.1 Value of the integral of \hat{K} over $\mathbb{R}_{\geq 0}$

We compute the integral of \hat{K} over $\mathbb{R}_{\geq 0}$:

$$\begin{aligned} \int_0^\infty \hat{K}(u) du &= \int_0^\infty \sum_{i=1}^n w_i \gamma_i e^{-\gamma_i(t-u)} du = \sum_{i=1}^n \left[w_i e^{-\gamma_i(t-u)} \right]_0^\infty \\ &= \sum_{i=1}^n w_i. \end{aligned}$$

5.D.2 Standard deviation of a BSS process

The variance of an integral of the form $\int_0^\infty \hat{K}(u) dW_u$ can be computed as follows:

$$\text{Var} \left(\sum_{i=1}^n \int_0^\infty w_i \gamma_i e^{-\gamma_i u} dW_u \right) = \int_0^\infty \left(\sum_{i=1}^n w_i \gamma_i e^{-\gamma_i u} \right)^2 du = \sum_{i=1}^n \sum_{j=1}^n \frac{w_i w_j \gamma_i \gamma_j}{\gamma_i + \gamma_j}.$$

It follows that:

$$\text{Std} \left(\sum_{i=1}^n \int_0^\infty w_i \gamma_i e^{-\gamma_i u} dW_u \right) = \sqrt{\sum_{i=1}^n \sum_{j=1}^n \frac{w_i w_j \gamma_i \gamma_j}{\gamma_i + \gamma_j}}.$$

5.D.3 The variance of the volatility process

The variance of the volatility process is equal to:

$$\begin{aligned} \text{Var}(\sigma_{T+\delta}) &= \text{Var} \left(\beta_0 + \beta_1 R_{1,T+\delta} + \beta_2 \sqrt{R_{2,T+\delta}} \right) \\ &= (\beta_1)^2 \text{Var} (R_{1,T+\delta}) + (\beta_2)^2 \text{Var} \left(\sqrt{R_{2,T+\delta}} \right) \\ &\quad + 2\beta_1 \beta_2 \rho_{T+\delta} \sqrt{\text{Var} (R_{1,T+\delta}) \text{Var} \left(\sqrt{R_{2,T+\delta}} \right)}. \end{aligned}$$

CHAPTER 6

The Factorial Path-Dependent Market Model

Abstract

This chapter introduces the factorial path-dependent market (FPDM) model, a multivariate asset price dynamics model in which these dynamics are determined by a set of elementary factors. In this framework, both the factorial drift and factorial volatilities are conditioned by the past dynamics of the factorial portfolios, resulting in a model mostly path-dependent. Building on this theoretical foundation, the chapter subsequently designs a market generator positioned midway between parametric models based on strong assumptions and purely data-driven approaches. The aim is to combine the best of both worlds, offering a model capable of faithfully reproducing the empirical financial dynamics while maintaining a clear understanding of the financial phenomena driven by the simulated price paths. To evaluate the effectiveness of the proposed approach, a thorough out-of-sample assessment of the market generator is conducted based on the S&P500 investment universe.

6.1 Introduction

"Backtesting against synthetic datasets should be the preferred approach for developing tactical investment algorithms" [115]. This statement from Marco Lopez De Prado underscores a key point: the importance of synthetic financial series in a modern quantitative finance approach. Beyond the interest for strategy backtesting, synthetic data are increasingly used by the industry for other applications such as stress testing ([115], [115], [177]), obtaining risk metrics ([73]), generating conditional scenarios ([19]), etc. However, the consistency of these synthetic data-based approaches depends to a large extent on the ability to accurately model multivariate asset price dynamics. Yet, this modeling represents one of the most challenging problems in quantitative finance.

The difficulty stems from a twofold complexity: a complexity inherent in marginal price dynamics (where each asset price is considered independently of the others) and a complexity in the correlation structure of asset price dynamics. On the first aspect, the distributions of asset returns (and log-returns) at different time scales are non-Gaussian, exhibit heavy tails, and are generally asymmetric

([146], [70]). Moreover, asset price dynamics are path-dependent, leading to significant time-series features such as volatility clustering and the Zumbach effect ([205], [106]). Regarding the correlation structure of asset price dynamics, it is itself dynamic and exhibits significant variability over time ([201]). Thus, a set of assets that are weakly correlated over a given period can be highly correlated over another period. Additionally, the joint distributions of returns for a significant portion of assets are non-elliptical ([64]), adding yet another layer of complexity to the modeling.

While classical multivariate dynamics models such as those based on multivariate geometric Brownian motion fail to capture these various empirical properties, more sophisticated approaches better suited to handle this complexity have emerged in recent years. In particular, models based on new machine learning methods have garnered particular interest. Among the most popular are models based on Restricted Boltzmann Machines ([133], [139]) and Generative Adversarial Networks ([139], [99], [164]), which have yielded very good results in terms of reproducing the various features characterizing financial time series. Nevertheless, these methods have also important limitations. Firstly, most of them are not suitable for handling the temporal dependence of large multi-dimensional data: as the data dimension increases, calibrating these models becomes impractical ([164]). However, in practice, asset portfolios are typically constructed from investment universes comprising several hundred to several thousand assets. Moreover, these models are typically black boxes. Therefore, while they can be very effective at reproducing the statistical properties of financial series, they do not provide an intelligible framework for the reproduced phenomena.

On the theoretical front, effectively modeling the dynamics of a price system goes beyond the technical question of "which model fits the data best?" Instead, it involves identifying the mechanisms through which the asset price system is formed and understanding the conditions that enable various historical market paths. In this respect, ambitious modeling should enable us to answer questions such as: "what causes a certain asset pair, which was weakly correlated at one time t , to become strongly correlated at another time t' ?" In broader terms, it must provide to some extent a hermeneutic of market dynamics. However, the value of such an approach goes beyond the simple search for knowledge for its own sake. It creates conditions that allow for the avoidance, or at least the mitigation, of generalization error better than a model based solely on the brute reproduction of historical sample characteristics, and thereby potentially anticipates the possibility of unprecedented but non-zero probability events.

It is in this perspective that the present chapter is situated. Its objective is to propose a general model for asset price dynamics that can be adapted into a market generator, bridging the gap between classical Monte Carlo approaches based on parametric models with strong assumptions and purely data-driven approaches without theoretical a priori. The aim is to combine the best of both worlds by offering a model capable of faithfully reproducing the empirical financial dynamics in all their complexity while maintaining a clear understanding of the financial phenomena driven by the simulated price paths. Furthermore, thanks to the theoretical framework induced by the model structure, this approach aims to disentangle contingent statistical phenomena, which are

merely expressions of specific historical realizations, from the structural mechanisms that generate the probability distribution of the asset price vector dynamics.

The article is structured as follows. Section 6.2 introduces the general framework of the Factorial Path-Dependent Market (FPDM) model and outlines the underlying logic of its structure. Section 6.3 derives a market generator from one specification of this model and presents a calibration method to the latter. Lastly, section 6.4 assesses this market generator and its calibration from various perspectives using market data.

6.2 General framework of the factorial Path-Dependent Market Model

6.2.1 General framework of the FPDM model

6.2.1.1 Elementary factor-based decomposition of asset price vector dynamics

Let us consider \mathbf{P} as the random vector of prices of dimension $n \times 1$ for an investment universe composed of n assets. Adopting a similar approach to that proposed in [171], we assume that the dynamics of \mathbf{P} are driven by a set of m factors denoted $\{\mathbf{F}_j\}_{1 \leq j \leq m}$, which is composed of m_C common factors and n idiosyncratic factors. We will refer to this set as the elementary factor set. For all t , $d\mathbf{P}_t$ is defined by the following multi-dimensional stochastic differential equation (SDE):

$$d\mathbf{P}_t = \mathbf{P}_t \odot (\mathbf{A}d\mathbf{F}_t), \quad (6.1)$$

where \mathbf{F} is the vector of elementary factors, and \mathbf{A} is an $n \times m$ factor loadings matrix that represents the sensitivity of the assets to the elementary factors, respectively given by:

$$\mathbf{F}_t^\top = \left(\left(\mathbf{F}_t^{(C)} \right)^\top, \left(\mathbf{F}_t^{(I)} \right)^\top \right) \quad \text{and} \quad \mathbf{A}^\top = \left(\left(\mathbf{A}^{(C)} \right)^\top, \mathbf{I}_n \right),$$

with $\mathbf{F}^{(C)}$ being the m_C -dimensional vector of common elementary factors, $\mathbf{F}^{(I)}$ the n -dimensional vector of idiosyncratic elementary factors, $\mathbf{A}^{(C)}$ the $n \times m_C$ matrix of the sensitivity of the assets to the common elementary factors, and \mathbf{I}_n is the n -dimensional identity matrix. Furthermore, we suppose that the dynamics of the elementary factors are given by:

$$d\mathbf{F}_t = \boldsymbol{\mu}_t dt + \sqrt{\boldsymbol{\Omega}_t} d\mathbf{W}_t, \quad (6.2)$$

with \mathbf{W} being an m -dimensional Brownian motion, $\boldsymbol{\mu}$ the drift vector of elementary factors of dimension $m \times 1$, and $\boldsymbol{\Omega}$ a $m \times m$ diagonal matrix with diagonal elements corresponding to the variance process of the elementary factors. In this framework, the solution to the asset price vector

is given by (see proof in appendix 6.B.1):

$$\mathbf{P}_t = \mathbf{P}_0 \odot \exp \circ \left(\int_0^t \mathbf{A} \boldsymbol{\mu}_u - \frac{1}{2} \cdot \text{diag} \left(\mathbf{A} \boldsymbol{\Omega}_u \mathbf{A}^\top \right) du + \int_0^t \mathbf{A} \sqrt{\boldsymbol{\Omega}_u} d\mathbf{W}_u \right).$$

Therefore, the instantaneous returns are expressed as a linear combination of elementary factor dynamics, with each factor associated with an independent source of randomness corresponding to the margin of \mathbf{W} . Furthermore, in this model, the drift vector and covariance matrix associated with instantaneous asset returns are entirely determined by the drift and volatility vector of the elementary factors. Indeed, at time t , the drift vector and covariance matrix of instantaneous asset returns are respectively defined as follows:¹:

$$\boldsymbol{\mu}_t = \mathbf{A} \boldsymbol{\mu}_t \quad \text{and} \quad \boldsymbol{\Sigma}_t = \mathbf{A} \boldsymbol{\Omega}_t \mathbf{A}^\top.$$

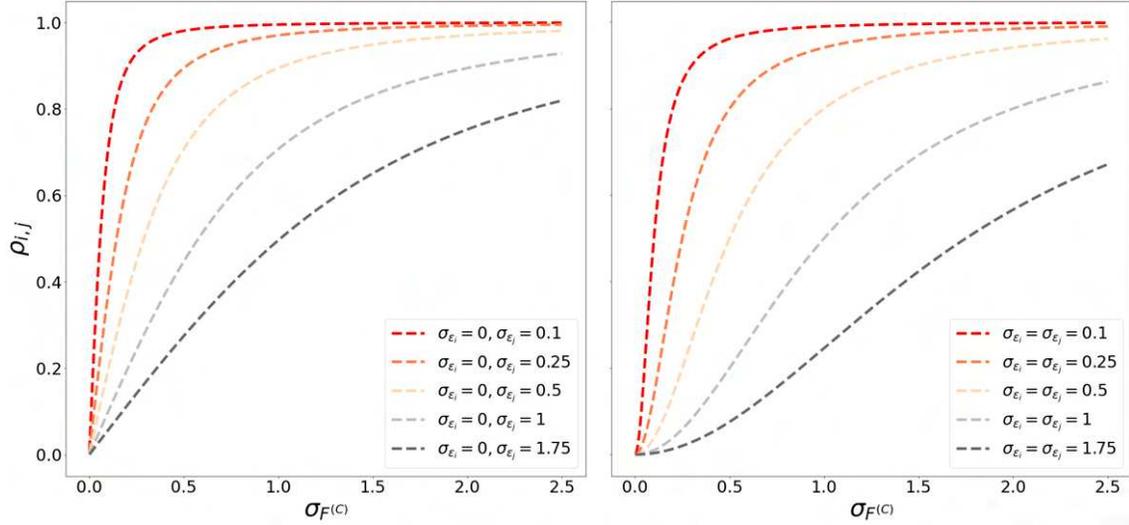


Figure 6.2: The values taken by $\rho_{i,j}$, the linear correlation coefficient between the instantaneous returns of two assets, as a function of the values taken by the factorial volatilities in the simple case where:

$$\mathbf{A} = \begin{bmatrix} 1 & 1 & 0 \\ 1 & 0 & 1 \end{bmatrix} \quad \text{and} \quad \text{diag}(\sqrt{\boldsymbol{\Omega}_t})^\top = \begin{bmatrix} \sigma_{F^{(C)}} & \sigma_{\epsilon_i} & \sigma_{\epsilon_j} \end{bmatrix}$$

In this example, since both assets, i and j , are positively exposed to a common risk factor $F^{(C)}$, their correlation $\rho_{i,j}$ follows an increasing relationship with the volatility of this factor. However, the profile of the relationship between these two quantities is also strongly influenced by the levels of volatility of the idiosyncratic factors.

From the expression of the matrix of instantaneous asset returns and given $\boldsymbol{\Omega}_t$ is a diagonal matrix, it follows that the instantaneous correlation between the returns of the i -th and j -th assets in \mathbf{P} is

¹Because $dt dt = 0$, $dt d\mathbf{W}_t = \mathbf{0}_m$ and $d\mathbf{W}_t d\mathbf{W}_t^\top = \mathbf{I}_m dt$:

$$(d\mathbf{P}_t \otimes \mathbf{P}_t) (d\mathbf{P}_t \otimes \mathbf{P}_t)^\top = \left(\mathbf{A} \boldsymbol{\mu}_t dt + \mathbf{A} \sqrt{\boldsymbol{\Omega}_t} d\mathbf{W}_t \right) \left(\mathbf{A} \boldsymbol{\mu}_t dt + \mathbf{A} \sqrt{\boldsymbol{\Omega}_t} d\mathbf{W}_t \right)^\top = \mathbf{A} \boldsymbol{\Omega}_t \mathbf{A}^\top dt.$$

defined by:

$$(\mathbf{C}_t)_{i,j} = \frac{\sum_{k=1}^m (\mathbf{A})_{i,k} \cdot (\mathbf{A})_{j,k} \cdot (\boldsymbol{\Omega}_t)_{k,k}}{\sqrt{\sum_{k=1}^m (\mathbf{A})_{i,k}^2 \cdot (\boldsymbol{\Omega}_t)_{k,k}} \sqrt{\sum_{k=1}^m (\mathbf{A})_{j,k}^2 \cdot (\boldsymbol{\Omega}_t)_{k,k}}}. \quad (6.3)$$

This expression highlights one of the main advantages of the factorial form of the model: generating a dynamic correlation structure between asset returns from the dynamics of the elementary factors. More precisely, since the exposure of assets to the factor given by the elements of the matrix \mathbf{A} is assumed to be constant, the correlation dynamics are solely driven by the changes in the factorial volatilities within the vector \mathcal{V} . Therefore, as illustrated in figure 6.1 through a simple example, the correlation structure between assets can undergo significant changes following movements in factorial volatilities. This property of the model is of major interest in explaining empirical phenomena, such as the abrupt increase in correlations between different assets that almost systematically accompanies an increase in market factor volatility.

6.2.1.2 The information driven the dynamics of factorial drift and volatilities

In the FPDM model introduced in section 6.2.1.1, the dynamics of the factors that drive the price vector \mathbf{P} depend on two major components: the vector of factorial drifts $\boldsymbol{\mu}$ and the factorial variances defined by the diagonal of $\boldsymbol{\Omega}$. However, these components themselves have dynamics that need to be defined. To this end, both can be viewed as functions that, given \mathcal{I}_t , a set of information available at time t , map to an m -dimensional vector. More specifically, $\boldsymbol{\mu} : \mathcal{I}_t \rightarrow \mathbb{R}^m$ and $\text{diag}_{M \rightarrow d}(\boldsymbol{\Omega}) : \mathcal{I}_t \rightarrow \mathbb{R}_+^m$. Adopting this approach, the identification of the information comprising \mathcal{I}_t becomes a central question. First and foremost, this information can be categorized into two main components: endogenous information and exogenous information. Here, endogenous information corresponds to the natural filtrations of \mathbf{P} and \mathbf{F} , i.e., $\{\mathbf{P}_u\}_{u \leq t}$ and $\{\mathbf{F}_u\}_{u \leq t}$. Exogenous information is defined, on the other hand, as the complement of endogenous information, representing the set of information on which $\boldsymbol{\mu}$ and $\boldsymbol{\Omega}$ depend but is not contained in $\{\mathbf{P}_u, \mathbf{F}_u\}_{u \leq t}$. The philosophy adopted by the FPDM model aligns with that of path-dependent volatility models ([101], [116], [117]), aiming to explain the dynamics of $\boldsymbol{\mu}$ and $\boldsymbol{\Omega}$ as much as possible through an endogenous manner. More specifically, we assume that at a given time t , all the relevant endogenous information on which $\boldsymbol{\mu}$ and $\boldsymbol{\Omega}$ depend is contained in the following set of state variables:

$$\mathcal{I}_{1,t} = \left\{ \tilde{\boldsymbol{\mu}}_t^{(k)}, \tilde{\boldsymbol{\Omega}}_t^{(k)} \right\}_{k=1}^{n_\tau},$$

where $\tilde{\boldsymbol{\mu}}_t^{(k)}$ and $\tilde{\boldsymbol{\Omega}}_t^{(k)}$ correspond to the following exponential weighting moving averages (EWMA):

$$\tilde{\boldsymbol{\mu}}_t^{(k)} = \frac{1}{\tau_k} \int_{-\infty}^t e^{-\frac{t-u}{\tau_k}} d\mathbf{F}_u \quad \text{and} \quad \tilde{\boldsymbol{\Omega}}_t^{(k)} = \frac{1}{\tau_k} \int_{-\infty}^t e^{-\frac{t-u}{\tau_k}} \boldsymbol{\Omega}_u du,$$

whose dynamics are respectively defined by (see details in the appendix 6.B.2):

$$d\tilde{\boldsymbol{\mu}}_t^{(k)} = \frac{1}{\tau_k} \cdot \left(d\mathbf{F}_t - \tilde{\boldsymbol{\mu}}_t^{(k)} dt \right), \quad \text{and} \quad d\tilde{\boldsymbol{\Omega}}_t^{(k)} = \frac{1}{\tau_k} \cdot \left(\boldsymbol{\Omega}_u - \tilde{\boldsymbol{\Omega}}_t^{(k)} \right) dt.$$

Given that $\mathcal{I}_{1,t}$ constitutes a set of information that impact $\boldsymbol{\mu}$ and $\boldsymbol{\Omega}$, we also have the following inclusion relation: $\mathcal{I}_{1,t} \subseteq \mathcal{I}_t$. However, in the considered approach, the EWMA estimators of the drift and covariance matrix of \mathbf{F} , represented by $\tilde{\boldsymbol{\mu}}_t^{(k)}$ and $\tilde{\boldsymbol{\Omega}}_t^{(k)}$ respectively, only impact $\boldsymbol{\mu}$ and $\boldsymbol{\Omega}$ indirectly. Indeed, the causal relationship operates through transmission channels that consist of metrics associated with a set of factorial portfolios denoted by \mathcal{Y} and defined as:

$$\mathcal{Y}(\mathcal{I}_t) = \{\mathbf{y}_{p,t}\}_{p=1}^{n_y},$$

with $\mathbf{y}_{p,t} \in \mathbb{R}^m$ and $\forall p, t : \|\mathbf{y}_{p,t}\|_1 = 1$. The modeling idea is that variations in factorial trends and volatilities arise from changes in anticipations about the future dynamics of factorial portfolios, anticipations which are formed based on the endogenous information $\mathcal{I}_{1,t}$. Furthermore, the composition of the factorial portfolios included in \mathcal{Y} is itself a deterministic function of the information \mathcal{I}_t . This general form includes the particular case where \mathcal{Y} ($\mathcal{Y}_t = \mathcal{Y} \forall t$) is invariant over time. In this specific case, it is clear that \mathcal{Y} is independent of the information \mathcal{I} . Another interesting special case is when the entire set of information on which \mathcal{Y} depends is contained in $\mathcal{I}_{1,t}$. In this configuration, \mathcal{Y} is purely path-dependent. Still, the metrics of the factorial portfolios that impact $\boldsymbol{\mu}$ and $\boldsymbol{\Omega}$ do not depend on the form taken by \mathcal{Y} . These will be of two types.

The first type, which we will refer to as *convolved dynamics features*, takes the form:

$$\hat{\boldsymbol{\mu}}_t(\mathbf{y}_{p,t}, \boldsymbol{\delta}_p) = \sum_{k=1}^{n_\tau} (\boldsymbol{\delta}_p)_k \cdot \mathbf{y}_{p,t}^\top \tilde{\boldsymbol{\mu}}_t^{(k)} = \mathbf{y}_{p,t}^\top \int_{-\infty}^t g_p(t-u) d\mathbf{F}_u, \quad (6.4)$$

where $\boldsymbol{\delta}_p \in \mathbb{R}^{n_y}$ and $g_p(s) = \sum_{k=1}^{n_\tau} \frac{(\boldsymbol{\delta}_p)_k}{\tau_k} e^{-\frac{s}{\tau_k}}$. These features thus correspond to stochastic convolutions with respect to $d\mathbf{F}$ over the interval $]-\infty, t]$. It should be noted that the kernels $\{g_p\}_{p=1}^{n_y}$ on which these factors depend can be highly diverse. In the case where $\boldsymbol{\delta}_p \in \mathbb{R}_+^{n_y}$ (resp. $\boldsymbol{\delta}_p \in \mathbb{R}_-^{n_y}$), g_p is a positive, decreasing, convex function (resp. negative, increasing, concave) over \mathbb{R}_+ . In contrast, when the coordinates of $\boldsymbol{\delta}_p$ are not of the same sign, g_p can be non-homogeneous and have variable sign over \mathbb{R}_+ .

The second type of features impacting $\boldsymbol{\mu}$ and $\boldsymbol{\Omega}$, that we will term as *historical volatility features*, takes the form:

$$\hat{\sigma}_t(\mathbf{y}_{p,t}, \mathbf{w}_p) = \sqrt{\sum_{k=1}^{n_\tau} (\mathbf{w}_p)_k \cdot \mathbf{y}_{p,t}^\top \tilde{\boldsymbol{\Omega}}_t^{(k)} \mathbf{y}_{p,t}} = \sqrt{\mathbf{y}_{p,t}^\top \left(\int_{-\infty}^t k_p(t-u) \boldsymbol{\Omega}_u du \right) \mathbf{y}_{p,t}} \quad (6.5)$$

where $\mathbf{w}_p \in \mathbb{R}_+^{n_y}$, $\|\mathbf{w}_p\|_1 = 1$ and $k_p(s) = \sum_{k=1}^{n_\tau} \frac{(\mathbf{w}_p)_k}{\tau_k} e^{-\frac{s}{\tau_k}}$. Therefore, the form of the kernels

$\{k_p\}_{p=1}^{n_y}$ is much more constrained than that of $\{g_p\}_{p=1}^{n_y}$: they are necessarily positive, decreasing, convex functions over \mathbb{R}_+ , and their integral is equal to 1. This distinction arises from the very nature of this second type of features, which correspond to averages of the realized volatility of factorial portfolios. Specifically, $\hat{\sigma}_t(\mathbf{y}_{p,T}, \mathbf{w}_p)^2$ is a moving average of the realized variance of the process $(\mathbf{y}_{p,T})^\top d\mathbf{F}_t$ (not of $(\mathbf{y}_{p,t})^\top d\mathbf{F}_t$).

These two types of features thus form the set

$$\mathcal{F}_t = \left\{ \hat{\mu}_t(\mathbf{y}_{p,t}, \boldsymbol{\delta}_p), \hat{\sigma}_t(\mathbf{y}_{p,t}, \mathbf{w}_p) \right\}_{p=1}^{n_y}, \quad (6.6)$$

which integrates as follows into the causal chain of the FPDM model: \mathcal{I}_t generates the set of factorial portfolios \mathcal{Y}_t , and the combination of \mathcal{I}_t and \mathcal{Y}_t defines the set of features \mathcal{F}_t on which $\boldsymbol{\mu}_t$ and $\boldsymbol{\Omega}_t$ depend. More precisely, as will be exposed in section 6.2.2, the set \mathcal{F}_t determines (either entirely or partially) $\boldsymbol{\Omega}_t$, and the triplet $(\boldsymbol{\Omega}_t, \mathcal{F}_t, \mathcal{Y}_t)$ in turn defines $\boldsymbol{\mu}_t$. The figure 6.3 summarizes the sequence of causal relationships in the FPDM model.

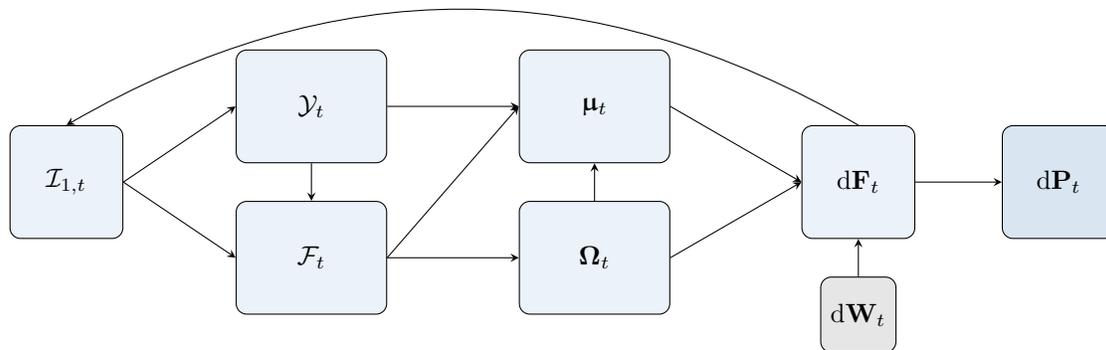


Figure 6.3: Causal diagram of the FPDM model in the case where \mathcal{Y} , $\boldsymbol{\mu}$, and $\boldsymbol{\Omega}$ depend solely on endogenous information. The endogenous information $\mathcal{I}_{1,t}$ defines the set of factorial portfolios \mathcal{Y}_t . The features constituting the set \mathcal{F}_t are computed from the pair $(\mathcal{I}_{1,t}, \mathcal{Y}_t)$, and this set determines the value of the matrix of factorial variances $\boldsymbol{\Omega}_t$. In turn, $\boldsymbol{\Omega}_t, \mathcal{F}_t$ and \mathcal{Y}_t determine the vector of factorial drifts $\boldsymbol{\mu}_t$. Finally, the pair of $\boldsymbol{\mu}_t$ and $\boldsymbol{\Omega}_t$, coupled with the random shocks modeled by $d\mathbf{W}_t$, generates the dynamics $d\mathbf{F}_t$ of the elementary factors, which ultimately leads to the variation in the vector of prices $d\mathbf{P}_t$. Furthermore, the dynamics of the elementary factors produces a feedback effect by increasing the endogenous information and, through the same causal chain, modifying the state of the system in turn.

6.2.2 Volatilities and drifts of elementary factors

6.2.2.1 The factorial volatilities

The modeling approach adopted for factorial volatilities $\sqrt{\boldsymbol{\Omega}}$ presents significant analogies with the PDV model proposed by Guyon and Lekeufack ([117]). Similar to this model, volatility dynamics are

considered as a primarily endogenous phenomenon with the addition and interaction of short-term exogenous dynamics. Accordingly, the matrix of factorial volatilities takes the following general form²:

$$\sqrt{\boldsymbol{\Omega}}_t = \text{diag}(\boldsymbol{\mathcal{V}}(\mathcal{F}_t) \odot \mathbf{X}_t \cdot \mathcal{S}_t), \quad (6.7)$$

where $\boldsymbol{\mathcal{V}}$ is a deterministic function of the set of features \mathcal{F}_t such as $\boldsymbol{\mathcal{V}} : \mathcal{F}_t \rightarrow \mathbb{R}_+^m$, \mathbf{X} is an m -dimensional stochastic process responsible for capturing the exogenous dynamics of factorial volatilities, and \mathcal{S} is a univariate stochastic process corresponding to a common factor sensitivity operator to random frictions.

More precisely, we define the path-dependent component of the vector of volatilities of the elementary factors $\boldsymbol{\mathcal{V}}(\mathcal{F}_t)$ as follows:

$$\boldsymbol{\mathcal{V}}(\mathcal{F}_t) = \underbrace{\mathbf{b}_0}_{(1)} + \underbrace{\sum_{p=1}^{n_y} \mathbf{b}_{1,p} \cdot \hat{\mu}_t(\mathbf{y}_p, \boldsymbol{\delta}_p^{(\mathcal{V})})}_{(2)} + \underbrace{\sum_{p=1}^{n_y} \mathbf{b}_{2,p} \cdot \hat{\sigma}_t(\mathbf{y}_p, \mathbf{w}_p^{(\mathcal{V})})}_{(3)}, \quad (6.8)$$

where:

- (1) is an intercept vector such as $\mathbf{b}_0 \in \mathbb{R}_+^m$.
- (2) is the component of $\boldsymbol{\mathcal{V}}(\mathcal{F}_t)$ attributable to the convolved dynamics features (equation 6.4), with $\mathbf{b}_{1,p} \in \mathbb{R}^m$.
- (3) is the component of $\boldsymbol{\mathcal{V}}(\mathcal{F}_t)$ attributable to the historical volatility features (equation 6.5), with $\mathbf{b}_{2,p} \in \mathbb{R}_+^m$ and $\sum_{p=1}^{n_y} \mathbf{b}_{2,p} \leq \mathbf{1}_m$.

It is first worth noting that, through the component $\boldsymbol{\mathcal{V}}(\mathcal{F}_t)$, the proposed model induces a nested factor structure analogous to that proposed by Chicheportiche and Bouchaud [65]. Indeed, similar to their model, the proposed framework implies that, on the one hand, asset returns depend on common risk factors, and, on the other hand, the volatility of these risk factors is itself driven by common factors. Regarding the form taken by $\boldsymbol{\mathcal{V}}(\mathcal{F}_t)$, this can be viewed as a generalization of the PDV model proposed by Guyon and Lekeufack ([117]) in a factorial and multivariate framework. Indeed, first, the 4-factor PDV ([117]) model is a specific case of 6.8 (see appendix 6.C.1), second, even the general form of 6.8 exhibits an analogous three-block structure. However, the generalization induced by 6.8 entails substantial specificities that need to be detailed.

Firstly, regarding component (2), where its counterpart in the Guyon and Lekeufack model corresponds to a moving average of past returns (up to a multiplicative constant) for the considered

²Or equivalently:

$$\boldsymbol{\Omega}_t = \mathcal{S}_t^2 \cdot (\boldsymbol{\mathcal{V}}(\mathcal{F}_t) \odot \mathbf{X}_t) (\boldsymbol{\mathcal{V}}(\mathcal{F}_t) \odot \mathbf{X}_t)^\top \odot \mathbf{I}_m.$$

asset, (2) is a linear combination of stochastic convolution of factorial portfolio dynamics. This aims to capture the impact of past dynamics of the factorial portfolios belonging to \mathcal{Y}_t on the level of factorial volatilities. The well-known leverage effect and the strong Zumbach effect ([205], [106]), respectively defined as the negative relationship between past returns and spot volatility, and the dependence of the volatility process on the historical price path, can thus be modeled through this component. In addition to these effects already captured in a univariate framework by the PDV model of Guyon and Lekeufack, (2) allows for the consideration of other effects specific to multidimensional and factorial modeling. In particular, the past dynamics of one elementary factor can impact the volatility level of another elementary factor. By extension, the volatility level of an asset can be conditioned by the past dynamics of other assets. Concretely, this transmission mechanism allows for the modeling of empirical phenomena such as the difference in amplitude of the leverage effect between individual assets and indices that group a set of assets ([44]).

This type of transmission mechanism is also enabled by component (3), which is a linear combination of historical volatility features associated with different factor portfolios. It is akin to the "historical volatility factor" in the Guyon and Lefeufack model. However, unlike the latter, (3) does not solely depend on the volatility of an elementary factor based on its own past trajectory (the past trajectory of the volatility of this elementary factor) but also depends on the past trajectories of volatilities of other elementary factors. This structure allows for capturing, in addition to the volatility feedback and volatility clustering phenomena already enabled by the Guyon and Lefeufack model, effects specific to multi-asset dynamics, such as volatility spillover effects ([123], [15], [66]). Thus, an increase in the volatility level of a subset of the investment universe can propagate to other assets through this relationship.

In addition to the path-dependent component $\mathcal{V}(\mathcal{F}_t)$ just defined, as per 6.7, the factorial volatilities are functions of the vector \mathbf{X} whose purpose is to capture dynamics of exogenous origin in volatility. The simplest case is, of course, when \mathbf{X} is invariant (typically $\forall t : \mathbf{X}_t = \mathbf{1}_m$), and therefore, the factorial volatilities are purely path-dependent. Beyond this particular case, we adopt and generalize once again the idea proposed by Guyon and Lekeufack by specifying \mathbf{X} as a mean-reverting process. More precisely, following the approach suggested in chapter 3 and adapted to the considered multivariate framework, we define \mathbf{X} as an m -dimensional exponential Ornstein-Uhlenbeck process, such that:

$$\mathbf{X}_t = \exp \circ (\mathbf{Y}_t) \quad \text{with} \quad d\mathbf{Y}_t = \mathbf{\Upsilon} (\bar{\mathbf{Y}} - \mathbf{Y}_t) dt + \mathbf{\Psi} d\mathbf{B}_t, \quad (6.9)$$

where $\mathbf{\Upsilon}$ is a $m \times m$ transition matrix, $\bar{\mathbf{Y}}$ is the unconditional expectation of \mathbf{Y} , $\mathbf{\Psi}$ is a $m \times m$ scatter matrix, and \mathbf{B} is a m -dimensional Brownian motion independent of \mathbf{W} . As shown in [156], the conditional distribution of \mathbf{Y}_{t+s} given \mathbf{Y}_t is normal at all times, such that:

$$\mathbf{Y}_{t+s} | \mathbf{Y}_t \sim \mathcal{N} \left(\bar{\mathbf{Y}} + e^{-\mathbf{\Upsilon} \cdot s} (\mathbf{Y}_t - \bar{\mathbf{Y}}), \text{vec}_{m \times m}^{-1} \left((\mathbf{\Upsilon} \oplus \mathbf{\Upsilon})^{-1} \left(\mathbf{I}_m - e^{-(\mathbf{\Upsilon} \mathbf{\Upsilon}^\top) \cdot s} \right) \text{vec} \left(\mathbf{\Psi} \mathbf{\Psi}^\top \right) \right) \right).$$

Therefore \mathbf{X}_{t+s} conditional on \mathbf{X}_t follows the log-normal distribution:

$$\mathbf{X}_{t+s}|\mathbf{X}_t \sim \mathcal{LN}\left(\bar{\mathbf{Y}} + e^{-\mathbf{\Upsilon} \cdot s} (\mathbf{Y}_t - \bar{\mathbf{Y}}), \text{vec}_{m \times m}^{-1}\left((\mathbf{\Upsilon} \oplus \mathbf{\Upsilon})^{-1} \left(\mathbf{I}_m - e^{-(\mathbf{\Upsilon} \mathbf{\Upsilon}^\top) \cdot s}\right) \text{vec}\left(\mathbf{\Psi} \mathbf{\Psi}^\top\right)\right)\right).$$

Furthermore, the asymptotic distribution of \mathbf{X} is given by:

$$\lim_{s \rightarrow +\infty} \mathbf{X}_{t+s}|\mathbf{X}_t \sim \mathcal{LN}\left(\bar{\mathbf{Y}}, \text{vec}_{m \times m}^{-1}\left((\mathbf{\Upsilon} \oplus \mathbf{\Upsilon})^{-1} \text{vec}\left(\mathbf{\Psi} \mathbf{\Psi}^\top\right)\right)\right).$$

Several remarks can be made regarding the specification of parameters defining the dynamics of \mathbf{X} . First, if we assume that the path-dependent component exhausts the structural relationships linking the volatilities of the elementary factors, it is consistent then for $\mathbf{\Upsilon}$ and $\mathbf{\Psi}$ to be both diagonal matrices. Additionally, the results presented in [117] and in chapter 3 suggest that, when using a primarily path-dependent calibration approach, the exogenous dynamics of volatility manifest as a short-term phenomenon (intraday). In this context, \mathbf{Y} mean-reverts very fast toward $\bar{\mathbf{Y}}$, implying high values for elements of $\mathbf{\Upsilon}$.

Finally, the last component of the factorial volatilities \mathcal{S} is a univariate process that defines a common sensitivity to all factors. This allows for capturing the homothetic dynamics of the matrix of factorial volatilities. Like \mathbf{X} , aside from the special case of constant \mathcal{S} , it is coherent to model it using a mean-reverting process. Thus, the choice of an exponential Ornstein-Uhlenbeck process remains meaningful in this context. An alternative is to model $\log(\mathcal{S})$ through a continuous-time ARMA (autoregressive-moving average) process ([52], [53], [61]). Due to the multitude of approaches included in this family of models, this choice provides a high degree of flexibility in capturing various types of dynamics, particularly in terms of temporal dependence relationships.

6.2.2.2 The drifts of the elementary factors

The modeling of the drift vector of the elementary factors proposed in this section aims to achieve a dual purpose. On one hand, it aims to incorporate the insights provided by financial literature, and on the other hand, to be flexible enough to capture potential market-specific patterns that may not be accounted for by standard approaches. In line with this objective, we propose to define this drift as the following linear combination:

$$\boldsymbol{\mu}_t = \sum_{p=1}^{n_y} \beta_{p,t} \cdot \Gamma_{p,t}, \quad (6.10)$$

where Γ_p is the performance factor associated with portfolio \mathbf{y}_p , and $\boldsymbol{\beta}_p$ is the vector of sensitivities of the elementary factors defined by:

$$\boldsymbol{\beta}_{p,t} = \frac{\text{Cov}\left(d\mathbf{F}_t, \mathbf{y}_{p,t}^\top d\mathbf{F}_t\right)}{\text{Var}\left(\mathbf{y}_{p,t}^\top d\mathbf{F}_t\right)} = \frac{\boldsymbol{\Omega}_t \mathbf{y}_{p,t}}{\mathbf{y}_p^\top \boldsymbol{\Omega}_t \mathbf{y}_{p,t}}. \quad (6.11)$$

The form 6.10, although very general in its current state, frames the modeling of $\boldsymbol{\mu}$ by defining the drifts of the elementary factors based on their respective exposures to the portfolios included in \mathcal{Y} . Therefore, the composition of \mathcal{Y} plays a crucial role in this framework. Furthermore, it is clear from 6.10 that the significance of the various portfolios contained in \mathcal{Y} on $\boldsymbol{\mu}$ is contingent upon the forms taken by the functions Γ_p . Here, according to our dual objective outlined in the introduction, we consider the following specification:

$$\Gamma_{p,t} = \underbrace{\lambda_p^{(S)}(\sigma_t(\mathbf{y}_{p,t}))}_{(1)} + \underbrace{\lambda_p^{(P)}(\hat{\sigma}_t(\mathbf{y}_{p,t}, \mathbf{w}_p))}_{(2)} + \underbrace{\hat{\mu}_t(\mathbf{y}_{p,t}, \boldsymbol{\delta}_p)}_{(3)} + \underbrace{\mathcal{E}_{p,t}}_{(4)}, \quad (6.12)$$

where:

- (1) corresponds to the instantaneous risk premium of the portfolio $y_{p,t}$, which is a function of the value taken by $\sigma_t(\mathbf{y}_{p,t})$, the spot volatility of this portfolio, such that $\lambda_p^{(S)} : \mathbb{R}_+ \rightarrow \mathbb{R}$.
- (2) corresponds to the premium compensating for the historical risk of the portfolio $y_{p,t}$ measured by $\hat{\sigma}_t(\mathbf{y}_{p,t}, \mathbf{w}_p)$ (equation 6.5), such that $\lambda_p^{(P)} : \mathbb{R}_+ \rightarrow \mathbb{R}$.
- (3) corresponds to the feedback effect of the past dynamics of $\mathbf{y}_{p,t}$ on its current drift defined by equation 6.4).
- (4) corresponds to the residual component of the performance factor associated with portfolio $\mathbf{y}_{p,t}$, not explained by the first three components.

This modeling approach has the double advantage of being highly flexible while preserving a clear understanding of the factors influencing the vector of factor drifts. For instance, as demonstrated in appendix 6.C.2, the proposed model can be specified to align with the theoretical coordinates of the capital asset pricing model (CAPM) ([190], [154]). In addition to a strict adoption of this type of specification based on a strong theoretical framework, those can serve as a baseline for integrating other components of $\boldsymbol{\mu}$. This allows for a combination of a theoretical model with a data-driven approach. Moreover, the combination of the four components that form a performance factor enables the modeling of complex drift patterns and different natures.

Firstly, components (1) and (2) allow linking the volatility levels of factorial portfolios - instantaneous volatilities in the case of (1) and past volatilities in the case of (2) - with drifts. However, the relationship between the level of volatility and expected returns is both common in financial modeling ([188], [118]) and extensively documented [103]. Furthermore, the very general definition

given to $\lambda_p^{(S)}(\cdot)$ and $\lambda_p^{(P)}(\cdot)$ enables the consideration of a variety of ways to model this relationship, including for instance linear and polynomial forms, as proposed in [118]. Regarding the dual risk premium structure employed, comprising a spot volatility premium and a historical volatility premium, it may seem unconventional at first glance. Nevertheless, in practice, this specific form allows for the modeling of various important empirical phenomena documented in financial literature. One such example is the market behaviors studied by French *et al.* ([103]), where the expected returns of stocks are positively related to predictable volatility through autoregressive (AR) models and negatively related to unexpected changes in volatility. In the proposed framework, considering y_p as a factorial portfolio representative of the equity market, this phenomenon can be captured by jointly specifying $\lambda_p^{(S)}$ as an increasing function on \mathbb{R}_+ and $\lambda_p^{(P)}$ as a positively increasing function on \mathbb{R}_+ . Indeed, in this context, on the one hand, the variation in (1) + (2) is negatively correlated with a volatility shock, and on the other hand, (1) + (2) follows a positive relationship with the level of volatility predicted by an AR model due to component (2). A second example of an empirical pattern that the dual form of the risk premium can capture is the well-documented "fly-to-quality" phenomenon ([28], [67], [16]). Indeed, let us suppose that y_q corresponds to a quality portfolio, and that the associated risk premiums take the form $\lambda_q^{(S)}(\sigma) = \bar{\lambda}_q^{(S)}\sigma^2$ and $\lambda_q^{(P)}(\sigma) = \bar{\lambda}_q^{(P)}\sigma^2$ with $0 < \bar{\lambda}_q^{(S)} = -\bar{\lambda}_q^{(P)}$. In this setup, during prolonged periods of stability in the volatility level, (1) + (2) is close to zero. However, if the instantaneous volatility suddenly spikes, (1) + (2) turns positive due to a positive delta between spot volatility and the historical volatility factor. Conversely, when volatility decreases, (1) + (2) turns negative, spot volatility becoming lower than historical volatility.

Regarding (3), this feedback feature enables the modeling of momentum and mean-reversion phenomena, two important determinants of empirical price dynamics ([172], [113], [189]). The nature of the feedback generated by this component depends on the value of δ_p , as clearly evident from its mathematical expression provided by 6.4. Thus, when $\delta_p \in \mathbb{R}_+^{n_y}$ and if at least one coordinate of $\delta_p > 0$, (3) may be viewed as a momentum factor. Conversely, if $\delta_p \in \mathbb{R}_-^{n_y}$ and at least one coordinate of $\delta_p < 0$, (3) can be seen as a reversal factor. Beyond these two polar cases, the form of (3) allows capturing more complex feedback structures, such as the coexistence of positive autocorrelation in returns over short horizons and negative autocorrelation over longer horizons highlighted by Poterba and Summers ([172]).

The component (4) differs from the first three components of 6.12 on several fronts. Firstly, it is not necessarily solely a function of endogenous information but may also depend on exogenous information contained in \mathcal{I}_t . Consequently, this component enables the consideration of exogenous determinants of factor drifts, allowing for the concrete incorporation of factors such as 'views' in the Black-Litterman sense on portfolios. Moreover, (4) can serve to ensure certain relationships between different components of the model, as in the case of the CAPM specification of μ discussed in appendix 6.C.2.

6.3 The Factorial Path-Dependent Market Generator

6.3.1 The market generator framework

6.3.1.1 The considered FPDM model

The theoretical framework proposed in section 6.2, defining the FPDM model, is intentionally very general and encompasses a large number of possible specifications. This section aims to define the FPDM model under consideration, which will be used, in a discretized version, as a market generator in the remainder of the article.

6.3.1.1.1 The set of factor portfolios \mathcal{Y}

Firstly, the set of factor portfolios \mathcal{Y} is defined as:

$$\mathcal{Y} = \{\mathbf{e}_p\}_{p=1}^m,$$

where \mathbf{e}_p is an $m \times 1$ vector, with the value of the p -th row being 1 and 0 for all other rows. This is one of the simplest possible specifications of \mathcal{Y} , which implies that the features on which the factor drifts and volatilities depend are each a function of a single elementary factor. Furthermore, through equations 4 and 5, the drift vector of the elementary factors can be simplified as follows:

$$\boldsymbol{\mu}_t = \sum_{p=1}^m \beta_{p,t} \cdot \Gamma_{p,t} = \sum_{p=1}^m \frac{\boldsymbol{\Omega}_t \mathbf{e}_p}{\mathbf{e}_p \boldsymbol{\Omega}_t \mathbf{e}_p} \cdot \Gamma_{p,t} = \sum_{p=1}^m \mathbf{e}_p \cdot \Gamma_{p,t} = \boldsymbol{\Gamma}_t.$$

Therefore, the drift of an elementary factor j depends solely on the performance factor $\Gamma_{j,t}$ associated with it (since $\beta_{j,t} = \mathbf{e}_j$).

6.3.1.1.2 The kernels

The specification adopted for the kernels associated with the features on which the factor drifts and volatilities depend is also made with a concern for parsimony. Thus, the considered model includes only two kernels: the first shared by all convolved dynamics features and the second shared by all historical volatility features, respectively defined by:

$$g^{(\hat{\mu})}(s) = \sum_{k=1}^{n_\tau} \frac{\boldsymbol{\delta}_k}{\tau_k} e^{-\frac{s}{\tau_k}} \quad \text{and} \quad g^{(\hat{\sigma})}(s) = \sum_{k=1}^{n_\tau} \frac{\mathbf{w}_k}{\tau_k} e^{-\frac{s}{\tau_k}}, \quad (6.13)$$

where $\boldsymbol{\delta}, \mathbf{w} \in \mathbb{R}_+^{n_\tau}$ and with

$$\tau_k = \exp \left(\log(\tau_-) + \frac{\log(\tau_+) - \log(t_-)}{n_\tau - 1} (k - 1) \right). \quad (6.14)$$

This method of defining the parameters $\{\tau_k\}_{k=1}^{n_\tau}$, introduced in chapter 5, allows for a good approximation for the majority of positive decreasing kernels on \mathbb{R}_+ as long as one opts for a specification

of n_τ, t_-, t_+ . In this case, we choose the following specification: $n_\tau = 10$ and $1/365$ and 5 (expressed in years).

6.3.1.1.3 The volatilities of elementary factors

The specification of the vector of elementary factors is based on the idea that one elementary factor, corresponding to the market factor, plays a specific role in the overall level of factorial volatilities. In this framework, the purely path-dependent component of the volatility of an elementary factor depends solely on its own past trajectory and the past trajectory of the market factor. Following this principle, by associating the market factor with the first (in terms of index) elementary factor, we define this purely path-dependent component for an elementary factor k as:

$$(\mathbf{V}(\mathcal{F}_t))_k = b_{0,k} + b_{1,k} \cdot \hat{\mu}_t(\mathbf{e}_k, \boldsymbol{\delta}) + b_{2,k} \cdot \hat{\sigma}_t(\mathbf{e}_k, \mathbf{w}) + b_{3,k} \cdot (\mathbf{V}(\mathcal{F}_t))_1, \quad (6.15)$$

where $\forall k : b_{0,k}, b_{2,k}, b_{3,k} \in \mathbb{R}_+$. In a less compact form, equation 6.15 can be rewritten as follows:

$$\begin{aligned} (\mathbf{V}(\mathcal{F}_t))_k &= b_{0,k} + b_{3,k} b_{0,1} + b_{1,k} \int_{-\infty}^t K^{(\hat{\mu})}(u) d(\mathbf{F}_u)_k + b_{3,k} b_{1,1} \int_{-\infty}^t K^{(\hat{\mu})}(u) d(\mathbf{F}_u)_1 \\ &\quad + b_{2,k} \sqrt{\int_{-\infty}^t K^{(\hat{\sigma})}(u)(\boldsymbol{\Omega}_u)_{k,k} du} + b_{3,k} b_{2,1} \sqrt{\int_{-\infty}^t K^{(\hat{\sigma})}(u)(\boldsymbol{\Omega}_u)_{1,1} du}. \end{aligned}$$

Certainly, for the specific case of the market factor (i.e. $k = 1$), $b_{3,1} = 0$. Consequently, the purely path-dependent component of the market factor volatility takes a form analogous to the PDV model by Guyon and Lekeudfack:

$$(\mathbf{V}(\mathcal{F}_t))_1 = b_{0,1} + b_{1,1} \int_{-\infty}^t K^{(\hat{\mu})}(u) d(\mathbf{F}_u)_1 + b_{2,1} \sqrt{\int_{-\infty}^t K^{(\hat{\sigma})}(u)(\boldsymbol{\Omega}_u)_{1,1} du}.$$

Regarding \mathbf{X} , the process responsible for capturing the exogenous dynamics of factorial volatilities, we adopt here the form suggested in section 6.2.2, namely an exponential OU process. More precisely, reusing the form 6.9, we consider the case where both $\boldsymbol{\Upsilon}$ and $\boldsymbol{\Psi}$ are diagonal matrices, such that the coordinates are independent. Furthermore, we assume that the mean-reversion parameters associated with this process are very high, corresponding to high or medium frequency phenomena. Thus, for the time step Δt considered in the market generator, the realizations of \mathbf{X}_t and $\mathbf{X}_{t+\Delta t}$ are approximately independent. This modeling assumes that, for low-frequency observations, the autocorrelation structure of volatilities is entirely captured by \mathbf{V} and \mathcal{S} .

To model the dynamics of the market sensitivity operator \mathcal{S} , we assume that $\log(\mathcal{S})$ follows a CARMA process. In practice, since the market generator is a discrete model, the selected model will be a standard ARMA(1,1). This choice, which may appear arbitrary at first glance, is actually determined by the empirical dynamics of \mathcal{S} , which are fairly well-modeled by this type of process (see appendix 6.G.3).

6.3.1.1.4 The factorial drifts

The specification of the factor drift vector follows a CAPM-like framework under the assumption of zero interest rates: only the market factor has a non-zero drift. Specifically, this vector is defined by the expression:

$$\boldsymbol{\mu}_t = \boldsymbol{\Gamma}_t = \mathbf{e}_1 \cdot \underbrace{(\bar{\mu} + \zeta \cdot \hat{\mu}_t(\mathbf{e}_1, \boldsymbol{\delta}) + \lambda \cdot \hat{\sigma}_t(\mathbf{e}_1, \mathbf{w}))}_{\Gamma_t^*}.$$

where Γ^* represents the market premium under the assumption of a zero risk-free rate. This premium comprises three components. The first, $\bar{\mu}$, is the invariant part of the market premium. The second component, $\zeta \cdot \hat{\mu}_t(\mathbf{e}_1, \boldsymbol{\delta})$, models the effect of past market factor dynamics on the current market premium. If ζ is positive, the market premium exhibits momentum: the market factor is subject to a positive feedback effect. Conversely, if ζ is negative, the market premium exhibits reversal: the market factor is subject to a negative feedback effect. The third component, $\lambda \cdot \hat{\sigma}_t(\mathbf{e}_1, \mathbf{w})$, represents a premium for historical volatility.

Several additional remarks can be made. Firstly, the market risk premium, and consequently the drift, is purely path-dependent. Only the feature variables determining the path-dependent component of the volatility of the market factors make the drifts time-variable. Furthermore, the market risk premium can be rewritten as follows:

$$\Gamma_t^* = \bar{\mu}' + \lambda^{\mathcal{V}} \cdot (\mathcal{V}_t)_1 + \zeta' \cdot \hat{\mu}_t(\mathbf{e}_1, \boldsymbol{\delta}) + \lambda' \cdot \hat{\sigma}_t(\mathbf{e}_1, \mathbf{w}),$$

$\bar{\mu}' = \bar{\mu} - \lambda^{\mathcal{V}} \cdot b_{0,1}$, $\zeta' = \zeta - \lambda^{\mathcal{V}} \cdot b_{1,1}$ and $\lambda' = \lambda - \lambda^{\mathcal{V}} \cdot b_{2,1}$. Therefore, it has as a special case $\Gamma_t^* = \bar{\mu} + \lambda^{\mathcal{V}} \cdot (\mathcal{V}_t)_1$. Another important point is that the drift vector of the assets is defined as:

$$\mathbf{A}\boldsymbol{\mu}_t = (\mathbf{A})_{[:,1]} \cdot \Gamma_t^*.$$

Consequently, in this framework, changes in the asset drifts are solely caused by movements in the market risk premium.

6.3.1.2 The market generator

The considered market generator is a discrete version of the FPDm model discussed in section 6.3.1.1. It is defined by algorithm 4. In addition to the time step Δt and the number of simulations n_s , it takes as input the set of parameters $\boldsymbol{\delta}, \mathbf{w}, \mathbf{A}, \boldsymbol{\tau}, \mathbf{b}_0, \mathbf{b}_1, \mathbf{b}_2, \mathbf{b}_3, \mathbf{S}, \bar{\mu}, \zeta, \lambda, a_0, a_1, a_2$, and the set of state variables $\mathbf{P}_0, \mathbf{M}_0^{(\bar{\mu})}, \mathbf{M}_0^{(\bar{\Omega})}, \epsilon_0$, whose determination is discussed in next section 6.3.2. These inputs are then used to generate a time series $\{\mathbf{P}_{u\Delta t}\}_{0 \leq t \leq u}$, which constitutes the output of the generator³. The objective of this section is to elucidate the various components of this market generator, including the discretization choices made and the rationale behind these choices.

³Of course, other elements used for simulating this time series can be added to the output, depending on the purpose of using the market generator. Typically, it may be beneficial to also retain the trajectories of elementary factors or volatility.

Algorithm 4 The Factorial Path-dependent Market Generator.

Input: $\mathbf{P}_0, \mathbf{M}_0^{(\hat{\mu})}, \mathbf{M}_0^{(\hat{\Omega})}, \epsilon_0, \delta, \mathbf{w}, \mathbf{A}, \boldsymbol{\tau}, \mathbf{b}_0, \mathbf{b}_1, \mathbf{b}_2, \mathbf{b}_3, \mathbf{S}, \bar{\mu}, \zeta, \lambda, a_0, a_1, a_2, \Delta t, n_s$.

For $t = 1$ to n_s **do**

1. Sample $\left(\mathbf{W}_{t+\Delta t}^\top, \mathbf{B}_{t+\Delta t}^\top, \epsilon_{t+\Delta t}\right)^\top$ from $\mathcal{N}(\mathbf{0}_{2m+1}, \mathbf{I}_{2m+1})$.
2. Update system equations:

$$\hat{\boldsymbol{\mu}}_{t+\Delta t} \leftarrow \delta^\top \mathbf{M}_t^{(\hat{\mu})}$$

$$\hat{\boldsymbol{\sigma}}_{t+\Delta t} \leftarrow \left(\mathbf{w}^\top \mathbf{M}_t^{(\hat{\Omega})}\right)^{\circ 1/2}$$

$$\mathcal{V}_{t+\Delta t} \leftarrow \mathbf{b}_0 + \mathbf{b}_1 \odot \hat{\boldsymbol{\mu}}_{t+\Delta t} + \mathbf{b}_2 \odot \hat{\boldsymbol{\sigma}}_{t+\Delta t} + \mathbf{b}_3 \cdot (\mathbf{b}_0 + \mathbf{b}_1 \odot \hat{\boldsymbol{\mu}}_{t+\Delta t} + \mathbf{b}_2 \odot \hat{\boldsymbol{\sigma}}_{t+\Delta t})_1$$

$$\mathbf{X}_{t+\Delta t} \leftarrow \exp \circ (\mathbf{S} \odot (\mathbf{B}_{t+\Delta t} - \mathbf{S}))$$

$$\log(\mathcal{S}_{t+\Delta t}) \leftarrow a_0 + a_1 \log(\mathcal{S}_{t+\Delta t}) + a_2 \epsilon_t + \epsilon_{t+\Delta t}$$

$$\boldsymbol{\mu}_{t+\Delta t} \leftarrow \mathbf{e}_1 \cdot (\bar{\mu} \cdot \mathbf{1}_m + \zeta \cdot \hat{\boldsymbol{\mu}}_{t+\Delta t} + \lambda \cdot \hat{\boldsymbol{\sigma}}_{t+\Delta t})$$

$$\sqrt{\boldsymbol{\Omega}}_{t+\Delta t} \leftarrow \text{diag}(\mathcal{V}_{t+\Delta t} \odot \mathbf{X}_{t+\Delta t} \cdot \mathcal{S}_{t+\Delta t})$$

$$\Delta \mathbf{F}_{t+\Delta t} \leftarrow \boldsymbol{\mu}_t \cdot \Delta t + \sqrt{\boldsymbol{\Omega}}_t \mathbf{W}_{t+\Delta t} \cdot \sqrt{\Delta t}$$

$$\mathbf{P}_{t+\Delta t} \leftarrow \left(\mathbf{P}_t + \mathbf{P}_t \odot (\mathbf{A}^\top \Delta \mathbf{F}_{t+\Delta t})\right)_+$$

$$\mathbf{M}_{t+\Delta t}^{(\hat{\mu})} \leftarrow \left(\mathbf{1}_{n_\tau}^\top \otimes \exp \circ (\boldsymbol{\tau}^{-1} \cdot \Delta t)\right) \odot \mathbf{M}_t^{(\hat{\mu})} + (\boldsymbol{\tau}^{-1})^\top \otimes \Delta \mathbf{F}_{t+\Delta t}$$

$$\mathbf{M}_{t+\Delta t}^{(\hat{\Omega})} \leftarrow \left(\mathbf{1}_{n_\tau}^\top \otimes \exp \circ (\boldsymbol{\tau}^{-1} \cdot \Delta t)\right) \odot \mathbf{M}_t^{(\hat{\Omega})} + (\boldsymbol{\tau}^{-1})^\top \otimes (\Delta \mathbf{F}_{t+\Delta t})^{\circ 2}$$

3. Update the set of price vector trajectories:

$$\{\mathbf{P}_{z\Delta t}\}_{0 \leq z \leq u} \leftarrow \{\mathbf{P}_{u\Delta t}\}_{0 \leq z \leq u-1} \cup \{\mathbf{P}_{u\Delta t}\}$$

end for.

Output: $\{\mathbf{P}_{k\Delta t}\}_{1 \leq k \leq n_s}$.

Firstly, the SDEs associated with the dynamics of the price vector and elementary factors are discretized using a simple Euler scheme. Interestingly, the price is subject to the positive part operator $(\cdot)_+$. This characteristic has two major advantages. Firstly, it ensures that prices remain positive or zero. Secondly, it allows for the possibility of bankruptcy or default: when the price reaches 0, its value remains null for all subsequent dates. In the case where the price vector corresponds to a set of stocks, this property allows, for example, to estimate by simulation the probability of default at a given time horizon for the different companies in the considered investment universe.

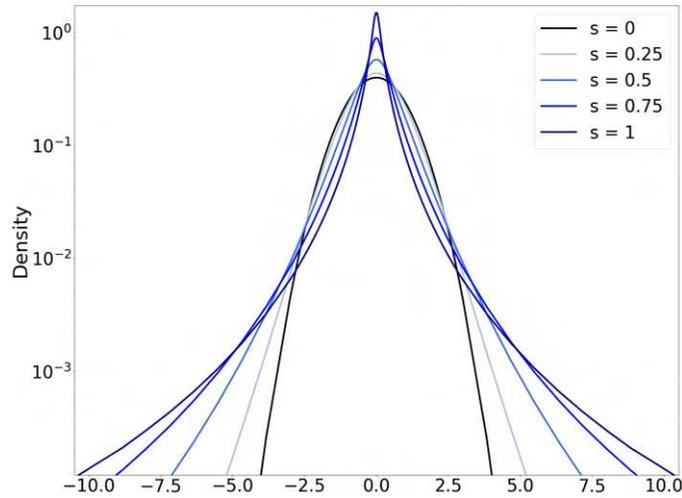
The EWMA of the variations and quadratic variations of the elementary factors are, in turn, aggregated respectively into matrices $\mathbf{M}^{(\tilde{\mu})}$ and $\mathbf{M}^{(\tilde{\Omega})}$ of dimension $m \times n_\tau$. Thus, the element at coordinate (j, p) of $\mathbf{M}^{(\tilde{\mu})}$ (resp. $\mathbf{M}^{(\tilde{\Omega})}$) corresponds to the EWMA with parameter τ_p of the variations (resp. quadratic variations) of the elementary factor j . Furthermore, the dynamics of these matrices are not directly modeled by discretizing the SDEs associated with them, but rather by discretizing the integrals that define these state variables. This approach, adopted for example in [182], has the beneficial property of ensuring the stability of the model regardless of the time step considered, which is not the case for classical discretization schemes of the SDEs for EWMA ([181]).

Now, concerning the drifts and factorial volatilities, only the components \mathbf{X} and \mathcal{S} need to be discretized. To start with, the process \mathbf{X} is not regarded in this discrete model as a stochastic process, in the sense that the different realizations of this random vector generated from this model form an i.i.d. sample. This stems from the assumption of very rapid mean reversion of the coordinates of \mathbf{X} as posited in section 6.3.1.1.3, which makes it consistent to model $\mathbf{X}_{t+\Delta t}$ from the asymptotic distribution of \mathbf{X} . Thus, for each simulation period:

$$\mathbf{X}_{t+\Delta t} \sim \mathcal{LN}(-\mathbf{S} \odot \mathbf{S}, \text{diag}(\mathbf{S})^2).$$

In practice, this choice is equivalent to opting for a non-Gaussian innovation process whenever $\mathbf{S} \neq \mathbf{0}_m$. Indeed, the marginal distributions of $\mathbf{X}_t \odot \mathbf{W}_t$ correspond (independent of each other) to Normal

Log-normal (NLN) mixture ([199]) of the form $W \cdot \exp(s(B - s))$, with $(W, B)^\top \sim \mathcal{N}(\mathbf{0}_2, \mathbf{I}_2)$. These marginal distributions have the important properties of sharing the first three centered moments regardless of the value taken by \mathbf{S} : zero mean, a standard deviation (variance) equal to 1, and zero skewness (see details in appendix 6.E.1). On the contrary, the kurtosis of the marginal distributions of $\mathbf{X}_t \odot \mathbf{W}_t$ depends on the specification of \mathbf{S} , since:



$$\mathbb{E} \left[\left(W e^{s(B-s)} \right)^4 \right] = 3e^{4s^2}.$$

Figure 6.4: NLN mixture distributions of the form $W e^{s(B-s)}$ with $(W, B)^\top \sim \mathcal{N}(\mathbf{0}_2, \mathbf{I}_2)$ for different values of s .

Thus, as illustrated in figure 6.4, the larger (the absolute value of) $(\mathbf{S})_j$, the thicker the tails of the distribu-

tions of $(\mathbf{X}_t \odot \mathbf{W}_t)_j$ exhibit. This characteristic enables capturing potential extreme variations of the elementary factors (and consequently, of the assets) occurring at short time scales. However, as mentioned earlier, the skewness of $(\mathbf{X}_t \odot \mathbf{W}_t)_j$ is necessarily zero due to the independence of $(\mathbf{W})_j$ and $(\mathbf{B})_j$, implying a symmetry of these factorial variations at the horizon Δt . Consequently, this modeling assumes that the potential asymmetry of the distributions of factorial increments is caused by the path-dependent component of volatility, and more specifically by the convolved dynamics features. For this reason, the choice of the simulation time step Δt is important. Indeed, in this framework, if the choice of a daily time step makes the model structurally unable to capture potential asymmetries in daily return distributions, opting for a smaller time step allows for a potential capture of these asymmetries through the impact of returns on the level of volatility.

Of course, alternative modeling choices are conceivable, either to capture a potential portion of the skewness unexplained by the dynamics of \mathcal{V}_t , or to maintain a relatively high simulation time step. One way is to relax the assumption of independence between \mathbf{W} and \mathbf{B} , so that the correlation between $(\mathbf{W})_j$ and $(\mathbf{B})_j$ can be different from 0. In this framework, the innovation process is still distributed according to an NLN mixture distribution but can exhibit a non-zero skew. However, the correlation between $(\mathbf{W})_j$ and $(\mathbf{B})_j$ results in a non-zero mean of the innovation process associated with factor j , which must be neutralized in order to maintain $\mathbb{E}[\Delta \mathbf{F}_{t+\Delta t}] = \boldsymbol{\mu}_t \cdot \Delta t$. Another possible approach to incorporate excess skewness is to replace $\mathbf{X}_t \odot \mathbf{W}_t$ with \mathbf{Z}_t , a vector whose marginals follow, for instance, skewed generalized t distributions or kernel density estimators (KDE). However, once again, this requires ensuring that \mathbf{Z}_t remains centered so that the factorial drifts stay fully captured by $\boldsymbol{\mu}$.

Remark 1 *The algorithm 4 implicitly assumes that the simulation time step is equal to the observation time step of the model on which the ARMA model parameters were calibrated. When this condition is not met, it is necessary first to simulate the trajectory of S using the time step with which the model was calibrated, and then perform interpolation between the simulation periods $\Delta t, 2\Delta t, \dots, n_s \Delta t$. In practice, linear interpolation will be used.*

6.3.2 Calibration of the market generator

6.3.2.1 Input information for model calibration

In the following sections, the calibration of 4 is performed based on a data matrix \mathbf{D} of the form:

$$\mathbf{D} = \begin{pmatrix} P_{1,t_0} & \cdots & P_{n,t_0} \\ P_{1,t_1} & \cdots & P_{n,t_1} \\ \vdots & \ddots & \vdots \\ P_{1,t_N} & \cdots & P_{n,t_N} \end{pmatrix} \quad (6.16)$$

where $t_0 < t_1 < \dots < t_N = T$. This matrix \mathbf{D} will be used to compute the central matrix for model calibration, namely the matrix of returns denoted \mathbf{R} , defined as:

$$\mathbf{R} = \begin{pmatrix} r_{1,1} & \dots & r_{n,1} \\ \vdots & \ddots & \vdots \\ r_{1,N} & \dots & r_{n,N} \end{pmatrix} \quad (6.17)$$

where:

$$r_{i,u} = \frac{P_{i,t_u} - P_{i,t_{u-1}}}{P_{i,t_{u-1}}}.$$

It is worth to note that if the formulation of \mathbf{D} does not imply a strict constancy of the observation frequencies of the vector \mathbf{P} , the estimation method proposed in the following sections is based on the assumption of moderate heterogeneity in \mathbf{P} frequencies. Thus, if moderate heterogeneity in daily observations caused by the presence of weekends or holidays is not problematic for the proposed estimation method, the same cannot be said for a mixture of daily and hourly observation frequencies, for instance.

6.3.2.2 The elementary factors decomposition

The price dynamics are driven in the FPDM model by the elementary factors. Accordingly, the first step in calibrating the market generator is to perform a decomposition of \mathbf{R} in the form:

$$\mathbf{R} = \widehat{\Delta}_{\mathbf{F}} \widehat{\mathbf{A}}^\top = \left(\widehat{\Delta}_{\mathbf{F}^c}^\top, \widehat{\Delta}_{\mathbf{F}^i}^\top \right)^\top \left((\widehat{\mathbf{A}}^c)^\top, \mathbf{I}_n \right) = \widehat{\Delta}_{\mathbf{F}^c} (\widehat{\mathbf{A}}^c)^\top + \widehat{\Delta}_{\mathbf{F}^i},$$

where $\widehat{\mathbf{A}}$ and $\widehat{\mathbf{A}}^c$ are respectively estimators of the loading matrix \mathbf{A} and \mathbf{A}^c , $\widehat{\Delta}_{\mathbf{F}}$ is the history of estimated variations of elementary factors:

$$\widehat{\Delta}_{\mathbf{F}} = \begin{pmatrix} \widehat{\mathbf{F}}_{t_1} - \widehat{\mathbf{F}}_{t_0} \\ \vdots \\ \widehat{\mathbf{F}}_{t_N} - \widehat{\mathbf{F}}_{t_{N-1}} \end{pmatrix}, \quad \widehat{\Delta}_{\mathbf{F}^c} = \begin{pmatrix} \widehat{\mathbf{F}}_{t_1}^c - \widehat{\mathbf{F}}_{t_0}^c \\ \vdots \\ \widehat{\mathbf{F}}_{t_N}^c - \widehat{\mathbf{F}}_{t_{N-1}}^c \end{pmatrix}, \quad \widehat{\Delta}_{\mathbf{F}^i} = \begin{pmatrix} \widehat{\mathbf{F}}_{t_1}^i - \widehat{\mathbf{F}}_{t_0}^i \\ \vdots \\ \widehat{\mathbf{F}}_{t_N}^i - \widehat{\mathbf{F}}_{t_{N-1}}^i \end{pmatrix}.$$

The method used for this purpose is defined by algorithm 5.

The first step of this proposed algorithm consists of using a matrix factorization commonly employed in various fields of data analysis: singular value decomposition. However, this factorization is not performed on \mathbf{R} , but on a standardized version of this matrix, $\bar{\mathbf{R}}$. The standardization process applied, defined in step 1, involves a double standardization: standardization per row followed by standardization per column. Therefore, it involves double standardization: standardization per row then standardization per column. The row standardization helps reduce the impact of volatility level variability on the singular value decomposition. This standardization makes particular sense

in the proposed model where the volatilities of elementary factors depend on the volatility level of the market factor and the sensitivity operator (\mathcal{S}). On the other hand, column standardization assigns equal importance to each asset, whereas a singular value decomposition performed directly on \mathbf{R} implicitly weights more volatile assets. Moreover, it normalizes the variance of the matrix $\bar{\mathbf{R}}$, denoted as v , to 1, which will allow for certain simplifications in subsequent steps.

Algorithm 5 Decomposition of the returns matrix \mathbf{R} into elementary factors.

Input: \mathbf{R}, n_δ

1. Compute the standardized matrix $\bar{\bar{\mathbf{R}}}$ from \mathbf{R} as follows:

$$\bar{\bar{\mathbf{R}}} = \mathbf{R}\mathbf{M}_{\text{st}} \quad \text{with} \quad \mathbf{M}_{\text{st}} = \text{diag}_{M \rightarrow D} \left(\frac{1}{N} \cdot \bar{\mathbf{R}}\bar{\mathbf{R}}^\top \right)^{-1/2},$$

$$\text{and where} \quad \bar{\mathbf{R}} = \text{diag}_{M \rightarrow D} \left(\frac{1}{n} \cdot \mathbf{R}^\top \mathbf{R} \right)^{-1/2} \mathbf{R}.$$

2. Perform a singular value decomposition of $\bar{\bar{\mathbf{R}}}$.

3. $v, q \leftarrow 1, \frac{n}{N}$

4. **Repeat until convergence:**

- (a) Compute

$$\phi_{\text{thr}} \leftarrow \arg \min_{\phi_i} \left| \phi_i - v(1 + \sqrt{q})^2 \right|$$

$$\tilde{\phi}_{\text{min}} \leftarrow \phi_{\text{thr}}(1 - t)$$

$$\tilde{\phi}_{\text{max}} \leftarrow \phi_{\text{thr}}(1 + t)$$

- (b) Estimate f_{KDE} , the density of $\{\phi_i \mid \phi_i \leq \phi_+\}$ using a kernel density method.

- (c) Find the parameters v and q that minimize:

$$v, q \leftarrow \arg \min_{v, q \in [0, 1]^2} \sum_{k=0}^{n_\delta} \left(\sqrt{f_{\text{KDE}}(\tilde{\phi}_k)} - \sqrt{f_{\text{MP}}(\tilde{\phi}_k | v, q)} \right)^2,$$

$$\text{where } \tilde{\phi}_k = \tilde{\phi}_{\text{min}} + \frac{\tilde{\phi}_{\text{max}} - \tilde{\phi}_{\text{min}}}{n_\delta} k.$$

5. Compute

$$\hat{\mathbf{A}}^c \leftarrow (\mathbf{U})_{[:, :m]}$$

$$\hat{\mathbf{\Delta}}_{\mathbf{F}^c} \leftarrow (\mathbf{D})_{[:, :m]} (\mathbf{V})_{[:, :m]}$$

$$\hat{\mathbf{\Delta}}_{\mathbf{F}^i} \leftarrow (\mathbf{D})_{[m_c :, m_c :]} (\mathbf{V})_{[:, :m]} (\mathbf{V})_{[m_c :, :]}^\top$$

Output: $\hat{\mathbf{A}}^c, \hat{\mathbf{\Delta}}_{\mathbf{F}^c}, \hat{\mathbf{\Delta}}_{\mathbf{F}^i}$

The second step of the algorithm 5 then proceeds to a singular value decomposition of $\bar{\bar{\mathbf{R}}}$, resulting

in the following factorization:

$$\bar{\mathbf{R}} = \mathbf{U}\mathbf{D}\mathbf{V}^\top,$$

where \mathbf{U} is a $N \times N$ matrix, \mathbf{V} a $n \times n$ matrix, and \mathbf{D} is a $N \times n$ matrix $\text{diag}_{M \rightarrow d}(\mathbf{D}) = (\sqrt{\phi_1}, \dots, \sqrt{\phi_n})^\top$ such as $\phi_1 > \dots > \phi_n$.

To build the estimators $\hat{\mathbf{A}}$ and $\hat{\Delta}_{\mathbf{F}}$ from the elements obtained through this decomposition, the next step involves distinguishing the dynamics of common elementary factors from those of idiosyncratic elementary factors. The m_c elements associated with the m_c largest singular values will define the component of $\bar{\mathbf{R}}$ generated by the common elementary factors, and the $n - m_c$ elements its component produced by the idiosyncratic elementary factors. More precisely, regarding the matrices associated with the common elementary factors:

$$\hat{\mathbf{A}}^c = \mathbf{M}_{\text{st}}^{-1/2}(\mathbf{V})_{[:,m_c]}, \quad \text{and} \quad \hat{\Delta}_{\mathbf{F}^c} = (\mathbf{U})_{[:,m_c]}(\mathbf{D})_{[m_c:,m_c]}. \quad (6.18)$$

As for the matrix of variations of idiosyncratic elementary factors:

$$\hat{\Delta}_{\mathbf{F}^i} = (\mathbf{U})_{[:,m_c+1:]}(\mathbf{D})_{[m_c+1:,m_c+1:]} \left(\mathbf{M}_{\text{st}}^{-1/2}(\mathbf{V})_{[:,m_c+1:]} \right)^\top. \quad (6.19)$$

The challenge is therefore to determine the value taken by m_c . In practice, this involves finding a threshold value ϕ_{thr} for the eigenvalues of $\frac{1}{N}\bar{\mathbf{R}}\bar{\mathbf{R}}^\top$ to separate the noise from the signal, an issue that has been extensively addressed in the academic literature ([135], [46], [85]). In these works based on results from random matrix theory, ϕ_{thr} is determined using the upper bound ϕ_+ of a Marchenko-Pastur distribution, whose density is defined by:

$$f_{\text{MP}}(\phi|v, q) = \begin{cases} \frac{\sqrt{(\phi_+ - \phi)(\phi - \phi_-)}}{2\pi q \phi} & \text{if } \phi \in [\phi_- : \phi_+], \\ 0 & \text{else,} \end{cases}$$

where $\phi_- = v(1 - \sqrt{q})^2$ and $\phi_+ = v(1 + \sqrt{q})^2$. However, if the values of the parameters $q = N/n$ and v correspond to the variance of $\frac{1}{N}\bar{\mathbf{R}}\bar{\mathbf{R}}^\top$ in the case where the entries of $\bar{\mathbf{R}}$ are independent identically distributed random variables, the choice of these values is no longer straightforward when this assumption does not hold (which is a priori the case here). For this reason, several approaches coexist in the literature for determining these parameters ([135], [85]).

The method adopted here is largely inspired by the one proposed by Lopez de Prado ([85]). This involves, after initializing v and q to 1 and $\frac{n}{N}$, the iterative execution of three successive steps. The first one assigns a provisional value to ϕ_{thr} as the nearest eigenvalue to $v(1 + \sqrt{q})^2$. The second step consist in estimating the density of the distribution of eigenvalues that are less than or equal to ϕ_{thr} using a kernel density method, density denoted f_{KDE} . The third step updates the parameters v and q to the values that minimize (a proxy of) the Hellinger distance between $f_{\text{KDE}}(\cdot)$

and $f_{\text{MP}}(\cdot|v, q)$. These three steps are then repeated until the convergence of ϕ_{thr} .

After this iterative process, m_c is defined by the number of eigenvalues strictly greater than ϕ_{thr} . The matrices $\widehat{\Delta}_{\mathbf{F}}^c$, $\widehat{\Delta}_{\mathbf{F}}^i$, and $\widehat{\mathbf{A}}^c$ are finally calculated using equations 6.18 and 6.19.

6.3.2.3 Estimation procedure

The objective now is to estimate all the parameters (other than the matrix \mathbf{A}) and state variables taken as input by the FPDM generator (algorithm 4), using the matrix of variations of the elementary factors $\widehat{\Delta}_{\mathbf{F}}$ obtained via algorithm 5. The method chosen for this purpose is defined by algorithm 6.

This one starts by calculating the various EWMA of the variations and quadratic variations of the elementary factors. These EWMA are initialized at period t_{rw} from the first rw periods (step 1.a. of algorithm 6) using a discretisation of

$$\begin{aligned}\tilde{\boldsymbol{\mu}}_{t_{rw}}^{(k)} &= \tilde{\boldsymbol{\mu}}_{t_0}^{(k)} \cdot e^{\frac{t_0 - t_{rw}}{\tau_k}} + \frac{1}{\tau_k} \int_{t_0}^{t_{rw}} e^{\frac{u - t_{rw}}{\tau_k}} d\mathbf{F}_u, \\ \text{and } \tilde{\boldsymbol{\Omega}}_{t_{rw}}^{(k)} &= \tilde{\boldsymbol{\Omega}}_{t_0}^{(k)} \cdot e^{\frac{t_0 - t_{rw}}{\tau_k}} + \frac{1}{\tau_k} \int_{t_0}^{t_{rw}} e^{\frac{t_0 - t_{rw}}{\tau_k}} (d\mathbf{F}_u \odot d\mathbf{F}_u).\end{aligned}$$

The values of $\tilde{\boldsymbol{\mu}}_{t_0}^{(k)}$ and $\tilde{\boldsymbol{\Omega}}_{t_0}^{(k)}$ are thus initialized using the means of the variations and quadratic variations of the elementary factors over the entire training period, which compensates for the lack of data for periods before t_0 . In practice, when rw is large enough, the impact of this initialization choice is small, as only the EWMA associated with a high τ_k are affected. For periods beyond t_{rw} , updating the values of EWMA is done in the same way as adopted in the market generator.

This initial step is followed by a second phase that involves considering the log-likelihood function associated with the equation governing the dynamics of the vector of elementary factors, under the assumption that $\mathbf{X}_t = \mathbf{1}_m$, such as:

$$\Delta \mathbf{F}_{t_u} \sim \mathcal{N}\left(\hat{\boldsymbol{\mu}}_{t_u} \cdot \Delta_u, \text{diag}(\mathcal{S}_u \cdot \boldsymbol{\nu}_u \cdot \Delta_u)^2\right),$$

where $\Delta_u = t_u - t_{u-1}$. This log-likelihood function is given by:

$$\begin{aligned}\mathcal{L}\left(\Theta | \widehat{\Delta}_{\mathbf{F}}\right) &= -\frac{1}{2} \sum_{u=rw+1}^N \left((\widehat{\Delta}_{\mathbf{F}})_u - (\boldsymbol{\mu})_u \cdot \Delta_u \right)^\top \left(\Delta_u \cdot (\mathcal{S}_u \cdot \boldsymbol{\nu}_u)^{\circ 2} \mathbf{I}_m \right)^{-1} \left((\widehat{\Delta}_{\mathbf{F}})_u - (\boldsymbol{\mu})_u \cdot \Delta_u \right) \\ &\quad + \log \left| \Delta_u \cdot (\mathcal{S}_u \cdot \boldsymbol{\nu}_u)^{\circ 2} \mathbf{I}_m \right| - m \log(2\pi), \quad (6.20)\end{aligned}$$

where Θ represents the set of parameters on which $\boldsymbol{\mu}$ and $\boldsymbol{\nu}$ depend. However, on the contrary to the classical maximum likelihood method, the proposed approach does not strictly maximize the likelihood of the model, due to its sequential structure. Indeed, step 2 of Algorithm 6 involves

iteratively maximizing 6.20 only over subsets of the model's parameter space. This first subset (step 2a) is defined by:

$$\Theta_1 = \{\boldsymbol{\delta}, \mathbf{w}, (\mathbf{b}_0)_1, (\mathbf{b}_1)_1, (\mathbf{b}_2)_1, (\bar{\boldsymbol{\mu}})_1, \zeta, \lambda\}.$$

Step 2a therefore consists in maximizing the likelihood of the univariate model of the dynamics of the first elementary factor, corresponding to the market factor. Since $\boldsymbol{\delta}$ and \mathbf{w} are estimated in this step, the kernels $K^{(\hat{\boldsymbol{\mu}})}$ and $K^{(\hat{\boldsymbol{\sigma}})}$ are determined solely from the data for this factor. This separate and primary optimization of the parameters of the market factor is motivated by the singular role it plays in the model. Step 2b involves maximizing 6.20 with respect to the parameter set Θ_2 , defined as:

$$\Theta_2 = \{(\mathbf{b}_0)_j, (\mathbf{b}_1)_j, (\mathbf{b}_2)_j, (\mathbf{b}_3)_j\}_{2 \leq j \leq m}.$$

Thus, this step determines the parameters that are functions of $\boldsymbol{\mu}$ and $\boldsymbol{\nu}$ associated with all elementary factors, with the exception of those associated with the market factor (estimated in step 2a). Step 2c, which constitutes the final step of phase 2, involves maximizing the log-likelihood with respect to \mathcal{S}_{t_u} for $rw + 1 \leq u \leq N$. In this step, the values of \mathcal{S} are directly obtained using the analytical solution of the problem presented in appendix 6.F.

Once the value of the log-likelihood function has converged after iterating through steps 2a, 2b, 2c, the algorithm estimates a_1, a_2, a_3 by fitting an ARMA(1, 1) model to the series $\{\log(\hat{\mathcal{S}}_{t_u})\}_{rw+1 \leq u \leq N}$, using the standard maximum likelihood estimation approach ([49]).

The final phase involves estimating the vector \mathbf{S} on which \mathbf{X} depends, utilizing a method-of-moments approach whose rationale is explained in detail in appendix 6.E.3. First, step 4a calculates the fourth-order moments of the m samples⁴:

$$\left\{ \frac{\left(\Delta \hat{\mathbf{F}}_u - \hat{\boldsymbol{\mu}}_u \cdot \Delta u \right)_j}{\left(\mathcal{S}_u \cdot \boldsymbol{\nu}_u \cdot \sqrt{\Delta u} \right)_j} \right\}_{rw+1 \leq u \leq N}.$$

The standardization of factor variations (i.e., $\hat{\mathbf{F}}_u$) serves two main purposes: firstly, it allows us to treat these samples as being i.i.d., and secondly, it enables us to isolate the impact of \mathbf{X} . Thus, in the specific case where the simulation time step Δt equals the average time step between observations, the computed empirical moments in step 4a directly serve as estimators of $\mathbb{E}[(\mathbf{X} \odot \mathbf{W})_j^4]$. In a broader context, these fourth-order moments enable us to estimate \mathbf{S} using the formula derived in appendix 6.E.3, which is precisely what step 4b accomplishes.

⁴The algorithm 6 uses the biased estimator of $\mathbb{E}[(\mathbf{X} \odot \mathbf{W})_j^4]$. A consistent alternative would be to use its unbiased estimator.

Algorithm 6 Estimation of the FPDM generator.

Input: $\widehat{\Delta}_{\mathbf{F}}, \Delta t$

1. **For** $k = 1$ **to** n_τ **do**

$$(a) \left\{ \tilde{\boldsymbol{\mu}}_{trw}^{(k)} \right\} \leftarrow \frac{e^{-\frac{t_0-trw}{\tau_k}}}{N} \cdot \sum_{u=1}^N \frac{(\widehat{\Delta}_{\mathbf{F}})_u}{t_u - t_{u-1}} + \frac{1}{\tau_k} \sum_{u=1}^{rw} e^{-\frac{t_u-trw}{\tau_k}} \cdot (\widehat{\Delta}_{\mathbf{F}})_u$$

$$\left\{ \tilde{\boldsymbol{\Omega}}_{trw}^{(k)} \right\} \leftarrow \frac{e^{-\frac{t_0-trw}{\tau_k}}}{N} \cdot \sum_{u=1}^N \frac{(\widehat{\Delta}_{\mathbf{F}} \odot \widehat{\Delta}_{\mathbf{F}})_u}{t_u - t_{u-1}} + \frac{1}{\tau_k} \sum_{u=1}^{rw} e^{-\frac{t_u-trw}{\tau_k}} \cdot (\widehat{\Delta}_{\mathbf{F}} \odot \widehat{\Delta}_{\mathbf{F}})_u$$

(b) **For** $u = rw + 1$ **to** N **do**

$$\left\{ \tilde{\boldsymbol{\mu}}_{tz}^{(k)} \right\}_{0 \leq z \leq u} \leftarrow \left\{ \tilde{\boldsymbol{\Omega}}_{tz}^{(k)} \right\}_{0 \leq z \leq u} \cup \left\{ e^{-\frac{t_u-t_{u-1}}{\tau_k}} \cdot \tilde{\boldsymbol{\mu}}_{t_{u-1}}^{(k)} + \frac{1}{\tau_k} \cdot (\widehat{\Delta}_{\mathbf{F}})_u \right\}$$

$$\left\{ \tilde{\boldsymbol{\Omega}}_{tz}^{(k)} \right\}_{0 \leq z \leq u} \leftarrow \left\{ \tilde{\boldsymbol{\Omega}}_{tz}^{(k)} \right\}_{0 \leq z \leq u-1} \cup \left\{ e^{-\frac{t_u-t_{u-1}}{\tau_k}} \cdot \tilde{\boldsymbol{\Omega}}_{t_u}^{(k)} + \frac{1}{\tau_k} \cdot (\widehat{\Delta}_{\mathbf{F}} \odot \widehat{\Delta}_{\mathbf{F}})_u \right\}$$

End for.
End for.

2. **Repeat until convergence:**

$$(a) \hat{\boldsymbol{\Theta}}_1^* \leftarrow \arg \min_{\boldsymbol{\Theta}_1 \in \mathcal{S}_\Theta} \mathcal{L} \left(\boldsymbol{\Theta} | \widehat{\Delta}_{\mathbf{F}} \right)$$

$$(b) \hat{\boldsymbol{\Theta}}_2^* \leftarrow \arg \min_{\boldsymbol{\Theta}_2 \in \mathcal{S}_\Theta} \mathcal{L} \left(\boldsymbol{\Theta} | \widehat{\Delta}_{\mathbf{F}} \right)$$

(c) **For** $u = rw + 1$ **to** N **do**

$$\hat{\mathcal{S}}_u^* \leftarrow \sqrt{\sum_{j=1}^m \frac{\left((\Delta \mathbf{F}_{t_u})_j - (\hat{\boldsymbol{\mu}}_{t_u})_j \cdot \Delta t_u \right)^2}{(\boldsymbol{\nu}_{t_u})_j^2 \cdot \Delta t_u}}$$

3. Fit ARMA(1,1) based on the time series $(\log(\hat{\mathcal{S}}_1^*), \dots, \log(\hat{\mathcal{S}}_N^*))$ to obtain $\hat{a}_0, \hat{a}_1, \hat{a}_2, \hat{\epsilon}_N$.

4. **For** $j = 1$ **to** m **do**

$$(a) \hat{c}_{j,4} \leftarrow \frac{1}{N - rw - 1} \sum_{u=rw+1}^N \left(\frac{(\Delta \mathbf{F}_u - \hat{\boldsymbol{\mu}}_u \cdot \Delta u)_j}{(\mathcal{S}_u \cdot \boldsymbol{\nu}_u \cdot \sqrt{\Delta u})_j} \right)^4$$

$$(b) (\hat{\mathbf{S}})_j \leftarrow 0.5 \sqrt{\log \left(\frac{1}{N-rw-1} \sum_{u=rw+1}^N (t_u - t_{u-1}) \left(\frac{\hat{c}_{j,4}}{3} - 1 \right)_+ + 1 \right)}.$$

End for.
Output: $\left\{ \tilde{\boldsymbol{\mu}}_{t_N}^{(k)} \right\}_{0 \leq k \leq n_\tau}, \left\{ \tilde{\boldsymbol{\Omega}}_{t_N}^{(k)} \right\}_{0 \leq k \leq n_\tau}, \hat{\boldsymbol{\Theta}}_1^*, \hat{\boldsymbol{\Theta}}_2^*, \hat{a}_0, \hat{a}_1, \hat{a}_2, \hat{\epsilon}_N$

6.4 Empirical assessment of the FPDM generator performance

6.4.1 Modalities of the conducted assessments

6.4.1.1 The considered market dataset

The objective of section 6.4 is to assess the capability of the FPDM generator defined in section 6.3.1.2 to produce realistic market scenarios, capturing various characteristics of the considered time series. We begin by defining the data under consideration.

Firstly, we consider an equity investment universe: the assets comprising the S&P500 index. More precisely, we focus on the 436 assets belonging to the S&P500 as of April 30, 2024, for which we have daily historical data from April 1, 2010, to April 30, 2024. The first 8 years of historical data, spanning from April 1, 2010, to April 30, 2018, constitute the training dataset used to calibrate the FPDM Generator. The remaining historical data, covering the period from May 1, 2018, to April 30, 2024, will be used as the test dataset for evaluating the model's performance. This division has the particularity that the training set does not include periods of very strong market turbulence similar to that of the 2020 stock market crash included in the test set, which enables to assess the model's ability to reproduce such events even without them appearing in the data used for its training.

6.4.1.2 The set of parameters of the generator

The various parameters of the market generator are determined by applying algorithms 5 and 6 successively on the training set.

To begin, the application of algorithm 5 results in obtaining 25 common elementary factors. Consequently, the FPDM Generator used comprises a total of 461 common elementary factors: 25 common factors and 436 idiosyncratic factors. If we focus more specifically on the first estimated common factor, the correlation between its daily variations and the returns of the S&P500 index over the same periods is 0.985 (see appendix 6.G.2). Therefore, considering this factor as the market factor is entirely coherent. It is also interesting to examine the separation achieved by algorithm 5 between the eigenvalues associated with the common elementary factors and the remaining eigenvalues, as illustrated in figure 6.5. Firstly, the Marchenko Pastur distribution estimated by the algorithm fits the data quite well. Additionally, the separation threshold between the 25 largest and the smaller eigenvalues coincides with several changes in the behavior of the spectrum of $\frac{1}{N} \bar{\mathbf{R}} \bar{\mathbf{R}}^\top$. Indeed, as illustrated by the plot in the top left (figure 6.5), the density of eigenvalues as a function of their magnitude appears to follow a power-law relationship for the 25 largest eigenvalues, which is not the case for eigenvalues below this threshold. Relatedly, except for the first eigenvalue associated with the market factor, the eigenvalues associated with the other common risk factors

follow the relationship: $\log(\phi_k) \approx a - b \cdot \log(\text{rank}(\phi_k))$. This is illustrated by the two plots at the bottom of figure 6.5, where the black line corresponds to the OLS regression line $a - b \cdot \log(\text{rank}(\phi_k))$ estimated from eigenvalues 2 to 25 (the parameters obtained here are $\hat{a} = 3.45$ and $\hat{b} = 0.92$). Even more remarkably, when performing the same regression on the oracle eigenvalues estimated using the approach proposed in [138], \hat{b} is almost exactly equal to 1, which could suggest the existence of an underlying fundamental financial relationship.

The algorithm 6 is then employed, subsequent to the application of algorithm 5, to estimate the remaining parameters of the model. This entails specifying $\{\tau_k\}_{k=1}^{n_\tau}$, Δt , and rw , in addition to the matrix obtained from algorithm 5. To determine the values of $\{\tau_k\}_{k=1}^{n_\tau}$, we utilize equation 6.14 (section 6.3.1.1.2) with $n_\tau = 20$, $\tau_- = 1/365$, and $\tau_+ = 5$, all expressed in years. The time step utilized for the simulations is $\Delta t = 1/3650$, also expressed in years. Moreover, we select a parameter $rw = 1008$, specified in the number of periods, which roughly corresponds to the initial 4 years of the training dataset used to initialize the various EWMA components of the model. Consequently, the estimation of the model parameters is practically conducted over a period of approximately 4 years, spanning from April 2014 to the end of April 2018.

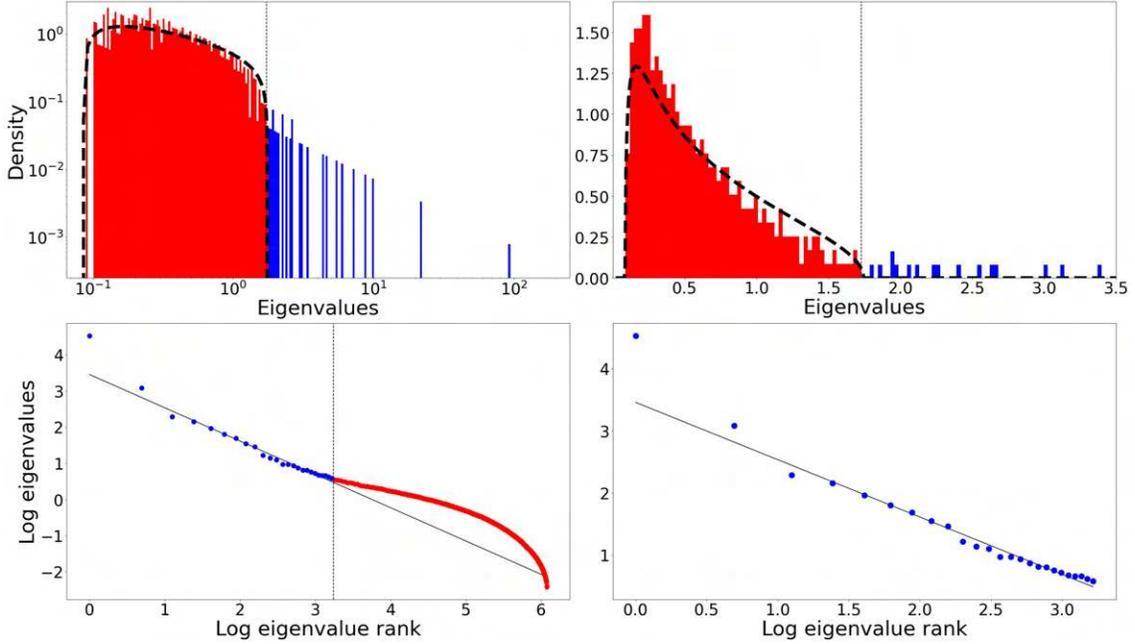


Figure 6.5: Histogram of the eigenvalue distribution of $\frac{1}{N} \bar{\mathbf{R}} \bar{\mathbf{R}}^\top$. The black dashed line that fits the red part of the histograms corresponds to the fitted Marchenko-Pastur distribution obtained from algorithm 5, used to separate common elementary factors from idiosyncratic elementary factors.

Several observations can be made following the application of algorithm 6. Firstly, as shown in figure 6.18 in appendix 6.G.1, the obtained kernels closely resemble time-shifted power law (TSPL) kernels, corroborating recent findings in the volatility literature ([117]). Regarding the parameters of

the drift and the path-dependent component of the volatility associated with the market factor, they are respectively of the same sign: $\bar{\mu}, \zeta, b_{0,1}, \lambda, b_{2,1}$ are positive, and $\zeta, b_{1,1}$ are negative. However, if the drift of the market factor is quite close to a positive affine relationship with the $(\mathbf{V}_t)_1$, then $\bar{\mu}/b_{0,1} > \zeta/b_{1,1} > \lambda/b_{2,1}$. Consequently, the constant part of the drift plays a more important role than in its approximation of the form $\Gamma_t^* \approx \bar{\mu} + \lambda^V \cdot (\mathbf{V}_t)_1$. Similarly, the reversal effect is more significant than the historical volatility risk premium effect if the affine relationship were perfectly respected. Another notable observation is the strong correlation of 0.875 between the purely path-dependent volatility component associated with the market factor (i.e., $(\mathbf{V})_1$) and the VIX, despite the latter not being included in the training set. This element suggests a high level of coherence between the volatility of the market factor estimated by the model results and the market’s priced volatility. Additionally, significant differences in the fourth-order moments are observed between the common elementary factors and the idiosyncratic elementary factors. For the 25 common factors, this moment ranges from 2.75 to 4.1, with mean and median values of 3.26 and 3.18, respectively, close to the Gaussian assumption (where this moment equals 3). In contrast, for the idiosyncratic factors, this moment ranges from 2.98 to 78, with mean and median values of 12.8 and 9.64, respectively.

6.4.2 General properties of generated datasets

The objective of this initial series of evaluations is to compare the overall properties of synthetic data generated by the FPDM generator model with market data.

6.4.2.1 Evaluation based on the marginal distributions

To begin, we delve into the individual dynamics of assets using simulated data, aiming to assess the consistency of their marginal return distributions with their empirical counterparts. Our analysis focuses on the first four moments of the daily, weekly, and monthly returns distributions for this purpose.

The comparison method used is as follows: we start by computing the various empirical moments considered, for both the historical realization and for each market scenario generated by the market generator. Therefore, for each asset i /moment p /temporal horizon l combination (e.g. the empirical mean of daily returns for Apple stock), we obtain a sample of 1000 empirical moments from the corresponding 1000 simulations. From the empirical cumulative distribution function $\tilde{F}_{i,p,l}$ estimated from this sample, we compute the estimated cumulative probability of the empirical moment obtained from the actual market data sample:

$$p_{i,p,l} = \tilde{F}_{i,p,l}(\hat{m}_{i,p,l}),$$

where $\hat{m}_{i,p,l}$ is the empirical p -th moment of the returns with a temporal horizon l of asset i calculated on real data. Subsequently, for each moment/time horizon pair, we compute the mean of these cumulative probabilities across the 436 assets, along with the three quartiles and the

proportion of cumulative probabilities between 0.05 and 0.95. The results are then report in the table 6.1.

		Mean $p_{i,p,l}$	Q1 $p_{i,p,l}$	Med. $p_{i,p,l}$	Q3 $p_{i,p,l}$	Prop. $p_{i,p,l} \in I_{90\%}$
Daily	Moment 1	0.535	0.349	0.589	0.749	0.913
	Moment 2	0.582	0.457	0.643	0.765	0.933
	Moment 3	0.462	0.203	0.43	0.718	0.966
	Moment 4	0.6	0.499	0.643	0.753	0.954
Weekly	Moment 1	0.426	0.233	0.412	0.574	0.954
	Moment 2	0.55	0.386	0.612	0.741	0.931
	Moment 3	0.488	0.265	0.462	0.698	0.961
	Moment 4	0.582	0.456	0.617	0.753	0.95
Monthly	Moment 1	0.427	0.236	0.41	0.571	0.954
	Moment 2	0.535	0.349	0.589	0.748	0.913
	Moment 3	0.398	0.205	0.315	0.561	0.97
	Moment 4	0.565	0.418	0.62	0.756	0.936

Table 6.1: Statistics related to the estimated cumulative probabilities set $\{p_{i,p,l}\}_{1 \leq i \leq 436}$. For example, the cell corresponding to the "Mean" column and the "Daily/Moment 3" row represents $\frac{1}{436} \sum_{i=1}^{436} p_{i,3,1} = 0.462$.

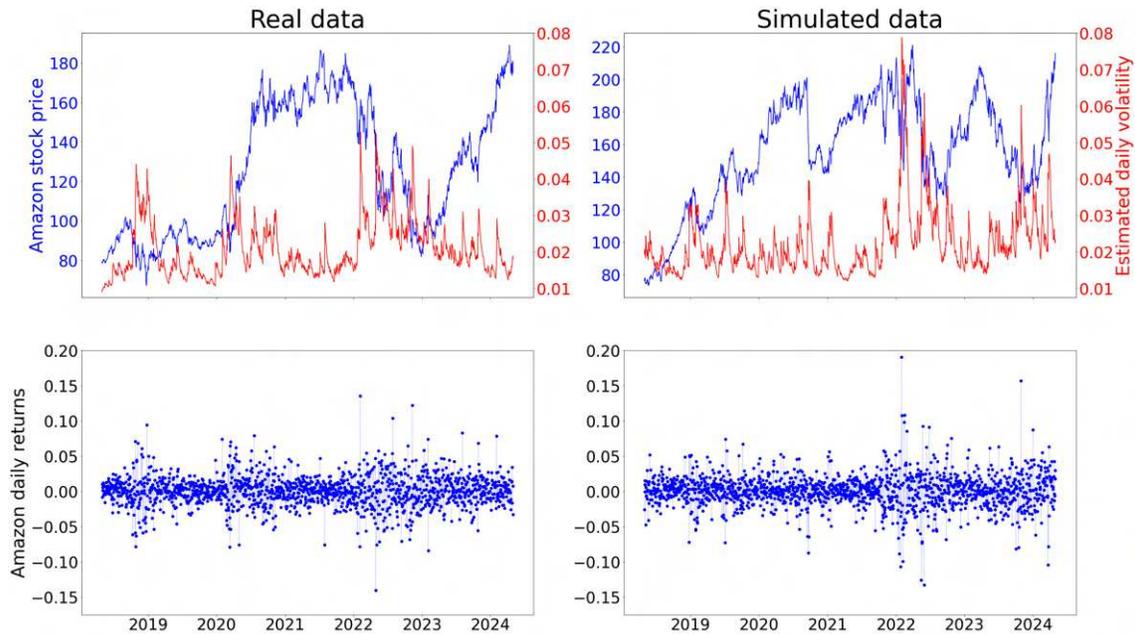


Figure 6.6: Price trajectory, volatility (estimated from a GARCH(1,1) model), and daily returns of Amazon stock: real data vs simulated data. The features characterizing the real data, such as the phenomenon of volatility spikes and volatility clustering, strong daily price swings, or the leverage effect, are well reproduced by the model.

The obtained figures tend to demonstrate a very good fit between the distributions of the returns generated by the model and their empirical counterparts. Indeed, the various empirical moments of the real data are generally close to the mean and median levels of their counterparts obtained

through simulations. Thus, except for the third-order moment of monthly returns, for all others pairs p and l , the mean and median of $\{\tilde{F}_{i,p,l}(\hat{m}_{i,p,l})\}_{i=1}^{436}$ fall between 0.4 and 0.6. Similarly, for each moment/time horizon pair, over 90% of the assets exhibit empirical moments within the 90% confidence interval of the model.

Moreover, the synthetic data generated by the FPDM Generator are not only coherent with market data in terms of asset return distributions, but also in terms of trajectories. Indeed, as illustrated in figure 6.6 with the example of Amazon stock⁵, the various characteristics of the price and return trajectories of individual stocks are well reproduced. It is particularly striking regarding the joint dynamics of price and volatility: in line with empirical data, on the one hand, volatility spikes coincide in the vast majority of cases with price drops; on the other hand, volatility tends to decrease relatively slowly following these shocks. This accurate modeling enables the reproduction of another related feature of financial series: the volatility clustering effect, which denotes the coexistence of periods of low and high volatility.

6.4.2.2 Evaluation based on the joint asset prices dynamics

The objective of this section is to assess the model's ability to capture the correlation structure of price dynamics.

To this end, the first type of evaluation will focus on the covariance matrix and the correlation matrix of daily returns. More specifically, we will evaluate estimators of the correlation matrix and the covariance matrix of the daily returns of the considered universe obtained from the simulations generated by the FPDM generator, with the sample estimators of these matrices computed for the test period. For this purpose, we will use as benchmarks, estimators of these matrices calculated on historical data over the period on which the FPDM generator is trained (i.e. from April 2014 to the end of April 2018). For both estimators calculated on historical and synthetic data, four types of estimators are considered⁶:

1. The unbiased sample estimator.
2. The Oracle approximating shrinkage (OAS) proposed in [62].
3. The first linear shrinkage estimator of Ledoit and Wolf introduced in [137].
4. The non-linear shrinkage estimator also proposed by Ledoit and Wolf, as detailed in [138].

To compare these estimators, two evaluation metrics are employed. The first one is the Frobenius norm of the difference between the sample test covariance (resp. correlation) matrix \mathbf{C} and the

⁵We take the example of Amazon here, as it is one of the largest capitalizations in the S&P500. However, the observations in this paragraph hold for all simulated assets.

⁶For each type of estimator, we first calculate the covariance matrix estimator, then from this, we compute the associated correlation matrix.

estimator covariance (resp. correlation) matrix $\hat{\mathbf{C}}$:

$$L_F(\hat{\mathbf{C}}, \mathbf{C}) = \|\hat{\mathbf{C}} - \mathbf{C}\|_F.$$

This is equivalent to calculating the root sum squared error between the elements of the two matrices. The second one is the minimum loss function proposed in [93], defined by:

$$L_{ML}(\hat{\mathbf{C}}, \mathbf{C}) = \frac{\text{Tr}(\hat{\mathbf{C}}^{-1} \mathbf{C} \hat{\mathbf{C}}^{-1})/n}{(\text{Tr}(\hat{\mathbf{C}}^{-1})/n)^2} - \frac{1}{\text{Tr}(\mathbf{C}^{-1})/n}.$$

As Ledoit and Wolf specify ([137], [138]), this cost function is designed to measure the quality of an estimator of the covariance matrix for use cases "where variance minimization decisions must be taken". The table 6.2 reports the obtained results.

		Corr. Matrix		Cov. Matrix	
		L_F	L_{ML}	L_F	L_{ML}
Sample	Hist. data	5.98×10^1	4.67×10^{-1}	5.68×10^{-2}	2.04×10^{-4}
	Sim. data	6.17×10^1	2.78×10^{-1}	4.01×10^{-2}	1.30×10^{-4}
OAS	Hist. data	6.10×10^1	3.94×10^{-1}	5.71×10^{-2}	1.72×10^{-4}
	Sim. data	6.17×10^1	2.78×10^{-1}	4.01×10^{-2}	1.30×10^{-4}
LW linear	Hist. data	6.15×10^1	3.74×10^{-1}	5.73×10^{-2}	1.64×10^{-4}
	Sim. data	6.17×10^1	2.78×10^{-1}	4.01×10^{-2}	1.30×10^{-4}
LW non-linear	Hist. data	6.15×10^1	3.67×10^{-1}	5.68×10^{-2}	1.65×10^{-4}
	Sim. data	5.64×10^1	2.70×10^{-1}	4.00×10^{-2}	1.24×10^{-4}

Table 6.2: Comparison of different estimators of the correlation and the covariance matrix of daily returns.

Firstly, the results concerning the correlation matrix are heterogeneous depending on the cost function considered. If we begin by focusing on the cost measured by the Frobenius loss function, the costs associated with correlation matrices calculated on historical data are generally close to their counterparts calculated on data simulated by the model. If we specifically center our analysis on the sample estimator, the estimator calculated on historical data slightly outperforms its counterpart calculated on synthetic data. While this result may initially appear somewhat disappointing regarding the model's ability to obtain better estimators of correlations, several factors must be considered to interpret it as accurately as possible. Firstly, the correlation structure is dynamic in the FPDM generator. However, outperforming the sample estimator on this metric and over this test period is much simpler with a constant correlation matrix model. In addition, due to this dynamic nature, the sample correlation matrix changes from one sample (generated by the model) to another. Thus, if we consider each synthetic sample separately, in 52.3% of cases, the sample correlation estimator obtained from these samples is associated with a Frobenius loss lower than that obtained with the sample correlation matrix calculated on the historical data. Furthermore, the most efficient estimator of the correlation matrix based on the criterion of Frobenius loss is the non-linear Ledoit-Wolf shrinkage estimator calculated on the simulated data. If we now focus on the

comparison between the estimators of the correlation matrix based on the criterion of the minimum loss function, the results clearly favor the simulated data. Thus, on this criterion, the cost of the sample correlation matrix calculated from this simulated data is more than 40% lower than the cost associated with its counterpart calculated on historical data, and about 25% lower than the cost of the best estimator calculated on historical data (the non-linear Ledoit-Wolf shrinkage estimator). The results obtained for the covariance matrices are similar. Indeed, for both considered cost functions, the sample estimator of the covariance matrix calculated from simulated data significantly outperforms all estimators of this covariance matrix calculated on historical data. Thus, considering the Frobenius loss, incorporating the effect of the standard deviations of returns of different assets improves the relative performance of the sample estimator calculated on simulated data compared to its associated correlation matrix. Therefore, in this context of use (that of calculating covariance matrices from a set of simulated scenarios over a medium to long-term horizon), the strength of the model seems to lie less in its ability to better capture linear correlations between asset returns than in improving the estimation of their individual standard deviations while maintaining a realistic correlation structure.

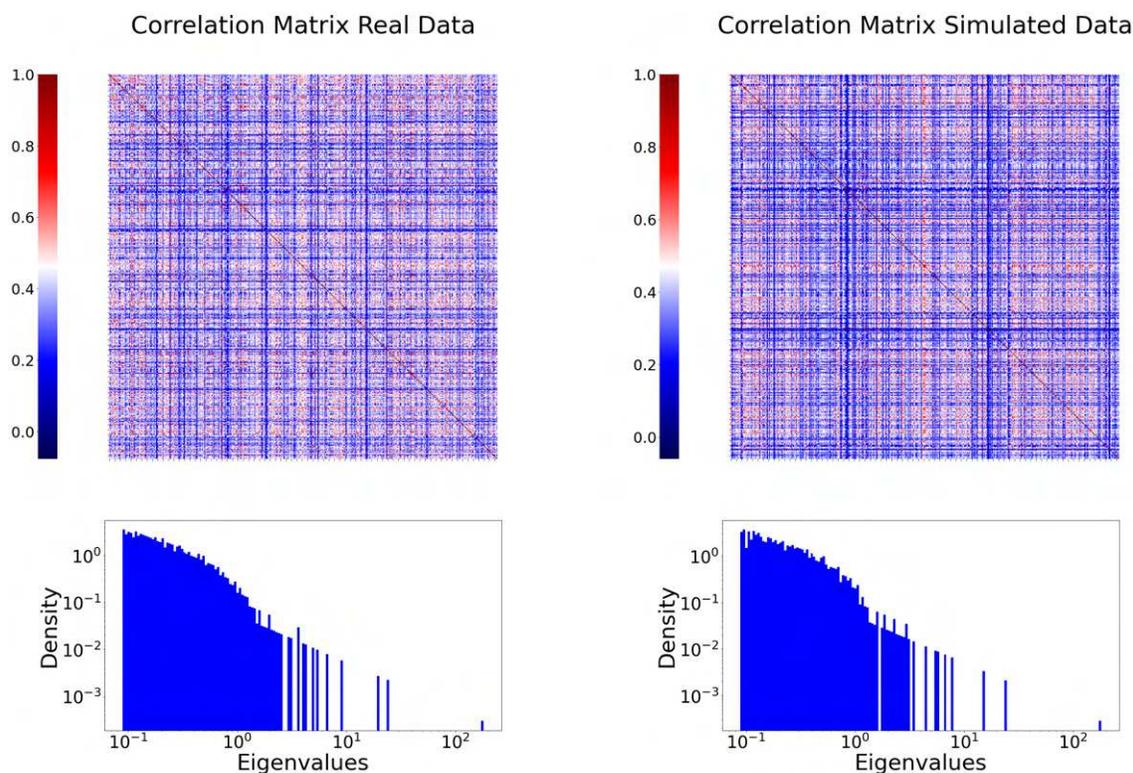


Figure 6.7: The sample correlation matrix and its spectrum: on the left obtained from the real test data, on the right obtained from a sample generated by the FPDM generator.

Other interesting findings emerge when considering, not the aggregation of the different generated scenarios, but each scenario separately. Specifically, while the sample correlation matrix varies

significantly from one generated sample to another, its spectrum retains features characteristic of asset return correlation matrices. That’s illustrated in figure 6.7. Thus, like its counterpart calculated on empirical data, the spectra of these matrices exhibit a similar structure, with, on one side, the largest eigenvalues, which are distributed following a power law distribution, and the other eigenvalues, which are distributed according to a Marchenko-Pastur law.

Due to the fact that stock return distributions are non-elliptical ([64]), it is interesting to complement this analysis of the data’s correlation structure by focusing on correlation measures capable of capturing more complex dependencies and tail behaviors. To this end, we now focus on two rank correlation measures: Spearman’s rho and Kendall’s tau. Similarly to the evaluation conducted on linear correlations, we then compare for each pair of assets (i.e. 94830 pairs, $436 \times (436 - 1)/2 = 94830$) the empirical Spearman’s rho and Kendall’s tau of the test sample with the empirical Spearman’s rho and Kendall’s tau of the simulated data. In addition, two benchmarks for these measures are used to put the obtained results into perspective: the empirical estimators and the estimators associated with the Gaussian copula calculated from the training sample⁷. The table 6.3 reports the obtained results.

	Spearman’s Rho		Kendall’s Tau	
	RMSE	MAE	RMSE	MAE
Emp. hist. data	9.006×10^{-2}	7.191×10^{-2}	6.510×10^{-2}	5.208×10^{-2}
Gaussian Copula	10.672×10^{-2}	8.653×10^{-2}	7.909×10^{-2}	6.426×10^{-2}
Emp. sim. data	9.069×10^{-2}	7.186×10^{-2}	6.603×10^{-2}	5.229×10^{-2}

Table 6.3: Comparison of different estimators of the correlation and the covariance matrix of daily returns.

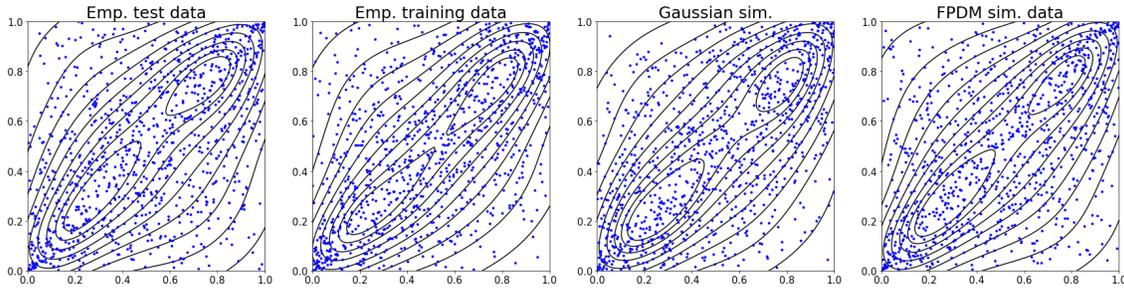


Figure 6.8: Dependograms of Johnson&Johnson and Pfizer stock daily returns.

As shown in table 6.3, the performance of the rank correlation measures estimated on simulated data is substantially identical to their empirical counterparts estimated on historical data. Furthermore, both of these outperform the estimators obtained under the assumption of a Gaussian correlation structure. Consequently, the model effectively captures the rank relationship that links the different pairs of returns present in the sample on which the generator is trained. While this may not be

⁷When two random variables are linked by a bivariate Gaussian copula with parameter ρ , the Spearman’s rho and Kendall’s tau are respectively equal to $\frac{6}{\pi} \arcsin(0.5\rho)$ and $\frac{2}{\pi} \arcsin(\rho)$ ([180]).

sufficient to obtain rank correlation estimators that are better than the empirical estimators, it does allow for more realistic joint distributions of returns compared to using a Gaussian copula, while employing a dynamic rather than a static approach.

6.4.3 Evaluation of strategy features replication

6.4.3.1 Strategies considered for evaluating the generator

One of the major interest of a market generator is its utility as a tool for backtesting strategies ([139]). Therefore, it makes sense to evaluate the RPDM generator based on its ability to reproduce the various time-series features characterizing different strategies.

To this end, we consider the following 5 strategies:

1. **EW**: The equally weighted portfolio, one of the most well-known strategies, that consists of assigning equal weights to each asset in the portfolio such as:

$$\mathbf{x} = n^{-1} \cdot \mathbf{1}_n.$$

2. **MV**: A constrained minimum variance portfolio, whose composition is the solution of the following optimization program ([179]):

$$\mathbf{x} = \arg \min_{\mathbf{x}} \mathbf{x}^\top \hat{\Sigma} \mathbf{x} \quad \text{s.t.} \quad \begin{cases} \mathbf{1}_n^\top \mathbf{x} = 1 \\ \mathbf{x} > -n^{-1} \cdot \mathbf{1}_n, \end{cases}$$

where $\hat{\Sigma}$ corresponds to the estimator of the asset returns covariance matrix. In practice, the estimator $\hat{\Sigma}$ used will be the daily returns covariance matrix over the last 4 years. Regarding the constraint of individual short selling on assets, this serves a dual purpose. First, it helps to stay within relatively realistic portfolios, as the sum of the absolute values of the weights is bounded by 3. Furthermore, it helps to restrict the range of the obtained metrics, as in the absence of constraints on the weights other than $\mathbf{1}_n^\top \mathbf{x}$, the behavior of the strategies can fluctuate extremely from one market scenario to another.

3. **TP**: A constrained tangency portfolio in the zero risk free-rate hypothesis, defined as the solution to ([179]):

$$\mathbf{x} = \arg \min_{\mathbf{x}} \frac{\mathbf{x}^\top \hat{\boldsymbol{\mu}}}{\sqrt{\mathbf{x}^\top \hat{\Sigma} \mathbf{x}}} \quad \text{s.t.} \quad \begin{cases} \mathbf{1}_n^\top \mathbf{x} = 1 \\ \mathbf{x} > -n^{-1} \cdot \mathbf{1}_n, \end{cases}$$

where $\hat{\boldsymbol{\mu}}$ is the estimated expected asset returns vector. Here, we use the vector of empirical means of the daily returns over the last 4 years as the estimator. Also, the constraint imposed on the weights serves the same purpose as that imposed on the minimum variance portfolio.

4. **TF**: A trend following strategy of cross-section momentum type ([129]) whose weight vector is defined by:

$$\mathbf{x} = \frac{\hat{\boldsymbol{\mu}}_t - Q_1(\hat{\boldsymbol{\mu}}_t)}{\mathbf{1}_n^\top (\hat{\boldsymbol{\mu}}_t - Q_1(\hat{\boldsymbol{\mu}}_t))} \quad (6.21)$$

where $Q_1(\hat{\boldsymbol{\mu}}_t)$ corresponds to the first quartile of $\hat{\boldsymbol{\mu}}_t$ itself defined by:

$$\hat{\boldsymbol{\mu}}_t = \mathbf{P}_t \circledast \mathbf{P}_{t-l} - \mathbf{1}_n.$$

Here, the window length l of the moving average estimator will be fixed to 1 year.

5. **RV**: A reversal strategy of cross-section momentum type, whose weights are defined in the same manner as for the TF strategy via equation 6.21, but using:

$$\hat{\boldsymbol{\mu}}_t = -\mathbf{P}_t \circledast \mathbf{P}_{t-l} - \mathbf{1}_n,$$

where l is also set to 1 year.

Each of these strategies will be examined in three distinct modes: buy-and-hold, fixed-weight, and dynamic.

In the buy-and-hold approach, the portfolio is constructed at the close of the last trading day of April 2018, and no rebalancing is performed thereafter. In this framework, only the price evolution affects the asset weights in the composition of the various portfolios. However, this mechanism alone is enough to significantly impact the behavior of the portfolio value considered as a random variable. For instance, suppose the market factor drops and then does not experience a significant rebound. In this scenario, all else being equal, the weights of securities with a very high beta tend to decrease relative to other securities, and this event tends to reduce the portfolio's exposure to the market factor. The interest of backtesting buy-and-hold strategies lies in evaluating the model's ability to accurately reproduce this type of path dependence.

The fixed-weight approach, for its part, involves daily rebalancing of the strategies to maintain the same composition as initially set. In this way, the effect of portfolio weight changes induced by the price dynamics to which buy-and-hold strategies are subject is neutralized. It is therefore the model's ability to replicate various constant linear combinations of the random variables constituting the S&P500 prices that is evaluated through this approach.

Finally, the dynamic approach involves rebalancing the portfolios monthly by updating the various parameters on which the composition of the portfolios associated with each strategy depends. Consequently, depending on the different market scenarios, the composition of the portfolios can vary significantly between real and synthetic data. Therefore, it is the model's ability to reproduce the market, conceptualized as an ecosystem, that is evaluated through these strategies. In other

words, the capacity to model the market not merely as a collection of assets with a simple correlation structure, but as a complex system whose mechanisms give rise to non-trivial statistical regularities.

Remark 2 *The parameters and estimators of the strategies considered are not necessarily optimal. Typically, to construct minimum variance and tangency portfolios, we use the sample covariance matrix, which constitutes a poor estimator of the covariance matrix from a portfolio optimization perspective. Furthermore, the "trend-following" and "reversal" portfolios with the buy-and-hold and fixed-weight approaches are not strictly speaking trend-following and reversal strategies since their composition is solely based on data as of April 30, 2018. However, these elements are not really problematic in this context since the goal is not to conduct a comparative study of the considered strategies, but rather to see if the RPDM generator is capable of reproducing the time-series features characterizing various types of investment portfolios.*

6.4.3.2 Results of numerical experiments

We start by individually considering the 5 types of strategies across the 3 rebalancing modes presented in the previous section. Tables 6.4, 6.5 and 6.6 compare a set of financial metrics associated with these strategies between real data and simulated data: their volatility and Sharpe ratio calculated from daily returns and then annualized, their maximum drawdown, and their value-at-risk and expected shortfall at the 95% and 99% thresholds, empirically calculated on a daily basis. Figure 6.9, 6.10, and 6.11 focus on the moments of the returns distributions of these strategies at different frequencies (from daily returns to 6-month returns).

Overall, the obtained results tend to prove that the model reproduces very well the different characteristics of the various strategies considered. Thus, for almost all distributions of the different strategies considered, the first four empirical moments of the real data fall within the 90% interval of the model calculated from the simulated data. Moreover, in a significant number of cases, these empirical moments are close to their median levels calculated from the simulations. Additionally, the non-monotonic relationship of the skewness of the returns of different strategies as a function of the considered time horizon is noteworthy. Thus, for both market data and simulated data, the distribution of returns initially becomes increasingly negatively skewed as the time horizon lengthens. However, past an inflection point located around 10 trading days, the skewness of the return distribution decreases and tends towards zero in the long term. Similarly, the negatively convex relationship between the kurtosis value and the time horizon produced by the model for the different strategies is also consistent with market data. In the same way as the moments of the distributions, the set of empirical financial metrics - reported in tables 6.4, 6.5, and 6.6 - for almost each strategy falls within the model's 90% confidence intervals and, in most cases, is fairly close to the median level obtained through simulations.

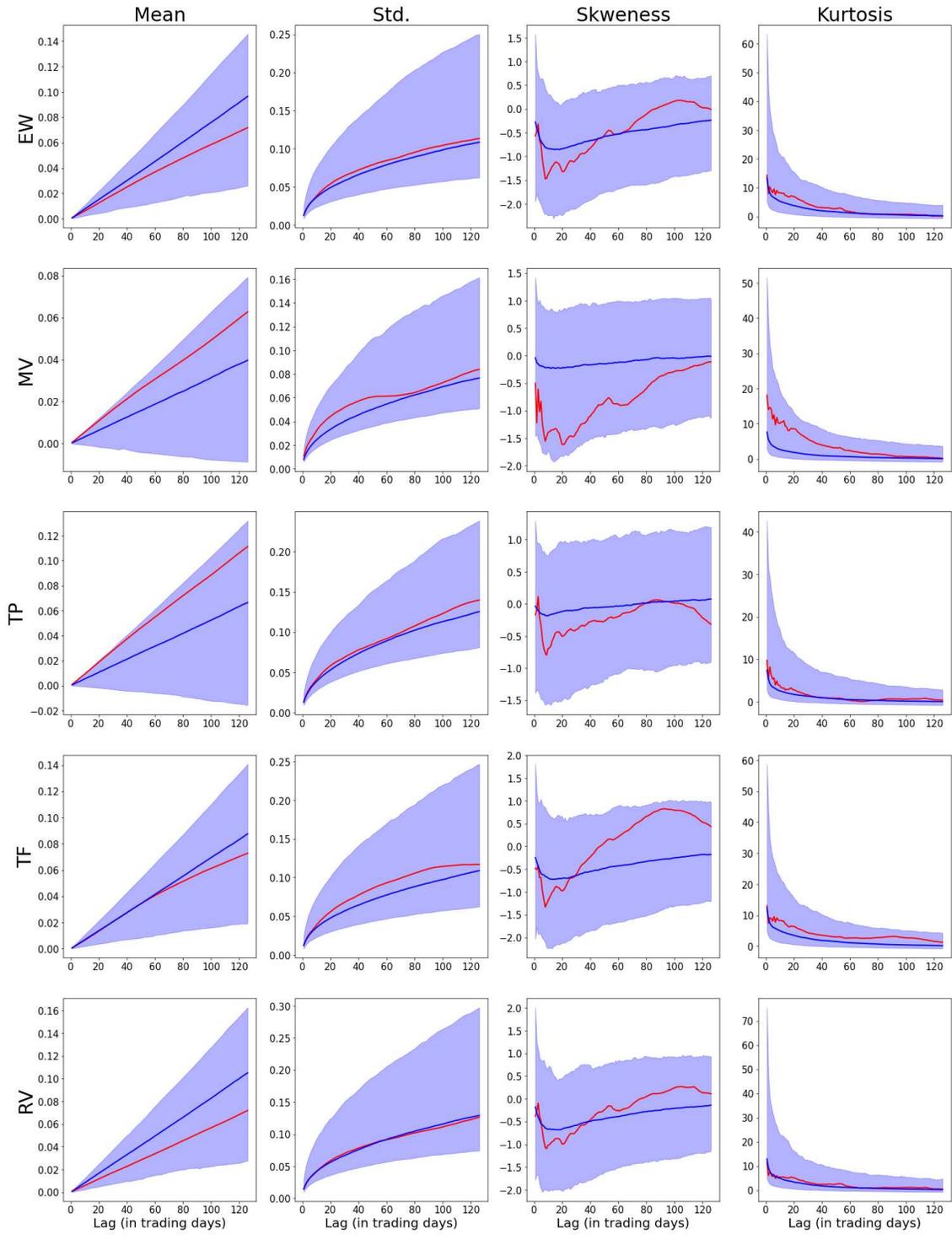


Figure 6.9: Evolution of the four first moments of the returns distributions for the different considered buy-and-hold strategies as a function of the time-horizon. The red curve corresponds to the real data, the blue curve to the median of the simulated data, and the blue area represents the 90% interval of the simulations.

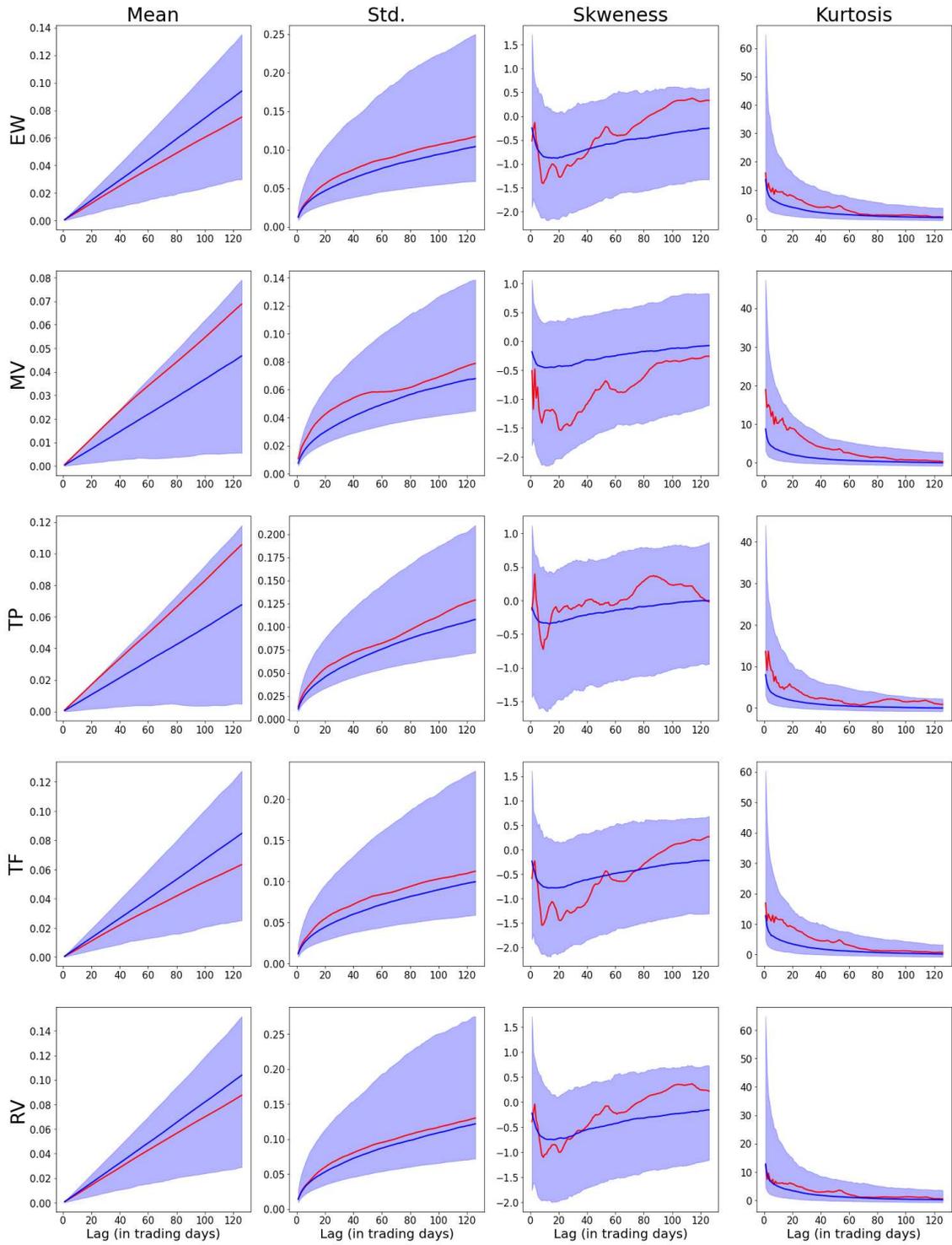


Figure 6.10: Evolution of the four first moments of the returns distributions for the different considered constant-weighted strategies as a function of the time-horizon. The red curve corresponds to the real data, the blue curve to the median of the simulated data, and the blue area represents the 90% interval of the simulations.

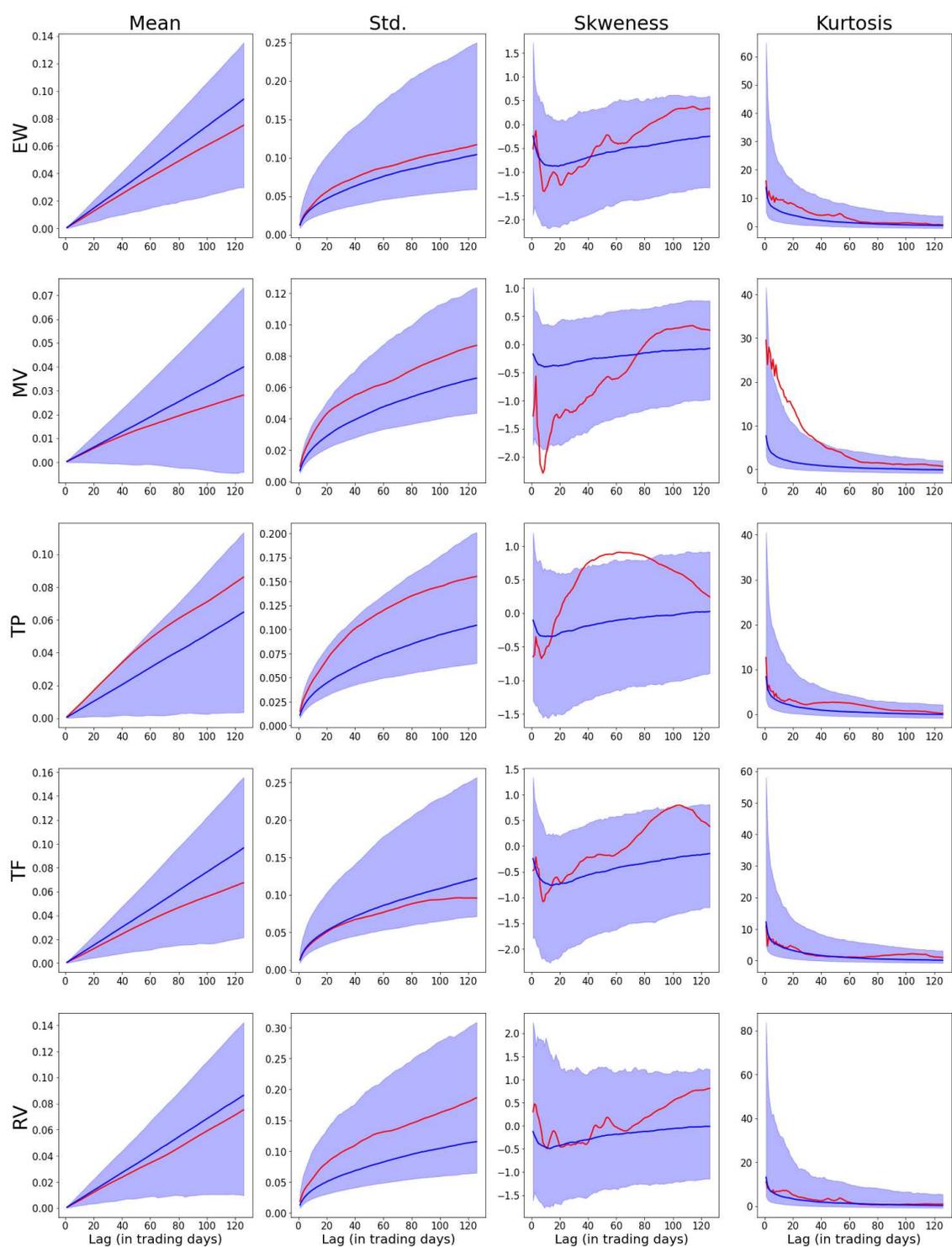


Figure 6.11: Evolution of the four first moments of the returns distributions for the different considered dynamic strategies as a function of the time-horizon. The red curve corresponds to the real data, the blue curve to the median of the simulated data, and the blue area represents the 90% interval of the simulations.

		Vol.	SR	MD	VaR _{95%}	VaR _{99%}	ES _{95%}	ES _{99%}
EW	Real data	20.45%	76.58%	36.38%	1.85%	3.22%	3.06%	5.51%
	Sim. data med.	20.26%	95.57%	30.05%	1.71%	3.63%	3.05%	5.60%
	Sim. data D1	14.63%	36.60%	16.06%	1.27%	2.51%	2.13%	3.60%
	Sim. data D9	34.40%	152.58%	62.29%	2.65%	6.37%	5.23%	10.43%
MV	Real data	17.40%	79.11%	32.88%	1.48%	2.89%	2.57%	5.12%
	Sim. data med.	12.87%	60.38%	21.92%	1.16%	2.19%	1.86%	3.16%
	Sim. data D1	9.94%	6.51%	12.23%	0.92%	1.65%	1.40%	2.18%
	Sim. data D9	22.14%	113.97%	47.75%	1.74%	3.97%	3.29%	6.50%
TP	Real data	22.25%	106.88%	36.72%	1.96%	3.64%	3.11%	5.28%
	Sim. data med.	20.34%	65.51%	31.57%	1.83%	3.49%	2.94%	4.93%
	Sim. data D1	15.57%	9.94%	18.59%	1.44%	2.54%	2.17%	3.43%
	Sim. data D9	32.02%	116.96%	60.27%	2.64%	5.85%	4.90%	9.35%
TF	Real data	20.45%	81.59%	36.15%	1.83%	3.36%	3.01%	5.73%
	Sim. data med.	19.45%	89.81%	29.88%	1.66%	3.52%	2.93%	5.32%
	Sim. data D1	14.19%	32.26%	15.22%	1.24%	2.43%	2.05%	3.42%
	Sim. data D9	33.24%	143.55%	61.58%	2.60%	6.20%	5.11%	10.22%
RV	Real data	22.86%	65.53%	30.80%	2.10%	3.76%	3.40%	5.77%
	Sim. data med.	22.82%	90.26%	33.79%	1.96%	4.12%	3.40%	6.24%
	Sim. data D1	16.64%	33.87%	18.36%	1.45%	2.84%	2.41%	4.05%
	Sim. data D9	40.69%	145.51%	69.33%	3.08%	7.34%	6.08%	12.35%

Table 6.4: Financial metrics associated with different buy-and-hold strategies: real vs simulated data.

		Vol.	SR	MD	VaR _{95%}	VaR _{99%}	ES _{95%}	ES _{99%}
EW	Real data	21.01%	76.12%	37.92%	1.83%	3.32%	3.13%	5.71%
	Sim. data med.	19.53%	77.26%	29.28%	1.52%	3.38%	2.88%	5.77%
	Sim. data D1	16.21%	3.82%	15.16%	1.29%	2.74%	2.36%	4.21%
	Sim. data D9	32.57%	151.14%	61.73%	2.61%	5.79%	4.82%	9.55%
MV	Real data	16.97%	87.00%	31.80%	1.41%	2.73%	2.51%	4.94%
	Sim. data med.	11.39%	64.09%	18.43%	0.95%	1.93%	1.65%	3.14%
	Sim. data D1	9.78%	-2.97%	10.44%	0.83%	1.56%	1.39%	2.32%
	Sim. data D9	19.22%	134.70%	39.57%	1.59%	3.18%	2.81%	5.25%
TP	Real data	21.35%	101.64%	27.63%	1.82%	3.71%	3.03%	5.31%
	Sim. data med.	17.08%	62.58%	27.52%	1.44%	2.86%	2.46%	4.61%
	Sim. data D1	14.73%	-5.01%	16.28%	1.26%	2.37%	2.07%	3.43%
	Sim. data D9	28.68%	134.11%	54.66%	2.42%	4.75%	4.12%	7.66%
TF	Real data	20.26%	68.83%	38.80%	1.79%	3.19%	2.98%	5.70%
	Sim. data med.	18.13%	76.19%	27.71%	1.43%	3.14%	2.66%	5.31%
	Sim. data D1	15.12%	5.63%	14.26%	1.22%	2.52%	2.19%	3.87%
	Sim. data D9	30.41%	148.46%	58.75%	2.42%	5.24%	4.48%	8.67%
RV	Real data	23.03%	79.16%	37.11%	2.08%	3.84%	3.42%	5.93%
	Sim. data med.	21.89%	76.56%	32.19%	1.72%	3.79%	3.22%	6.40%
	Sim. data D1	18.21%	4.34%	17.22%	1.46%	3.04%	2.64%	4.69%
	Sim. data D9	36.40%	150.06%	65.93%	2.92%	6.39%	5.35%	10.53%

Table 6.5: Financial metrics associated with different fixed-weight strategies: real vs simulated data.

		Vol.	SR	MD	VaR _{95%}	VaR _{99%}	ES _{95%}	ES _{99%}
EW	Real data	21.01%	76.12%	37.92%	1.83%	3.32%	3.13%	5.71%
	Sim. data med.	19.5%	96.04%	29.28%	1.65%	3.53%	2.93%	5.39%
	Sim. data D1	13.95%	37.85%	15.16%	1.2%	2.42%	2.03%	3.46%
	Sim. data D9	34.73%	152.87%	61.73%	2.59%	6.39%	5.23%	10.45%
MV	Real data	15.6%	44.99%	33.55%	1.19%	2.54%	2.29%	4.82%
	Sim. data med.	11.29%	70.18%	17.91%	1.01%	1.94%	1.64%	2.79%
	Sim. data D1	9.05%	12.03%	10.25%	0.83%	1.49%	1.27%	2.02%
	Sim. data D9	16.58%	126.55%	36.8%	1.39%	2.9%	2.49%	4.78%
TP	Real data	23.75%	89.95%	30.05%	2.17%	3.73%	3.53%	6.41%
	Sim. data med.	17.38%	75.3%	26.49%	1.54%	3.01%	2.53%	4.32%
	Sim. data D1	13.45%	19.95%	15.19%	1.21%	2.2%	1.89%	3.02%
	Sim. data D9	26.56%	129.78%	52.43%	2.22%	4.7%	4.01%	7.64%
TF	Real data	21.23%	75.43%	31.23%	1.89%	3.63%	3.09%	5.42%
	Sim. data med.	21.36%	91.4%	31.77%	1.83%	3.83%	3.19%	5.79%
	Sim. data D1	15.74%	36.51%	17.23%	1.39%	2.74%	2.28%	3.78%
	Sim. data D9	34.45%	144.22%	63.48%	2.74%	6.42%	5.27%	10.61%
RV	Real data	29.64%	50.69%	46.69%	2.55%	4.82%	4.17%	7.44%
	Sim. data med.	20.41%	85.35%	31.17%	1.72%	3.64%	3.05%	5.54%
	Sim. data D1	14.19%	30.52%	16.02%	1.24%	2.4%	2.02%	3.37%
	Sim. data D9	41.3%	141.02%	68.33%	2.97%	7.49%	6.2%	12.7%

Table 6.6: Financial metrics associated with different dynamic strategies: real vs. simulated data.

Beyond this generally very positive big picture regarding the FPDM generator’s ability to generate realistic market scenarios, there is, however, a relative disparity in performance when it comes to reproducing certain time-series features depending on the strategies considered.

To begin, the data generated by the model seems, on average, to slightly underestimate the risk associated with the various MV portfolios in these different approaches (buy-and-hold, fixed-weighted, and dynamic) compared to real data. Thus, the different financial risk metrics of these strategies are significantly higher than their median model level. Similarly, the empirical distributions of returns of these MV strategies are more skewed than the median level of the return distribution of these strategies on simulated data. This potential underestimation of risk could be explained by the fact that this strategy is built from the empirical estimator of the covariance matrix. Indeed, in the case of buy-and-hold and fixed-weighted approaches, and even for the initialization of the dynamic approach, this matrix is calculated over the exact period on which the model is fitted. Consequently, this strategy tends to maximize the model risk. However, this potential underestimation of risk requires nuance. Firstly, the various empirical risk metrics associated with the MV strategies remain within the 1st and 9th deciles of the simulated data. As a result, the empirical features of these strategies are not outliers when viewed through the lens of the model. Furthermore, it is noteworthy that the proposed model seems to avoid, or at least partially mitigate, the phenomenon of overfitting. This shows up in the fact that, while the empirical estimator of the daily annualized volatility of the constant-weighted MV portfolio computed from the learning period is 7.89% (even

when using a shrinkage estimator of the covariance matrix, this estimated volatility remains below 8%), the first decile of this volatility calculated from the simulations exceeds 9%.

Additionally, the rebalancing approach plays a somewhat significant role in the proximity between metrics on real data and their median levels on simulated data. Thus, while with the buy-and-hold approach, the empirical risk metrics of the strategies TP and RV are close to their median levels on simulated data, they are more off-center towards higher deciles when considering dynamic rebalancing. However, once again, like for the MV strategy, the financial metrics associated with real data remain within the 80% confidence interval of the simulated data. Therefore, it is difficult to conclusively assert based on these results that the model underestimates the risk of these strategies.

Beyond effectively capturing the features of the individual strategies, the model also reproduces the correlation between these strategies quite well overall, as shown in figures 6.13, 6.15, and 6.17. On this point, however, the proximity between real and simulated data depends significantly on the rebalancing approach adopted. For the buy-and-hold and constant-weighted approaches, the correlations between the different strategies in the simulated data are very close to those in the real data (figures 6.13 and 6.15). This correlation structure is however less similar to the real data for the dynamic rebalancing approach (figure 6.17). Furthermore, within this approach, the disparity between real correlations and correlations in simulated data depends on the pairs of strategies considered. Thus, the correlation of the pair of strategies EW/MV is well reproduced by the model. Conversely, the empirical correlations between pairs involving the TP, TF, and RV strategies differ significantly from the correlations obtained from simulated data. However, these three strategies share a common characteristic: their composition depends on an estimator of the drift vector, specifically a moving average estimator (over the last 4 years for the TP strategy and over the last year for the TF and RV strategies). This observation thus suggests that the disparity between the correlations in real and simulated data could be attributable to an imperfect modeling of drift dynamics by the generator. Therefore, the assumption of the market factor as the unique determinant of asset drift is likely too simplistic and warrants amendment.

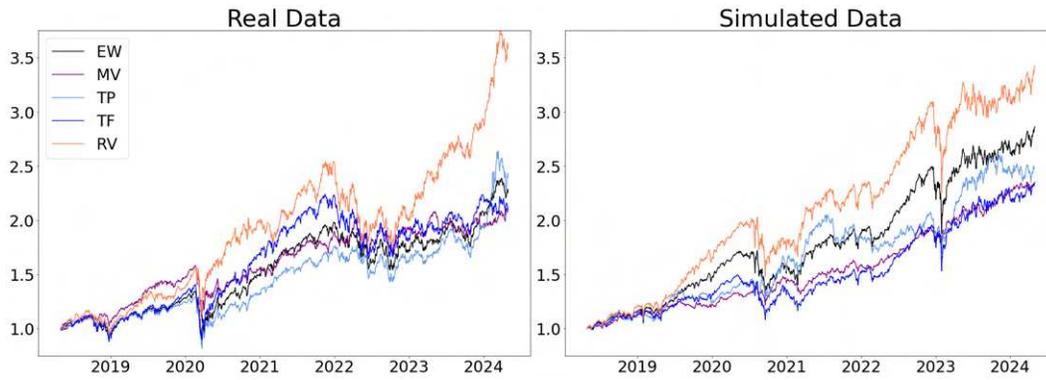


Figure 6.12: Cumulative returns of the considered buy-and-hold strategies: real vs simulated data.

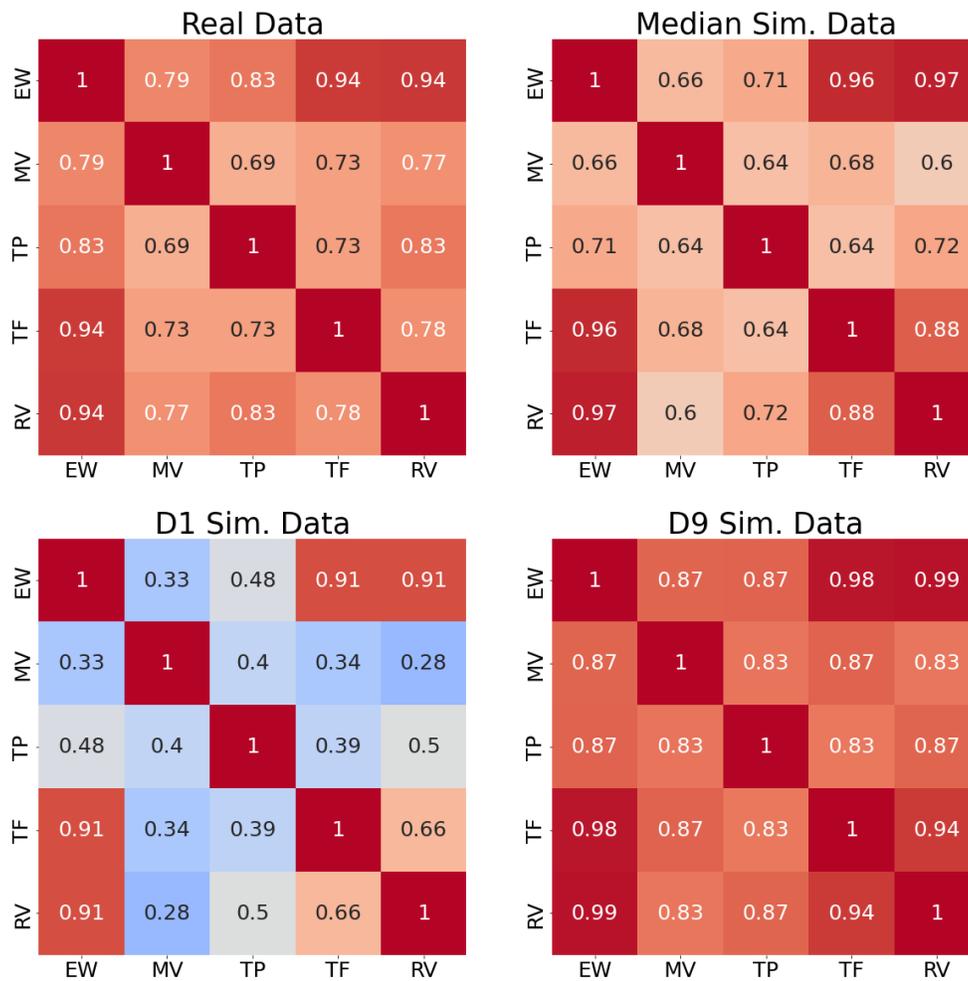


Figure 6.13: Correlation matrix of daily returns of the different buy-and-hold strategies: real vs. simulated data.

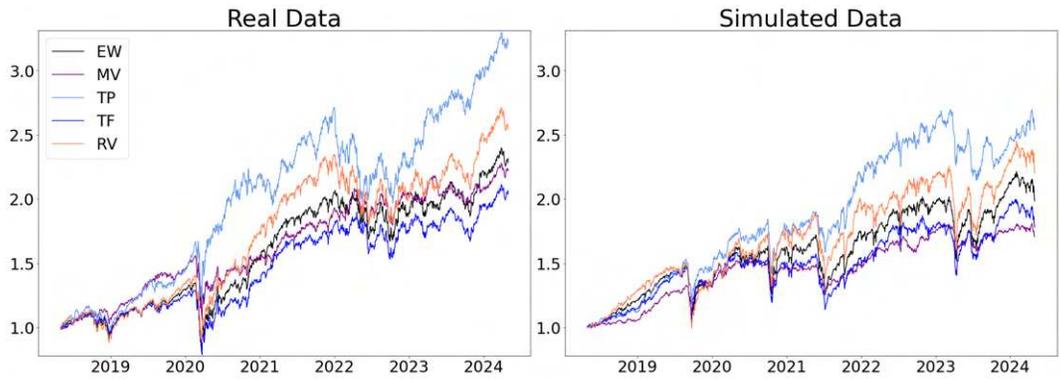


Figure 6.14: Cumulative returns of the considered fixed-weighted strategies: real vs simulated data.

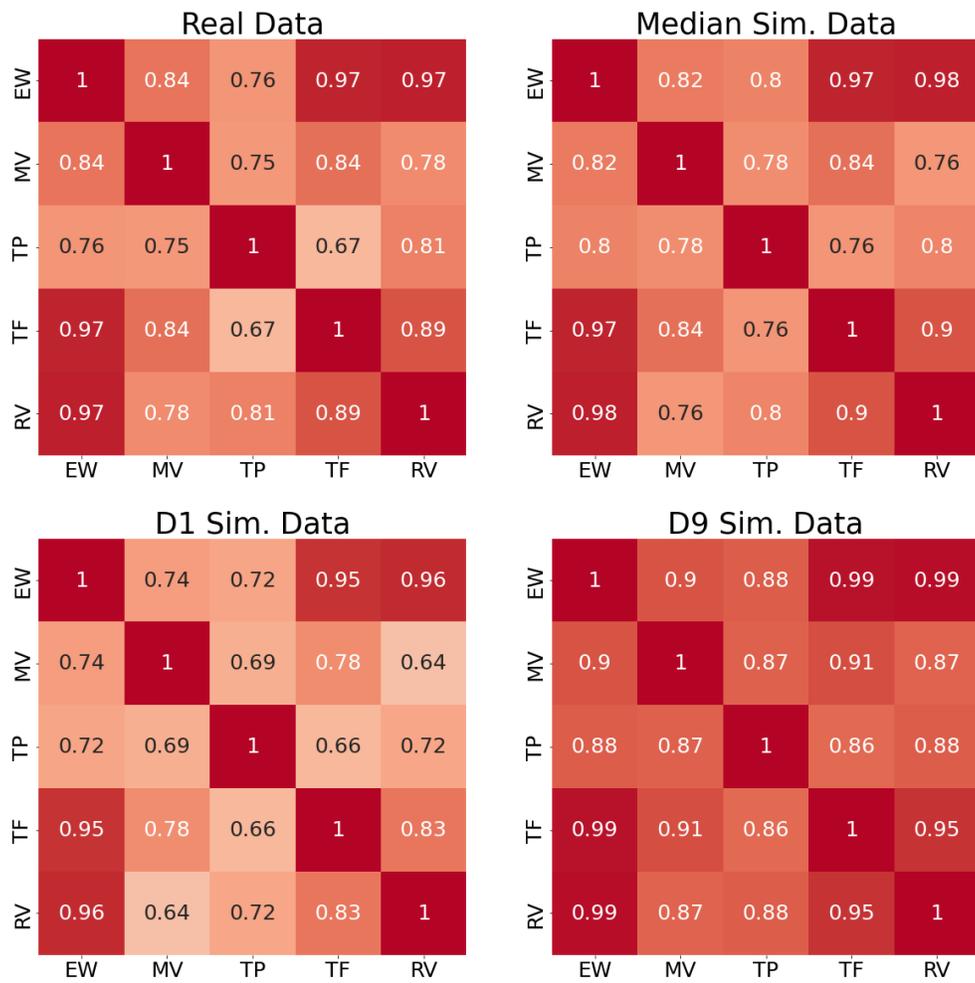


Figure 6.15: Correlation matrix of daily returns of the different fixed-weighted strategies: real vs simulated data.

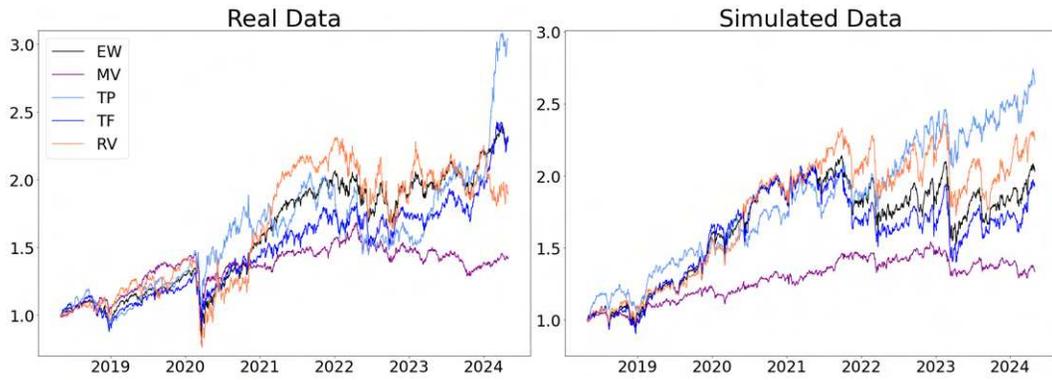


Figure 6.16: Cumulative returns of the considered dynamic strategies: real vs simulated data.

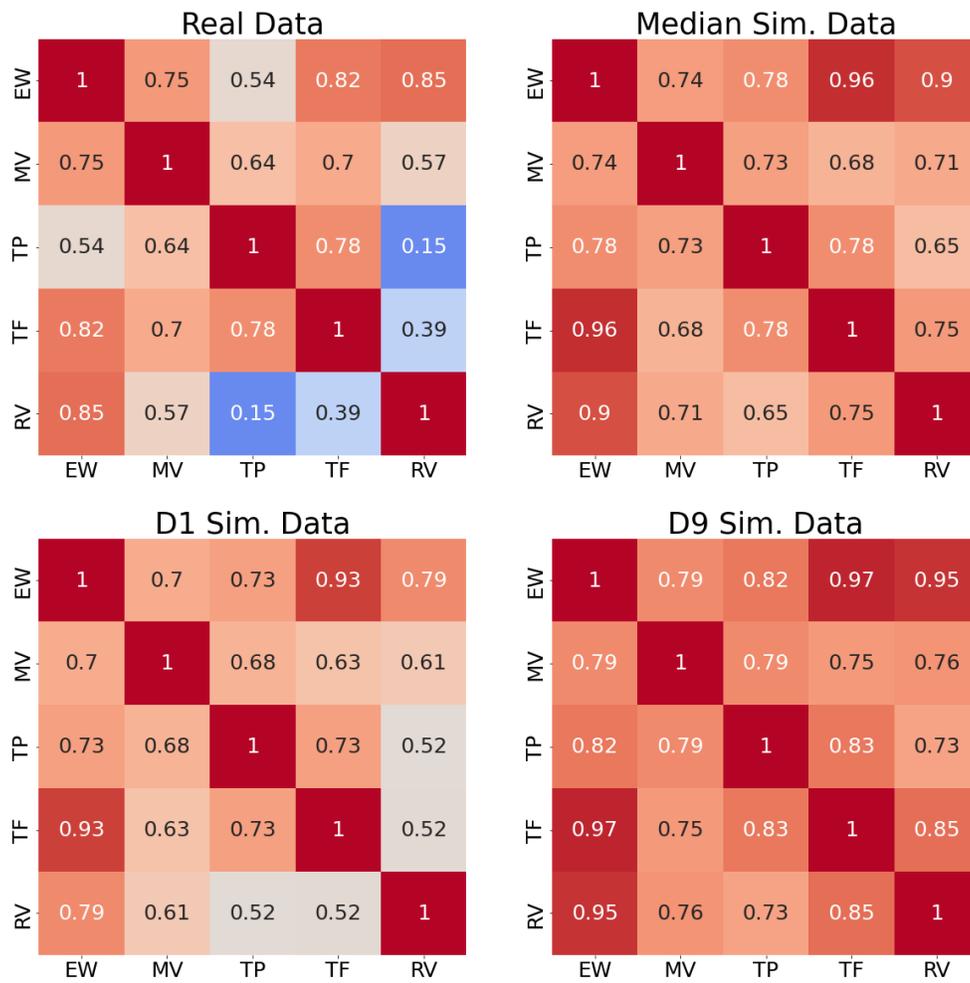


Figure 6.17: Correlation matrix of daily returns of the different dynamic strategies: real vs simulated data.

6.5 Conclusion

In this article, we have introduced the Factorial Path-Dependent Market (FPDM) model, a new theoretical framework for modeling multivariate asset price dynamics. In this framework, asset prices are driven by a set of elementary factors of two kinds: common elementary factors that link the dynamics of assets together and generate the correlation structure between individual price variations, and idiosyncratic elementary factors specific to each asset. The dynamics of the vector of these elementary factors are described by a multidimensional Itô SDE, whose drifts and volatility vectors depend either partially or completely on its own past trajectory. As a result, the proposed approach falls within the realm of path-dependent modelings. This choice aims to account for the fundamentally endogenous nature of financial markets, which manifests in various forms, ranging from the volatility formation process to the emergence of price trends caused by feedback effects from previous price dynamics. In practice, the path-dependence in the proposed model is implemented by modeling factor drifts and volatilities as functions of features encoding the past trajectory of factor portfolios.

After defining this theoretical framework in section 6.2, section 6.3 introduced a market generator derived from a specific version of the FPDM model. In this one, the path-dependent component of the volatility of each elementary factor depends on both the past trajectory of the elementary factor to which it is associated and the past trajectory of the elementary factor that constitutes the market factor. The market factor thus plays a singular role in this framework, firstly in the overall volatility of the investment universe under consideration, and secondly in shaping the level and dynamics of the correlation structure among the assets it comprises. Furthermore, in the considered specification, the market factor is the only elementary factor with a non-zero drift, thus producing a CAPM-like modeling.

Once this market generator has been defined, a calibration method for it has been proposed. This process consists of two main phases. The first involves factorizing a historical dataset of returns from the investment universe under consideration, such that each return is expressed as a linear combination of variations corresponding to what, in the model, are the elementary factors. The second phase takes as input the elements resulting from this factorization to estimate the various model parameters using an adapted form of maximum likelihood estimation.

Section 6.4 provided an extensive evaluation of this market generator using an investment universe of 436 assets from the S&P500. For this purpose, the data set was divided into two parts: the first, spanning from April 1, 2010, to April 30, 2018, was used to fit the model, while the second part, covering data from May 1, 2018, to April 30, 2024, was used to conduct the various tests. All evaluations of the market generator were thus conducted out-of-sample.

A first set of tests showed that the moments of the marginal distributions of the asset returns

generated by the model were highly consistent with their empirical counterparts calculated from market data. A second type of evaluation, based on the joint dynamics of the assets, was then conducted. A first key result from these tests is that using synthetic data generated by the FPDM generator enables us to obtain an estimator of the covariance matrix of returns that is significantly better than standard or shrinkage-based estimators derived from historical data. Supplementary results from these tests suggest that this improvement seems to stem less from the model's ability to better capture linear correlations between asset returns and more from its enhancement in estimating their individual standard deviations while maintaining a realistic correlation structure. This, however, does not imply that the model's ability to capture correlations is poor. For example, the estimated Kendall's tau and Spearman's rho calculated on simulated data exhibit performance on test data similar to the empirical estimators of these correlation measures calculated on historical data. Furthermore, these results must also be seen in the light of the fact that, on the one hand, the correlation structure in the model is dynamic, and on the other, the test horizon is relatively long (six years), which constitutes rather unfavorable evaluation conditions for the model.

The final set of tests aimed to evaluate the model's ability as a tool for backtesting investment strategies. For this purpose, five types of strategies were considered, each through three different approaches: buy-and-hold, constant-weight with daily rebalancing, and dynamic with monthly rebalancing. The results initially demonstrated that the various time series features characterizing the individual trajectories of these strategies were generally well replicated by the generator. This includes moments of return distributions calculated at different frequencies, as well as various risk metrics such as value-at-risk and expected shortfall at different confidence levels, along with maximum drawdown. In this respect, then, the FPDM generator seems to be a relevant and powerful tool for backtesting various strategies on synthetic data. The results regarding the model's capture of the correlation between these strategies are slightly more nuanced. Thus, while the correlations of returns among strategies under the buy-and-hold and constant-weighted modes are well reproduced by the model, the difference between the correlations among dynamic strategies from simulated data differs quite significantly from their counterparts obtained from empirical data. However, this latter observation only pertains to the strategies whose composition depends on an estimator of the drift vector, specifically a moving average estimator. This implies that the difference in correlations between real and simulated data may stem from an imperfect modeling of drift dynamics by the generator. Therefore, the assumption that the market factor is the sole determinant of asset drift is likely overly simplistic and may require adjustments, such as allowing other elementary factors to have a non-zero drift.

More generally, the market generator presented in this article could be improved in multiple ways. As an example, the model could incorporate "views" in the Black-Litterman sense ([37]) to integrate exogenous information into the model. Furthermore, the market generator proposed is just one possible (discrete) specification of the FPDM model introduced in section 6.2. Hence, for example, more sophisticated specifications that directly incorporate sectoral decomposition via elementary factor portfolios could further refine the proposed framework. Additionally, the model can be used

to generate conditional scenarios, such as given a certain trajectory of the market factor, which also deserves further investigation.

Regardless, the results obtained from the market generator proposed in this article already demonstrate that the FPDM model constitutes an extremely rich framework that can be used as a white box to generate highly realistic simulations of price trajectories in a high-dimensional investment universe.

Appendix 6.A List of notations

The following table presents the matrix operators and standard matrices/vectors used in this article.

Symbol	Description
$\mathbf{1}_n$	$n \times 1$ vector of ones
\mathbf{e}_i	$n \times 1$ vector whose value is 1 for coordinate i and 0 elsewhere
\mathbf{I}_n	Identity matrix of dimension $n \times n$
$(\mathbf{V})_i$	Element i of the vector \mathbf{V}
$(\mathbf{M})_{[i,j]}$	Entry at row i and column j of the matrix \mathbf{M}
$(\mathbf{M})_{[i,:]}$	Row i of the matrix \mathbf{M}
$(\mathbf{M})_{[:,j]}$	Column j of the matrix \mathbf{M}
\mathbf{M}^\top	Transpose of the matrix \mathbf{M}
\mathbf{M}^{-1}	Inverse of the square matrix \mathbf{M}
$\mathbf{M}^{\circ p}$	Element-wise application of the power p to each coordinate of \mathbf{M}
$f \circ (\mathbf{M})$	Element-wise application of the function f to each coordinate of \mathbf{M}
\odot	Hadamard product
\oslash	Hadamard division
\otimes	Kronecker product
\oplus	Kronecker sum
$\text{diag}(\mathbf{V})$	Converts the vector \mathbf{V} into a diagonal matrix
$\text{diag}_{M \rightarrow d}(\mathbf{M})$	Transformation of the square matrix \mathbf{M} into a vector from the diagonal of \mathbf{M}
$\text{diag}_{M \rightarrow D}(\mathbf{M})$	Transformation of the square matrix \mathbf{M} into a diagonal matrix by retaining only the diagonal of \mathbf{M}
$\text{Tr}(\mathbf{M})$	Trace of the square matrix \mathbf{M}
$\text{vec}(\mathbf{M})$	The vec operator that transforms a matrix into a column vector
$\text{vec}_{m \times m}^{-1}(\mathbf{V})$	The inverse vec operator that transforms the $m \times 1$ vector \mathbf{V} into a square matrix $m \times m$
$\ \mathbf{M}\ _F$	The Frobenius norm of the matrix \mathbf{M}

Appendix 6.B Stochastic differential equations involved in the FPDM model

6.B.1 Calculation of the asset price vector solution

The SDE describing the dynamics of the price vector \mathbf{P} is given by:

$$d\mathbf{P}_t = \mathbf{P}_t \odot (\mathbf{A}\boldsymbol{\mu}_t dt) + \mathbf{P}_t \odot (\mathbf{A}\sqrt{\boldsymbol{\Omega}_t} d\mathbf{W}_t).$$

However, if f is a C^2 function such that $\mathbb{R} \times \mathbb{R}^m \rightarrow \mathbb{R}^n$, using the general Itô formula ([166]), we have:

$$df_k(t, \mathbf{P}_t) = \frac{\partial f_k(t, \mathbf{P}_t)}{\partial t} dt + \sum_i \frac{\partial f_k(t, \mathbf{P}_t)}{\partial P_i} d(\mathbf{P}_t)_i + \frac{1}{2} \sum_{i,j} \frac{\partial^2 f_k(t, \mathbf{P}_t)}{\partial P_i \partial P_j} d(\mathbf{P}_t)_i d(\mathbf{P}_t)_j.$$

Considering $f(t, \mathbf{P}_t) = \ln \circ \mathbf{P}_t$, we have:

$$\begin{aligned} \frac{\partial f_k(t, \mathbf{P}_t)}{\partial t} dt &= 0, \\ \sum_i \frac{\partial f_k(t, \mathbf{P}_t)}{\partial P_i} d(\mathbf{P}_t)_i &= \frac{d(\mathbf{P}_t)_k}{(\mathbf{P}_t)_k} = (\mathbf{A}\boldsymbol{\mu}_t dt + \mathbf{A}\sqrt{\boldsymbol{\Omega}_t} d\mathbf{W}_t)_k, \\ \frac{1}{2} \sum_{i,j} \frac{\partial^2 f_k(t, \mathbf{P}_t)}{\partial P_i \partial P_j} d(\mathbf{P}_t)_i d(\mathbf{P}_t)_j &= -\frac{1}{2(\mathbf{P}_t)_k} d(\mathbf{P}_t)_k d(\mathbf{P}_t)_k = -\frac{1}{2} \cdot (\mathbf{A}\boldsymbol{\Omega}_t \mathbf{A}^\top)_{k,k} dt. \end{aligned}$$

It follows that:

$$d \ln \circ \mathbf{P}_t = \left(\mathbf{A}\boldsymbol{\mu}_t - \frac{1}{2} \cdot \text{diag} \left(\mathbf{A}\boldsymbol{\Omega}_t \mathbf{A}^\top \right) \right) dt + \mathbf{A}\sqrt{\boldsymbol{\Omega}_t} d\mathbf{W}_t$$

Consequently, the solution to the asset price vector is given by:

$$\mathbf{P}_t = \mathbf{P}_0 \odot \exp \circ \left(\int_0^t \mathbf{A}\boldsymbol{\mu}_u - \frac{1}{2} \cdot \text{diag} \left(\mathbf{A}\boldsymbol{\Omega}_u \mathbf{A}^\top \right) du + \int_0^t \mathbf{A}\sqrt{\boldsymbol{\Omega}_u} d\mathbf{W}_u \right).$$

6.B.2 EWMA estimators and their stochastic differential equations

Consider the following vector stochastic differential equation (SDE):

$$d\tilde{\boldsymbol{\mu}}_t^{(j)} = \frac{1}{\tau_j} \cdot \left(d\mathbf{F}_t - \tilde{\boldsymbol{\mu}}_t^{(j)} dt \right).$$

By setting $f\left(t, \tilde{\boldsymbol{\mu}}_t^{(j)}\right) = e^{\frac{t}{\tau_j}} \tilde{\boldsymbol{\mu}}_t^{(j)}$, and applying Ito's formula, we obtain:

$$df\left(t, \tilde{\boldsymbol{\mu}}_t^{(j)}\right) = \frac{e^{\frac{t}{\tau_j}} \tilde{\boldsymbol{\mu}}_t^{(j)}}{\tau_j} dt + e^{\frac{t}{\tau_j}} d\tilde{\boldsymbol{\mu}}_t^{(j)} = \frac{e^{\frac{t}{\tau_j}}}{\tau_j} \cdot d\mathbf{F}_t$$

It follows that the solution corresponds to the following EWMA:

$$\tilde{\boldsymbol{\mu}}_t^{(j)} = \tilde{\boldsymbol{\mu}}_0^{(j)} e^{\frac{-t}{\tau_j}} + \frac{1}{\tau_j} \int_0^t e^{\frac{u-t}{\tau_j}} \cdot d\mathbf{F}_u.$$

In the same manner, when considering

$$d\tilde{\boldsymbol{\Omega}}_t^{(j)} = \frac{1}{\tau_j} \cdot \left(d\mathbf{F}_t d\mathbf{F}_t^\top - \tilde{\boldsymbol{\Omega}}_t^{(j)} dt \right) = \frac{1}{\tau_j} \cdot \left(\tilde{\boldsymbol{\Omega}}_t - \tilde{\boldsymbol{\Omega}}_t^{(j)} \right) dt,$$

and setting $f\left(t, \tilde{\boldsymbol{\Omega}}_t^{(j)}\right) = e^{\frac{t}{\tau_j}} \tilde{\boldsymbol{\Omega}}_t^{(j)}$, we obtain:

$$\tilde{\boldsymbol{\Omega}}_t^{(j)} = \tilde{\boldsymbol{\Omega}}_0^{(j)} e^{\frac{-t}{\tau_j}} + \frac{1}{\tau_j} \int_0^t e^{\frac{u-t}{\tau_j}} \cdot \tilde{\boldsymbol{\Omega}}_u^{(j)} du.$$

Appendix 6.C Important specifications of the FPDM model

6.C.1 The 4-factor PDV as a specific case of the FPDM model

Suppose that $m = n = 1$, $\mathbf{A} = 1$, and $n_\tau = 2$. In this univariate single-factor framework, there exists a unique factor portfolio, which is $y = 1$. Therefore $\mathcal{Y}_t = 1 \forall t$. If we additionally assume that the drift is zero, we then have the following FPDM model, which corresponds to the 4-factor PDV model by Guyon and Lekeufack:

$$\left\{ \begin{array}{l} \frac{dP_t}{P_t} = dF_t = \sigma_t dW_t, \\ \sigma_t = \bar{\sigma} + \mathbf{b}_1 \cdot \hat{\boldsymbol{\mu}}_t(1, \boldsymbol{\delta}) + \mathbf{b}_2 \cdot \hat{\sigma}_t(1, \mathbf{w}), \\ \hat{\boldsymbol{\mu}}_t(1, \boldsymbol{\delta}) = (\boldsymbol{\delta})_1 \cdot \tilde{\boldsymbol{\mu}}_t^{(1)} + (\boldsymbol{\delta})_2 \cdot \tilde{\boldsymbol{\mu}}_t^{(2)}, \\ \hat{\sigma}_t(1, \mathbf{w}) = \sqrt{(\mathbf{w})_1 \cdot \tilde{V}_t^{(1)} + (\mathbf{w})_2 \cdot \tilde{V}_t^{(2)}}, \\ d\tilde{\boldsymbol{\mu}}_t^{(j)} = \frac{1}{\tau_j} \cdot \left(\frac{dP_t}{P_t} - \tilde{\boldsymbol{\mu}}_t^{(j)} dt \right), \\ d\tilde{V}_t^{(j)} = \frac{1}{\tau_j} \cdot \left(\left(\frac{dP_t}{P_t} \right)^2 - \tilde{V}_t^{(j)} dt \right). \end{array} \right.$$

6.C.2 Definition of the factorial drift vector in the context of the CAPM

Let π represent the value of the market portfolio, and assume that at any time, there exists a vector \mathbf{y}_t^* such that, $\forall t$:

$$\frac{d\pi_t}{\pi_t} = (\mathbf{y}_t^*)^\top d\mathbf{F}_t.$$

Let us now suppose that the vector of drifts for elementary factors depends on the $n + 1$ factor portfolios $y_t^*, \mathbf{e}_1, \dots, \mathbf{e}_n$ (where \mathbf{e}_j represents the factor portfolio composed entirely of factor j), such that:

$$\boldsymbol{\mu}_t = \boldsymbol{\beta}_t(\mathbf{y}_t^*) \cdot \Gamma_t^* + \sum_{p=1}^n \boldsymbol{\beta}_t(\mathbf{e}_1) \cdot \Gamma_{p,t}^{(I)} = \boldsymbol{\beta}_t(\mathbf{y}_t^*) \cdot \Gamma_t^* + \sum_{p=1}^n \Gamma_{p,t}^{(I)},$$

where

$$\Gamma_{p,t}^{(I)} = \mathcal{E}_{p,t}^{(I)} = (1 - (\mathbf{A}\boldsymbol{\beta}_t(\mathbf{y}_t^*))_i) \cdot r_t + (\boldsymbol{\beta}_t(\mathbf{y}_t^*))_i \cdot r_t.$$

From the expression of $\boldsymbol{\mu}$ and since for $i > m$, $(\mathbf{y}_t^*)_i = 0$, the drift of \mathbf{y}^* is given by:

$$\begin{aligned} \mu(\mathbf{y}_t^*) &= (\mathbf{y}_t^*)^\top \boldsymbol{\mu}_t \\ &= (\mathbf{y}_t^*)^\top \Gamma_t^{(I)} + (\mathbf{y}_t^*)^\top \boldsymbol{\beta}_t(\mathbf{y}_t^*) \cdot \Gamma_t^* \\ &= (\mathbf{y}_t^*)^\top \frac{\boldsymbol{\Omega}_t \mathbf{y}_t^*}{(\mathbf{y}_t^*)^\top \boldsymbol{\Omega}_t \mathbf{y}_t^*} \cdot r_t + (\mathbf{y}_t^*)^\top \frac{\boldsymbol{\Omega}_t \mathbf{y}_p^*}{(\mathbf{y}_p^*)^\top \boldsymbol{\Omega}_t \mathbf{y}_p^*} \cdot \Gamma_t^* \\ &= r_t + \Gamma_t^* \end{aligned}$$

It follows that the vector of the asset drift is defined by:

$$\begin{aligned} \boldsymbol{\mu}_t &= \mathbf{A}\boldsymbol{\mu}_t \\ &= \mathbf{A}\mathcal{E}_t + \mathbf{A}\boldsymbol{\beta}_t(\mathbf{y}_t^*) \cdot \Gamma_t^* \\ &= (\mathbf{1}_n - \mathbf{A}\boldsymbol{\beta}_t(\mathbf{y}_t^*)) \cdot r_t + \mathbf{A}\boldsymbol{\beta}_t(\mathbf{y}_t^*) \cdot r_t + \mathbf{A}\boldsymbol{\beta}_t(\mathbf{y}_t^*) \cdot \Gamma_t^* \\ &= \mathbf{1}_n \cdot r_t + \frac{\text{Cov}(d\mathbf{P}_t \otimes \mathbf{P}_t, d\pi_t/\pi_t)}{\text{Var}(d\pi_t/\pi_t)} \cdot \Gamma_t^* \\ &= \mathbf{1}_n \cdot r_t + \boldsymbol{\beta}_t^{(A)} \cdot (\mu(\mathbf{y}_t^*) - r_t). \end{aligned}$$

This then results in a relationship analogous to that of the CAPM.

Appendix 6.D Approximation of the Wasserstein distance for eigenvalues

If the support of the distributions f_{MP} and f_{KDE} is included in $[\phi_{\min} : \phi_{\max}]$, then

$$\int_{\mathbb{R}} \left(\sqrt{f_{\text{KDE}}(\phi)} - \sqrt{f_{\text{MP}}(\phi|v, q)} \right)^2 d\phi = \int_{\phi_{\min}}^{\phi_{\max}} \left(\sqrt{f_{\text{KDE}}(\phi)} - \sqrt{f_{\text{MP}}(\phi|v, q)} \right)^2 d\phi.$$

It follows that

$$\int_{\mathbb{R}} \left(\sqrt{f_{\text{KDE}}(\phi)} - \sqrt{f_{\text{MP}}(\phi|v, q)} \right)^2 d\phi \approx \frac{1}{n_x} \sum_{i=1}^{n_x} \left(\sqrt{f_{\text{KDE}} \left(\phi_{\min} + \frac{\phi_{\max} - \phi_{\min}}{n_\delta} i \right)} - \sqrt{f_{\text{MP}}(\delta_x i|v, q)} \right)^2.$$

Appendix 6.E Normal log-normal mixture

6.E.1 First four moments of a normal log-normal mixture of the form

$$W e^{s(B-s)}$$

Using the results of [199], if $X = W e^{sB}$ with $(W, B)^\top \sim \mathcal{N}(\mathbf{0}_2, \mathbf{I}_2)$: $\mathbb{E}[Z] = 0$, $\mathbb{E}[Z^2] = e^{2s^2}$, $\mathbb{E}[Z^3] = 0$, and $\mathbb{E}[Z^4] = 3e^{8s^2}$. Therefore, setting $X = W e^{s(B-s)}$, we have:

$$\begin{aligned} \mathbb{E}[Z] &= e^{-s^2} \mathbb{E}[X] = 0, \\ \mathbb{E}[Z^2] &= e^{-2s^2} \mathbb{E}[X^2] = 1, \\ \mathbb{E}[Z^3] &= e^{-3s^2} \mathbb{E}[X^3] = 0, \\ \mathbb{E}[Z^4] &= e^{-4s^2} \mathbb{E}[X^4] = 3e^{4s^2}. \end{aligned}$$

6.E.2 First four moments of an i.i.d. sample from a normal log-normal mixture

We consider the random variable Z defined by:

$$Z = \frac{1}{\sqrt{n}} \sum_{i=1}^n Y_i,$$

where Y_1, \dots, Y_n is an i.i.d sample such that $\mathbb{E}[Y] = \mathbb{E}[Y^3] = 0$, $\mathbb{E}[Y^2] = 1$, and $\mathbb{E}[Y^4] = c_4$. Using the multinomial theorem, we can express the moments of Z of order q as follows:

$$\mathbb{E}[Z^q] = n^{-\frac{q}{2}} \cdot \mathbb{E} \left[\left(\sum_{i=1}^n Y_i \right)^q \right] = n^{-\frac{q}{2}} \cdot \sum_{k_1 + \dots + k_n = q} \mathbb{E} \left[\binom{q}{k_1, \dots, k_n} Y_1^{k_1} Y_2^{k_2} \dots Y_n^{k_n} \right]$$

From this expression, it is easy to verify that $\mathbb{E}[Z] = \mathbb{E}[Z^3] = 0$ and $\mathbb{E}[Z^2] = 1$. To determine the 4th moment of Z , we can start by noting that only two types of combinations, such that $k_1 + \dots + k_n = 4$, are associated with a non-zero expectation due to $\mathbb{E}[Y] = 0$ and $\mathbb{E}[Y^3] = 0$. The first combination implies that one of the terms k_j is equal to 4, while the remaining $n - 1$ terms are equal to 0. The second combination involves two of the terms k_j being equal to 2, with the rest being equal to 0. Therefore:

$$\begin{aligned} \mathbb{E} \left[\left(\frac{1}{\sqrt{n}} \sum_{i=1}^n Y_i \right)^4 \right] &= \frac{1}{n^2} \left(\frac{4!}{4!(0!)^{n-1}} \cdot \binom{n}{1} \cdot \mathbb{E}[Y^4] + \frac{4!}{(2!)^2(0!)^{n-2}} \cdot \binom{n}{2} \cdot \mathbb{E}[Y^2] \right) \\ &= \frac{1}{n^2} \left(n \cdot c_4 + 6 \cdot \frac{n(n-1)}{2} \right) \\ &= \frac{c_4 + 3(n-1)}{n} \end{aligned} \tag{6.22}$$

Consequently, if $Y_i = W_i e^{s(B_i - s)}$ where W_i and B_i are i.i.d. standard normal realizations, using the expression for the third moment of Y_i calculated in section 6.E.1, we have:

$$\mathbb{E} \left[\left(\frac{1}{\sqrt{n}} \sum_{i=1}^n Y_i \right)^4 \right] = \frac{3e^{4s^2} + 3(n-1)}{n}.$$

6.E.3 Estimation of the vector \mathbf{S} using the method of moments

We aim to estimate the vector \mathbf{S} on which \mathbf{X} depends. To begin with, we assume that the time step between observations is constant, and that the ratio of the sample time step to the simulation time step is equal to n , a strictly positive natural number:

$$\frac{1}{N-rw-1} \sum_{u=rw+1}^N (t_u - t_{u-1}) = \frac{\bar{\Delta}_u}{\Delta t} = n, \quad \text{with } n \in \mathbb{N}^*.$$

In this context, and assuming $\hat{\boldsymbol{\mu}}_u, \hat{\mathbf{S}}_u, \hat{\boldsymbol{\nu}}_u$ are constant between t_{u-1} and t_u , the equation for the variation of elementary factors over this period is given by:

$$\Delta \hat{\mathbf{F}}_u = \sum_{i=1}^n \hat{\boldsymbol{\mu}}_u \cdot \Delta t + \hat{\mathbf{S}}_u \cdot \hat{\boldsymbol{\nu}}_u \cdot \mathbf{X}_i \cdot \mathbf{W}_i \cdot \sqrt{\Delta t}$$

We then introduce the process \hat{Z}_j , such that:

$$\hat{Z}_{j,u} = \frac{(\Delta \mathbf{F}_u - \hat{\boldsymbol{\mu}}_u \cdot \Delta_u)_j}{(\mathcal{S}_u \cdot \boldsymbol{\nu}_u \cdot \sqrt{\Delta_u})_j} = \frac{\sum_{i=1}^n \mathbf{X}_i \cdot \mathbf{W}_i \cdot \sqrt{\Delta t}}{\sqrt{\Delta_u}} = \frac{1}{\sqrt{n}} \sum_{i=1}^n \left(\exp \circ (\mathbf{S} \odot (\mathbf{B}_i - \mathbf{S})) \odot \mathbf{W}_i \right)_j.$$

Therefore $\hat{Z}_{j,rw+1}, \dots, \hat{Z}_{j,N}$ forms an i.i.d. sample drawn from an NLN mixture distribution of the form considered in section 6.E.1. To estimate $(\mathbf{S})_j$, we use a method-of-moments approach by equating \hat{c}_4 as the empirical estimator of $\mathbb{E}[\hat{Z}_j^4]$ associated with the sample $\{\hat{Z}_{j,u}\}_{rw+1 \leq u \leq N}$ with its theoretical counterpart obtained in section 6.E.2. This gives us:

$$\hat{c}_4 = \frac{3e^{4(\mathbf{S})_j^2} + 3(n-1)}{n}$$

$$(\mathbf{S})_j = 0.5 \sqrt{\log \left(n \left(\frac{\hat{c}_4}{3} - 1 \right) + 1 \right)}$$

However, in practice, \hat{c}_4 may be less than 3, resulting in the absence of a real solution for $(\mathbf{S})_j$. Additionally, the choice of Δt may lead to the ratio of the sample time step to the simulation time step not being an integer. Therefore, to address these issues, it is appropriate to use the following estimator for $(\mathbf{S})_j$:

$$(\hat{\mathbf{S}})_j = 0.5 \sqrt{\log \left(\frac{1}{N-rw-1} \sum_{u=rw+1}^N (t_u - t_{u-1}) \left(\frac{\hat{c}_4}{3} - 1 \right)_+ + 1 \right)}.$$

Appendix 6.F Estimation of the market sensitivity operator by maximum likelihood

The solution to this optimization program is reached when the partial derivatives of 6.20 with respect to \mathcal{S}_{t_u} are equal to 0. These partial derivatives are given by:

$$\frac{\partial \mathcal{L}(\hat{\boldsymbol{\Delta}}_{\mathbf{F}}; \boldsymbol{\Theta}_j)}{\partial \mathcal{S}_{t_u}} = \frac{1}{\mathcal{S}_{t_u}^3} \sum_{j=1}^m \left(\frac{\left((\Delta \mathbf{F}_{t_u})_j - (\hat{\boldsymbol{\mu}}_{t_u})_j \cdot \Delta_{t_u} \right)^2}{(\boldsymbol{\nu}_{t_u})_j^2 \cdot \mathcal{S}_{t_u}} - \mathcal{S}_{t_u}^2 \right).$$

We can then deduce the solution value of \mathcal{S}_{t_u} :

$$\mathcal{S}_{t_u} = \sqrt{\sum_{j=1}^m \frac{\left((\Delta \mathbf{F}_{t_u})_j - (\hat{\boldsymbol{\mu}}_{t_u})_j \cdot \Delta_{t_u} \right)^2}{(\boldsymbol{\nu}_{t_u})_j^2 \cdot \Delta_{t_u}}}.$$

Appendix 6.G Additional results of the numerical experiment

6.G.1 Fitted kernels

The table 6.7 shows the weights of the different exponential kernels in the composition of the kernels $K^{(\hat{\mu})}$ and $K^{(\hat{\sigma})}$ obtained after calibration by the algorithm 6⁸. Figure 6.18 shows the shape of these kernels and compares them to their respective approximations by a TSPL kernel.

	τ_1	τ_2	τ_3	τ_4	τ_5	τ_6	τ_7	τ_8	τ_9	τ_{10}
$(\delta)_k$	0	0.2%	39.7%	55.9%	0.3%	0	3.8%	0	0	0
$(w)_k$	23.2%	20.8%	28.7%	23.4%	0	3.2%	0.6%	0	0	0.1%

Table 6.7: Weight of the different exponential kernels in the composition of the kernels $K^{(\hat{\mu})}$ and $K^{(\hat{\sigma})}$.

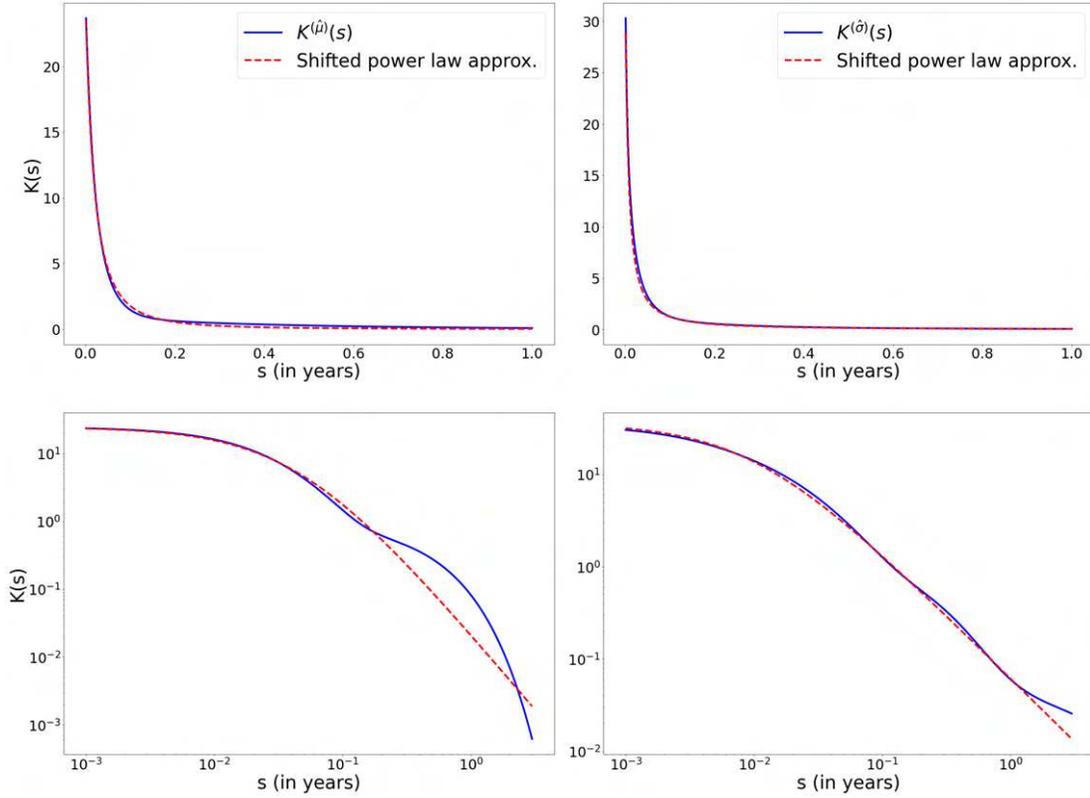


Figure 6.18: Kernels $K^{(\hat{\mu})}$ and $K^{(\hat{\sigma})}$ obtained as a result of the calibration performed by algorithm 5 and their respective approximations by a TSPL kernel.

⁸As exposed in section 6.3.1.1.2, the values of $\{\tau_k\}_{k=1}^{10}$ are defined by:

$$\tau_k = \exp\left(\log(1/365) + \frac{\log(5) - \log(1/365)}{9}(k-1)\right).$$

6.G.2 The market factor

The first common elementary factor obtained from the algorithm output 5 is hypothesized to be the market factor. Figure 6.19 allows for evaluating the coherence of this hypothesis by comparing the data related to this factor with the data related to the S&P500 index, which is a good proxy for the market portfolio. In addition, the path-dependent component estimated for this market factor, obtained after calibrating the market generator form algorithm 6, i.e. $(\mathcal{V})_1$, is compared with the VIX over the same period.

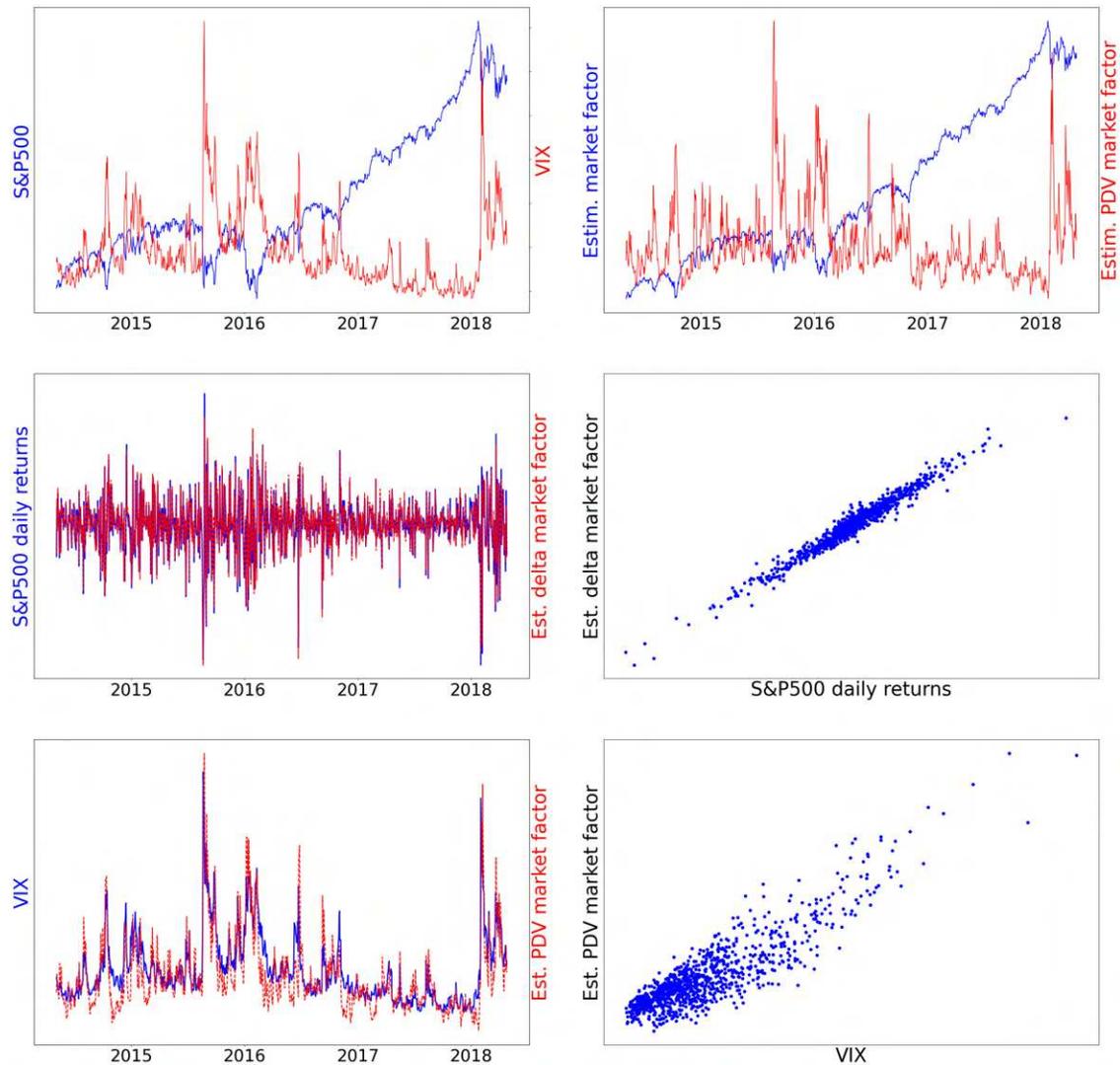


Figure 6.19: Comparison between the market factor estimated by the algorithm 5 and the S&P500 index, as well as between $(\mathcal{V})_1$ and the VIX.

6.G.3 The market sensitivity operator

The figure 6.20 allows us to compare the estimated trajectory of the market sensitivity operator \mathcal{S} obtained from the algorithm 6 with the trajectory simulated by the model over the same period. The figure 6.21 allows us to compare the empirical distribution and its autocorrelation of this market sensitivity operator with those generated by the model. Overall, the ARMA(1,1) model used to simulate the trajectory of $\log(\mathcal{S})$ reproduces the characteristics of the time series of this operator quite well. However, it can be noted that the empirical autocorrelation of \mathcal{S} tends to become more negative compared to the simulated data. Additionally, the estimated empirical value of the market sensitivity operator appears to exhibit a seasonal component. Specifically, it seems to reach its lowest values during the latter half of December, then rise again in the first days of January. Consequently, taking this potential seasonality into account could improve the modeling quality of the proposed model.

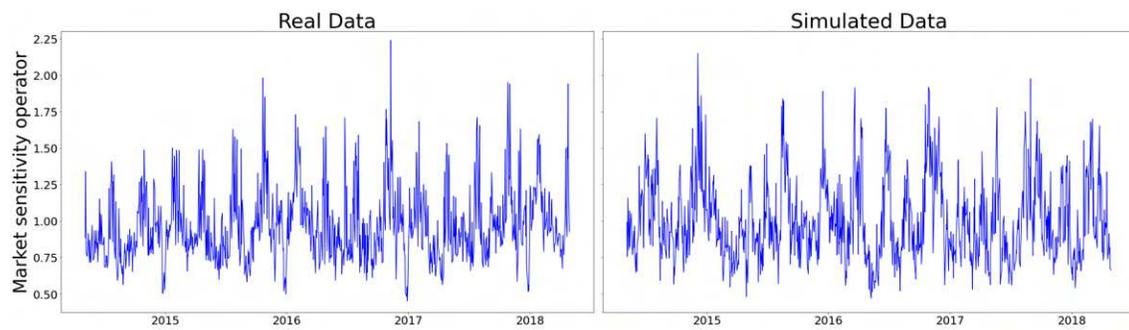


Figure 6.20: On the left, the estimated trajectory of the market sensitivity operator; on the right, a trajectory simulated from the model fitted on this path.

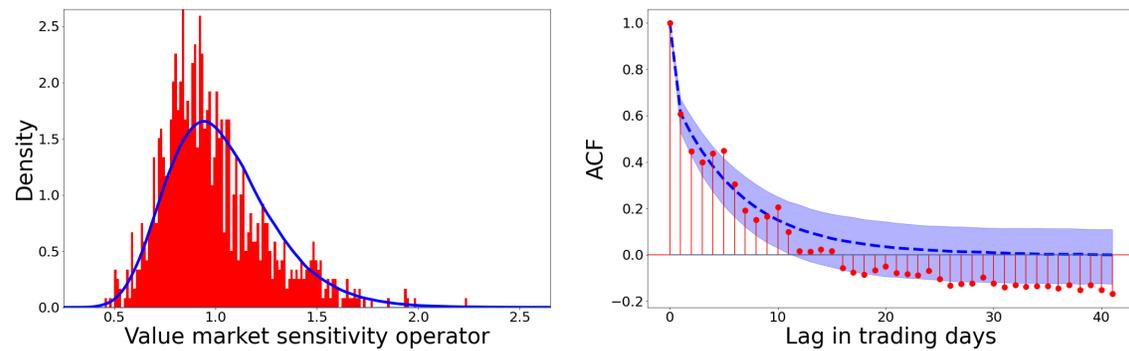


Figure 6.21: Distribution and autocorrelation of the market sensitivity operator: real vs. simulated data.

6.G.4 Some distributions of individual stock returns

Figures 6.22, 6.23, and 6.24 illustrate and compare, through three examples, the stock return distributions generated by the model with empirical data for daily, weekly, and monthly horizons.

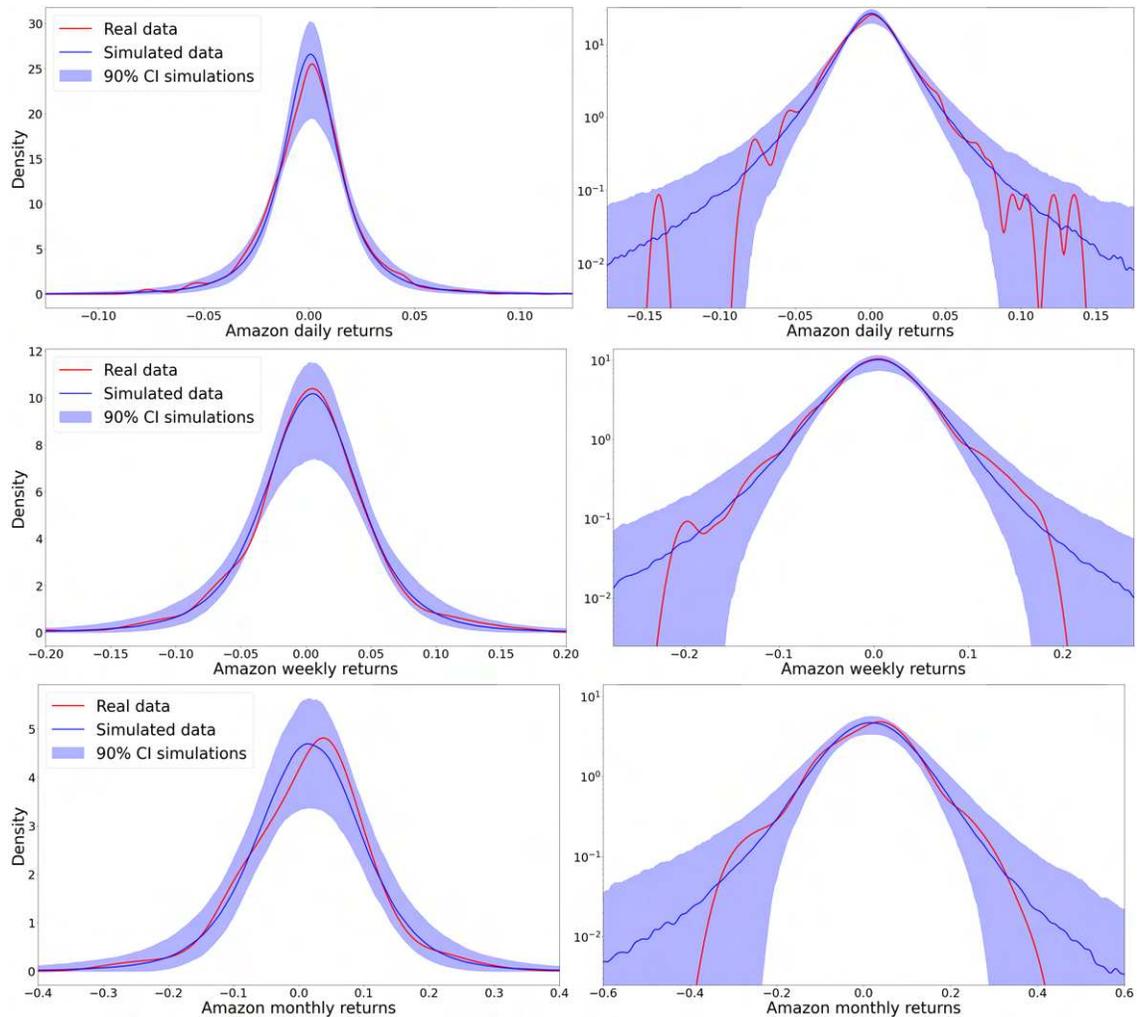


Figure 6.22: Comparison of the distribution of daily returns of Amazon stock between real data and data simulated using the FPDM generator. The KDE distribution of simulated data is estimated using the entire daily data set from the 1000 conducted simulations. The confidence interval is obtained by estimating the KDE distribution of daily returns for each simulation and then calculating the 5th and 95th percentiles of densities for each considered return level.

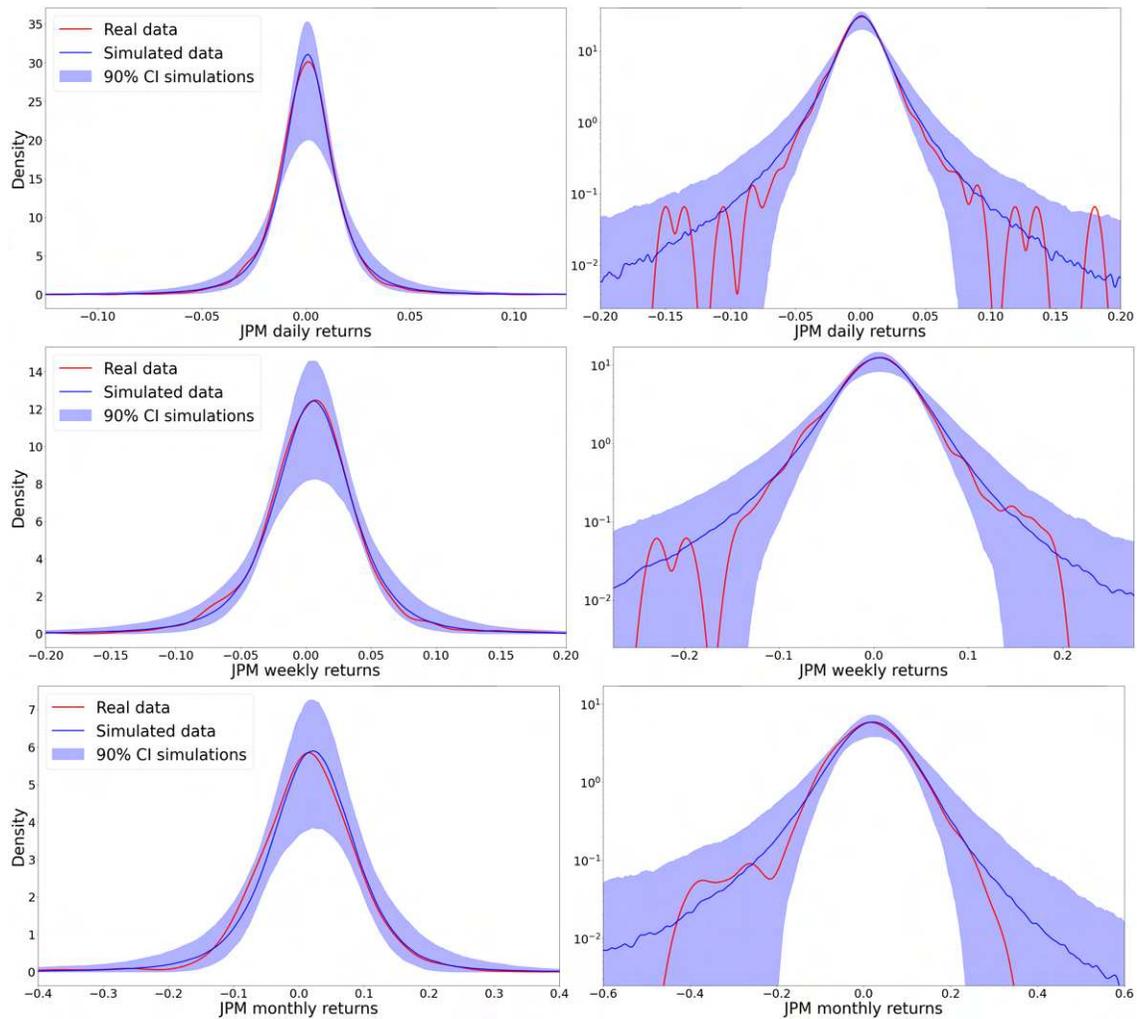


Figure 6.23: Comparison of the distribution of daily returns of JPMorgan Chase & Co. stock between real data and data simulated using the FPDM generator. The KDE distribution of simulated data is estimated using the entire daily data set from the 1000 conducted simulations. The confidence interval is obtained by estimating the KDE distribution of daily returns for each simulation and then calculating the 5th and 95th percentiles of densities for each considered return level.

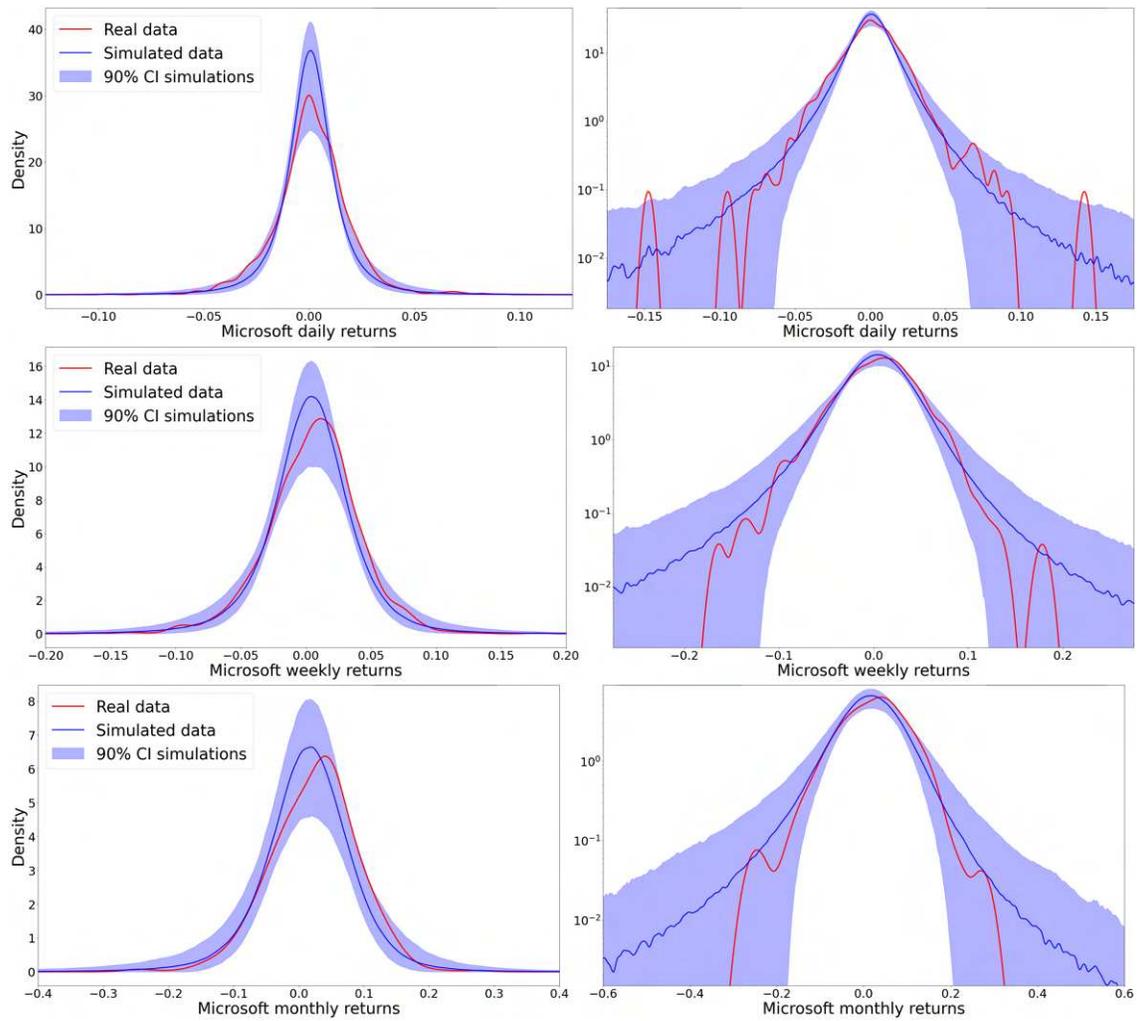


Figure 6.24: Comparison of the distribution of daily returns of Microsoft stock between real data and data simulated using the FPD generator. The KDE distribution of simulated data is estimated using the entire daily data set from the 1000 conducted simulations. The confidence interval is obtained by estimating the KDE distribution of daily returns for each simulation and then calculating the 5th and 95th percentiles of densities for each considered return level.

6.G.5 Cumulative returns of the considered investment strategies

The following graphs compare the dynamics of different strategies between real data and data simulated from the FPDM generator.

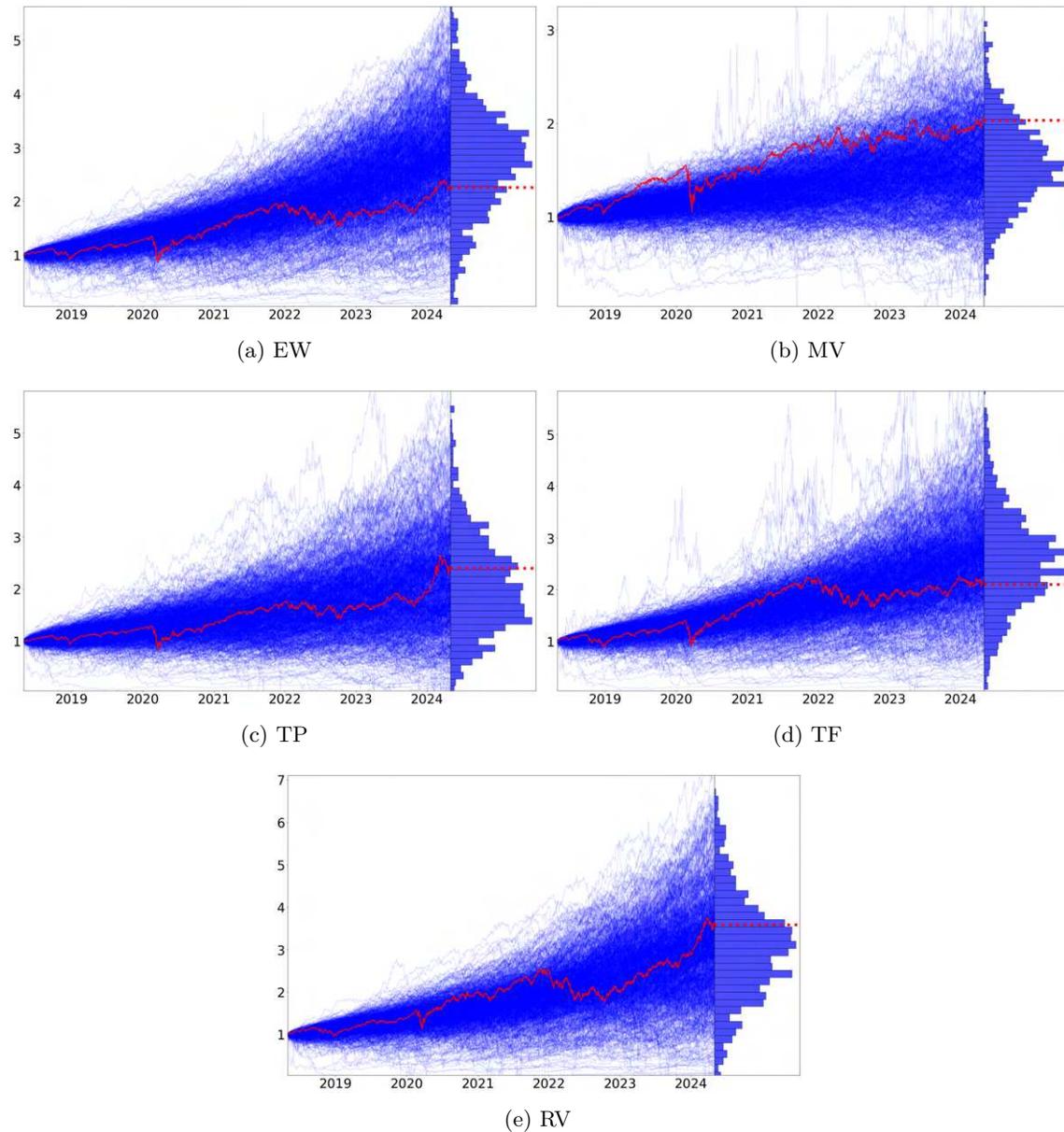


Figure 6.25: Cumulative returns of the different considered strategies with the buy-and-hold rebalancing approach: real data vs simulated data. In each plot, the red line corresponds to the cumulative returns of the strategy computed from the real data, while the blue lines represent those computed from the 1000 simulations considered in section 6.4.3.1.

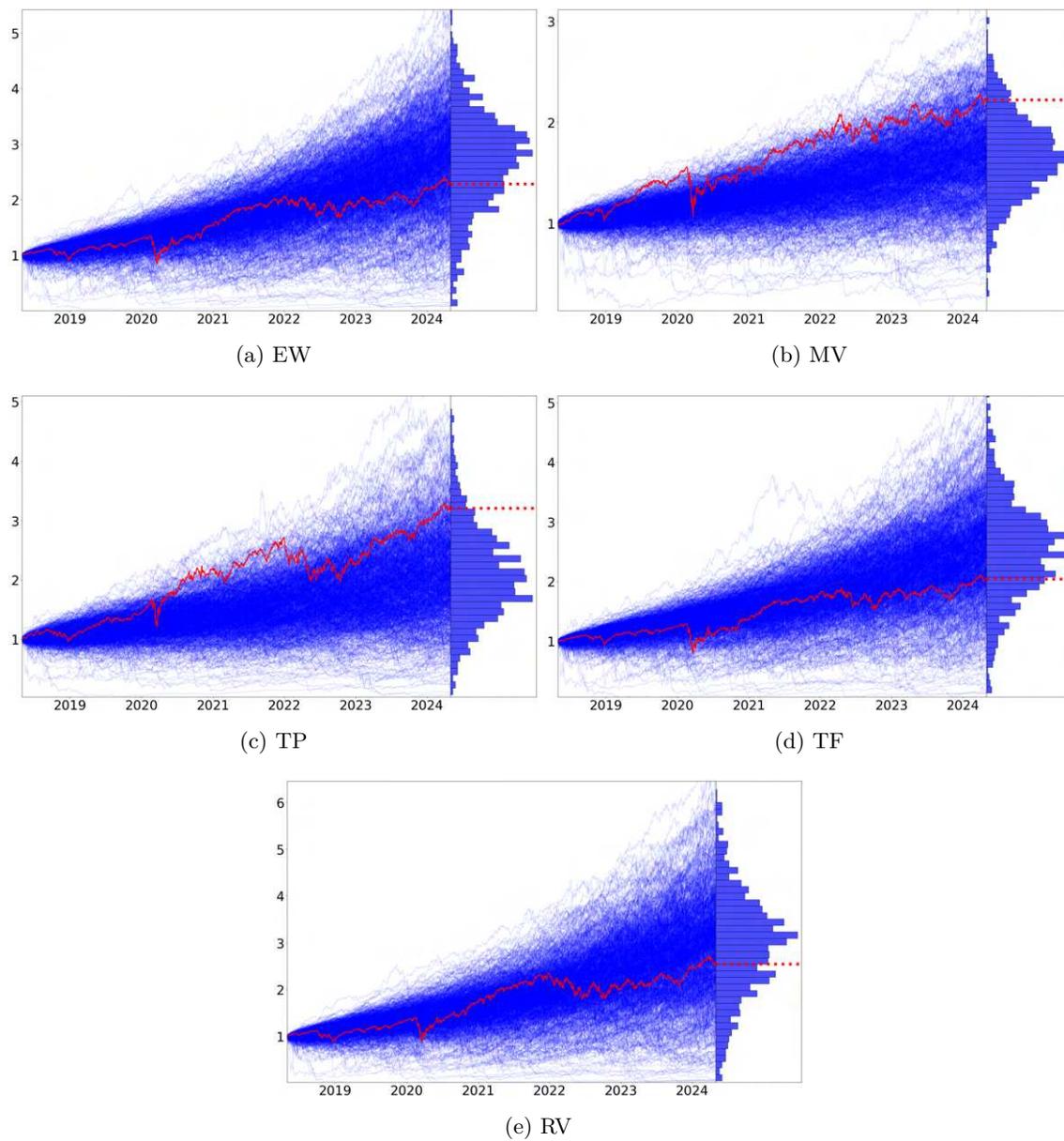


Figure 6.26: Cumulative returns of the different considered strategies with the constant-weighted rebalancing approach: real data vs simulated data. In each plot, the red line corresponds to the cumulative returns of the strategy computed from the real data, while the blue lines represent those computed from the 1000 simulations considered in section 6.4.3.1.

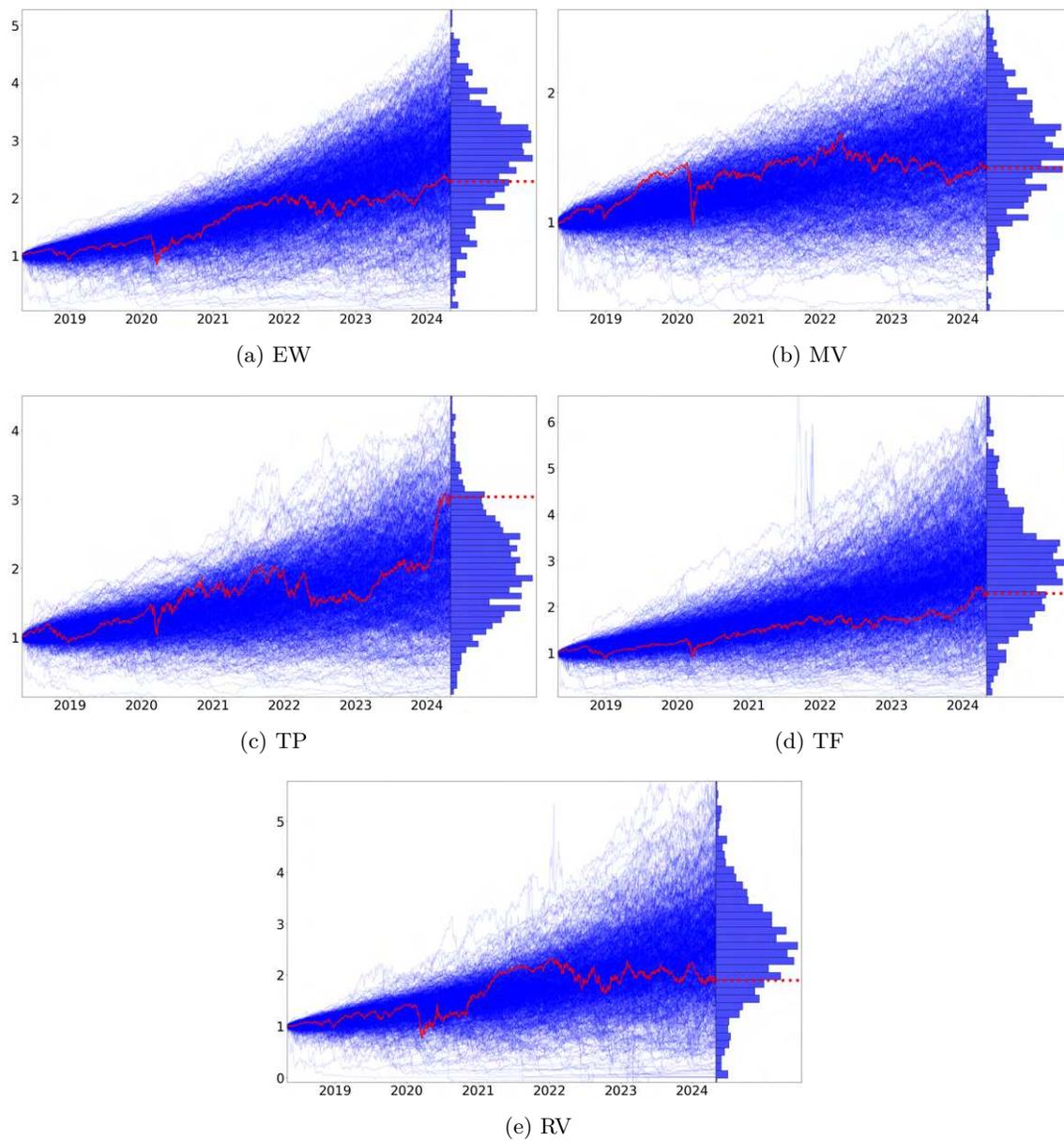


Figure 6.27: Cumulative returns of the different considered strategies with the dynamic rebalancing approach: real data vs simulated data. In each plot, the red line corresponds to the cumulative returns of the strategy computed from the real data, while the blue lines represent those computed from the 1000 simulations considered in section 6.4.3.1.

CHAPTER 7

Conclusion

Cette thèse a permis de proposer de nouveaux modèles de dynamiques de prix d'actifs capables de rendre compte d'une part substantielle des propriétés empiriques des séries financières. Si les formes de modélisation qui ont été mobilisées sont multiples, toutes ont en partage d'intégrer une rétroaction forte des dynamiques passées sur les mouvements de prix présents et futurs.

Cette rétroaction s'opère principalement par le truchement de la volatilité appréhendée comme un processus essentiellement path-dependent. L'information contenue dans ces trajectoires est en pratique encodée dans deux types de variables d'état : des variables d'état de tendance des prix ou de tendance factorielle, et des variables d'état de volatilité historique. Cette formalisation permet ainsi la capture de phénomènes empiriques clés comme le syndrome de Joseph, le leverage effect ou encore l'effet Zumbach, par des mécanismes explicites de rétroaction. Outre son intérêt théorique, le chapitre 5 a montré que ce cadre de modélisation pouvait être utilisé comme un outil puissant pour la prédiction de la volatilité. Aussi, la capture de la relation entre dynamique de volatilité et dynamique de prix induit un ensemble d'autres propriétés, elles aussi vérifiées empiriquement. Parmi celles-ci figurent l'évolution du skewness des distributions de rendements en fonction de l'échelle de temps considérée, et de manière plus générale, l'évolution des distributions de rendements elles-mêmes.

L'influence mutuelle du drift et de la volatilité dans ce cadre, et son impact sur les dynamiques de prix ont également été mises en évidence. Tout d'abord, la propriété de path-dependence de la volatilité rend celle-ci tributaire du niveau du drift. Ainsi, les conditions de détermination du drift peuvent modifier les dynamiques et les distributions du processus de volatilité, et par extension celles du prix. À cet égard, plusieurs hypothèses de modélisation du drift comme fonction du niveau de volatilité ont été explorées. Sous ce type d'hypothèses, le drift est alors à la fois un déterminant de la volatilité tout en étant déterminé par celle-ci. Il a été ainsi montré que cette relation, articulée à la propriété de path-dependence, peut produire l'émergence de mécanismes d'auto-ajustement des dynamiques de prix. Exemplement, la relation négative entre tendance passée des prix et niveau de volatilité, conjointe à une relation croissante entre niveau du drift et

niveau de la volatilité, engendre un effet de réversion du prix de l'actif.

Autre élément clé, si les dynamiques de volatilités sont explicables dans une large mesure par ces mécanismes de rétroaction, il subsiste néanmoins une part résiduelle qui semble s'originer dans des causes exogènes. Les résultats empiriques obtenus dans les divers articles suggèrent que ces variations représentent une part significative (au sens statistique) des mouvements de la volatilité à court terme, dès lors que les prix d'actifs sont appréhendés comme des processus purement diffusifs. Cependant, à une échelle de temps plus longue, ces variations sont soit absorbées, soit indifférenciables d'un effet de rétroaction positif de la volatilité sur elle-même. Ainsi, dans le cadre d'une modélisation de la volatilité incorporant cet effet, la volatilité peut être exprimée comme le produit d'un modèle purement path-dépendent avec un processus à mémoire courte indépendant des dynamiques de prix. Bien que l'autocorrélation de ce processus, modélisant les frictions aléatoires exogènes de volatilité, décroisse rapidement, des chocs exogènes issus de ce processus peuvent avoir un impact de longue durée par le biais de la composante path-dépendent de la volatilité, qui agit alors comme courroie de transmission. Ce phénomène de persistance tient à la lente diminution de l'autocorrélation des variables d'état dont est fonction la composante path-dépendent de la volatilité, variables d'état dont la mémoire présente une structure proche d'une loi puissance décalée (shifted power law) modélisable par une combinaison linéaire de plusieurs moyennes mobiles exponentielles.

Outre des contributions à la modélisation univariée, la présente thèse a également introduit un modèle multivarié des dynamiques des prix d'actifs, s'inscrivant dans une approche factorielle. Sur le plan théorique d'abord, cette modélisation a permis de fournir une grille d'intelligibilité aux dynamiques multidimensionnelles des prix d'actifs. À l'instar des modèles univariés, les mouvements de prix sont là encore hautement dépendants des dynamiques passées, mais la path-dépendance s'opère ici sur les trajectoires des facteurs dont dépendent les prix. Ce modèle a notamment permis d'expliquer des phénomènes empiriques tels que les mouvements de corrélation ou les effets de contagion de la volatilité (volatility spillover effect) à partir de dynamiques factorielles endogènes. Sur le plan pratique, l'utilisation de ce cadre de modélisation comme outil de simulation a également démontré une efficacité remarquable pour générer des séries financières multidimensionnelles réalistes. Cette formalisation permet donc de rendre compte de la majeure partie des propriétés caractéristiques des trajectoires de prix d'actifs en modélisant un marché, non comme une simple collection d'actifs munie d'une structure de corrélation simple, mais comme un système complexe où s'enchevêtrent différents mécanismes de rétroaction.

Plus globalement, les résultats obtenus au fil de cette thèse ont démontré l'intérêt des modélisations dans lesquelles les dynamiques de prix d'actifs sont surdéterminées par les évolutions passées du marché. Il ressort ainsi l'idonéité de ces approches à saisir la nature foncièrement endogène des marchés financiers.

Bibliography

- [1] Abi Jaber E., (2019). Lifting the Heston model. *Quantitative Finance*, 19(12), pp.1995-2013.
- [2] Abi Jaber E., Cuchiero C., Larsson M., and Pulido S. (2021). A weak solution theory for stochastic Volterra equations of convolution type. *The Annals of Applied Probability*, 31(6), 2924-2952.
- [3] Abi Jaber E., and De Carvalho N. (2023). Reconciling rough volatility with jumps. Available at *SSRN* 4387574.
- [4] Abi Jaber E., and El Euch O. (2019). Multifactor approximation of rough volatility models. *SIAM Journal on Financial Mathematics*, 10(2), 309-349.
- [5] Abi Jaber E., Larsson M., and Pulido S. (2019). Affine volterra processes.
- [6] Aït-Sahalia Y., and Kimmel R. (2007). Maximum likelihood estimation of stochastic volatility models. *Journal of financial economics*, 83(2), 413-452.
- [7] Albrecher H., Mayer P., Schoutens W., and Tistaert J. (2007). The Little Heston Trap. *Wilmott*, (1), 83-92.
- [8] Ali U., and Hirshleifer D. (2020). Shared analyst coverage: Unifying momentum spillover effects. *Journal of Financial Economics*, 136(3), 649-675.
- [9] Alizadeh S., Brandt M. W., and Diebold F.X. (2002). Rangebased estimation of stochastic volatility models. *The Journal of Finance*, 57(3), 1047-1091.
- [10] Andersen T. G., Bollerslev T., Diebold F. X., and Labys P. (2003). Modeling and forecasting realized volatility. *Econometrica*, 71(2), 579-625.
- [11] Ang A., and Chen J. (2002). Asymmetric correlations of equity portfolios. *Journal of financial Economics*, 63(3), 443-494.
- [12] Arratia A., Cabaña A., and Cabaña, E.M. (2018). Embedding in law of discrete time ARMA processes in continuous time stationary processes. *Journal of Statistical Planning and Inference*, 197, 156-167.
- [13] Bachelier L. (1900). Théorie de la spéculation. In *Annales scientifiques de l'École normale supérieure* (Vol. 17, pp. 21-86).

- [14] Bacry E., Delour J., and Muzy J. F. (2001). Multifractal random walk. *Physical review E*, 64(2), 026103.
- [15] Baele L. (2005). Volatility spillover effects in European equity markets. *Journal of Financial and Quantitative Analysis*, 40(2), 373-401.
- [16] Baele L., Bekaert G., Inghelbrecht K., and Wei M. (2020). Flights to safety. *The Review of Financial Studies*, 33(2), 689-746.
- [17] Baillie R.T., Bollerslev T., and Mikkelsen, H. O. (1996). Fractionally integrated generalized autoregressive conditional heteroskedasticity. *Journal of Econometrics*, 74(1), 3-30.
- [18] Bakshi G., and Kapadia N. (2003). Delta-hedged gains and the negative market volatility risk premium. *The Review of Financial Studies*, 16(2), 527-566.
- [19] Babura M., Giannone D., and Lenza M. (2015). Conditional forecasts and scenario analysis with vector autoregressions for large cross-sections. *International Journal of forecasting*, 31(3), 739-756.
- [20] Barndorff-Nielsen O. E., and Schmiegel J. (2009). Brownian semistationary processes and volatility/intermittency. *Advanced financial modelling*, 8, 1-26.
- [21] Barra S., Carta S.M., Corrigan A., Podda A.S., and Recupero D.R. (2020). Deep learning and time series-to-image encoding for financial forecasting. *IEEE/CAA Journal of Automatica Sinica*, 7(3), 683-692.
- [22] Bayer C., and Breneis S. (2023). Weak Markovian approximations of rough Heston. *arXiv preprint arXiv:2309.07023*.
- [23] Bayer C., Friz P., and Gatheral J. (2016). Pricing under rough volatility. *Quantitative Finance*, 16(6), 887-904.
- [24] Bayer C., Horvath B., Muguruza A., Stemper B., and Tomas M. (2019). On deep calibration of (rough) stochastic volatility models. *arXiv preprint arXiv:1908.08806*.
- [25] Bekaert G., Engstrom E. C., and Xu N. R. (2022). The time variation in risk appetite and uncertainty. *Management Science*, 68(6), 3975-4004.
- [26] Belkhouja M., and Boutahary M. (2011). Modeling volatility with time-varying FIGARCH models. *Economic Modelling*, 28(3), 1106-1116.
- [27] Bergomi L. (2015). Smile dynamics II, *Risk*.
- [28] Bernanke B. S., Gertler M., and Gilchrist S. (1994). The financial accelerator and the flight to quality.
- [29] Bennedsen M., Lunde A., and Pakkanen M.S. (2016). Decoupling the short-and long-term behavior of stochastic volatility. *arXiv preprint arXiv:1610.00332*.

- [30] Berger J.O. (2013). *Statistical decision theory and Bayesian analysis*. Springer Science & Business Media.
- [31] Bianchi M.L., Hitaj A., and Tassinari G.L. (2020). Multivariate non-Gaussian models for financial applications. *arXiv preprint arXiv:2005.06390*.
- [32] Bickel P.J., and Doksum K.A. (2015). *Mathematical statistics: basic ideas and selected topics, volumes I* CRC Press.
- [33] Bildirici M., and Ersin Ö. (2014). Asymmetric power and fractionally integrated support vector and neural network GARCH models with an application to forecasting financial returns in ise100 stock index. *Economic Computation and Economic Cybernetics Studies and Research*, 48, 1-22.
- [34] Björk T. (2009). *Arbitrage theory in continuous time*. Oxford university press.
- [35] Black F. (1976). Studies of Stock Price Volatility Changes. *Proceedings of the 1976 Meetings of the Business and Economics Section, American Statistical Association*.
- [36] Black F. (1986). Noise. *The journal of finance*, 41(3), 528-543.
- [37] Black F., and Litterman R. (1992). Global portfolio optimization. *Financial analysts journal*, 48(5), 28-43.
- [38] Black F., and Scholes M. (1973). The pricing of options and corporate liabilities. *Journal of political economy*, 81(3), 637-654.
- [39] Blanc P., Donier J., and Bouchaud J.P. (2017). Quadratic Hawkes processes for financial prices. *Quantitative Finance*, 17(2), 171-188.
- [40] Bochud T. and Challet D. (2007). Optimal approximations of power laws with exponentials: application to volatility models with long memory. *Quantitative Finance*, 7(6), 585-589.
- [41] Bollerslev T., Chou, R.Y., and Kroner K. F. (1992). ARCH modeling in finance: A review of the theory and empirical evidence. *Journal of econometrics*, 52(1-2), 5-59.
- [42] Bollerslev T., Engle R.F., and Nelson D.B. (1994). ARCH models. *Handbook of Econometrics*, 4, 2959-3038.
- [43] Borland L., and Bouchaud J. P. (2005). On a multi-timescale statistical feedback model for volatility fluctuations. *arXiv preprint physics/0507073*.
- [44] Bouchaud J.P., Matacz A., Potters M. (2001). Leverage effect in financial markets: The retarded volatility model. *Physical review letters*, 87(22), 228701.
- [45] Bouchaud J.P., and Potters M. (2003). *Theory of financial risk and derivative pricing: from statistical physics to risk management*. Cambridge university press.
- [46] Bouchaud J. P., and Potters M. (2009). Financial applications of random matrix theory: a short review. *arXiv preprint arXiv:0910.1205*.

- [47] Bouchaud J.P. (2010), The endogenous dynamics of markets: price impact and feedback loops. *arXiv preprint arXiv:1009.2928*.
- [48] Bouchaud J.P. (2021). De la physique statistique aux sciences sociales: les défis de la pluridisciplinarité. *Chaire d'innovation technologique Liliane Bettencourt-Collège de France*, 108.
- [49] Box G.E., Jenkins G.M., Reinsel G.C., and Ljung G.M. (2015). *Time series analysis: forecasting and control*. John Wiley & Sons.
- [50] Brandt M.W. and Jones C.S. (2006). Volatility forecasting with range-based EGARCH models. *Journal of Business & Economic Statistics*, 24(4), 470-486.
- [51] Brockwell P., Chadraa E., and Lindner A. (2006). Continuous-time GARCH processes. *The Annals of Applied Probability*, 16(2), 790-826.
- [52] Brockwell P. J. (2001). Continuous-time ARMA processes. *Handbook of statistics*, 19, 249-276.
- [53] Brockwell P. J. (2004). Representations of continuous-time ARMA processes. *Journal of Applied Probability*, 41(A), 375-382.
- [54] Brooks L. D., and Buckmaster D. A. (1976), Further Evidence on the Time Series Properties of Accounting Income. *Journal of Finance*, vol. 31, issue 5, 1359-73.
- [55] Brooks C., and Persaud G. (2001). Volatility forecasting for risk management. *Journal of Forecasting*, 20(5), 341-356.
- [56] Bruder B., and Gaussel N. (2011). Risk-Return Analysis of Dynamic Investment Strategies. Available at SSRN 2465623.
- [57] Campbell J.Y., and Shiller R.J. (1988). The dividend-price ratio and expectations of future dividends and discount factors. *The review of financial studies*, 1(3), 195-228.
- [58] Campbell J.Y. et Shiller R.J. Valuation ratios and the long-run stock market outlook: An update. *Advances in Behavioral Finance*, 2005, vol. 2, p. 173-201.
- [59] Carhart M.M. (1997), On Persistence in Mutual Fund Performance, *Journal of Finance*, 52(1), pp. 57-82.
- [60] Carr P., Geman H., Madan D.B., and Yor M. (2002). D *The journal of Business*, 75(2), 305-332.
- [61] Chambers M.J., and Thornton M. A. (2012). Discrete time representation of continuous time ARMA processes. *Econometric Theory*, 28(1), 219-238.
- [62] Chen Y., Wiesel A., Eldar Y. C., and Hero A. O. (2010). Shrinkage algorithms for MMSE covariance estimation. *IEEE transactions on signal processing*, 58(10), 5016-5029.
- [63] Chicheportiche R., and Bouchaud J.P. (2011). Goodness-of-fit tests with dependent observations. *Journal of Statistical Mechanics: Theory and Experiment*, 2011(09), P09003.

- [64] Chicheportiche R., and Bouchaud J.P. (2012), The joint distribution of stock returns is not elliptical. *International Journal of Theoretical and Applied Finance*,15(03), 1250019.
- [65] Chicheportiche R., and Bouchaud J.P. (2015). A nested factor model for non-linear dependencies in stock returns. *Quantitative Finance*, 15(11), 1789-1804.
- [66] Christiansen C. (2007). Volatilityspillover effects in European bond markets. *European Financial Management*, 13(5), 923-948.
- [67] Ciner C., Gurdgiev C., and Lucey B.M. (2013). Hedges and safe havens: An examination of stocks, bonds, gold, oil and exchange rates. *International Review of Financial Analysis*, 29, 202-211.
- [68] Cizeau P., Potters M., and Bouchaud J.P. (2001). Correlation structure of extreme stock returns. *Quantitative Finance*, 1(2), 217222.
- [69] Colacito R., Engle R.F., and Ghysels E. (2011). A component model for dynamic correlations. *Journal of Econometrics*, 164(1), 45-59.
- [70] Cont R. (2001). Empirical properties of asset returns: stylized facts and statistical issues. *Quantitative finance*, 1(2), 223.
- [71] Cont R. (2011). Benoit Mandelbrot et la modélisation mathématique des risques financiers.
- [72] Cont R., and Das P. (2022). Rough volatility: fact or artefact?. *arXiv preprint arXiv:2203.13820*.
- [73] Cont R., Cucuringu M., Xu R., and Zhang, C. (2022). Tail-gan: Learning to simulate tail risk scenarios. *arXiv preprint arXiv:2203.01664*.
- [74] Corsi F. (2009). A simple approximate long-memory model of realized volatility. *Journal of Financial Econometrics*, 7(2), 174-196.
- [75] Coutin L., and Decreusefond L. (2001). Stochastic Volterra equations with singular kernels. *Stochastic analysis and mathematical physics* (pp. 39-50). Boston, MA: Birkhäuser Boston.
- [76] Cox J.C. (1975). Notes on Option Pricing I: Constant Elasticity of Variance Diffusions. *Working Paper*, Stanford University.
- [77] Cox J.C., Ingersoll Jr J. E., and Ross S. A. (1985). An intertemporal general equilibrium model of asset prices. *Econometrica: Journal of the Econometric Society*, 363-384.
- [78] Csáji B.C. (2001). Approximation with artificial neural networks. *Faculty of Sciences, Eötvös Lornd University, Hungary*, 24(48), 7.
- [79] Cui Z., Chen W., and Chen Y. (2016). Multi-scale convolutional neural networks for time series classification. *arXiv preprint arXiv:1603.06995*.

- [80] Dandapani A., Jusselin P., and Rosenbaum M. (2021). From quadratic Hawkes processes to super-Heston rough volatility models with Zumbach effect. *Quantitative finance*, 21(8), 1235-1247.
- [81] De Bondt W. F., and Thaler R. H. (1987). Further Evidence on Investor Overreaction and Stock Market Seasonality. *Journal of Finance*, 42(3), 557-581.
- [82] Diebold F.X. (1998). Elements of forecasting. Cincinnati, OH, USA: South-Western College Pub.
- [83] Ding Z., Granger C.W., and Engle R.F. (1993). A long memory property of stock market returns and a new model. *Journal of Empirical Finance*, 1(1), 83-106.
- [84] Dissanaik G. (1997). Do Stock Market Investors Overreact? *Journal of Business Finance & Accounting*, 24(1), 27-50.
- [85] De Prado M.L., (2018). *Advances in financial machine learning*. John Wiley & Sons.
- [86] Drgulescu A.A., and Yakovenko V.M. (2002). Probability distribution of returns in the Heston model with stochastic volatility. *Quantitative finance*, 2(6), 443-453.
- [87] Duffie D. (1991). The theory of value in security markets. *Handbook of mathematical economics*, 4, 1615-1682.
- [88] Duffie D., Pan J., and Singleton K. (2000). Transform analysis and asset pricing for affine jumpdiffusions. *Econometrica*, 68(6), 1343-1376.
- [89] El Euch O., and Rosenbaum M. (2018). Perfect Hedging in Rough Heston Models. *Annals of Applied Probability*, 28(6), 3813-3856.
- [90] El Euch O., and Rosenbaum M. (2019). The Characteristic Function of Rough Heston Models. *Mathematical Finance*, 29(1), 3-38.
- [91] El Euch O. (2018). *Quantitative Finance under rough volatility*, PhD diss., Sorbonne Université.
- [92] Engle R.F. (1982). Autoregressive conditional heteroscedasticity with estimates of the variance of United Kingdom inflation. *Econometrica: Journal of the Econometric Society*, 987-1007.
- [93] Engle R.F., Ledoit O., and Wolf M. (2019). Large dynamic covariance matrices. *Journal of Business & Economic Statistics*, 37(2), 363-375.
- [94] Engle R.F., Lilien D. M., and Robins R. P. (1987). Estimating time varying risk premia in the term structure: The ARCH-M model. *Econometrica: journal of the Econometric Society*, 391-407.
- [95] Engle R.F., and Patton A.J. (2007). What good is a volatility model?. *Forecasting volatility in the financial markets* (pp. 47-63). Butterworth-Heinemann.

- [96] Fama E.F. (1970). Efficient capital markets: A review of theory and empirical work. *The Journal of Finance*, 25(2), 383-417.
- [97] Figlewski S., and Wang X. (2000). Is the 'Leverage Effect' a leverage effect?. Available at *SSRN* 256109.
- [98] Filimonov V., and Sornette D. (2012). Quantifying reflexivity in financial markets: Toward a prediction of flash crashes. *Physical Review*, E 85.5: 056108.
- [99] Flaig S., and Junike G. (2022). Scenario generation for market risk models using generative neural networks. *Risks*, 10(11), 199.
- [100] Fortin I., and Kuzmics C. (2002). Taildependence in stockreturn pairs. systemic risk (No. 9-230). *Intelligent Systems in Accounting, Finance & Management*, 11(2), 89-107.
- [101] Foschi P., and Pascucci A. (2008). Path dependent volatility. *Decisions in Economics and Finance*, 31, 13-32.
- [102] Franses P.H., and Van Dijk D. (1996). Forecasting stock market volatility using (nonlinear) Garch models. *Journal of forecasting*, 15(3), 229-235.
- [103] French K. R., Schwert G. W., and Stambaugh R. F. (1987). Expected stock returns and volatility. *Journal of financial Economics*, 19(1), 3-29.
- [104] Gatheral J., Bouchaud J.P. (2021). The Complex Dynamics of Financial Prices.
- [105] Gatheral J., Jaisson T., and Rosenbaum M. (2018). Volatility is Rough. *Quantitative Finance*, 18(6), 933-949.
- [106] Gatheral J., Jusselin P., and Rosenbaum M. (2020). The Quadratic Rough Heston Model and the Joint S&P500/VIX Smile Calibration Problem. *arXiv* preprint arXiv:2001.01789.
- [107] Gil M., Alajaji F., and Linder T. (2013). Rényi divergence measures for commonly used univariate continuous distributions. *Information Sciences*, 249, 124-131.
- [108] Gloter A. (2007). Efficient estimation of drift parameters in stochastic volatility models. *Finance and Stochastics*, 11(4), 495-519.
- [109] Gomes C., and Waelbroeck H. (2015). Is market impact a measure of the information value of trades? Market response to liquidity vs. informed metaorders. *Quantitative Finance*, 15(5), 773-793.
- [110] González-Hermosillo M. B. and Hesse H. (2009). Global market conditions and systemic risk (No. 9-230). *International Monetary Fund*.
- [111] Granger C. W., and Newbold, P. (1974). Spurious regressions in econometrics. *Journal of econometrics*, 2(2), 111-120.

- [112] Grinblatt M., Titman S. and Wermers R. (1995), Momentum Investment Strategies, Portfolio Performance, and Herding: A Study of Mutual Fund Behavior, *American Economic Review*, 85(5), pp. 1088-1105.
- [113] Gropp J. (2004). Mean reversion of industry stock returns in the US, 1926-1998. *Journal of Empirical Finance*, 11(4), 537-551.
- [114] Gulisashvili A. (2012). *Analytically tractable stochastic stock price models*. Springer Science & Business Media.
- [115] Gustafsson J., and Jonsson C. (2023). *Scenario Generation for Stress Testing Using Generative Adversarial Networks: Deep Learning Approach to Generate Extreme but Plausible Scenarios*.
- [116] Guyon J. (2014). Path-dependent volatility. *Risk Magazine*.
- [117] Guyon J., and Lekeufack J. (2023). Volatility is (mostly) path-dependent. *Quantitative Finance*, 23(9), 1221-1258.
- [118] Han Y. (2011). On the relation between the market risk premium and market volatility. *Applied financial economics*, 21(22), 1711-1723.
- [119] Harvey A.C. (1990). *Forecasting, Structural Time Series Models and the Kalman Filter*. Institute of Mathematical Statistics.
- [120] Hernandez A. (2016). Model calibration with neural networks. Available at *SSRN* 2812140.
- [121] Heston S.L. (1993). A closed-form solution for options with stochastic volatility with applications to bond and currency options. *The review of financial studies*, 6(2), 327-343.
- [122] Hobson D.G., and Rogers L.C. (1998). Complete Models with Stochastic Volatility. *Mathematical Finance*, 8(1), 27-48.
- [123] Hong Y. (2001). A test for volatility spillover with application to exchange rates. *Journal of Econometrics*, 103(1-2), 183-224.
- [124] Hornik K., Stinchcombe M., and White H. (1989). Multilayer feedforward networks are universal approximators. *Neural networks*, 2(5), 359-366.
- [125] Horvath B., Muguruza A., and Tomas M. (2021). Deep learning volatility: a deep neural network perspective on pricing and calibration in (rough) volatility models. *Quantitative Finance*, 21(1), 11-27.
- [126] Hubalek F., Kallsen J., and Krawczyk L. (2006). Variance-optimal hedging for processes with stationary independent increments. *The Annals of Applied Probability*, 16(2), 853-885
- [127] Jegadeesh N., and Titman S. (1993). Returns to buying winners and selling losers: Implications for stock market efficiency. *The Journal of finance*, 48(1), 65-91.

- [128] Jorion, P. (2000). Risk management lessons from longterm capital management. *European financial management*, 6(3), 277-300.
- [129] Jusselin P., Lezmi E., Malongo H., Masselin C., Roncalli T., and Dao T. L. (2017). Understanding the momentum risk premium: An in-depth journey through trend-following strategies. Available at SSRN3042173.
- [130] Kelly B., Malamud S., and Zhou K. (2024). The virtue of complexity in return prediction. *The Journal of Finance*, 79(1), 459-503.
- [131] Kimoto T., Asakawa K., Yoda M., and Takeoka M. (1990). Stock market prediction system with modular neural networks. In 1990 *IJCNN international joint conference on neural networks* (pp. 1-6). IEEE.
- [132] Kim S., Shephard N., and Chib S. (1998). Stochastic volatility: likelihood inference and comparison with ARCH models. *The review of economic studies*, 65(3), 361-393.
- [133] Kondratyev A., and Schwarz C. (2019). The market generator. Available at SSRN3384948.
- [134] Kou S. G. (2002). A jump-diffusion model for option pricing. *Management science*, 48(8), 1086-1101.
- [135] Laloux L., Cizeau P., Bouchaud J.P., Potters M. (1999). Noise dressing of financial correlation matrices. *Physical review letters*, 83(7), 1467.
- [136] Lanne M., and Saikkonen P. (2005). Nonlinear GARCH models for highly persistent volatility. *The Econometrics Journal*, 8(2), 251-276.
- [137] Ledoit O., and Wolf M. (2004). A well-conditioned estimator for large-dimensional covariance matrices. *Journal of multivariate analysis*, 88(2), 365-411.
- [138] Ledoit O., and Wolf M. (2020). Analytical nonlinear shrinkage of large-dimensional covariance matrices. *The Annals of Statistics*, 48(5), 3043-3065.
- [139] Lezmi E., Roche J., Roncalli T., and Xu J. (2020). Improving the Robustness of Trading Strategy Backtesting with Boltzmann Machines and Generative Adversarial Networks. Available at SSRN 3645473.
- [140] Lopez de Prado M. (2016). A robust estimator of the efficient frontier. Available at SSRN 3469961.
- [141] Lopez de Prado M. (2019). Tactical investment algorithms. Available at SSRN 3459866.
- [142] Lord R., Koekkoek R., and Dijk D.V. (2010). A comparison of biased simulation schemes for stochastic volatility models. *Quantitative Finance*, 10(2), 177-194.
- [143] Lynch P.E., and Zumbach, G.O. (2003). Market heterogeneities and the causal structure of volatility. *Quantitative Finance*, 3(4), 320.

- [144] Malevergne Y., Pisarenko V., Sornette D. (2005). Empirical distributions of stock returns: between the stretched exponential and the power law?. *Quantitative Finance*, 5(4), 379-401.
- [145] Maller R.A., Müller G. and Szimayer A. (2008). GARCH modelling in continuous time for irregularly spaced time series data. *Bernoulli*, 14(2), 519-542.
- [146] Mandelbrot B.B. (1967), The Variation of Some Other Speculative Prices, *Journal of Business* 40, 393-413
- [147] Mandelbrot B. B. (1973). Le syndrome de la variance infinie et ses rapports avec la discontinuité des prix. *Économie appliquée*, 26(2), 321-348.
- [148] Mandelbrot B.B., Hudson R.L. (2010). *The (mis)behaviour of markets: a fractal view of risk, ruin and reward*. Profile books.
- [149] Mandelbrot B.B. (2003). Heavy tails in finance for independent or multifractal price increments. In *Handbook of Heavy Tailed Distributions in Finance* (pp. 1-34). North-Holland.
- [150] Marchioro M. (2017). A risk decomposition framework consistent with performance measurements. *Risk and Performance Attribution, Quantitative Research Series. Stat-Pro website*.
- [151] Maruyama G. (1955). Continuous Markov processes and stochastic equations. *Rendiconti del Circolo Matematico di Palermo*, 4:4890.
- [152] Mechkov S. (2015). Fast-reversion limit of the Heston model. Available at SSRN 2418631.
- [153] Mejía Vega C.A. (2018). Calibration of the exponential OrnsteinUhlenbeck process when spot prices are visible through the maximum log-likelihood method. Example with gold prices. *Advances in Difference Equations*, 2018, 1-14.
- [154] Merton R.C. (1973). An intertemporal capital asset pricing model. *Econometrica: Journal of the Econometric Society*, 867-887.
- [155] Merton R. C. (1976). Option pricing when underlying stock returns are discontinuous. *Journal of financial economics*, 3(1-2), 125-144.
- [156] Meucci A. (2009). Review of statistical arbitrage, cointegration, and multivariate Ornstein-Uhlenbeck.
- [157] Morel R., Rochette G., Leonarduzzi R., Bouchaud J. P., and Mallat S. (2022). Scale dependencies and self-similar models with wavelet scattering spectra. *arXiv preprint arXiv:2204.10177*.
- [158] Morrill J., Fermanian A., Kidger P., and Lyons T. (2020). A generalised signature method for multivariate time series feature extraction. *arXiv preprint arXiv:2006.00873*.
- [159] Moskowitz T.J., and Grinblatt M. (1999). Do industries explain momentum?. *The Journal of finance*, 54(4), 1249-1290.

- [160] Mrázek M., and Pospíšil J. (2017). Calibration and simulation of Heston model. *Open Mathematics*, 15(1), 679-704.
- [161] Munk C. (2013). *Financial asset pricing theory*. OUP Oxford.
- [162] Muzy J. F., Delour J., and Bacry E. (2000). Modelling fluctuations of financial time series: from cascade process to stochastic volatility model. *The European Physical Journal B-Condensed Matter and Complex Systems*, 17, 537-548.
- [163] Nelson D.B. (1991). Conditional heteroscedasticity in asset returns: A new approach. *Econometrica: Journal of the Econometric Society*, 347-370.
- [164] Ni H., Szpruch L., Wiese M., Liao S. and Xiao B. (2020). Conditional Sig-Wasserstein GANs for Time Series Generation. *arXiv preprint arXiv:2006.05421*.
- [165] Oksendal B., and Zhang T.S. (1993). The stochastic Volterra equation. In *Barcelona Seminar on Stochastic Analysis: St. Feliu de Guíxols, 1991* (pp. 168-202). Birkhäuser Basel.
- [166] Oksendal B. (2013). *Stochastic differential equations: an introduction with applications*. Springer Science & Business Media.
- [167] Pardoux É., and Protter P. (1990). Stochastic Volterra equations with anticipating coefficients. *The Annals of Probability*, 18(4), 1635-1655.
- [168] Parmigiani G., and Inoue L. (2009). *Decision theory: Principles and approaches*. John Wiley & Sons.
- [169] Plerou V., Gopikrishnan P., Amaral L.A.N., Meyer M., and Stanley H.E. (1999). Scaling of the distribution of price fluctuations of individual companies. *Physical review e*, 60(6), 6519.
- [170] Poon S. H., and Granger C.W.J. (2003). Forecasting volatility in financial markets: A review. *Journal of economic literature*, 41(2), 478-539.
- [171] Pelger M. (2019). Large-dimensional factor modeling based on high-frequency observations. *Journal of Econometrics* 208.1 (2019): 23-42.
- [172] Poterba J. M., and Summers L. H. (1988). Mean reversion in stock prices: Evidence and implications. *Journal of financial economics*, 22(1), 27-59.
- [173] Potluru V. K., Borrajo D., Coletta A., *et al.* (2023). Synthetic Data Applications in Finance. *arXiv preprint arXiv:2401.00081*.
- [174] Protter P. (1985). Volterra equations driven by semimartingales. *The Annals of Probability*, 13(2), 519-530.
- [175] Rajan U., Seru A., and Vig V. (2015). The failure of models that predict failure: Distance, incentives, and defaults. *Journal of financial economics*, 115(2), 237-260.
- [176] Richard E. (2018). L'Argent de Zola et le krach de l'Union générale.

- [177] Rizzato M., Wallart J., Geissler C., Morizet N., and Boumlaik N. (2023). Generative Adversarial Networks applied to synthetic financial scenarios generation. *Physica A: Statistical Mechanics and its Applications*, 623, 128899.
- [178] Rogers L.C.G. (2019). Things we think we know. Preprint, available at <https://www.skokholm.co.uk/lcgr/downloadable-papers>.
- [179] Roncalli T. (2011). Understanding the impact of weights constraints in portfolio theory. Available at SSRN 1761625.
- [180] Roncalli T. (2020). *Handbook of Financial Risk Management*. CRC Press.
- [181] Rosenbaum M., and Zhang J. (2021). Deep calibration of the quadratic rough Heston model. *arXiv preprint arXiv:2107.01611*.
- [182] Rosenbaum M., and Zhang J. (2022). On the universality of the volatility formation process: when machine learning and rough volatility agree. *arXiv preprint arXiv:2206.14114*.
- [183] Ross S. (1976), The arbitrage theory of capital pricing, *Journal of Economic Theory* 13, 341-360.
- [184] Samuelson P. A. (1965). Rational theory of warrant pricing. In *Henry P. McKean Jr. Selecta* (pp. 195-232). Cham: Springer International Publishing.
- [185] Samuelson P. A. (1965). Proof that Properly Anticipated Prices Fluctuate Randomly. *Industrial Management Review*, 6(2), 41-49.
- [186] Schöbel R., and Zhu J. (1999). Stochastic volatility with an OrnsteinUhlenbeck process: an extension. *Review of Finance*, 3(1), 23-46.
- [187] Scholes, M. S. (2000). Crisis and risk management. *American Economic Review*, 90(2), 17-21.
- [188] Scruggs J. T. (1998). Resolving the puzzling intertemporal relation between the market risk premium and conditional market variance: A twofactor approach. *The Journal of Finance*, 53(2), 575-603.
- [189] Serletis A., Rosenberg A.A. (2009). Mean reversion in the US stock market. *Chaos, Solitons & Fractals*, 40(4), 2007-2015.
- [190] Sharpe W.F. (1964). Capital asset prices: A theory of market equilibrium under conditions of risk. *The journal of finance*, 19(3), 425-442.
- [191] Stein E.M., and Stein J.C. (1991). Stock price distributions with stochastic volatility: an analytic approach. *The review of financial studies*, 4(4), 727-752.
- [192] Tasche D. (1999). Risk contributions and performance measurement. *Report of the Lehrstuhl für mathematische Statistik, TU München*.

- [193] Tegnér M., Poulsen R. (2018). Volatility is log-normal But not for the reason you think. *Risks*, 6(2), 46.
- [194] Usmani M., Adi S.H., Raza K., and Ali S.S.A. (2016). Stock market prediction using machine learning techniques. In 2016 *3rd international conference on computer and information sciences (ICCOINS)* (pp. 322-327). IEEE.
- [195] Van Vliet P., Blitz D., and Bart van der Grient B. (2011). Is the relation between volatility and expected stock returns positive, flat or negative?. Flat or Negative.
- [196] Wang Z., and Oates T. (2015). Imaging time-series to improve classification and imputation. In *Twenty-Fourth International Joint Conference on Artificial Intelligence*.
- [197] Wei Y., Wang Y., and Huang D. (2010). Forecasting crude oil market volatility: Further evidence using GARCH-class models. *Energy Economics*, 32(6), 1477-1484.
- [198] Wu P., Muzy J.F., and Bacry E. (2022). From rough to multifractal volatility: The log S-fBM model. *Physica A: Statistical Mechanics and its Applications*, 604, 127919.
- [199] Yang M. (2004). Normal log-normal mixture: Leptokurtosis, skewness and applications. *Econometric Society*.
- [200] Zakoian J. M. (1994). Threshold heteroskedastic models. *Journal of Economic Dynamics and control*, 18(5), 931-955.
- [201] Zeevi A., and Mashal R. (2002). Beyond correlation: Extreme co-movements between financial assets. Available at *SSRN* 317122.
- [202] Zola É. (1891). *L'Argent*. Charpentier.
- [203] Zumbach G., and Lynch P. (2001). Heterogeneous Volatility Cascade in Financial Markets. *Physica A: Statistical Mechanics and its Applications*, 298(3-4), 521-529.
- [204] Zumbach G. (2009). Time Reversal Invariance in Finance. *Quantitative Finance*, 9(5), 505-515.
- [205] Zumbach G. (2010), Volatility Conditional on Price Trends. *Quantitative Finance* 10.4 (2010): 431-442.