

Conditioning the Information in Asset Pricing

Carlo Sala

Submitted for the degree of Ph.D. in Economics

Swiss Finance Institute

Università della Svizzera Italiana (USI), Switzerland

swiss:finance:institute



Advisor: *Prof. Barone Adesi Giovanni. Swiss Finance Institute at USI*

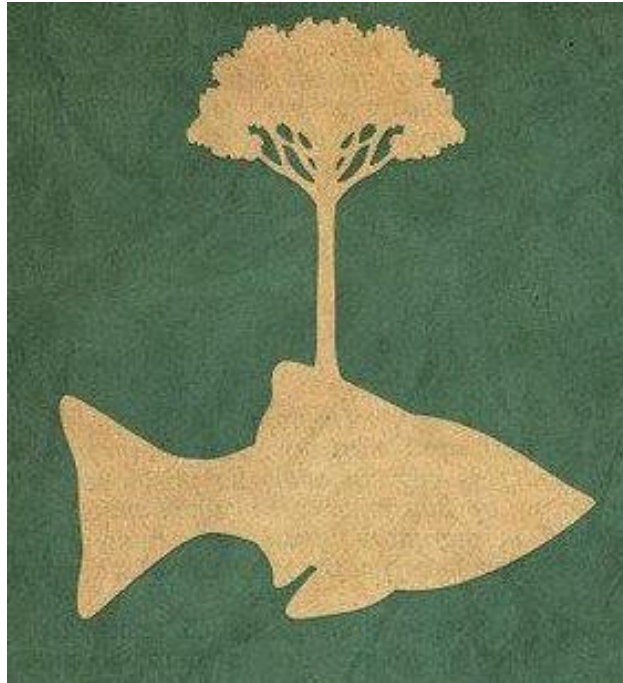
June 1, 2016

I certify that except where due acknowledgment has been given, the work presented in this thesis is that of the author alone; the work has not been submitted previously, in whole or in part, to qualify for any other academic award; and the content of the thesis is the result of work which has been carried out since the official commencement date of the approved research program.

Carlo Sala
Lugano, June 1, 2016

Acknowledgments

These graduating years have been very busy, exciting and rewarding. I would love to thank first and foremost my advisor, Prof. Barone Adesi for supporting and helping me since the 2009. Thank you for all your attention and assistance, I have thoroughly enjoyed and benefited from being a student of yours. A very special thank also goes to Prof. Mira for our inspiring and insightful initial discussions on how to estimate the pricing kernel from a Bayesian perspective and for her support during the past four years, and to Prof. Mancini for helping me with the MatLab code (this really helped me a lot!) and with his suggestions during the PhD and the job market. I also thank the members of my PhD committee, Professors Mele and Platen, for their helpful career advice and for their great suggestions about my thesis. Last but not least, I want to thank all professors and colleagues of the institute of finance. I've learned from you every day.



“Everybody is a genius. But if you judge a fish by its ability to climb a tree, it will live its whole life believing that it is stupid.”

Albert Einstein

Contents

Table of Contents	vi
List of Tables	vii
List of Tables	vii
List of Figures	viii
List of Figures	ix
1 Extended Abstract	1
2 The Measure's Problem, the Pricing Kernel and its Puzzles	7
2.1 The physical and the risk-neutral measures in finance	8
2.1.1 The pricing kernel	11
2.2 Introduction to the pricing kernel puzzle	12
2.2.1 More on the measures' problem	14
2.3 Joining the measures	19
3 Theoretical Set Up And The Literature So Far	22
3.1 The empirical pricing kernel (EPK)	23
3.1.1 PK as the conditional Radon-Nikodym derivative	24
3.1.2 The Riesz representation theorem	27
3.1.3 Macroeconomic derivation	28
3.1.4 No-arbitrage derivation	31
3.2 The investor's risk aversion	33
3.3 The literature so far	35

4	Conditioning for the information in asset pricing	38
4.1	Information and market efficiency	39
4.2	The importance of the Information in Asset Pricing	43
4.2.1	Econometric problems behind the estimation of the real-world measure	46
4.2.2	Defining the financial pricing kernel as a Radon-Nikodym derivative	50
4.3	The behaviours of local and strict local martingales in finance	52
4.3.1	Projecting the Radon-Nikodym derivative onto a smaller filtration set	54
4.3.2	Filtration shrinkage and information inaccessibility for strict local martingales	57
4.4	Empirical Application	59
4.4.1	Extending the findings to the risk-neutral measure	61
4.5	Fully and partially dynamic models, a comparison	63
4.5.1	A conditional version of the basic consumption based model	64
4.6	From consumption to portfolio optimization problems, the conditional case	67
4.6.1	Presenting the problem	67
4.6.2	The Lévy-Itô model and the portfolio optimization problem	69
4.6.3	The theoretical optimal choice	73
4.6.4	The real-world optimal choice	75
4.6.5	The power utility case	79
4.6.6	Connection with the PK	81
4.6.7	Expressing the information premium as the Kullback-Leibler divergence	83
4.7	Future ideas	87
5	The Methodology - part I	89
5.1	GJR GARCH-FHS and Monte Carlo simulation	90
5.1.1	The estimation of the physical parameters	90
5.1.2	The estimation of the risk-neutral parameters	92
5.2	The main limitation: the incompleteness of the physical measure	94
6	The Bayesian approach	95
6.1	Why a Bayesian non-parametric proposal?	96
6.2	The Dirichlet Process	97
6.2.1	Posterior distribution	99
6.2.2	The DP for the physical estimation	100
6.2.3	The precision parameter	101

6.3	Possible extensions	102
7	Making the Physical Measure Forward Looking	104
7.1	The Methodology - part II	104
7.1.1	The equity risk premium adjustments	105
7.2	A frequentist justification of the physical measure correction	108
8	The Datasets	112
8.1	The S&P 500 Index Options - The SPX	112
8.2	The S&P 500 Index	115
8.3	The risk-free rate and the dividend-yield	115
9	Empirical Results	117
9.1	The classical and the revised EPK, a comparison	118
9.1.1	Daily estimates	120
9.1.2	Linking the precision parameter to the options market	122
9.1.3	Option moments and the conditional physical measure	125
9.1.4	Focus on a single time t and time-to-maturity τ	129
9.2	Robustness checks	137
9.2.1	GJR GARCH robustness check	139
A	Appendix	141
A.1	Absolute Continuity and the Radon-Nikodym Derivative	141
A.2	Approximating to the mean and the variance of a ratio using Taylor	142
A.3	The Riesz Representation Theorem	145
A.4	Properties of Lévy processes	147
A.5	Costraints for stationarity	149
A.6	Consistency of DP marginals	150
A.7	Moments of DP	152
A.8	DP posterior moments	153
A.9	Derivation of the expected rate of return of investing in contingent claims	154
	Bibliography	155
	Tables	167

CONTENTS

vi

Figures

176

List of Tables

A.1	Summary statistics of the S&P 500 options index (SPX)	167
A.2	Summary statistics of the 1988-2002, 2002-2004 S&P 500 Index prices and returns .	168
A.3	EPK horizontal support: summary statistics so far	169
A.4	Summary statistics of the precision parameter	169
A.5	GJR GARCH physical parameters - θ	170
A.6	GJR GARCH risk neutral parameters - $\tilde{\theta}$ - Simplex method	171
A.7	GJR GARCH risk-neutral parameters - $\tilde{\theta}$ - Quasi-Newton	172
A.8	Summary statistics of $p_{t,T}^\dagger$, $p_{t-\Delta_t,T}$ and $q_{t,T}$	173
A.9	Yearly pointwise difference between $q_{t,T} - p_{t,T}^\dagger$ and $q_{t,T} - p_{t-\Delta_t,T}$ with FHS	174
A.10	GJR GARCH Risk Neutral parameters - Robustness Check	175

List of Figures

A.1 Measurable partitions of Ω	150
A.3 Intuition behind the model	177
A.4 1988-2002, 2002-2004 time series of S&P 500 daily closing prices and volumes	178
A.5 2002-2004 weekly S&P 500 Index closing prices (Wednesday only) and 1988-2002 daily log-returns	179
A.6 1988-2002 daily S&P 500 squared log returns and FHS	180
A.7 2002-2004 daily S&P 500 log and squared log returns	181
A.8 1988-2002, 2002-2004 Autocorrelation (ACF) of S&P 500 log and squared log returns	182
A.9 1988-2002, 2002-2004 S&P 500 log returns empirical and normal histograms	183
A.10 1988-2002 and 2002-2004 Normal and QQ plots of S&P 500 log returns	184
A.11 Single day ($t = 61$) all range of times-to-maturity probability density functions and Conditional EPKs with decreasing $\alpha_{t,T}^*$ and R.P.= 4%	185
A.12 Single day ($t = 3$) all range of times-to-maturity PDFs, Conditional EPKs and partially-conditional EPKs with decreasing $\alpha_{t,T}^*$ and R.P.= 8%	186
A.13 Time series of DOTM (OTM) put/call, total amount and relative $\alpha_{t,T}^*$	187
A.14 Daily Δ_t moments between fully and partially-conditional physical measure with respect to the risk-neutral measure for short time-to-maturity ($\tau < 60$)	188
A.15 Single day ($t = 4$), short and medium times-to-maturity ($\tau = 24/57/82$) left tails with decreasing $\alpha_{t,T}^*$ and R.P.= 4%	189
A.16 Single day ($t = 4$), medium and long times-to-maturity ($\tau = 150/241/332$) left tails with decreasing $\alpha_{t,T}^*$ and R.P.= 4% and 8%.	190
A.17 Single day ($t = 4$), short and medium times-to-maturity ($\tau = 24/57/82$) right tails with decreasing $\alpha_{t,T}^*$ and R.P.= 4%	191

A.18 Single day ($t = 4$), medium and long times-to-maturity ($\tau = 150/241/332$) right tails with decreasing $\alpha_{t,T}^*$ and R.P.= 4% and 8%	192
A.19 Daily log Δ_t moments	193
A.20 2002-2003-2004 yearly estimates of the FHS fully and partially-conditional physical measure and risk-neutral measure with $\alpha_{t,T}^* = 2$ and R.P.= 4%	194
A.21 Single day ($t=90$) two times-to-maturity ($\tau = 31$ and $\tau = 94$) conditional and partially-conditional EPKs, with $\alpha_{t,T}^*=1.75$ and 0.75 and R.P.=4%	195
A.22 Single day ($t = 90$) all range of times-to-maturity fully and partially-conditional expected return with $\alpha_{t,t+31}^* = 1.75$ and decreasing and R.P.=4%	196
A.23 Single day ($t = 63$) all range of time-to-maturity fully and partially-conditional expected return with $\alpha_{t,t+31}^* = 2.5$ and decreasing and R.P.=8%	197
A.24 Single day ($t = 41$) all range of time-to-maturity PDFs (left column), Conditional EPKs ($M_{t,t+\tau}^\dagger$) and partially-conditional EPKs($M_{t-\Delta_t,t+\tau}$) (right column) with $\alpha_{t,T}^* = 2.5$ and decreasing, 50.000 simulations and R.P.=8%	198
A.25 Single day ($t = 41$) with focus on the single time-to-maturity ($\tau_s= 38/73/164/255/346$) Conditional EPKs ($M_{t,t+\tau}^\dagger$) and partially-conditional EPKs ($M_{t-\Delta_t,t+\tau}$) with decreasing $\alpha_{t,T}^* = 2.5$ and decreasing, 50.000 simulations and R.P.=8%	199
A.26 Single days ($t = 41, 90$) single time-to-maturity conditional, $M_{t,t+\tau}$, and unconditional, $\widehat{M}_{t-\Delta_t,t+\tau}$, EPKs with $\alpha_{t,t+\tau}^* = 1.5$	200
A.27 2002-2003-2004 yearly estimates of the FHS (green) and Gauss. (green) Conditional EPKs ($M_{t,t+\tau}^\dagger$) and FHS (black) and Gauss. (blue) partially-conditional EPK ($M_{t-\Delta_t,t+\tau}$) with $\alpha_{t,T}^* = 1$ and R.P.= 4%	201
A.28 Ljung-Box test for 2000/3500/5000/9818 observations - normal returns	202
A.29 Ljung-Box test for 2000/3500/5000/9818 observations - squared returns	203
A.30 Ljung-Box test for 2000/3500/5000/9818 observations - absolute returns	204
A.31 Ten legs Lagrange Multiplier ARCH test for 2000/3500/5000/9818 observations	205
A.32 GJR GARCH parameters (θ) robustness check	206
A.33 Pointwise daily parameters (θ) difference	207

CHAPTER 1

Extended Abstract

Estimating the market's subjective distribution of future returns by means of backward looking historical data leads to uninformative and, at best, partially-conditional measures. What is missing are the investors' forward looking beliefs. This long-lasting problem affects a huge amount of literature and leads to puzzles and suboptimal results.

The goal of this thesis is to propose a new flexible and model-free methodology for the estimation of the conditional physical measure which is then used to investigate different empirical and theoretical applications in financial economics.

Differently than what is required by the general theory, is the scarcity and not the completeness of information the norm for a financial investor in the real world. Recalling that the units of measure of an investor that uses real data are the \mathbb{P} -measure and the relative real-world probability, this thesis builds a bridge between what is theoretically required by fundamental theorems of asset pricing and what is instead effectively achievable by an econometrician that uses real data. If the latter is not aligned with the theory, it may lead to unexisting puzzles and inefficiencies. This may be particularly true from the viewpoint of the types of information used in estimation, a characteristic to often violated in practice. Furthermore, as a second step, this thesis analyzes how the impacts of this discrepancy may spread over the risk neutral valuation theory.

Given the above research questions it emerges that the main limitation of the approaches present in literature lies in the assumption that all the information about the real-world probabilities of future returns can be fully extrapolated from their historical record. This assumption may lead

to large errors, that can be reduced if significant information on the physical distribution of security returns is available from other sources. Arguing that the majority of the econometric models present in literature are not able to exploit at its best the abundance of information presents in the market, and knowing that the all the investor's relevant and necessary information can be found in the financial markets, I solve this informational issue in a very natural way: letting the data from different sources of information speak at the utmost. To be sure to capture the highest amount of information embedded in different asset prices, I flexibly bridge in a new way two strands of the neoclassical literature: the one related to the risk-neutral distribution, extracted from the cross-section of option data, and the one related to the physical distribution extracted from the time series of stock returns. The innovation of the model is that the two measures, once estimated independently, are then blended into a new one to obtain a fully homogeneous ratio of the two. Homogeneity focuses on the degree of conditionality of the information of the two measures, a characteristic often violated in literature. A natural approach to exploit simultaneously and flexibility multiple data and provide statistical inference is the Dirichlet process (Ferguson (1973)[69]). Using the precision parameter of the Dirichlet process as a proxy for the missing information, I calibrate it with respect to the daily liquidity of the option market. The revised measure, exploiting its high flexibility and being free of any constraints - but the ones required by the Fundamental Theorems of Asset Pricing (henceforth: FTAP) - is the best candidate to test and answers to puzzles concerning the consistence of the distribution implicit in option prices and the time series properties of the underlying asset prices (Bates (1996)[19]). The most convenient way to test for the consistency of the proposed measure is to use it for the estimation of the pricing kernel (henceforth: PK).

Defined as a ratio of state prices per unit probabilities, the neoclassical theory requires the PK to be a monotonically decreasing function. This follows naturally from the definition of the economical marginal rate of substitution. The PK is in fact the financial counterpart of the marginal rate of substitution where different financial assets proxy consumption. Counterintuitively with the theory, recent empirical studies found several violations on different parts of the PK: at the extremes as well as in the central part of the functional. Since Jackwerth (2000)[93] and Rosenberg and Engle (2002)[64], this is known as the "pricing kernel puzzle"¹.

While the risk-neutral moments extracted from option surfaces are by construction forward looking, the ones inferred from historical returns are only partially informative, thus suboptimal with

¹Under the neoclassical setting a non puzzling PK must be non-increasing in wealth. Projecting the PK onto market returns, a non-decreasing PK leads to the puzzling existence of a contingent claim that stochastically dominates the market index (see Dybvig (1988)[58] for the main theory and Beare (2011)[20], Beare and Schmidt (2015)[21] for possible extensions).

respect to investors' future beliefs. This underestimation of the physical filtration produces a misalignment with respect to the full conditioning of the information thus violating the neoclassical theory. Empirically it turns out that most of papers present in literature are then affected by a non-homogeneity bias.

The ultimate goal of finance is how to determine the risk-return trade-off. To achieve it, all of asset pricing comes down to two basic principles: first, any asset value is nothing but an expected discounted payoff; second, a dollar when there is scarcity of money provides higher utility than the same dollar when there is plenty of money around. Therefore, states that provide one dollar when money is scarce have today higher prices. Estimating how the latter principle - that is, how to determine the PK - enters into the former, is the above cited ultimate goal of finance.

While the introduction of the risk-neutral probabilities and their use in asset pricing becomes a triumph of the modern financial community, theory has left us without a way to recover both the real-world probability distribution and the PK that enter into the fundamental pricing equation. The estimation of the PK and its constituents is then of paramount importance in financial economics.

My line of research is motivated by the novel recovery theorem of Ross (2015)[140] that boosted the attention on the economic and econometric issues related to the estimation of the market's subjective assessment of the real-world probabilities and by the empirical studies of the PK which, starting from Jackwerth (2000)[93], produced conflicting and often puzzling results.

Given the central role of the real-world probabilities for many financial operations, the proposed problem naturally embraces different theoretical and empirical fields of financial economics. Although wide in principle, I solve the problem econometrically with a focus on the proper modellisation of the information present in the market. Throughout the paper, I explore and exploit three main interconnected macro areas, namely: the higher informative content provided by the joint use of stocks and options data with respect to just using historical stock returns, the econometrical difficulties encountered in the estimation of a fully-conditional and time-varying physical measure and the non-homogeneity bias that affects the two measures thus, by consequence, the PK.

First: options are by construction informative and forward looking financial assets. Since the seminal paper of Arrow it is known that the necessary information about investor's present and future beliefs are encoded in these assets. Along with their higher availability on the market, the attention of academics and practitioners on the predictive content of option data is increased substantially

over the last few years. Options get even more informative if combined with other sources of information, i.e.: stock data. Chernov and Ghysel (2000)[42], Pastorello, Renault and Touzi (2000) [128], Polson and Stroud (2003) [133] and Eraker (2004) [65] demonstrate how a unified frameworks that incorporates the information of option surfaces and stock returns assures a higher degree of statistical precision in the estimation of the parameters than only using historical stock data. Since then, the literature relative to the joint estimation techniques has found considerable attention. Nevertheless, I claim that the overall informative power, presents naturally in the market, has not yet been exploited at its best for the estimation of the physical measure. Starting from this point and knowing that a possible reason for the PK non-monotonicity is the lack of investors' forward-looking information, I propose a new methodology that, updating its values given the observations in the market, joins naturally and flexibly the risk-neutral and physical measures. Tested on a daily basis and for different times-to-maturity, the model allows us to show how option prices mirror the missing investors' risk preferences, errors and beliefs.

Second: the literature offers different parametric and non-parametric methodologies to estimate the physical distribution. As a common denominator, almost all of them are based upon a stream of backward-looking stock returns, hence of past information thus ignoring the investors' forward looking beliefs. Encoding most of the risk appetite of the investment, the obtained measure is then, at best, only partially-conditional thus not fully informative². Enlarging the time length of the estimation leads only apparently to more conditional measures. The last bite of information, even if would be informative of the future, is in fact washed out by the huge amount of information used in estimation. Conditionality is of key importance with respect to today volatility and higher moments. I estimate them by means of an GJR GARCH - FHS model.

As a direct consequence of these two points, I find that the vast majority of papers that investigate empirically the PK properties compare a fully-conditional numerator extracted from options data with an unconditional or, at best, a partially-conditional denominator extracted from past stock returns. The disalignment between a backward-looking, hence uninformative, historical measure with respect to a naturally forward-looking risk-neutral measure that is able to express the rich admixture of the instors beliefs about the present and the future, loads heavily onto the relative PK. Market microstructures, possible small sample biases (Leisen (2015)[109]), the natural extra fragility intrinsic in any ratio of measures (Jobson and Korkie (1980)[101]; Sala and Barone-Adesi (2015)[145]) and the frequency of the estimation can further amplify the misestimation. As a direct

²Among others, Jackwerth and Brown (2012)[37], Ziegler (2007)[159] and Beare (2011)[20] point out this issue but remain silent with respect to a possible resolution.

consequence, the obtained functional is non-homogeneous with respect to the conditionality of the information. The consequent results are then invariably spurious and misleading.

I test my model on US index and index options data. The obtained empirical results are robust and outperform other methodologies. Interesting results are achieved in periods of strong market up/down turns, when the market movements produce high liquidity in the option market and the impact of many day-by-day operations is amplified. Given my results, I claim that some of the puzzles present in literature are due to an econometric bias with respect to an improper modelling of the information.

My findings are of values for academics, central bankers and other decision takers who wish to infer market beliefs about future distributions from traded asset prices. They also provide a way of testing the rationality of option prices. The implications of my research are likely to be relevant for risk management³, regulation and financial policy. Asset management will also benefit from improved understanding of the PK (i.e. Sala and Barone-Adesi (2015)[144]).

Moving from the empirical to a more theoretical viewpoint, I show how the suboptimal use of a too small information set may impact strongly onto the nature and on the properties of the functional governing the empirical PK and its constituents. In turn these affect also the stability of the fundamental theorems of asset pricing. More specifically I show how, projecting the empirical pricing kernel onto a too small filtration set, the missing information is translated into jumps. Under this framework in fact the process passes from being absolutely continuous - as required by the theory - to being absolutely continuous with jumps. Given the strong and unique interconnectivity among the measures and the PK, this change of nature of the PK impacts strongly onto the validity of the risk neutral measure that is considered a true probabilistic measure only if the relative measure is a true martingale process. Following Jarrow and Larsson (2012)[98] I propose a model-free way to test the market efficiency.

Finally, both for the discrete and the continuous cases I show, by means of a portfolio optimization, how a small and rational investor who wants to maximize her wealth by trading in simple economy made of a risk-less and a risky assets, achieves smaller profits if uses suboptimally the market's information. What lowers the investor's profits is the information premium, which (for the continuous case) is nothing but the Kullback-Leibler distance between the optimal and the sub-optimal PKs. The proposed models can be of interest for asset managers. The case of a financial modeller that has less than the required amount of information for pricing is in fact much closer to

³Kelly and Jiang (2014)[105] propose a new measure of tail risk but from cross sections of stock returns only

the reality than the optimal amount case.

The thesis is so organized: after a brief introduction of the above cited problems, chapters one and two present a theoretical review of the estimation of the real-world measure, of the pricing kernel and the relative empirical and theoretical violations of its misestimation, i.e.: the pricing kernel puzzle. The remaining chapters, which are my main contributions to the literature, propose the new estimation technique and its theoretical and empirical applications.

CHAPTER 2

The Measure's Problem, the Pricing Kernel and its Puzzles

In this chapter I introduce the physical measure's problems, the financial pricing kernel (henceforth: PK) and its main violations, also known in literature as PK puzzles.

The chapter starts presenting the risk-physical and the risk-neutral measures, their use in financial economics and how they are estimated. Focusing on the importance of the use of the information for the estimation of these quantities, I present some of the main theoretical and empirical pitfalls present in literature. Most of these problems will then be recalled in the subsequent chapters, where I will show how to possibly overcome these issues.

Although the PK is a byproduct of the two densities and a discount factor, I focus most of the attention on the issues that affect the risk-physical measure⁴ with a particular emphasis on the main theoretical and then econometrical problems present in literature. As it will be made out clear along this chapter and all over the thesis, it is in fact the real-world measure the one more prone to have informational biases. Accordingly I focus the attention on how a suboptimal physical measure may impact strongly on the estimation of the PK.

Finally I introduce how the joint use of different asset prices may provide a solution for these information issues. Along the chapter it will emerge how the main problem between most of puzzles present in literature is not the econometric model used, but instead the inputs used in the econometric model. As long as the input used cannot be informative (by structure), there cannot in fact exist any model that could fix the issue in a natural and valid way.

⁴The abundant financial literature describes the above mentioned "physical" measure with many different names, i.e.: "real-world", "risk-adjusted", "historical", "statistical", "actual", "objective" or "subjective" depending on the framework.

2.1 The physical and the risk-neutral measures in finance

Although I will focus the attention on the physical measure, its definition is somehow blurred and it will be understood more easily by defining the risk-neutral measure first. Justified by the three fundamental theorems of asset pricing and defined under some technical requirements, the risk-neutral measure is a fully theoretical probability measure built on the assumption that the today value of an unknown financial asset is equal to its future expected payoffs discounted at the today risk-free rate. Given this framework, as a main advantage, all investors are neutral with respect to all exogenous and endogenous risk. This strong but economically accepted assumption makes many pricing tasks incredibly simpler.

Unlike the extremely powerful but somehow fictitious risk-neutral measure, the real-world probability of future states⁵ takes into consideration also the investor's subjective beliefs thus assuming the eventual existence of a subjective risk premium. Dealing with the often intangible past, present and future subjective beliefs of the investors, the real-world measure estimation is by construction a not a trivial task to perform. This econometrical issue is what made and makes the risk neutral measure appealing, both in academia and in the industry thus leaving the theoretical and the empirical literature mostly silent on how to estimate the less straightforward but more correct real-world one.

Estimating the real-world measure, the main limitation of the different approaches present in literature is on the assumption that all the information about the real-world probabilities of future returns, can be fully extrapolated from the historical record. This assumption may lead to large errors, that can be reduced if significant information on the physical distribution of security returns is available from other sources.

Although the literature is almost silent on its theoretical and empirical estimation, the use of the real-world measure in finance is huge. Given the goal of the thesis I will only focus on those papers that use the real-world measure as a "tool" for the estimation of the PK.

Jackwerth (2000)[93] and Jackwerth and Brown (2013)[37] propose a risk-adjusted non-parametric kernel density estimation over 48 non-overlapping monthly returns. Ait Sahalia and Lo (1998)[3] use the same technique but using 1,008 overlapping daily returns. Even if markets would be strongly efficient and fully transparent, so that the last value of the time series would be fully representative and informative with respect to the investors beliefs, the huge amount of data used by these tech-

⁵The future state can be linked to any financial risky action whose unknown outcome can be computed by means of a conditional expectation i.e.: buy a stock, start an investment, enter into a pension plan and so on.

niques washes out its impact thus making the final result totally uninformative with respect to the future views of the market. It follows that both methods are essentially unconditional with respect to the today observations thus providing the least conditional approach to estimate the real-world future probabilities. A more sophisticated technology is present in Christoffersen, Heston and Jacobs (2013)[45] where the physical parameters are extracted from 240 monthly returns by optimizing the Heston and Nandi GARCH model with filtered historical innovations. The methodology is in line with Rosenberg and Engle (2002)[64] and Barone-Adesi, Engle and Mancini (2008)[15] which extract the real-world probabilities through an asymmetric GJR-GARCH (1,1) with empirical innovations over a past stream of daily observations. Although these GARCH techniques improve the degree of conditionality of the estimation, given the structure of the underlying used - stock returns - a full inference as well as a full recovery of the investors future beliefs is impossible without using more informative dataset. Bliss and Panigirtzoglou (2004)[28] relies on parametric PKs to extract the real-world densities from the risk-neutral ones. Unfortunately modelling the PK parametrically may lead to substantial errors in estimation. An antithetic approach is Ross (2015)[140] which uses only options data to infer the real-world probabilities. Among the other limitations, the Ross recovery theorem is inconsistent with some evidences from the stochastic volatility modeling literature - e.g. Bollerslev, Chou, and Kroner (1992)[30] - indicating that future state probabilities depend more on the recent events than long-ago events, but that long-ago events still have some predictive power.

Christoffersen et al.(2012)[47] and Bollerslev and Todorov(2011)[31] confirm that most of the risk and risk pricing information of the underlying asset can be extracted from derivatives product. Once adjusted for a risk premium⁶, the risk-neutral measure extracted from the option surface may reflect all the information publicly available to investors thus providing most of the missing information. Given this powerful feature of the option surface, my approach is to solve the presented problem empirically and in a very natural way: letting the market data speak as much as possible. As a consequence I go for a novel intermediate approach that uses both historical and forward-looking data to provide a time-varying and fully-conditional real-world measure. As a starting

⁶The risk premium adjustment is needed for consistency with respect to the objective density. The adjustment however implies that state prices are function of index returns only. In the presence of other priced risk, such as the variance risk premium discussed by Heston and Nandi (2000)[87], or higher premium like the skewness or kurtosis risk premium it is interesting to investigate whether better objective estimates may be obtained rescaling the risk-neutral distribution to remove the volatility premium. However, Chorro et al. (2012)[43] suggest that the improvement induced by the variance risk premium, should be marginal. That should hold even more if parameters are re-calibrated.

point and benchmark I use the already advanced econometric model of Barone-Adesi, Engle and Mancini (2008) and I improve it toward its full conditionality. This already high starting point makes my exercise more complex as well as more valuable for the literature.

Econometrically, the main tools needed in estimation are: the GJR-GARCH-FHS model to estimate the daily moments and the Dirichlet process for mixing the measures.

I do not plan to modify the numerator, that is the risk-neutral density. In fact that would imply inefficiencies in the option market that, although rare given my dataset, are of course possible but beyond the scope of this thesis.

If from one side the use of the risk-neutral measure to extract the entire real-world probability is gaining attention (Ross (2015)[140]), the much more natural approach of using the risk-neutral density to improve the estimation of the physical density is not common in the finance literature. As Ross (2015)[140] shows, under particular assumptions, it is possible to recover the entire physical density from the risk-neutral one. Given its structure, the model is not affected by the homogeneity-bias but, the price to pay to obtain a complete recovery of the entire physical density from the risk-neutral one is pretty high in terms of assumptions needed. A full recovery is in fact possible only in a special case, namely when the state transition matrix has full rank, the underlying state vector is bounded and the utility function is state-independent. Option prices and investors in practice are unlikely to satisfy these requirements thus confirming Borovicka, Hansen and Scheinkman (2015)[33]. Carr and Yu (2012)[40] and Audrino, Huitema and Ludwig (2015)[9] propose possible extensions to relax some of the initial assumptions of the Ross recovery theorem. Differently with respect to these papers I do not extract the entire physical measure from the risk-neutral one but I only use it to complete the objective. As a primary advantage, I don't need to put any structure neither on the functional form of the PK, nor on the preferences of the investors. It is reasonable to conjecture that limited departures from the conditions that ensure full recovery may still lead to the possibility of using the option surface to improve significantly the assessment of the physical distribution. My conjecture is supported by the widespread use of physical probabilities based on option prices in the business community. These probabilities are generally based on the Black-Scholes model, in spite of its inability to fit well empirical option prices. It appears therefore that the usefulness of option prices to predict physical probabilities is widely recognized and it is quite a common practice in the business community.

2.1.1 The pricing kernel

Defined as a discounted ratio of state prices per unit of objective probability, the neoclassical theory requires the PK to be a monotonically decreasing function in wealth. Acting among them, the PK is what links the risk-neutral probability to the real-world one. How the information flows between the two worlds is thus determined by the PK and viceversa if one uses the measures themselves to determine the PK. Mathematically, the PK is nothing but a discounted operator that allows us to move between the two measures. It follows that a correct PK must convey all the present and future expectations, beliefs and errors of the investor. In a continuous world with no-arbitrage⁷, the today PK, $(M_{t,T})$, is defined as the Radon-Nikodym derivative of the risk-neutral measure with respect to the physical measure of security returns. The time window, represented with ${}_{t,T}$, underlies the forward looking nature of the variables and applies to all inputs of the PK thus to the PK itself. Given a finite economy, $0 < t < T < \infty$, if the two measures satisfy mild regularity conditions, the PK is defined as the present value of the ratio of the risk-neutral density of returns, $q_{t,T}(R)$, divided by the physical density, $p_{t,T}(R)$:

$$M_{t,T} = PV_t \left[\frac{q_{t,T}(R)}{p_{t,T}(R)} \right] \quad \forall t \in T \quad (2.1)$$

From the FTAP, equation (2.1) must be fully-conditional with respect to the time t expectation and the relative PK a strictly positive martingale process: $M_{t,T} > 0$. By no-arbitrage constraints, the ratio (2.1) is related to the expected gross return, $R_{t,T}$, from investing in simple state contingent claims:

$$R_{t,T} = \frac{1}{M_{t,T} \cdot r_t T} \quad (2.2)$$

where r_t is the gross risk-free rate and T is the chosen time interval. If the time interval is allowed to change, the PK becomes a stochastic process that can be used to price options of different maturities. Despite its high component of randomness, most of the parameters of the PK are chosen a priori in much of the existing literature. The consequences of these arbitrary choices are quite dramatic, especially for pricing states far away from the current state of the market. A deterministic PK may lead to severe mispricing and uneffective hedging strategies. It follows that a proper modellisation of the PK can only be achieved by means of a time-varying stochastic model.

⁷Requiring the existence of no-arbitrage I am implicitly assuming the existence and validity of the free portfolio formation (FPF) and the law of one price (LOP) or, equivalently, the existence of a linear subset of the whole sample space and the linearity of the function.

2.2 Introduction to the pricing kernel puzzle

Despite its key role in asset pricing, there is still not a cut and clear agreement among financial researchers and practitioners on the best procedure to properly estimate the PK.

Broadly speaking, I refer to the PK puzzle when the estimated PK is not general enough to properly explain the whole cross section of option data ⁸. Graphically this is expressed as a non-monotonic decreasing function, as required by the neoclassical theory.

Depending on the completeness of the market, the non-monotonicity of the PK implies, for a complete market framework, the existence of a trading strategy in contingent claims that a.s. first-order stochastically dominates the underline returns (Dybvig (1988)[58]). In presence of market incompleteness, the monotonicity of the PK is the key ingredient for the literature that deals with the incomplete-market stochastic dominance option pricing bounds (Perrakis and Ryan (1984)[130], Levy (1985)[111] and Constantinides, Jackwerth and Perrakis (2002)[50]). Moreover, a misestimation of the PK is often directly related to the apparently over/underpricing of out-of-the money (OTM) and deeply out-of-the money (DOTM) options. Empirically the main problems usually arise at the extremes of the PK or in the area of zero returns where, by a flex, the obtained results violate the non-arbitrage arguments. The former issue is primarily caused by two problems of inadequacy: one theoretical and one empirical. The first is mainly due to a bad modellisation of extreme events so that is not possible to fully capture the rare but existing episodes that lies into the deepest tails of the distribution. As a consequence the model produces bad estimate at the extremes of the PK that may lead to apparent mispricing. The second one is possibly due to the low liquidity of deeply out-of-the-money put and (above all) call⁹ option which poses an important empirical challenge. In between there is the U-shaped PK puzzle which may be caused by the missing information from the call options or from the lack of options themselves.

The theoretical problems could be overcome by a better modelling, while the empirical ones could only be reduced by applying numerical artifacts or by setting strong a-priori model assumptions ¹⁰.

⁸In this thesis I refer primarily to index option prices but, with no loss of generality, the same conclusion can be extended to other financial asset classes.

⁹It is by now and empirically known fact that, mainly for hedging reasons, put options are more traded than call options.

¹⁰Being an empirical issue linked to the low tractability of the model inputs, due to the missing or low liquidity of the data, it might be possible to reduce the bias of the estimated final results by i.e.: imposing non-increasing/decreasing bounds in the tails of the final function, smoothing the extreme outcomes of the tails synthetically, producing a higher amount of data by simulation, fitting extreme distributions for deepest values and so on. Some of these corrections have been already applied in literature (see 3.3)

I will show that a better modelling can be achieved as long as the distribution of the innovations is properly chosen. The between problem is instead solvable by properly exploiting the market options information.

Also the puzzling results in the central area, which among the others may be due to an incorrect theoretical model or by a misestimation of the risk premium, can be alleviated through the presented model or by better calibrating the risk premium. It turns out that also the central violation is strongly connected to the apparent underpricing of call options. Due to its high flexibility and ability of being properly adapted to different market environments, the proposed model is able to answer to both problems.

The literature dealing with the PK estimation is huge. Concerning its violations in estimation, in literature is possible to find more than a dozen of different explanations regarding the existence or the non existence of the PK puzzle. The different approaches of [85]Heston (1993), Jackwerth (2000)[93], Heston and Nandi (2000)[87], Engle and Rosenberg (2002)[64], Bliss and Panigirtzoglou (2003)[28], Barone-Adesi, Engle and Mancini (2008, henceforth BEM) [15], Chabi-Yo (2008)[157], Chabi-Yo, Garcia and Renault (2012)[158], Christoffersen, Heston and Jacobs (2012)[44], Song and Xiou (2012) [149] to mention only some of the methods present in literature, lead to very different conclusions on security pricing and their related empirical puzzles. Lately, the debate has also been enriched by behavioral finance models, such as the ones proposed by Ziegler (2007)[159], Shefrin (2008)[148], Hens and Reichlin (2011)[84], and Barone-Adesi, Mancini and Shefrin (2013)[18].

No matter what is the story behind the paper, they all have a common strong drawback: the denominator of the PK - the physical measure - is unconditional, thus improperly estimated. This theoretical bias lead to dramatic empirical problems. It turns out that the main reasons behind the misestimation lies in the type of assets used for the estimation of the measure.

As noted by Bollerslev, Chou, and Kroner (1992)[30]: “future state probabilities depend more on the recent events than long-ago events, but long-ago events still have some predictive power. Misspecification of state probabilities induces error in the estimation of the PK since the denominator of the state-price-per unit probability is incorrectly measured”; hence, the good quality of final results, which depend by the past history, the today scenario and by the forward-looking data used in estimations, are heavily linked to the information content embedded into the measures used for the final estimations. Bollerslev et al. (1992)[32] propose an asymmetric GARCH model with empirical innovation densities to go around the problem. The proposed denominator, although improved, is still poor from an informational point of view thus producing poor PK estimates.

Following their insight and partially in line with Ross (2015)[140], in this thesis I propose a new

non-parametric methodology that I apply to the denominator to overcome the issue. By empirically mixing the two measure, I aim to improve the physical measure, hence the overall PK.

One of the main goal of this thesis is to focus on the joint informational content of option prices and stock prices with respect to stock prices alone. I do not attempt to answer the question, “Is the model I assume correct?” Instead, I test whether, under the model I present, inference based on option and stock prices can lead to better small-sample properties for the PK.

2.2.1 More on the measures' problem

Being the central concept of the thesis, in this section I stress and I go more in deep with respect to the measures' problems in the estimation of the empirical PK. The PK is a the discounted ratio of measures. It goes by consequence that the single measures and the PK are among them uniquely and tightly interconnected. As advantage, the knowledge of two of them automatically implies the knowledge of the third one. As a disadvantage, the misestimation of one of the two also loads naturally onto the third one. The literature is abundant of different methodologies, both parametric and nonparametric to estimate the two measures and the PK¹¹, either jointly or separately.

While theoretically the PK is a decreasing function of aggregate resources, many empirical papers found violations in different areas of the functionals (Aït Sahalia and Lo (1998)[3], Jackwerth (2000)[93], Brown and Jackwerth (2001)[37], Rosenberg and Engle (2002)[64], Yatchew and Härdle[156], Ziegler (2007)[159]). The persistency and robustness of these violations put interest on the investigation of the so-called pricing kernel puzzle. Since then, researchers have taken great interest in proposing different econometric techniques to estimate the PK and its measures trying to answer to the puzzle.

The Real-World Measure

From the literature it emerges that, among the others, the most problematic econometric task to perform is to assure full conditionality to the time-varying estimation of the market's subjective distribution of future returns.

Empirically it turns out that to propose a day-by-day estimation methodology of the measures that compose the PK is as much important as econometrically non-trivial. The importance of a correct estimation of these measures lies in the wide use of the PK for many daily operations (i.e.: asset pricing and risk management). The econometric issues, as also partially commented in Bliss

¹¹The nonparametric ones have to be preferred. From Fama (1965)[68] and Mandelbrot (1966)[118] it is well-known that none of these quantities are neither normal nor have a closed parametric form.

and Panigirtzoglou (2004)[28], lie in the nature of the underlying. Using historical data many estimation methodologies put unreal and theoretically not required stationary assumption on the estimation of the measures or on the PK itself. Therefore, it is not by chance that many works only propose monthly or yearly estimations remaining silent for the daily ones. It turns out that, from this viewpoint, the main reason behind a biased estimation is not the technique used for the estimation but rather the data used as input.

Needless to say, models' inputs are of key importance to determine the type and the quality of the final outputs. Market option prices provide a naturally forward-looking measure; in fact, by contract, the owner of an option has the right, but not the obligation, to exercise it at expiration (or before it, if it is an American or exotic option). This feature is reflected into the option value, which is in fact a non-decreasing function of volatility. Therefore, the market prices of options, through the implied volatilities, encode important forward-looking information about the future distribution of prices of the underlying asset. Also the higher moments of the distribution embed important information. This is particularly true into the tails: where lies most of sentiment.

The same richness cannot be achieved by stock and index prices. By their contractual nature these assets are options-free and unique, hence poorer from an informative viewpoint.

Estimating the PK and extracting the risk aversion from stock prices is a well-known problem in the literature. Despite their unambiguous superiority in estimation, it is only from the beginning of the millennium that scholars have begun using options data for estimations (Chernov and Ghysels (2000)[42]). The superiority in estimation of options with respect to stocks (and also futures) is manifold.

First, stock prices have discounting as well as time-horizon problems. By contract definition, stocks do not expire but live infinitely. They are defined over an indefinite time horizon; therefore the discounting process becomes non-trivial. As a consequence, additional assumptions (which often times are unreal, i.e.: on the characteristics of future dividends), are needed to determine the discounted cash flow. On top of that, the obtained final outcome is statistically not much informative being the discounted final cash flow just a single value; this means that no inference about variations in preferences over different time horizons is possible.

On the contrary, option contracts have by definition a bounded life that is defined by a fixed time-to-maturity, T , and is known from the inception of the contract¹². Moreover, for each time t , we

¹²Only perpetuity options differ from this characteristic, but are an exceptional case, more theoretical than really used in practice. American and exotic options instead, have the flexibility to be exercised at any time or at fixed

have the so-called option surface: a broad spectrum of times-to-maturity, T_i and strikes K_j ¹³ that covers different states of the world. These characteristics allow for natural inference on preferences over specific horizons and simultaneously over different horizons and strikes.

These features make options qualitatively superior also to futures and forward contracts which, by their nature, do not share the discounting but only the time-horizons problem. In fact, even though these contracts have finite maturities, they do not differentiate across states of the world, thus providing only a single statistic for each expiry date/observation date pair. As for the stocks, having a single data a direct density estimation is not possible without further assumptions¹⁴.

As a consequence, we can directly estimate a time-varying risk-neutral density from a cross section of options with no need to bound the structure of the data we sample from, but the same is not possible from the time series of the underlying. For the physical density in fact, to obtain good results when making inference from a time series of stock returns, we need to put a priori unrealistic bounds on the time structure of the data. In conclusion: inferring densities from the option surfaces does not share the above cited stationarity problems. While the degree of assumed stationarity can be of different length, in no case can be justified economically

Given the above characteristics, it is natural that just using historical stock data it is not possible to properly capture the investors future beliefs. The difficulty in estimating the objective measure is that it directly depends on the evolution of the underlying process, which time series is only partially informative. Estimations can be further complicated by possible data-problems (i.e.: data scarcity) and market frictions. For these reasons, some authors fully avoid the density estimation and propose a ratio of estimated risk-neutral measure over an unknown physical measure (Golubev et al. (2008))[77].

As (a non-)alternative, some papers propose to increase the degree of conditionality statistically by increasing the rolling window of the estimations. Jackwerth (2000)[93] tries to “enrich” the informative content of the physical measure by working on a longer time series. He proposes to use ten instead of the classical two to five years of data for estimation. As expectable, the obtained results are qualitatively almost unchanged. Due to the intrinsic backward nature of the dataset, to increase the length of the rolling window used in estimation is not a solution. At some extremes it might also lead to even more misleading results. In fact, by using more and more data, there is

time before T . Although non-trivial, a reliable discounting is still possible.

¹³Both indices are finite, i.e. $i = 1, \dots, M$ and $j = 1, \dots, N$.

¹⁴As noted by Jackwerth and Rubinstein (2006)[94], starting from at least 8 option prices we have enough information to determine the general shape of the implied distribution. Although possible, to have too a few data in estimation may lead to possibly dangerous small sample bias.

a natural statistical improvement coming from the higher conditionality of the measure but at the same time we reduce the overall informative content of the last data. Also in the case in which this single stock value would be fully informative with respect to the future scenario of the market, it would become almost negligible and prevailed by the high amount of past data used. Using time series of historical returns thus make the single values almost fully unconditional. Needless to say, the same problem is then naturally translated onto the relative estimations.

The Risk-Neutral Measure

While the real-world densities presents several difficulties, it is by now an empirical and theoretical fact that the most important information embedded in financial instruments is the state price density (SPD), or the Arrow-Debreu state prices.

The time-state preference model of Arrow (1964)[6] and Debreu (1959)[7], which proposes the now-named Arrow-Debreu security, models a very basic financial instrument (called pure or primitive security) that pays one unit of numeraire (like a currency or a commodity) on one specific state of nature and zero elsewhere. Passing from discrete to continuous states, Arrow-Debreu securities are defined by the so-called state price density (SPD). Under the continuous framework the security pays one unit of numeraire x if the state falls between x and $x + dx$ and zero elsewhere. As a consequence of their high informative content, the Arrow-Debreu securities become one of the key element to work with and understand the general economic equilibrium under uncertainty and to determine the price of any contingent claims. For these reasons the estimation of such SPDs has been a very important topic of research within the financial economics community. Once that a complete set of option prices for a specific time-to-maturity is available, there are many parametric and non-parametric methods to recover the risk-neutral measure¹⁵.

To summarize: to obtain a time-varying estimation of the real-world density it is of key importance, more than to pick the right extraction model, to use the right source of information. Being historical stock returns only backward looking, hence only partially informative, we need to somehow complete the measure by using other sources of information, i.e.: the investors' future sentiment extracted from the implied moments of the option surface.

Parametric methods define the functional form of the risk-neutral distribution which can then be defined by estimating the relative set of parameters.

¹⁵See Bahara (1997)[11] and Jackwerth (1999)[92] for a complete review of the topic.

These methods can be divided into three broad categories:

- Expansion methods: which correct and expand the basic distribution to make it more flexible;
- Generalized distribution methods: which add extra moments to generalize the basic normal and log-normal distributions;
- Mixture methods: which blend different distributions.

Although parametric methods have the advantage of being less computationally intensive and more precise once the parametrization is correctly done, it is well known that they do not perform well with financial market data ¹⁶. Any parametric misspecification leads to possibly highly inconsistent estimates which in finance can lead to extreme mispricing and to large uncovered risks. Even more, testing a fully parametric model is always a joint test of the model and the (arbitrary chosen) parameters. Changing the latter could lead to completely different outcomes.

Better results can be obtained via non-parametric methods. Not requiring any specific parametric form and thus achieving greater flexibility ¹⁷ in fitting the measure on option prices. Although not in principle, also non-parametric models have to follow some assumptions to better model the economy (i.e.: the no arbitrage principle.). In any case, these assumptions, leaving freedom to the final functional form to estimate are surely weaker than any parametric model and less likely to be violated in practice (see: Äit-Sahalia and Lo (1998)[3] and Äit-Sahalia and Duarte (2003)[2]). Non-parametric models are so classified:

- Kernel methods: which are comparable to regression methods but without specifying the parametric form of the function;
- Maximum entropy methods: which, satisfying some minor constraints, achieve the fit of the distribution by minimizing some specific loss function;
- Curve fitting methods: which are a broad class of methods where the objective of the estimation is approximated by some general function.

If the estimation of the risk-neutral measure has been largely studied with the production of satisfactory results the same is not true for the estimation of the physical measure.

¹⁶There are no generally accepted parametric forms of asset prices, volatility surfaces, or put/call price functions (Campbell, Lo and MacKinlay (1997), chapter 2[38]).

¹⁷The main drawbacks of the non-parametric models are: a bad convergence in small samples, which is also amplified when the derivatives of the function are estimated, and the necessity of a usually higher than available data since are usually very data-intensive methods.

The higher reliability and precision of the risk-neutral measure with respect to the physical one can also be implicitly deduced from a particular case of the PK puzzle which arises frequently in literature. By a flex in the central area of the function some papers (i.e. Jackwerth (2000)) show the existence of a puzzling PK in the area of zero or nearly zero returns. This is by far the area with the highest amount of options prices available, thus where the risk-neutral measure is at its highest precision.

Therefore, even though the misestimation of the risk-neutral measure could still be among the possible causes of some PK puzzles, it is considered as the more stable, easier to estimate and reliable between the two measure.

To conclude: to obtain a time-varying estimation of the real-world density it is of key importance, more than to pick the right extraction model, to use the right source of information. Being historical stock returns only backward looking, hence only partially informative, I need to somehow complete the measure by using other sources of information, i.e.: the investors' future sentiment extracted from the implied moments of the option surface.

2.3 Joining the measures

As documented by Eraker (2004)[65], and following Chernov and Ghysel (2000)[42] and Pan (2000)[126], the use of the risk-neutral and physical measures became very powerful in estimation once properly mixed.

The advantages are multiple. First, if the joint measure is properly calibrated, it is easier to disentangle the risk premium that arises from volatility and jumps. Second, the big quantity of data required to produce and an accurate analysis using stock market data makes the overall estimation technique extremely time consuming. Thanks to the one-to-one relation of options to the conditional returns distributions, I can exploit the richness of options and use shorter samples for estimation. Since non parametric methods are usually extremely data-intensive this feature could overcome the problem by reducing the amount of data inputs required thus making computations possible and less time consuming (Rubinstein (1994)[142], Jackwerth and Rubinstein (1996)[94], Dumas Fleming and Whaley (1998)[57] and 'Ait-Sahalia and Lo (1998)[3]). Last but not least the strong relation among the two asset classes can be exploited for a model misspecification diagnostic since, by theory, the estimated parameters implicit in derivatives prices have to be consistent with those used in asset prices data alone

Encouraged by the documented high estimation power of the joint methodology and as a consequence of the high difficulty in obtaining a valid estimation methodology for a pure physical distribution, in this thesis I propose a new non-parametric methodology to estimate the physical measure by joining the two measures.

As a main idea, I solve the problem using part of the naturally forward looking information provided by the risk-neutral measure, and I mix it with the objective one to complete it so that the new measure becomes conditional with respect to all the information available thus leading to a fully conditional, hence homogeneous PK. In particular I should condition to the present level of volatility. I reconstruct it by means of an asymmetric GJR GARCH process with empirical innovations. As a main result I obtain a more informative and flexible measure that impacts positively the estimations of the PKs up to answer to PK puzzle.

Statistically, a natural approach to exploit simultaneously different sources and provide statistical inference is the Bayesian approach (see, among the others Berger (1985)[23], Bernardo and Smith (1994)[23]).

The forward looking information that can be extracted from option prices would be an economically coherent measure only if investors were risk-neutral. Generally investors are rational and provide physical (subjective) estimates such that the two measures can no longer be the same¹⁸. What would fill this gap so that the two measures turn back to be equal is the risk aversion adjustment:

$$\text{Risk-Neutral prob.} = \text{Risk Physical prob.} \cdot \text{Risk Aversion} \quad (2.3)$$

By the FTAP, under no arbitrage conditions, the risk-neutral returns are “risk-adjusted” physical returns which cannot earn more than the risk free rate:

$$\mathbb{E}^{\mathbb{Q}}(R_{t,T}) = \mathbb{E}^{\mathbb{P}}(R_{t,T}) - \text{Risk Premium}_{t,T} \quad (2.4)$$

$$= \text{rf} \quad (2.5)$$

As a consequence, being the proposed new physical probabilities a mixture of measures, I need to properly correct both measures by a risk premium¹⁹. A full treatment of the risk-neutral adjustment

¹⁸In this case, the use of the risk-neutral probability approach would produce highly biased forecasted values and would only be good to express future market expectations.

¹⁹Without the risk premium adjustment, the use of the risk-neutral distribution for forecasting would lead to heavily biased forecasts. For example, if used by financial regulators as diagnostic tool for future financial distress, the unadjusted measure would lead to an increased future market turbulence rather than a reduced one. See Anagnou et al. (2002)[4] for a review of the literature on the topic.

in chapter 6. The risk adjustment indicates investors' preference for risk. What stated is true in one-period models; for a generalized inter-temporal n-periods models also higher moments premium matters (i.e.: volatility premium, kurtosis premium, skewness premium and so on).

CHAPTER 3

Theoretical Set Up And The Literature So Far

After presenting the empirical pricing kernel, which will then be extensively used and rearranged in the next chapters of this thesis I will do a step back to present, from a theoretical viewpoint, four different approaches to derive the financial PK. These methods are part of an extensive list, I don't pretend to be exhaustive. All derived under no-arbitrage or in equilibrium, the first two approaches presented are more closely related to measure theory and functional analysis. The last two methods instead have stronger economical foundations in their initial assumptions. Analyzed from a larger viewpoint, all models boil down to the same final result thus showing how all approaches can be seen as somehow all linked together.

As a first and more convenient definition, I defined in chapter 2 the PK as a ratio of measures, which is conveniently defined as the Radon-Nikodym[124] derivative of two measures; in section (3.1.1), I properly define which are the technical requirements and the economical meanings of these two measures. I then analyze the martingale properties of the PK and its relation with the first FTAP²⁰.

Following Harrison and Kreps (1979)[80], I show how the PK is nothing but a direct application of the Riesz representation theorem[138].

Finally, starting from stronger economical foundations, I derive the PK under a general equilibrium model and under no arbitrage conditions. The former is linked to a set of microeconomic assumptions (i.e.: Lucas (1978)[115] and Rubinstein (1976)[141]) while the latter to probabilistic ones (i.e.: Black and Scholes (1973)[27] and Merton (1973)[120]).

²⁰No-arbitrage, market completeness and market efficiency are the three elements upon which, from a theoretical viewpoint, is based most of the financial economic theory. The treatment of these elements compose the three FTAP.

In section (3.2) I briefly investigate the one-to-one relation that links the investor's risk aversion with the PK.

I close the chapter with a literature review concerning the parametric and non-parametric methodologies used to estimate the PK.

3.1 The empirical pricing kernel (EPK)

Relative to the PK, the final goal of this thesis is to study the effect of a proper conditioning of the information on the inputs that compose the PK intended, as defined in (2.1), as a state price per unit of objective probability.

My work is divided in two parts: the first (chapters (3) and (4)), with a stronger theoretical foundation, puts the basis for the second, treated in the remaining chapters, where I test empirically what previously proposed. Throughout the thesis I refer to the PK when the analysis is mostly theoretical and to empirical PK (henceforth: EPK) for the empirical one.

Playing and somehow abusing with words, the PK is the “characteristic function” of any asset pricing model; in fact, in it, we find all the relevant and necessary information required for pricing any type of financial asset class. By the same token but from a statistical viewpoint, it can also be seen as the “sufficient statistic” of any asset pricing model.

Given the focus of the paper, before I present the model for the estimation of the Empirical Pricing Kernel, a more rigorous definition of the filtration set used is needed. Unless differently stated, these specifications apply to all models throughout the thesis. Defined in a fixed and finite planning horizon $t \in T$, where $T < \infty$ a filtration is nothing but an increasing family of σ -algebras $\{\mathcal{F}_t : t \in T\}$. It follows that:

$$\mathcal{F}_s \subset \mathcal{F}_t \subset \mathcal{F}_T \subset \mathcal{F} \quad \text{for} \quad 0 \leq s \leq t \leq T$$

represents the information flow that generates $\mathcal{F} = \sigma(\bigcup_t \{\mathcal{F}_t : t \in T\})$.

Defined on a rich enough filtered probability space $(\Omega, \mathcal{F}, P, \mathbb{F})$, the filtration $\mathbb{F} = (\mathcal{F}_t)_{t \in T}$ is assumed to satisfy the usual hypothesis thus implying a complete and right continuous filtered probability space. Completeness is achieved if the probability space is P complete and \mathcal{F}_0 contains all the P -null sets of \mathcal{F} . Right continuity is defined as:

$$\mathcal{F}_t = \mathcal{F}_{t+} = \bigcap_{u>t} \mathcal{F}_u \quad \text{for all } t \geq 0 \quad (3.1)$$

A correct modellization of the information flow is crucial for a proper measurability of any random process. In fact, a stochastic process $\{Z_t : t \in T\}$ is said to be adapted (hence measurable) to a filtration $\{\mathcal{F}_t : t \in T\}$ if, for each $t \in T$, the process is \mathcal{F}_t measurable.

Definition 3.1.1. *Defined in a no-arbitrage economy, where for each time $t \geq 0$ the probability space is described by $(\Omega, \mathcal{F}, P, \mathbb{F})$, the time t conditional EPK for $T = t + \tau$ is defined as:*

$$M_{t,T} = e^{-r_t(T-t)} \frac{q_{t,T}(S_T|S_t)}{p_{t,T}(S_T|S_t)} \Big|_{\mathcal{F}_t} > 0 \quad (3.2)$$

where $q_{t,T}$ represents the conditional risk-neutral or state price density (SPD), $p_{t,T}$ the conditional real-world density, r_t the continuously compounded daily risk-free rate and S_t is a proxy for the market portfolio.

For my empirical exercise, which will be deeply presented in chapters 5, 7 and 9, S_t represents the S&P 500 index thus projecting the EPK onto the extended positive real line occupied by all possible values taken by the index.

Subscript $_{t,T}$ emphasizes the forward looking orientation of the value to which applies and denotes that all parameters are fully-conditional to all information available at date t with respect to a future time T . This conditioning, too often violated by many models, plays a key role for a correct definition of a homogeneous EPK. To lighten the notations I assume $\tau = (T - t)$ as fixed and equal to one year.

Function (4.17) can be estimated in different ways, depending on how is defined and how the ingredients that compose it are obtained. In all cases, since the EPK represents the fully non parametric investors behaviors, a proper modellisation can only be achieved by placing as much less structure as possible to the functional that describes it. My empirical approach is to let the data speak at the utmost hence to be fully non-parametric with respect to the structure of the EPK and only partially parametric for the estimation of the empirical moments. Once the conditional measures $q_{t,T}$ and $p_{t,T}$ are estimated, the EPK is recovered by simply taking their discounted ratio thus not imposing any constraints on EPK but the ones underpinning the FTAP.

3.1.1 PK as the conditional Radon-Nikodym derivative

Following most of the literature and from a measure theory viewpoint, the PK is usually defined as the Radon-Nikodym derivative of two measures, once properly definite. Exploiting the Radon-Nikodym theorem is a straightforward and elegant way to quickly and fully define the PK. As long as the assumptions behind the model are satisfied, the model has also strong economical foundation. I

start presenting a generic unconditional form of the Radon-Nikodym derivative and its connections with the economy then, in chapter (3.1.1), I make it conditional to better study how the daily information flows among the two measures. Although not immediate in principle, the link of the obtained PK with the FTAP is strong and self-explanatory. I explore it as well.

As anticipated in chapter 2 pricing in a risk-neutral world has the extraordinary advantage of using a unique probabilistic measure which is economically neutral and so applicable to all investors. Unfortunately, if the representative investor's risk attitude is not neutral, the obtained values may provide misleading results. As a link among the two measures, there is the PK which is the collector of all beliefs, errors and premiums of the investor: in a nutshell it embeds all the relevant information required to convert the risk-neutral measure into a real-world measure and vice-versa. Mathematically, given its role, it is convenient to express the PK as a discounted Radon-Nikodym derivative. So represented the PK is nothing but a discounted kernel: a time adapted operator which, working as a transition function of a stochastic process, allows us to move from a neutral to a subjective world.

Proposition 3.1.1. *Defined on a measurable space $(\Omega, \mathcal{F}, \mathbb{F})$, the unconditional version of the Radon-Nikodym derivative requires that:*

1. P is a σ -finite measure on $(\Omega, \mathcal{F}, \mathbb{F})$ and is atomless;
2. Q is a σ -finite measure on $(\Omega, \mathcal{F}, \mathbb{F})$ and is absolutely continuous with respect to P on \mathbb{F} :
 $Q \ll P^{21}$;
3. the Radon-Nikodym derivative of Q with respect to P , is a nonnegative Borel measurable function M defined on the extended real line ($M : \mathbb{R}^+ \rightarrow \mathbb{R}^+$) and satisfies $P\{x \in \mathbb{R}^+ : M(x) = y\} = 0$ for all $y \in \mathbb{R}^+$.

If assumptions 1 and 2 are satisfied, then there exists an a.s. (with respect to measure Q) random variable M which follows directly from the application of Appendix (A.1) and fits into the last point of proposition 4.2.1. It follows that the unconditional version of the Radon-Nikodym derivative, is defined as:

$$M = \frac{dQ}{dP} \tag{3.3}$$

and satisfies the following:

²¹More on the absolute continuity of the measures in Appendix (A.1).

- $Q(A) = \mathbb{E}^P[M \cdot \mathbb{1}_A], \quad \forall A \in \mathcal{F}$
- $\mathbb{E}^P[M] = 1,$
- $M > 0,$ P - a.s.
- $M \in \mathcal{F}$

The obtained value, is a kernel, an operator. Once discounted, equation (4.22) leads to the financial PK. Mathematically, equation (4.22) is also-called the likelihood ratio between Q and P on \mathcal{F} or the change of measure.

The technical requirements of proposition (4.2.1) have all natural economic interpretations and the extension to their time-dependent counterparts is, at least theoretically, immediate. The violations of all of these assumptions lead to puzzling results. Related to the analysis of the PK puzzle, the importance of the atomlessness²² of the underlying probability space is also analyzed by Hens and Reichlin (2011)[84]. They show how a complete market equipped with an atomic probability space have some unaccessible bets. As a consequence the whole spectrum of payoffs is not achievable. Graphically this is translated into a flex on the PK functional, thus to a violation. The absolute continuity requirement of the measures can be translated in economic term as an absence of arbitrage. If the assumption about the equality of the two measures with respect to the negligible events would be violated, there would exist a set $A \in \mathcal{B}(\mathbb{R}^+)$ with $P(A) = 0$ and $Q(A) > 0$. Although the inequality might apply only to rare events, a violation of the absolute continuity it is all that is necessary for having a sure arbitrage in the market. In such a scenario, a contingent claim with payoff equal to the indicator function for this set would have an a.s. strictly positive price together with an a.s. zero payoff. Shorting such a product would lead to a sure profit with no risk (Dybvig (1988)[58]).

The third requirement ensures no flatness for the PK, hence a unique ordered ranking. In graphical terms this is nothing but the convexity of the PK.

Before I close the section I can exploit the above concepts to shade some lights on the difference in the the use of the terms PK and state price density (SPD) which have been sometimes interchangeably used in literature to define similar (but not equal) quantities. The former, used in the

²²Roughly speaking, atomlessness is linked with continuity and atomicity with discreteness of the probability space. An atomless space put zero mass to a given unit, i.e.: the probability of any particular value is equal to zero. Formally: given a measurable space and two measurable sets: $A \in \Omega$ and $B \subset A$, a set $A \in \Omega$ is called an atom if:

$$P(A) > 0 \quad \text{and} \quad P(B) = 0 \quad \text{where} \quad P(B) < P(A) \quad (3.4)$$

Is atomless when:

$$P(A) > P(B) > 0 \quad \text{where} \quad P(B) < P(A) \quad (3.5)$$

this lead to a continuity of values for the non-atomic measure.

more recent literature, is intended as the Radon-Nikodym derivative of the risk neutral measure with respect to the physical one the latter instead used i.e. in Dybvig (1988)[58] is defined as the Radon-Nikodym derivative of the risk neutral measure with respect to the Lebesgue measure. I refer to the recent literature and use the term state-price-density to define the risk neutral measure (or Arrow-Debreu).

3.1.2 The Riesz representation theorem

From the viewpoint of the functional analysis, the existence of the PK is a just an application of the Riesz representation theorem: any linear functional on a vector space can be represented by a vector in that space. This is indeed what has been proposed by Hansen and Richard (1987)[79] extending and refining the original findings of Harrison and Kreps (1979)[80] making them conditional (hence dynamic).

As a main idea they define, for each time t , a conditional pricing function F as a functional form that maps a time T terminal payoff x with a time t price. Both are random variables. The beauty of the finding as well as its connection with the more economic equilibrium models is in its generality: once valid the theorem can be in fact applied with no loss of generality to any asset pricing function.

Following LeRoy et al. (2000)[110] the unconditional and most generic form of the Riesz representation theorem states that:

Theorem 3.1.1. *if $F : \mathcal{H}^{23} \rightarrow \mathbb{R}$ is a bounded, linear and continuous functional on a Hilbert space \mathcal{H} over \mathbb{R}^{24} , then there exists a uniquely determined vector $k_f \in \mathcal{H}$ s.t.;*

$$F(x) = k_f \cdot x \quad \forall x \in \mathcal{H} \quad (3.6)$$

or, in dual form:

$$F(x) = \langle k_f, x \rangle \quad \forall x \in \mathcal{H} \quad \text{and} \quad \|F\| = \|k_f\| \quad (3.7)$$

where the vector k_f is the Riesz kernel of F .

Conversely, any vector $k \in \mathcal{H}$ defines a bounded linear functional F_k on \mathcal{H} by:

$$F_k(x) = \langle k, x \rangle \quad \forall x \in \mathcal{H} \quad \text{and} \quad \|F_k\| = \|k\| \quad (3.8)$$

For a proof of theorem (3.1.1), see Appendix (A.3).

²³Although we used \mathcal{H} to identify the suboptimal filtration set, for this section it only represents Hilbert spaces.

²⁴The Riesz representation theorem can also be extended to complex values. Here I only consider Hilbert spaces over \mathbb{R} .

Its application to asset pricing follows naturally. Hansen and Richard (1987)[79] consider a pricing functions F that maps \mathcal{F}_T -adapted payoffs into \mathcal{F}_t -adapted prices. It follows that the canonical Hilbert space $L^2(\mathcal{F}_T) = L^2(\Omega, \mathcal{F}_T, P)$ is replaced by its conditional version:

$$L^2_{\mathcal{F}_t}(\mathcal{F}_T) = \{x \in L^0(\mathcal{F}_T) : \mathbb{E}[x^2|\mathcal{F}_t] < \infty\} \quad (3.9)$$

where $L^0(\mathcal{F}_T) = L^0(\Omega, \mathcal{F}_T, P)$ is the space of all \mathcal{F}_T -measurable random variables and the conditional second moment is a.s. finite. To be valid, the pricing functional F needs to satisfy a conditional continuity restriction $F : L^2_{\mathcal{F}_t}(\mathcal{F}_T) \rightarrow L^0(\mathcal{F}_T)$ s.t.:

$$F(ax + by) = aF(x) + bF(y) \quad \text{for all } a, b \in L^0(\mathcal{F}_T) \quad \text{and all } a, b \in L^2_{\mathcal{F}_t}(\mathcal{F}_T) \quad (3.10)$$

this assure linearity of the pricing functional.

Moreover, the functional needs to be bounded s.t.:

$$|F(x)| \leq c\sqrt{\mathbb{E}[x^2|\mathcal{F}_t]} \quad \text{for all } x \in L^2_{\mathcal{F}_t}(\mathcal{F}_T) \quad \text{and } c \in L^0(\mathcal{F}_t) \quad (3.11)$$

If the above assumptions hold, the unconditional duality $\langle x, y \rangle = \mathbb{E}^P[xy] \in \mathbb{R}$ of the Hilbert space $L^2(\mathcal{F}_T)$ can be replaced by its conditional counterpart $\langle x, y \rangle^{\mathcal{F}_t} = \mathbb{E}^P[xy|\mathcal{F}_t] \in L^0(\mathcal{F}_t)$ of the Hilbert space $L^2_{\mathcal{F}_t}(\mathcal{F}_T)$ ²⁵.

Applying the generic equation (3.6) to asset pricing:

$$F(x) = \mathbb{E}^P(k_f \cdot x) \quad \forall x \in \mathcal{M} \quad (3.12)$$

where $\mathcal{M} \in \mathcal{H}$ represents the asset span and k_f is the financial PK. Being the expectation the sum across all possible states (s) of the world $F(x) = \sum_s k_s x_s$, then, setting $k_{f_s} = k_s/\phi_s$:

$$F(x) = \sum_s \phi_s k_{f_s} x_s = \mathbb{E}^P(k_f \cdot x) \quad (3.13)$$

where the expectation under P is due to ϕ_s which are the investors' subjective densities for the possible states.

3.1.3 Macroeconomic derivation

The finance PK is a nothing but a generalization of the marginal rate of substitution (henceforth: MRS). While the latter is related to consumption goods, the former is related to financial assets.

²⁵The square integrability for pricing is often too restrictive. It is easily possible to relax it and have similar results under L^1 .

In both cases are the optimal solution of a generic consumer problem (CP). What changes is the vector of goods chosen by the investor to maximize her welfare.

Following Chocrane (2001)[48], Ingersoll (1987)[89] and Mas-Colell and Green (1995)[49] I present the necessary steps to derive the inter-temporal investor's first order conditions that give the basic consumption-based model. The MRS is often see as the most important result from the overall consumer theory. What I present, although obtained from a very simple model, is easily extendible to more complex settings and is at the hearth of many economic theories.

Mathematically, the MRS is nothing but the solution of an optimization problem, more specifically of a constrained maximization problem. A rational consumer chooses her inter-temporal optimal consumption bundle: the one that maximizes her welfare (her utility $U(\cdot)$) on the set of her feasible consumption bundles (her budget set w):

$$\max_{x \in \mathbb{R}_t^+} U(x) \tag{3.14}$$

$$\text{s.t. } p \cdot x \leq w \tag{3.15}$$

where $x = (x_1, \dots, x_t)$ is a vector of goods and the budget set depends to the goods prices p . The existence of a valid representative investor with a utility function U depends on a few basic conditions which must be met at each time:

- The preference set of the investor must obey four preordering properties, namely: completeness, transitivity, relaxivity and continuity. Completeness requires that $u(s_1) \geq u(s_2)$ ²⁶ or $u(s_1) \leq u(s_2)$, so that all alternatives can be compared, hence evaluated.

For $u(s_i)$ transitivity means that if $u(s_1) \geq u(s_2)$ and $u(s_2) \geq u(s_3)$, then $u(s_1) \geq u(s_3)$.

Relaxivity is a self-explaining property which states that i.e.: $u(s_1) \geq u(s_1)$.

Finally, to guarantee the existence of a utility function, the continuity property is determinant. The property requires the openness of the strictly preferred and worse subsets. Only if this last assumption satisfied a valid continuous utility function can exist²⁷.

- The sum of individual wealth functions, which are continuous and homogeneous of degree 1, is the aggregate wealth.

²⁶Where $u(\cdot)$ is the utility function and (s_i) for $i = 1, \dots, T$ are the finite states of the world

²⁷The interested reader is referred to Hildebrand and Kirman (1988)[88] or Luce and Raiffa (1957)[116] for detailed explanations and to Ingersoll (1987)[89] for a summary.

- The aggregate demand function is as well continuous and homogeneous function, but of degree 0.
- Following the Walrasian property, the entire endowment is fully consumed at the end of each time horizon.

The utility function $U(\cdot)$ is assumed to be a twice differentiable function of wealth w , for $w > 0$ and measures the investor's relative preference for different level of aggregate wealth (w). All investors share the non-satiation property: more consumption is always preferred to less: $U'(w) > 0$. On the contrary, the investors risk attitude is not unique and is represented by the concavity of the utility function:

1. if $U''(w) < 0$ the investor is risk averse,
2. if $U''(w) = 0$ is risk neutral and
3. if $U''(w) > 0$ is risk seeking.

Under the most general settings, the investor is risk averse and has a marginal utility of wealth that decreases as the total wealth increases, thus producing a concave utility function.

Introducing δ , the subjective discount factor, a two periods inter-temporal utility function is defined as:

$$U(c_t, c_{t+1}) = u(c_t) + \delta \mathbb{E}_t^P [u(c_{s_{t+1}})] \quad (3.16)$$

$$= u(c_t) + \delta \sum_s (u(c_{s_{t+1}}) p_t(s_{t+1}|s_t)) \quad (3.17)$$

where: c represents consumption such that c_t is the today consumption and $c_{s_{t+1}}$ is the consumption for a given state s at the future time $t + 1$. The randomness of the future consumption is estimated by \mathbb{E}_t^P which is the subjective expectation conditional to the time t information (under this setting the conditionality of the expectation with respect to the information set is graphically identified by t). The same expectation can be translated as the sum across states of all feasible conditional future probabilities: $p_t(s_{t+1}|s_t)$.

Introducing ξ_t as the amount of asset the investor chooses to buy²⁸ at time t , e as the agent's original wealth (endowment) and P_t as the price of the asset at time t with payoff $\psi_{s_{t+1}}$ ²⁹, then the two periods investor's optimization problem is:

$$\max_{\xi_t} \{u(c_t) + \delta \mathbb{E}_t^P [u(c_{s_{t+1}})]\} \quad (3.18)$$

²⁸It is assumed that at each time t investors can buy/sell as much as they want.

²⁹The payoff is the price at time $t + 1$ plus any dividend, if exist.

subject to the budget constraint at time (t) and the Walrasian property at time ($t + 1$):

$$c_t = e_t - P_t \xi_t \quad (3.19)$$

$$c_{s_{t+1}} = e_{s_{t+1}} + \psi_{s_{t+1}} \xi_{s_{t+1}} \quad (3.20)$$

If u is continuous, and $P_t \gg 0$, the consumer's budget set is compact (is closed and bounded) and the Weierstrass theorem implies that the CP has a solution. Convexity determines its unicity. Plugging the constraints into the objective, the first order condition of the problem yields the investor's optimal consumption-investment choice:

$$P_t u'(c_t) = \mathbb{E}_t^P [\delta u'(c_{s_{t+1}}) \psi_{s_{t+1}}] \quad (3.21)$$

then, solving for P_t :

$$P_t = \mathbb{E}_t^P \left[\underbrace{\delta \frac{u'(c_{s_{t+1}})}{u'(c_t)}}_{MRS} \psi_{s_{t+1}} \right] \quad (3.22)$$

Equation (3.22) underlies all of asset pricing and is a valid equation to price any asset. MRS is the time t Marginal Rate of Substitution, also known in finance as stochastic discount factor (SDF) or PK.

The MRS or PK is a random variable that generates today's prices from future payoffs reflecting the investors' preference for payoffs over different states of the world (Chocrane (2001)[48]). Setting $m = \text{MRS}$, I split the above equation in two parts to express the price as a discounted payoffs:

$$P_t^{\mathbb{P}} = \mathbb{E}_t^P [m_{s_{t+1}} \psi_{s_{t+1}}] \quad (3.23)$$

where the conditional expectation is taken with respect to the physical measure and $m_{s_{t+1}}$ is the inter-temporal MRS, the rate at which the investor is willing to substitute one unit of consumption today t for one unit of consumption tomorrow s_{t+1} .

3.1.4 No-arbitrage derivation

The results obtained in the previous section are based upon "absolute" or "equilibrium" pricing models. I now turn to "relative" or "no-arbitrage" pricing models³⁰. Also under this framework

³⁰In "absolute" models the value of cash flows are based upon their exposure to fundamental sources of microeconomic risks, while in "relative" pricing I can learn of one asset's value given the prices of other related assets. The former are more general and applicable to all assets but produces approximated results, the latter instead are easier to implement but of limited practical applicability. Their main drawbacks are: no-arbitrage models lack of economic foundation, while equilibrium models are often inaccurate, this is usually due to the fact that consumption data are not always available and often difficult to estimate precisely other than highly model-dependent.

the key question is how to construct a coherent discount factor.

From the FTAP, the no arbitrage (NA)³¹ condition together with market completeness and its efficiency imply the existence of unique and positive Arrow-Debreu state prices: a unique risk neutral measure under which the process is a Q-martingale (hence the measure is an Equivalent Martingale Measure) such that the risk free rate is the future expected return on any asset³².

It also implies the existence of a strictly positive PK under which any asset price can be expressed as the subjective expected value of the product of the asset payoff and the PK.

Under this setting, the PK provides a complete and unique description of expected returns, asset prices and risk premia of the investor.

Merton (1973)[120] and Lucas (1978)[115] are the first to apply this settings to generalize the asset pricing problem. The utility function is now linked to the agent's wealth e_t and the payoff's function depends on the underlying asset value S_t .

In continuous time, the fundamental value equation (3.23) can be rewritten as:

$$P_t^{\mathbb{P}} = e^{-r_{t,T}\tau} \int_0^{\infty} \psi_T(S_T) \lambda \frac{U'(S_T)}{U'(S_t)} p_{t,T}(S_T|S_t) dS_T \quad (3.24)$$

where $\tau = T - t$ is the time-to-maturity, $\lambda e^{-r_{t,T}\tau} = \delta$ the discount factor³³, $\psi_T(S_T)$ is the asset's payoff and $p_{t,T}(S_T|S_t)$ is the time t physical probability with respect to the future state S_T .

Under some mild assumptions required by the FTAP, the price of any asset can be uniquely computed under the risk neutral measure as a discounted expected payoff:

$$P_t^{\mathbb{Q}} = e^{-r_{t,T}\tau} \int_0^{\infty} \psi_T(S_T) q_{t,T}(S_T|S_t) dS_T \quad (3.25)$$

$$= e^{-r_{t,T}\tau} E_t^{\mathbb{Q}}[\psi_T(S_T)|\mathcal{F}_t] \quad (3.26)$$

where $q_{t,T}(S_T|S_t)$ is the time t state price density (the risk neutral measure or SPD) and the conditional expectation is now no longer taken under the subjective probability measure but under the risk neutral one, thus reflecting an objective belief about possible future (T) states of the world.

³¹NA condition: given a probability space $(\Omega, \mathcal{F}, \mathbb{P})$, no arbitrage (NA) \implies there is a probability measure Q with $Q(\omega) > 0 \forall \omega \in \Omega$, such that every discounted asset price process $S_n = \{S_n(t), t = 0, 1, \dots, T\}$ is a Q -martingale, $n = 1, \dots, N$. The measure Q is an equivalent martingale measure (EMM).

³²For details see: Harrison, Kreps and Pliska (1979, 1981)[80][81], Dybvig and Ross (1987, 2003)[59] and Arrow (1964)[6] Debreu(1959)[7]

³³ λ is independent to the index level and is just for scaling purposes

The information set I_t is now represented by the filtration set \mathcal{F}_t

Joining equations (3.24) and (3.25) I derive the today t PK relative to the future time T , $M_t(S_T)$ as:

$$M_{t,T}(S_T) = \frac{q_{t,T}(S_T|S_t)}{p_{t,T}(S_T|S_t)} \quad (3.27)$$

$$= \lambda \frac{U'(S_T)}{U'(S_t)} \quad (3.28)$$

Therefore, the time (t) marginal rate of substitution is a discounted conditional expected value:

$$MRS_{t,T} = e^{-r_t, T \cdot \tau} E_t^{\mathbb{P}}[M_{t,T}(S_T)|\mathcal{F}_t] \quad (3.29)$$

The relation between the risk neutral pricing equation (3.26) and Lucas asset pricing equation can be derived by using equation (3.27) in (3.25) such that:

$$P_t = e^{-r_t, T \cdot \tau} E_t^{\mathbb{Q}}[\psi_T(S_T)|\mathcal{F}_t] \quad (3.30)$$

$$= e^{-r_t, T \cdot \tau} \int_0^\infty M_{t,T}(S_T) \cdot \psi_T(S_T) p_{t,T}(S_T|S_t) dS_T \quad (3.31)$$

$$= e^{-r_t, T \cdot \tau} E_t^{\mathbb{P}}[M_{t,T}(S_T) \cdot \psi_T(S_T)|\mathcal{F}_t] \quad (3.32)$$

From (3.32) the definition and role of the PK is visible. The above equation is in fact a state-dependent function which discounts payoffs just using $M_{t,T}(S_T)$ - the collector of all the investor's risk preferences - and a time parameter.

Finally, since at time t the value $U(S_t)$ is known, by integration of (3.27) and (3.28) it is possible to extract the today market utility function through the PK:

$$U(S_T) = U(S_t) + \int_{S_t}^{S_T} U'(S_t) \frac{1}{\lambda} \frac{q(x)}{p(x)} dx \quad (3.33)$$

$$= U(S_t) + \int_{S_t}^{S_T} U'(S_t) \frac{1}{\lambda} K(x) dx \quad (3.34)$$

where K is $\frac{q(x)}{p(x)}$.

3.2 The investor's risk aversion

The PK, the investors' utility functions and her risk attitudes in the markets are all tightly related and all dependent on each others.

By defining the measure of absolute risk aversion (ARA) as the negative ratio of the derivative of the PK over the PK, Arrow (1964)[6] and Pratt (1965)[134] show a connection between the PK and the measure of risk aversion of a representative agent.

The risk attitudes in the market are often described in terms of relative risk aversion (RRA), which is defined as:

$$RRA(S_T) = \zeta(S_T) = -S_T \underbrace{\frac{U''(S_T)}{U'(S_T)}}_{ARA} \quad (3.35)$$

The agent's risk aversion is a measure of curvature of the agent's utility function; the more an investor is risk averse and the more pronounced is the curvature. In this case, the Arrow-Pratt measure of risk aversion is normalized by the level of stock price to keep a fixed scale in measuring the risk aversion.

Starting from the definition of the PK given in (3.28):

$$M_{t,T}(S_T) = \lambda \frac{U'(S_T)}{U'(S_t)} \quad (3.36)$$

and taking its first derivative:

$$M'_{t,T}(S_T) = \lambda \frac{U''(S_T)}{U'(S_t)} \quad (3.37)$$

We obtain the definition of RRA:

$$RRA(S_T) = \zeta(S_T) = -S_T \frac{\lambda M'_{t,T}(S_T) U'(S_t)}{\lambda M_{t,T}(S_T) U'(S_t)} \quad (3.38)$$

$$= -S_T \frac{M'_{t,T}(S_T)}{M_{t,T}(S_T)} \quad (3.39)$$

Exploiting (3.27) the investor RRA can also be defined as:

$$\zeta(S_T) = -S_T \frac{[q_{t,T}(S_T|S_t)/p_{t,T}(S_T|S_t)]'}{q_{t,T}(S_T|S_t)/p_{t,T}(S_T|S_t)} \quad (3.40)$$

$$= -S_T \frac{[q'_{t,T}(S_T|S_t)p_{t,T}(S_T|S_t) - p'_{t,T}(S_T|S_t)q_{t,T}(S_T|S_t)]/p_{t,T}^2(S_T|S_t)}{q_{t,T}(S_T|S_t)/p_{t,T}(S_T|S_t)} \quad (3.41)$$

$$= -S_T \frac{q'_{t,T}(S_T|S_t)p_{t,T}(S_T|S_t) - p'_{t,T}(S_T|S_t)q_{t,T}(S_T|S_t)}{q_{t,T}(S_T|S_t)p_{t,T}(S_T|S_t)} \quad (3.42)$$

$$= S_T \left[\frac{p'_{t,T}(S_T|S_t)}{p_{t,T}(S_T|S_t)} - \frac{q'_{t,T}(S_T|S_t)}{q_{t,T}(S_T|S_t)} \right] \quad (3.43)$$

It follows that once the risk neutral and the physical distributions are available, the estimate of the investor's RRA is just a straightforward application of equation (3.43). It goes by logic that the higher the precision of the measure estimates, the better the overall modellisation of the investor's

RRA. The ARA value turns to a decreasing function if the PK value is increasing in market return. Due to their strong interconnectedness, if the PK is estimated suboptimally it may also lead to further puzzles like the risk aversion puzzle. Although closely related the ARA puzzle is still different from the PK one (Jackwerth (2000)[93]). The above violation happens only if the PK puzzle is in play, thus leading also to the ARA puzzle, but the opposite is not necessarily true.

3.3 The literature so far

Being a key element of any asset pricing model, an in deep theoretical and empirical comprehension of the PK and its related puzzles are of paramount importance both from an academic and from a practical point of view. This high interest is reflected in the big amount of literature that analyzes these subjects.

Among them, the paper of Jackwerth (2000)[93] is the first one that, using separately option and stock data, highlights the existence of a puzzling behaviour of the PK. By searching for the smoothest risk neutral distribution and performing a kernel density estimation for the subjective probability he finds an opposite behaviour of the PK and the relative risk aversion after and before the 1987 crisis. While before the crisis both quantities are in line with the neoclassical economic theory, after the crisis there are strong violations in the area of no returns (ATM, the central part of the PK). Using a more advanced but still insufficient technique for the estimation of the conditional physical measure, the same puzzling result has been empirically obtained also by Rosenberg and Engle (2002)[64]. Their paper is enriched by the empirical analysis that finds a link between the puzzle and the counter cyclicity of the risk aversion.

Through a GJR-GARCH-FHS model for both the risk neutral and the physical measures Barone-Adesi, Engle and Mancini (2008)[15] partially solved the PK puzzle empirically. Although the model is able to account for the the equity and the variance risk premium and can provide daily estimates of the densities, the physical measure estimation is still mostly backward looking, hence incomplete. This leads to puzzling PK, above all for highly volatile days and for days with high trading of call options. An interesting alternative to the BEM approach to option pricing has been proposed by Chorro et al. (2012)[43]. They also fit the option volatility surface through filtered historical simulation ³⁴. However they keep GARCH parameters constant between the physical and the risk neutral distribution. They test several families of GARCH models, choosing the one that provides

³⁴More precisely they use generalized hyperbolic distributions (GHD).

the best fit to option prices. Their choice of risk adjustment is therefore implicit in the filtering that different GARCH provide, leading to simulated distributions that are more or less heavy tailed. Their approach is less computationally intensive than BEM. However it is not directly amenable to the study of the PK, because it does not estimate the physical and the pricing distribution separately.

Other papers deal with the PK estimation problem using quasi non-parametric models. Äit-Sahalia and Duarte (2003)[2] use a combination of constrained least squares (CLS) regression and smoothing techniques to estimate the PK on simulated and real data. Birke and Piltz (2009)[25] approach the problem with a fully kernel based method using only the strike as a predictor. Hardle and Hlavka (2009)[80] model the PK as function of the strike and time, then use a constrained non-linear least squares procedure to estimate its dynamics. Äit-Sahalia and Lo (2000)[3] extend the classical use of a non-parametrically estimated PK and risk aversion to analyze an economically linked value-at-risk (VaR) measure that is then compared to a statistical VaR. Christoffersen, Heston and Jacobs (2010)[44] analyze the impact of the variance risk premium for the PK. Generalizing the Heston and Nandi (2000)[87] GARCH model and allowing for the variances to differ by incorporating a negative risk premium, they state that the negativity of this premium is the reasons of a U shaped PK. Other papers treat the variance as the key elements for the puzzling behaviour of PK, among them Chabi-Yo, Garcia and Renault (2007)[158], Chabi-Yo (2008)[157], Song and Xiu (2012)[149].

From a behavioural viewpoint, Barone-Adesi, Mancini and Shefrin (2012)[18], provide the first rigorous definition of investor sentiment as the difference between the objective distribution and the one perceived by the representative investor that sets prices. They extend BEM to a behavioural setting, by modelling investors' excess optimism and overconfidence, that is the differences in the means and variances of the two distributions, and relating them to the difference between the PKs perceived by two investors. One of them is the sophisticated investor introduced in the BEM paper above, the other one has always a monotonically non-increasing PK. The resulting estimates of optimism and overconfidence are validated against results from investors' surveys and other sentiment indicators available in the literature. Finally, Barone-Adesi, Mancini and Shefrin, in their contribution to the Handbook of Systemic Risk (2013)[17], explore the potential systemic risks posed by the above sentiment estimates, analyzed at the market-wide and individual firm level. Thanks to its flexibility and model-free features, the PK estimation methodology proposed in this thesis can easily be extended to these papers and make the overall empirical analysis even more behavioural. In fact, the investor utility maximization could be achieved by properly choosing the mixing distri-

bution of p and q depending on the market environments, thus producing more subjective and more refined results. Ziegler (2007)[159] and Hens and Reichlin (2011)[84] as well analyze the subject under a behavioural point of view and state that the main reason of a possible PK puzzle may be due to misspecified and heterogeneous investors' behaviours.

Relying only on the historical record, the main drawback of all these papers is the deficitary estimation of the conditional physical measure.

The approach proposed in this thesis to solve the above presented problem is somehow comparable in spirit to Jackwerth and Brown (2001)[37], Ziegler (2007)[159], Chaudhuri and Schroder[41] and Barkhagen et al. (2016)[13]. The first two papers try to explain the puzzle by estimating a denominator which is a weighted average. The former is a weighted average of two PKs associated with high and low wealth states. The latter explains the puzzle by playing with the heterogeneous beliefs of investors and performing a mixture of two lognormal distributions that represent optimism and pessimism. These seemingly ad hoc proposals are directly related to the level of variance, while I play with them indirectly by properly mixing the two measures relative to the options market which turns out to be a more stable and less model-dependent methodology. Also Chaudhuri and Schroder mix options and stock information to enrich the denominator of the PK hence of the PK overall. Barkhagen et al. (2016) uses options data and numeraire portfolio for the estimation of the real world measure. Also Beare and Schmidt (2012) [21] pointed out the importance of having a fully conditional measure at the denominator to produce a homogeneous measure. I differ from them in a general sense, since the aim of their paper is to check for the statistical robustness of the PK puzzle, and in more detailed sense since they fix the denominator by using non informative historical return data. The paper of Ross (2015)[140] as well as other similar or relative papers will be discussed throughout the chapters of the thesis.

CHAPTER 4

Conditioning for the information in asset pricing

While the theoretical and also the big majority of the empirical asset pricing literature assumes the investor to be fully informed, or at least to be able to perfectly manage all the markets information, the reality is very much different. There is in fact a big gap between what is required by the financial literature and what is instead achievable by i.e. and econometrician that uses real data. The norm is in fact that normal investors are not able to properly capture all the required and necessary information to price any assets or at least to use these information to act rationally with respect to any investment decisions. To lesser extents, the same is true also for more sophisticated investors. At the hearth of this issue there is usually a byproduct of a theoretical convention and an econometric problem. The former is the by now widespread accepted convention of using the risk-neutral measure as if it would be the real one. Being probably the most important innovation of the last years in asset pricing, the risk-neutral economy is both a bless and a curse for any pricing model. Is a bless because its application allows us to obtain a often reliable price. Is a curse because it assumes the risk-neutral measure to be the real one. In reality investors are quite never neutral with respect to the risks. This produces a gap, which might be big in particular conditions (i.e.: days of high volatility).

The latter is instead an econometric problem: to capture the fully nonlinear and usually unobservable past, present and above all future beliefs that affect the investors decision with respect to future outcome is usually not a straightforward task to perform. Aside from the technicalities of the econometric models used, it turns out that the biggest issue lurks in the inputs used in estimation. The use of only backward looking assets is in fact not enough to produce a forward looking measure, not matter which model is used. It follows that the estimation of the real-world measure is still an

unanswered problem both theoretically and econometrically.

In this chapter we review and adapt some of models presented in chapter (3) to show the impacts of a suboptimal use of the information in asset pricing.

4.1 Information and market efficiency

Market efficiency is a strictly necessary condition for the existence of an equivalent martingale measure (Jarrow (2012)[96]). This makes the market's information a key factor for the validity of the fundamental theorems of asset pricing.

Acting among them, the pricing kernel (henceforth: PK) is what links the fictitious risk-neutral probability to the real-world one. How the information flows between the two worlds is thus determined by the PK and viceversa if one uses the measures themselves to determine the PK.

It follows that the market's efficiency, the PK and the three fundamental theorems of asset pricing are among them tightly related by the market's information. How this information is defined and used is then of key importance for many investment decisions.

After presenting the econometrical issues associated to the estimation of a fully conditional and time-varying real-world measure this chapter analyzes, from a probabilistic viewpoint, the consequences of using a suboptimal information set for the estimation of the empirical pricing kernel (henceforth: EPK)³⁵ and how this may have an impact on the fundamental theorems of asset pricing.

Showing how the improper use of market's information may lead to unexisting puzzles, this papers links together two apparently unrelated streams of literature: one more empirical, pertaining to the analysis of the PK puzzles and one more theoretical, that analyzes the importance of the market efficiency for the validity of the fundamental theorems of asset pricing. While the link between these two disciplines has been already analyzed from the viewpoint of the equity and option markets alone³⁶, its link with the PK, that is often the byproduct of the two markets, is still unclear. Three are the main innovations of the chapter. First, it proposes for the first time a probabilistic investigation on the improper use of the market information for the estimation of the real-world

³⁵We generally refer to the pricing kernel (PK) as the theoretical one and to the empirical pricing kernel (EPK) for its empirical version.

³⁶Since the seminal paper of Fama (1970)[?] the analysis of the markets efficiency is a long lasting and still unanswered problem in finance. Fama (1991),(1998)[?][?] provide a great review with a focus on the equity market. The link between market efficiency and the option market is much less debated but presented theoretically by Jarrow (2013)[97].

measure that composes the EPK. Its link with the violations of the fundamental theorems of asset pricing will be made immediate. Second, it applies and extends the recent findings of Protter (2015)[136] to the analysis of the PK. Finally, it is a first attempt to shows in a model-free way how to detect market inconsistencies, hence market inefficiencies (Jarrow and Larsson (2012)[98]) through the analysis of the pricing kernel.

Linking together three different disciplines - finance theory, empirical and mathematical finance - the nature of the analysis is mainly theoretical and sometimes mathematically rigorous, but always supported by real world empirical examples.

As a consequence of the three fundamental theorems of asset pricing (Delbaen and Schachermayer, (1994), (1998)[52][53], Harrison and Pliska, (1981), (1983)[81][82], Jarrow, (2012)[96], Jarrow and Larsson (2012)[98]) it is only if a market is arbitrage-free³⁷, complete and sensitive that there exists a numeraire such that the risk-neutral measure is effectively a unique equivalent martingale measure, $\mathbb{Q} = \mathbb{P}^{38}$ (Frahm (2016)[74]).

While the first and the second fundamental theorems of asset pricing only guarantees respectively the existence and uniqueness of an equivalent martingale measure, these theorems cannot determines the nature of the martingale itself. Defined as the third fundamental theorem of asset pricing, a unique equivalent martingale measure (EMM) exists only in presence of a no free lunch with vanishing risk (NFLVR) complete market with no dominance (ND). The NFLVR is in fact instable under information reduction and, if is not supported by no dominance, leads at most to the existence of an equivalent local martingale measure (ELMM). From an economic viewpoint, what lies between an ELMM price and an EMM price is a bubble or a risk-less profit. Not only, most of the theory underpinning the risk neutral valuation assumes the existence of an efficient market. It follows that there is an intimate relation between option pricing and efficiency and that most of the theory may fail in practice if the market is not efficient. To prevent the possible big consequences of adopting a wrong pricing technique, riding an unexisting bubble or setting up an effective hedge it is thus of key importance to be able to properly³⁹ use the market information.

The reason of the importance of the information for the nature of the prices lies in the structure of the prices themselves. A semimartingale is defined as the sum of two components: a local martin-

³⁷More precisely, the market has to be free of lunches with vanishing risk (NFLVR), which is an asymptotic extension of the no arbitrage principle.

³⁸Following much of the literature \mathbb{Q} and \mathbb{P} represent respectively the risk-neutral and the real-world measures.

³⁹By properly we simply mean as required by the different neoclassical theories.

gale and an adapted finite-variation process. The lack of crucial information, hence the projection of the semimartingale onto a coarser filtration set, may not only lead to lose the first component but also to change the nature of the finite variation part of the process thus impacting on the nature of the process itself. This produces non negligible effects for arbitrage theory. What would bring things back to an equilibrium would be the existence of a finite variation part that acts as a compensator. In economic terms, this compensator takes the role of the risk premium among the two measures. If it is properly calibrated, then the economy would be in an equilibrium with the existence of a valid pricing kernel (a true martingale) and a unique EMM. It turns out that in reality this is rarely true.

In fact, although crystal clear in theory, to channel the flow of information under the two measures is not a straightforward task empirically. From the viewpoint of the market information, the empirical issues relative to the estimation of the same variable under the two measures may lead to the market insensitiveness⁴⁰ such that the two measures are by consequence non-equivalent.

It is important to recall that as long as the deviations are random, market efficiency does not require the market prices to be equal to the theoretical or true values. This is not guaranteed under a suboptimal filtration set. In fact, if the existence of heterogeneous beliefs is not a problem, this is not so if the two measures do not share the same nullsets. The lack of absolute continuity among the two measures have thus strong impacts on the nature of the change of measure. As a remark, also the uniqueness of the risk neutral measure, that is the second fundamental theorem of asset pricing, is often at risk in reality.

Recalling that the units of measure of an investor that uses real data are the \mathbb{P} -measure and the relative real-world probability, this chapter builds a bridge between what is theoretically required by fundamental theorems of asset pricing and what is instead effectively achievable by an econometrician that uses real data. If the latter is not aligned with the theory, it may lead to unexisting puzzles and inefficiencies. This may be particularly true from the viewpoint of the types of information used in estimation, a characteristic to often violated in practice. Furthermore, as a second step, this chapter analyzes how the impacts of this discrepancy may spread over the risk neutral valuation theory. Touching both the empirical and the theoretical aspect of the problem, the obtained findings, although theoretical in principle, are relevant also from a practical perspective.

⁴⁰Following the terminology that underpins the H-hypothesis of Brémaud and Yor (1978)[35] the market is non sensitive or, equivalently, not immersed.

Acting among the measures, the best way to analyze how the information moves among the two is to analyze the PK. Defined as a discounted ratio of state prices per unit of objective probability, the neoclassical theory requires the PK to be a monotonically decreasing function in wealth. It is by now common practice, both in academia and for the industry, to use stocks and options market data as the usual inputs to estimate respectively the real-world⁴¹ and risk-neutral measures that compose the EPK. However, while stock returns are by construction backward-looking hence - at best - only partially informative market data, option surfaces are naturally forward-looking financial assets. This mismatch violates the general theory that requires both measures to pertain to the same filtration set. Due to this disalignment, it turns out that most of papers present in literature that study the EPK are then affected by a non-homogeneity bias: they compare a fully conditional forward-looking risk-neutral measure extracted from option surfaces with a partially conditional real-world measure extracted from time series of historical returns (i.e.: Ait Sahalia and Lo (1998)[3]; Jackwerth (2000)[93]; Brown and Jackwerth (2001)[37]; Rosenberg and Engle (2002)[64]; Barone-Adesi et al. (2008)[15]; Yatchew and Härdle (2006)[156]).

Sala (2015)[143] overcomes this issue “completing” the real-world measure by exploiting the naturally forward-looking information extracted from option prices. Options’ moments are estimated by means of an asymmetric GJR-GARCH-FHS model and then blended together with the physical information non-parametrically. The obtained physical measure, encoding past, present and future information is now fully conditional hence in line with the neoclassical theory. As a demonstration, the full conditionality of the ratio leads to valid EPK compared to the partially conditional ones that appear to be strongly invalid, above all during highly volatile days⁴².

Following a different econometric technique but applying the same principle, Chaudhuri and Schroder (2015)[41] follows the same logic and use options data to analyze the PK violations. Both papers show how, most of the times, the so called PK puzzle⁴³ is just an illusory violation due to the improper use of the market information. This chapter doesn’t aim to extend the proposed models

⁴¹The so-called real-world measure has many different names in literature. Among the others it is also known as the physical, the objective, the personal, the statistical measure just to cite some of the known names. Throughout the chapter we refer to it as either the real-world or the physical measure.

⁴²More details in Sala (2015)[143].

⁴³While the neoclassical theory requires the financial PK to be a monotonically decreasing function in wealth, starting from Jackwerth (2000)[92] and Rosenberg and Engle (2002)[64] some empirical studies found strong violations on different areas of the functional. The so called PK puzzles is a direct consequence of this disagreement between theory and practice. An increasing function in fact leads to a puzzling representation of the investors risk aversion.

empirically, instead it analyzes the theoretical framework that underpins and justifies the use of a fraction of risk-neutral measure at denominator or, equivalently, it shows how the use of a suboptimally informed real-world measure is not enough to produce both a theoretically and an empirically correct EPK. While the literature relative to different empirical violations is vast, the same is not true for the theoretical side. As far as we know, this is the first attempt to analyze how the gap between what is theoretically required and empirically achievable impacts on the EPK and more in general on the fundamental theorems of asset pricing.

4.2 The importance of the Information in Asset Pricing

How to econometrically estimate the real-world measure is still an unanswered problem. At the root of most of problems there is the effective difficulty in capturing the investors' subjective beliefs relative to the investment uncertainty. These beliefs, playing a crucial role in any investment decision, must be part of any investor's filtration. While this is not a problem under the risk-neutral measure (details will follow), it is a big issue under the real-world one. The literature about this problem is huge and growing faster after Ross (2015)[140].

By construction, the today investors' beliefs of a future decision are naturally forward-looking. An investor decides if and how to trade depending, among the others, on her personal beliefs. Therefore, the evaluation of any risky investment must take into account the forward nature of the investors' subjective beliefs. But what is crystal clear theoretically under a mathematical viewpoint is not the same achievable in practice by an econometrician. For example:, counterintuitively with respect to their nature, the markets subjective distribution of future returns are estimated by means of backward looking data (i.e.: Ait Sahalia and Lo (1998)[3]; Jackwerth (2000)[93]; Brown and Jackwerth (2001)[37]; Rosenberg and Engle (2002)[64]; Barone-Adesi et al. (2008)[15]; Yatchew and Härdle (2006)[156]). Relying on these data, an important fraction of the investor's risk and preferences are systematically lost. As a consequence, a discrepancy between what is empirically obtainable and what is theoretically required by the neoclassical asset pricing literature arises. While this is just a fraction of the huge literature that deals with the estimation of the real-world measure and the relative pricing kernel, it clearly emerges that the estimation of a complete filtration set is econometrically a non-trivial problem. In fact, it not only requires an econometrically advanced model in estimation, but also (and more importantly) the availability of particular data. There is in fact no econometrical model that can make backward-looking data forward-looking⁴⁴.

⁴⁴Unless one does not apply strong and usually unacceptable a priori assumptions.

The larger the forward-looking information bias - usually due to higher market volatility- , the larger is the subsequent mispricing.

From a probabilistic viewpoint, this gap is translated into a smaller filtration set:

$$\mathcal{H}_t \subset \mathcal{F}_t \quad \forall t > 0 \quad (4.1)$$

Both sets are increasing⁴⁵ in time and contain all available and potentially usable information:

$$\mathcal{H}_t = \{x_{-\infty}, \dots, x_{t-1}, x_{t-\Delta_t}\} \quad \text{and} \quad \mathcal{F}_t = \{x_{-\infty}, \dots, x_{t-1}, x_t\} \quad (4.2)$$

What makes \mathcal{H}_t smaller is Δ_t : the fraction of missing forward-looking information that involves any risky decision to undertake from today with respect to a future time.

From here on \mathcal{F}_t is used to represent the theoretical information set and \mathcal{H}_t the suboptimal one. In connection with the above example \mathcal{H}_t would be the filtration set obtainable from a stream of past stock returns while \mathcal{F}_t the one obtainable from a cross section of options data. Abusing with words we may claim that, econometrically and from the viewpoint of the information extractable, options data are at least semi-strong form efficient asset classes while stocks are just weak-form efficient. It is well known that while the semi-strong implies the weak form the reverse is not true.

Being the driver of any forecasting model, a suboptimal filtration set has a sure impact on the estimation of any random variable. Future stochastic events are usually modelled by means of conditional expectations under the real-world⁴⁶:

$$\mathbb{E}^P(X|\mathcal{F}_t) = \int_{\mathcal{F}} x dP \quad \text{for each } F \in \mathcal{F} \quad (4.5)$$

where the conditionality is with respect to what is known at time t : the information set. Applying the above generic case to a finance related problem, i.e.: the today prediction of the unknown

⁴⁵Two assumptions underpin this statement: the first is that information is time-varying, the second is that decision makers keep memory of the entire past data.

⁴⁶Let's assume that at time t one wants to forecast the tomorrow's value of a random variable X_{t+1} given the set of available information I_t . This is an optimization problem; more precisely the problem is solved by picking the best predictor among all possible predictors by choosing the one that minimizes the expected quadratic prediction error:

$$\mathbb{E}^P[(x_{t+1} - \hat{x}_{t+1|t})^2|I_t] \quad (4.3)$$

Given the problem, the best (minimum mean squared error: MMSE) predictor is the conditional expectation given the information set:

$$\hat{x}_{t+1|t} \equiv \mathbb{E}^P(x_{t+1}|I_t) \quad (4.4)$$

tomorrow price $P_{t,t+1}$ given, I_t , the information available at time t :

$$P_{t,t+1} = \mathbb{E}^P(P_{t,t+1}|I_t) = \int_{-\infty}^{+\infty} P_{t,t+1} f(P_{t,t+1}|I_t) dP_{t,t+1} \quad \text{for } t < t+1 \quad (4.6)$$

Given the time horizon of the prediction, here $t, t+1$, the above integral⁴⁷ represents the weighted sum of the averaged possible future values, $P_{t,t+1}$, under the physical probability. From (4.6) emerges clearly the importance of a right use of the information set. The average is in fact computed using the I_t conditional probabilities, $f(P_{t,t+1}|I_t)dP_{t,t+1}$. All information available thus enter into the $P_{t,t+1}$ forecast. The poorer the information set, the worse the final outcome.

Given the different informative content of the inputs it follows that while theoretically, under some technical assumptions, the same expectation can be estimated indifferently using either the physical \mathbb{E}^P or the risk-neutral \mathbb{E}^Q measure, in reality this equivalence is often violated. For example, applying and extending under the two measures the above fundamental asset pricing equation (4.6) to compute the today price, $G_{t,T}$, of a generic contingent claim that expires at time T :

$$G_{t,T} = e^{-r_{t,T} \cdot \tau} \mathbb{E}^Q[\varphi_{t,T}(S_T)|\mathcal{F}_t] \quad (4.7)$$

$$= \int_0^\infty M_{t,T}(S_T) \cdot \varphi_{t,T}(S_T) p_{t,T}(S_T|S_t) dS_T \quad (4.8)$$

$$= \mathbb{E}^P[M_{t,T}(S_T) \cdot \varphi_{t,T}(S_T)|\mathcal{F}_t] \quad (4.9)$$

$$= \mathbb{E}^P[M_{t,T}(S_T) \cdot \varphi_{t,T}(S_T)] \quad (4.10)$$

$$\neq \mathbb{E}^{\hat{P}}[\widehat{M}_{t-\Delta_t,T}(S_T) \cdot \varphi_{t,T}(S_T)] \quad (4.11)$$

$$= \mathbb{E}^{\hat{P}}[M_{t-\Delta_t,T}(S_T) \cdot \varphi_{t,T}(S_T)|\mathcal{H}_t] \quad (4.12)$$

$$= \int_0^\infty \widehat{M}_{t-\Delta_t,T}(S_T) \cdot \varphi_{t,T}(S_T) \widehat{p}_{t-\Delta_t,T}(S_T|S_t) dS_T \quad (4.13)$$

$$= e^{-r_{t,T} \cdot \tau} \mathbb{E}^{\widehat{Q}}[\varphi_{t,T}(S_T)|\mathcal{H}_t] \quad (4.14)$$

$$= \widehat{G}_{t-\Delta_t,T} \quad (4.15)$$

where, for each $t \in T$:

$$x = \mathbb{E}^P(x|\mathcal{F}_t) \quad \text{and} \quad \widehat{x} = \mathbb{E}^{\widehat{P}}(x|\mathcal{H}_t) \quad (4.16)$$

to represents the conditional expectation under the optimal (left) and suboptimal (right) information set. For all equations $\tau = T - t$ represents the time-to-maturity, $\varphi_{t,T}(S_T)$ the terminal payoff of the product given the value of the underlying S_T and the PK is defined as:

$$M_{t,T} = e^{-r_t(T-t)} \frac{q_{t,T}(S_T|S_t)}{p_{t,T}(S_T|S_t)} \Big|_{\mathcal{F}_t} > 0 \quad (4.17)$$

⁴⁷The extremes of the integral may also be defined i.e.: put and call options are bounded by their strike prices either above or below.

The inequality in (4.11) is due to the missing but necessary information (relative to T) of the today (time t) risk-physical measure:

$$p_{t,T}(S_T|S_t) \neq p_{t-\Delta_t,T}(S_T|S_t). \quad (4.18)$$

Although the missing information is referred to the future, its impact is absorbed today and propagated onto the tomorrow price forecast. As a consequence, while the left hand side is fully conditional to all of today values, the right hand side is not. Not only, if the PK is extracted from the discounted ratio of the pricing over the physical measure, it is immediate that an improper use of the information set propagates a.s. from the real-world to the risk-neutral pricing equation through the PK, $\widehat{M}_{t-\Delta_t,T}$. The difference in (4.18) is then the focus of the analysis.

It follows naturally that a suboptimal information set has also an impact on the relative probability measures:

$$\mathcal{F} \Rightarrow P \quad \text{and} \quad \mathcal{H} \Rightarrow \widehat{P} \quad (4.19)$$

$$\mathcal{F} \Rightarrow Q \quad \text{and} \quad \mathcal{H} \Rightarrow \widehat{Q} \quad (4.20)$$

where P (resp. \widehat{P}) represents the physical probability measure for a the optimal \mathcal{F} (resp. suboptimal \mathcal{H}) information set. The same logic applies for the risk-neutral measures. The difference between $G_{t,T}$ and $\widehat{G}_{t-\Delta_t,T}$ thus mirrors the distance between an empirically biased asset pricing with respect to the theoretical one. This distance represents the sub-optimality of the information set: is an information premium that distorts the overall risk premium.

In a nutshell: conditional expectations reflect the change in unconditional probabilities given some auxiliary information; the definition and use of this information is then of fundamental importance to correctly evaluate this change. If the filtration used is missing of relevant information, hence is suboptimal, the projection onto the smaller set leads a.s. to an inequality:

$$\mathbb{E}^P(X|\mathcal{F}_t) \neq \mathbb{E}^{\widehat{P}}(X|\mathcal{H}_t) = \int_{\mathcal{H}} x dP \quad \text{for each } H \in \mathcal{H} \quad (4.21)$$

Conditional expectations are of great importance in finance. The definition of a martingale, upon which are based the FTAP, is nothing but a conditional expectation.

4.2.1 Econometric problems behind the estimation of the real-world measure

The PK is a the discounted ratio of measures. It goes by consequence that the single measures and the PK are among them uniquely and tightly interconnected. As advantage, the knowledge of two

of them automatically implies the knowledge of the third one. As a disadvantage, the misestimation of one of the two also loads naturally onto the third one. The literature is abundant of different methodologies, both parametric and nonparametric to estimate the two measures and the EPK⁴⁸, either jointly or separately.

While theoretically the EPK is a decreasing function of aggregate resources, many empirical papers found violations in different areas of the functionals (Aït Sahalia and Lo (1998)[3], Jackwerth (2000)[93], Brown and Jackwerth (2001)[37], Rosenberg and Engle (2002)[64], Yatchew and Härdle[156], Ziegler (2007)[159]). The persistency and robustness of these violations put interest on the investigation of the so-called pricing kernel puzzle. Since then, researchers have taken great interest in proposing different econometric techniques to estimate the EPK and its measures trying to answer to the puzzle. From the literature it emerges that, among the others, the most problematic econometric task to perform is to assure full conditionality to the time-varying estimation of the market's subjective distribution of future returns.

Empirically it turns out that to propose a day-by-day estimation methodology of the measures that compose the EPK is as much important as econometrically non-trivial. The importance of a correct estimation of these measures lies in the wide use of the EPK for many daily operations (i.e.: asset pricing and risk management). The econometric issues, as also partially commented in Bliss and Panigirtzoglou (2004)[28], lie mostly in the nature of the inputs than in the econometric model used for the estimation. Using historical data many estimation methodologies put unreal and theoretically not required stationary assumption on the estimation of the measures or on the EPK itself. Therefore, it is not by chance that many works only propose monthly or yearly estimations remaining silent for the daily ones. It turns out that, from this viewpoint, the main reason behind a biased estimation is not the technique used for the estimation but rather the data used as input. Needless to say, models' inputs are of key importance to determine the type and the quality of the final outputs. Market option prices provide a naturally forward-looking measure; in fact, by contract, the owner of an option has the right, but not the obligation, to exercise it at expiration (or before it, if it is an American or exotic option). This feature is reflected into the option value, which is in fact a non-decreasing function of volatility. Therefore, the market prices of options, through the implied volatilities, encode important forward-looking information about the future distribution of prices of the underlying asset. Also the higher moments of the distribution embed important information. This is particularly true into the tails: where lies most of sentiment.

⁴⁸The nonparametric ones have to be preferred. From Fama (1965)[68] and Mandelbrot (1966)[118] it is well-known that none of these quantities are neither normal nor have a closed parametric form.

The same richness cannot be achieved by stock and index prices. By their contractual nature these assets are options-free and unique, hence poorer from an informative viewpoint.

Estimating the EPK and extracting the risk aversion from stock prices is a well-known problem in the literature. Despite their unambiguous superiority in estimation, it is only from the beginning of the millennium that scholars have begun using options data for estimations (Chernov and Ghysels (2000)[42]). The superiority in estimation of options with respect to stocks (and also futures) is manifold.

First, stock prices have discounting as well as time-horizon problems. By contract definition, stocks do not expire but live infinitely. They are defined over an indefinite time horizon; therefore the discounting process becomes non-trivial. As a consequence, additional assumptions (which often times are unreal, i.e.: on the characteristics of future dividends), are needed to determine the discounted cash flow. On top of that, the obtained final outcome is statistically not much informative being the discounted final cash flow just a single value; this means that no inference about variations in preferences over different time horizons is possible.

On the contrary, option contracts have by definition a bounded life that is defined by a fixed time-to-maturity, T , and is known from the inception of the contract⁴⁹. Moreover, for each time t , the so-called option surface is available: a broad spectrum of times-to-maturity, T_i and strikes K_j ⁵⁰ that covers different states of the world. These characteristics allow for natural inference on preferences over specific horizons and simultaneously over different horizons and strikes.

These features make options qualitatively superior also to futures and forward contracts which, by their nature, do not share the discounting but only the time-horizons problem. In fact, even though these contracts have finite maturities, they do not differentiate across states of the world, thus providing only a single statistic for each expiry date/observation date pair. As for the stocks, having a single data a direct density estimation is not possible without further assumptions⁵¹.

As a consequence, it is possible to directly and naturally estimate a time-varying risk-neutral density from a cross section of options with no need to bound the structure of the data we sample from, but the same is not possible from the time series of the underlying. For the physical density in

⁴⁹Only perpetuity options differ from this characteristic, but are an exceptional case, more theoretical than really used in practice. American and exotic options instead, have the flexibility to be exercised at any time or at fixed time before T . Although non-trivial, a reliable discounting is still possible.

⁵⁰Both indices are finite, i.e. $i = 1, \dots, M$ and $j = 1, \dots, N$.

⁵¹As noted by Jackwerth and Rubinstein (2006)[94], is starting from at least 8 option prices that there are enough information to determine the general shape of the implied distribution. Although possible, to have too a few data in estimation may lead to possibly dangerous small sample bias.

fact, to obtain good results when making inference from a time series of stock returns, one needs to put a priori unrealistic bounds on the time structure of the data. In conclusion: inferring densities from the option surfaces does not share the above cited stationarity problems. While the degree of assumed stationarity can be of different length, in no case can be justified economically

Given the above characteristics, it is natural that just using historical stock data it is not possible to properly capture the investors future beliefs. The difficulty in estimating the objective measure is that it directly depends on the evolution of the underlying process, which time series is only partially informative. Estimations can be further complicated by possible data-problems (i.e.: data scarcity) and market frictions. For these reasons, some authors fully avoid the density estimation and propose a ratio of estimated risk-neutral measure over an unknown physical measure (Golubev et al. (2008))[77].

As (a non-)alternative, some papers propose to increase the degree of conditionality statistically by increasing the rolling window of the estimations. Jackwerth (2000)[93] tries to “enrich” the informative content of the physical measure by working on a longer time series. He proposes to use ten instead of the classical two to five years of data for estimation. As expectable, the obtained results are qualitatively almost unchanged. Due to the intrinsic backward nature of the dataset, to increase the length of the rolling window used in estimation is not a solution. At some extremes it might also lead to even more misleading results. In fact, by using more and more data, there is a natural statistical improvement coming from the higher conditionality of the measure but at the same time this reduces the overall informative content of the last data. Also in the case in which this single stock value would be fully informative with respect to the future scenario of the market, it would become almost negligible and prevailed by the high amount of past data used. Using time series of historical returns thus make the single values almost fully unconditional. Needless to say, the same problem is then naturally translated onto the relative estimations.

While the real-world densities presents several difficulties, it is by now an empirical and theoretical fact that the most important information embedded in financial instruments is the state price density (SPD), or the Arrow-Debreu state prices.

The time-state preference model of Arrow (1964)[6] and Debreu (1959)[7], which proposes the now-named Arrow-Debreu security, models a very basic financial instrument (called pure or primitive security) that pays one unit of numeraire (like a currency or a commodity) on one specific state of nature and zero elsewhere. Passing from discrete to continuous states, Arrow-Debreu securities are defined by the so-called state price density (SPD). Under the continuous framework the security

pays one unit of numeraire x if the state falls between x and $x + dx$ and zero elsewhere. As a consequence of their high informative content, the Arrow-Debreu securities become one of the key element to work with and understand the general economic equilibrium under uncertainty and to determine the price of any contingent claims. For these reasons the estimation of such SPDs has been a very important topic of research within the financial economics community. Once that a complete set of option prices for a specific time-to-maturity is available, there are many parametric and non-parametric methods to recover the risk-neutral measure⁵².

To summarize: to obtain a fully time-varying conditional estimation of the real-world density it is of key importance, more than to pick the right econometric model, to use the right source of information. Being historical stock returns only backward-looking, hence only partially informative, it is needed to somehow complete the measure by using other sources of information, i.e.: the investors' future sentiment extracted from the implied moments of the option surface.

4.2.2 Defining the financial pricing kernel as a Radon-Nikodym derivative

The risk-neutral pricing has the extraordinary advantage of relying on a unique probabilistic measure that is by construction neutral and so applicable to all investors. Unfortunately, if the representative investor's risk attitude is not neutral, the obtained values may provide approximative results. Linking the two measures the PK is then the collector of all beliefs, errors and premiums of the investor. Theoretically it embeds all the relevant and necessary information required to convert the risk-neutral measure into a real-world one and vice-versa. Mathematically it is convenient to express the PK as a discounted Radon-Nikodym derivative. As a consequence the PK is nothing but a discounted kernel: a time adapted operator which, working as a transition function of a stochastic process, allows us to move from a neutral to a subjective world. This section defines the unconditional version of the PK as a Radon-Nikodym derivative, in the next one we make it time-dependent hence more usable in finance.

Proposition 4.2.1. *Defined on a measurable space $(\Omega, \mathcal{F}, \mathbb{F})$, the unconditional version of the Radon-Nikodym derivative requires that:*

1. P is a σ -finite measure on $(\Omega, \mathcal{F}, \mathbb{F})$ and is atomless;

⁵²See Bahara (1997)[11] and Jackwerth (1999)[92] for a complete review of the topic.

2. Q is a σ -finite measure on $(\Omega, \mathcal{F}, \mathbb{F})$ and is absolutely continuous with respect to P on \mathbb{F} :
 $Q \ll P$ ⁵³;
3. the Radon-Nikodym derivative of Q with respect to P , is a nonnegative Borel measurable function M defined on the extended real line ($M : \mathbb{R}^+ \rightarrow \mathbb{R}^+$) and satisfies $P\{x \in \mathbb{R}^+ : M(x) = y\} = 0$ for all $y \in \mathbb{R}^+$.

If assumptions 1 and 2 are satisfied, then there exists an a.s. (with respect to measure Q) random variable M which follows directly from the application of Appendix (A.1) and fits into the last point of proposition 4.2.1. It follows that the unconditional version of the Radon-Nikodym derivative, is defined as:

$$M = \frac{dQ}{dP} \quad (4.22)$$

and satisfies the following:

- $Q(A) = \mathbb{E}^P[M \cdot \mathbb{1}_A]$, $\forall A \in \mathcal{F}$
- $\mathbb{E}^P[M] = 1$,
- $M > 0$, P - a.s.
- $M \in \mathcal{F}$

The obtained value, is a kernel, an operator. Once discounted, equation (4.22) leads to the financial PK. Mathematically, equation (4.22) is also-called the likelihood ratio between Q and P on \mathcal{F} or the change of measure.

The technical requirements of proposition (4.2.1) have all natural economic interpretations and the extension to their time-dependent counterparts is, at least theoretically, immediate. The violations of all of these assumptions lead to puzzling results. It turns out that the second assumption is the most fragile one with respect to a suboptimal use of the information.

Related to the analysis of the PK puzzle, the importance of the atomlessness⁵⁴ of the underlying probability space is also analyzed by Hens and Reichlin (2011)[84]. They show how a complete market equipped with an atomic probability space have some unaccessible bets. As a consequence

⁵³More on the absolute continuity of the measures in Appendix (A.1).

⁵⁴Roughly speaking, atomlessness is linked with continuity and atomicity with discreteness of the probability space. An atomless space put zero mass to a given unit, i.e.: the probability of any particular value is equal to zero. Formally: given a measurable space and two measurable sets: $A \in \Omega$ and $B \subset A$, a set $A \in \Omega$ is called an atom if:

$$P(A) > 0 \quad \text{and} \quad P(B) = 0 \quad \text{where} \quad P(B) < P(A) \quad (4.23)$$

Is atomless when:

$$P(A) > P(B) > 0 \quad \text{where} \quad P(B) < P(A) \quad (4.24)$$

this lead to a continuity of values for the non-atomic measure.

the whole spectrum of payoffs is not achievable. Graphically this is translated into a flex on the PK functional, thus to a violation. The absolute continuity requirement of the measures can be translated in economic term as an absence of dominating trading strategies. If the assumption about the equality of the two measures with respect to the negligible events would be violated, there would exist a set $A \in \mathcal{B}(\mathbb{R}^+)$ with $P(A) = 0$ and $Q(A) > 0$. Although the inequality might apply only to rare events, a violation of the absolute continuity it is all that is necessary for having a sure arbitrage in the market. In such a scenario, a contingent claim with payoff equal to the indicator function for this set would have an a.s. strictly positive price together with an a.s. zero payoff. Shorting such a product would lead to a sure profit with no risk (Dybvig (1988)[58]). If would be possible to exploit such a scenario, the no dominance assumption would be non existing thus showing market inefficiency. It turns out that a proper or improper use of the market information, leading to real or illusory puzzles may lead to misleading conclusion with respect to the efficiencies of the market. The empirical analysis of these scenarios is left for future researches. The third requirement ensures no flatness for the PK, hence a unique ordered ranking. In graphical terms this is nothing but the convexity of the PK.

4.3 The behaviours of local and strict local martingales in finance

The projection of a martingale onto a smaller filtration set is still a martingale, the same holds for super and quasimartingales (Stricker (1977)[150]). Things change for local martingales (Follmer and Protter (2010) [73]) and strict local martingales (Protter (2015)[136]). This is especially true when the process is no longer adapted (hence not measurable) to the smaller filtration.

Due to their widespread use, both local and strict local martingales are of key importance in mathematical finance⁵⁵. Following Johnson and Helms (1963)[103] which refined the concept of “class D” supermartingales and extending the Doob-Meyer decomposition (1954 -1963)[54][122], the notion of local martingale has been created by Itô and Watanabe (1965)[91]. The concept has then been used extensively in finance through the different theories involving the stochastic integration.

Strict local martingales have been introduced in finance by Herrison and Kreps, Harrison and Pliska

⁵⁵Strict local martingales play an important role in different field of the recent finance literature, among the others: for the Benchmark approach of Platen and Heat (2006)[132], for the analysis of relative arbitrage of Fernzhold and Karatzas (2010)[70] and for the understanding of financial bubbles of Heston et al.(2007)[86] and Jarrow et al. (2010)[100].

(1978), (1981)[80][81] and Delbaen and Schachermayer (1994), (1998)[52][53] on their works on the absence of arbitrage and the no free lunch with vanishing risk (NFLVR) condition. More recently, the concept has been extended to models of financial bubbles and to test the validity of stochastic volatility models.

While since the seminal work of Itô (1978) the literature on the expansions of filtrations has grown considerably⁵⁶, the mirroring shrinkage of filtration has not attracted the same degree of attention. It is only recently that the subject has gained more consideration. In mathematical finance, most of the debate is on the existence of financial bubbles⁵⁷ and in credit risk (see, among the others: Jarrow et al. (2007)[99]). As far as we know, the literature is still silent on the behaviour of the PK to filtration shrinkage.

Following the terminology of Elworthy, Li and Yor (1998)[61] a strict local martingale is locally a martingale but not a true martingale while a local martingale is a martingale equipped with a sequence of stopping times.

Loosely speaking, a strict local martingale remains such on a smaller filtration if and only if there is a sequence of reduced (projected) stopping times that is common to both filtrations. In case this is not possible and the process passes from being a totally continuous to a mixed process made of a continuous part and a set of totally inaccessible stopping times. These inaccessibility represents the information lost due to the smaller filtration. This is exactly what happens when one neglects necessary information from the filtration set of the physical distribution. To turn back to a true martingale one needs of an absolutely continuous compensator.

As anticipated above, the PK is nothing but a change of measure. The passage from the physical to the risk-neutral measure is mainly a change of drift which passes from being subjective (usually represented with μ) to be an objective and common value (represented with r). In this case, assuming the risk-neutral measure to be unbiased, it follows that what one observes under a smaller filtration set is “less” than a subjective drift, what is missing are the subjective beliefs that affect the expected returns of the investor. The proper determination of the risk premium is then the key issue. More precisely, the determination of the premium for extreme events (nullsets) will turn out to be the main discriminant for the change of nature of the process.

Having at denominator of the EPK an unconditional physical distribution that is mostly uninformative into the tails not only leads to puzzling results but also to much smaller supports. This

⁵⁶For a review on the subject: see Chapter 6 of Protter (2001)[135].

⁵⁷From a theoretical viewpoint, the existence of financial bubbles is due to the different behaviours of martingales and local martingales to possible filtrations' shrinkage.

is translated in EPKs that, instead of being absolutely continuous processes are required by the neoclassical theory, are continuous process with jumps at the extremes of the functional⁵⁸. These jumps represent the degree of missing information at the extremes of the EPK. As anticipated, while theoretically the PK should guarantee a smooth flow of information among the two measures the same is not so easily achievable in the real-world.

While a theoretically fully correct EPK is not achievable in the real-world, options data give a drastical help in alleviating the informative problems. Completing the information set by means of the implied options moments, it is not only possible to gain from the point of view of the non-monotonicity problem, but it is also possible to reduce (and sometimes eliminate) the number of jumps at the extremes of the distribution. This is particularly true for left tail of the distribution, the one more sensitive to the change of investors sentiment⁵⁹. Linking the empirical analysis to the fundamental theorem of asset pricing we can say that while the absence of market completeness impacts on the validity of the EPK, the flow of information from the two measures gets better including options data into the information set.

4.3.1 Projecting the Radon-Nikodym derivative onto a smaller filtration set

This and in the next sections present the main results of the analysis. As a first step, the previously defined unconditional Radon-Nikodym derivative is made conditional hence more appropriate for different financial applications. The second part of the section analyzes how different semimartingales behave once projected onto coarser filtration sets. Restricting the analysis to the conditional version of the above results, also the link with the fundamental theorems of asset pricing gets more natural.

⁵⁸It is a fact that the PK is moving and, by the Hansen and Jagannathan bounds (1991)[78], it needs of a high enough variance to be valid. Anyhow, this is not to be confused with irrational jumping values which lead to economically invalid results.

⁵⁹Empirically, the right tails of the distribution often suffer of low liquidity due to the lower trading of call options. Therefore, explosive values in the right tails is not due to a lack of forward-looking information into the information set but is a lack of the entire information set. The same happens for the left tails, but with a lower frequency.

From (4.22) we recall without proving that:

$$Q_t(A) = \mathbb{E}^P[M_t \cdot \mathbf{1}_A], \quad \forall A \in \mathcal{F}_t \quad (4.25)$$

$$= \int_A M_t dP_t \quad (4.26)$$

then, certainly $Q_t \ll P_t$ for all $t \in T$. The Radon-Nikodym theorem goes on the opposite direction:

Theorem 4.3.1. *If P_t and Q_t are two σ -finite measures on $(\Omega, \mathcal{F}_t, P, \mathbb{F})$ and $(\Omega, \mathcal{F}_t, Q, \mathbb{F})$ such that $Q_t \ll P_t$ on \mathcal{F}_t for all $t \in T$, then there exists a sequence of strictly positive M_t :*

$$M_t = \left. \frac{dQ}{dP} \right|_{\mathcal{F}_t} > 0 \quad \forall t > 0 \quad (4.27)$$

such that $Q_t(A) = \int_A M_t dP_t$ for all $A \in \mathcal{F}_t$.

From theorem (4.3.1), it follows that:

Proposition 4.3.1. *If $(M_t)_{t>0}$ is an \mathcal{F}_t -adapted sequence of strictly positive martingales on a complete and filtered probability space, with i.e. $\Omega = \mathbb{R}^T$ and $\mathbb{E}^P[M_t] = 1$, then $(M_t dP_t)_{t>0}$ is a consistent family of probability measures on $(\mathbb{R}_t)_{t \in T}$.*

From Harrison and Kreps (1979) [80], the new measure satisfies the necessary conditions to be an equivalent martingale measure (EMM), namely the probabilistic equivalence of the measures, the existence of a non-negative value from the Radon-Nikodym derivative and the martingale property of the price process under the change of measure. This leads to the first FTAP which states that a market has no free lunches with vanishing risk (NFLVR) if and only if there exists an EMM. For a proof, see [52] [53]. While the existence of the measure is not at risk under information reduction, the same is not true for the nature of the martingale that may become a martingale only locally.

Given their structure, it is known that some categories of semimartingales remain such once projected onto a coarser and adapted filtration hence are robust with respect to a lesser amount of information:

Theorem 4.3.2. *Let M_t be a martingale/supermartingale/quasimartingale on $\mathbb{F} = (\mathcal{F}_t)_{t \in T}$, then their optional projections⁶⁰ onto the smaller filtration $\mathbb{H} = (\mathcal{H}_t)_{t \in T}$ are still a martingale/supermartingale/quasimartingale for the filtration \mathbb{H} .*

⁶⁰Any nonanticipating and right continuous (càdlàg) process is an optional process. If $M = (M_t)_{t \geq 0}$ is a bounded and measurable stochastic process, then the optional projection \widehat{M} of M is the a.s. unique (and bounded) process such that for any stopping time τ :

$$\mathbb{E}^P(M_\tau \mathbf{1}_{\tau < \infty}) = \mathbb{E}^{\widehat{P}}(\widehat{M}_\tau \mathbf{1}_{\tau < \infty}) \quad (4.28)$$

Following [73] and exploiting the interconnection among the three classes of semimartingales, the proofs of the above results are presented all together.

Proof. Martingale: If $s \leq t$:

$$\mathbb{E}^{\widehat{P}}(\widehat{M}_t | \mathcal{H}_s) = \mathbb{E}^{\widehat{P}}(\mathbb{E}^{\widehat{P}}(M_t | \mathcal{H}_t) | \mathcal{H}_s) = \mathbb{E}^{\widehat{P}}(M_t | \mathcal{H}_s) \quad (4.29)$$

$$= \mathbb{E}^{\widehat{P}}(\mathbb{E}^P(M_t | \mathcal{F}_s) | \mathcal{H}_s) = \mathbb{E}^{\widehat{P}}(M_s | \mathcal{H}_s) \quad (4.30)$$

$$= \widehat{M}_s \quad (4.31)$$

where the a.s. equality of $\mathbb{E}^{\widehat{P}}(M_t | \mathcal{H}_t) = \widehat{M}_t$ is a direct consequence of the uniform integrability of the “class D” processes $(M_t)_{0 \leq t \leq n}$ where n is arbitrarily extended to ∞ .

Supermartingale: Here the key element is the non-negativity of M_t . From the FTAP, a necessary condition for the existence of the PK is its strict non-negativity. It follows that:

$$\mathbb{E}^{\widehat{P}}(\widehat{M}_t | \mathcal{H}_s) = \mathbb{E}^{\widehat{P}}(\mathbb{E}^{\widehat{P}}(M_t | \mathcal{H}_t) | \mathcal{H}_s) = \mathbb{E}^{\widehat{P}}(M_t | \mathcal{H}_s) \quad (4.32)$$

$$= \mathbb{E}^{\widehat{P}}(\mathbb{E}^P(M_t | \mathcal{F}_s) | \mathcal{H}_s) \leq \mathbb{E}^{\widehat{P}}(M_s | \mathcal{H}_s) \quad (4.33)$$

$$= \widehat{M}_s \quad (4.34)$$

Quasimartingale: A quasimartingale is just the difference of two supermartingales:

$$A_t = B_t - C_t \quad (4.35)$$

where B_t and C_t are positive and right continuous supermartingales and A_t is a quasimartingale iff there exists a constant D_t such that:

$$\sup \sum_{1 \leq i \leq n} \mathbb{E}(|M_{t_i} - \mathbb{E}(M_{t_{i+1}} | \mathcal{F}_{t_i})|) \leq D_t \quad (4.36)$$

where the supremum is taken over all finite sets of the partition $t_1 < t_2 \cdots < t_n < t_{n+1}$ ⁶¹. As a consequence of the previous proof the result follows. \square

Martingales, supermartingales and quasimartingales are all examples of semimartingales. Theorem (4.3.2) extends the Stricker’s theorem (1977)[150] for non-adapted semimartingales.

⁶¹Dealing with quasimartingales the partition is usually over $[0, \infty]$ such that: $0 = t_1 < t_2 \cdots < t_n < t_{n+1} = \infty$. The inclusion of ∞ in the index set makes it homeomorphic to $[0, t]$ for $0 < t \leq \infty$.

4.3.2 Filtration shrinkage and information inaccessibility for strict local martingales

Due to the empirical challenges presented in the previous sections (4.1 and 4.2.1) if one uses just stock data for the estimation of the real-world that governs the EPK, it has not only to deal with two different filtrations but also with the possibility of having illusory puzzles. Both are a direct consequence of the violation of the neoclassical theory that requires to have the same set of information for both the real-world and the risk-neutral measures.

Not sharing the same nullsets, the missing information makes the relative EPK unstable, particularly at the extremes. Instability that is caused by the new nature of EPK. In fact, differently than the previous classes of semimartingales, local and strict local martingales do not share the same degree of stability. Projected onto a smaller filtration set the finite variation part may acquire a drift which could be singular (with respect to $d(\widehat{M}, \widehat{M})_t$). While the local part represents the randomness of the process, the finite variation represents its “drift” thus, by consequence, the risk premium that governs the change of measure. It follows that the optional projections of local and strict local martingales need not be local and strict local martingales anymore and the stability of the process is no longer guaranteed but mined by the difficulty of estimating properly the risk premium in the real-world under information reduction. This shows why if an econometrician uses wrongly a suboptimal information set it may be affected by possible mispricing and “relative” arbitrages.

The different behaviours of the diverse classes of semimartingales and their relation with the EPK can be demonstrated applying and extending the recent findings of Protter (2015)[136]. Starting from the concept of total inaccessibility, let’s firstly define when a filtration set is significantly smaller. From a probabilistic viewpoint, a totally unaccessible stopping time is nothing but the flipping side of a a predictable stopping time:

Definition 4.3.1. *A map $\tau : \Omega \rightarrow \mathbb{R}^+$ is a predictable stopping time if there exists an announcing sequence of stopping times τ_n where a.s. and for each n :*

$$\tau_{n+1} \geq \tau_n, \quad \tau > \tau_n, \quad \lim_{n \rightarrow \infty} \tau_n = \tau \quad (4.37)$$

We define a totally unaccessible stopping time, TU , if for any predictable stopping time τ :

$$P(\tau = TU) = 0 \quad (4.38)$$

Examples of totally inaccessible stopping times include the jump times of the Poisson processes.

Theorem 4.3.3. *A sub-filtration is significantly (and sufficiently) poorer with respect to its larger counterpart if it has less stopping times and/or if some of them are present but totally inaccessible.*

Keeping the interconnection between theory and empirics alive, an empirical technical detail may help to clarify. Given this framework, the smaller filtration set is not necessarily as such because it contains less data, but because these data (i.e.: stock returns) are qualitatively less informative than the ones forming the larger set (i.e.: options data).

The effect of not capturing the forward-looking information of the investor, as instead required by the neo classical theory, is that \mathbb{H} has a significantly and/or qualitatively lower amount of information. With respect to the above definition it means that the smaller filtration set loses some or many \mathbb{F} stopping times and some of the remaining become totally inaccessible. Most of them lie into the tails of the distribution⁶², where sentiment matters the most. A smaller filtration set is then defined as sub-optimal whenever it misses something required by the neoclassical theory i.e.: the forward looking beliefs for the real-world measure and the relative EPK.

Theorem (4.3.3) clarifies and better characterizes the value of Δ_t defined in (4.2) and brings us naturally to the following theorem:

Theorem 4.3.4. *If M_t is a sequence of absolutely continuous strict local martingales defined on $(\Omega, \mathcal{F}, \mathbb{F})$ for each $t \in T$, its projection onto a significantly and sufficiently poorer smaller filtration \mathbb{H} became a mixed process with random jumps:*

$$M_t = \mathbb{E}^P(M|\mathcal{F}_t) = \frac{dQ}{dP} \Big|_{\mathcal{F}_t} \neq \frac{d\widehat{Q}}{d\widehat{P}} \Big|_{\mathcal{H}_t} = \mathbb{E}^{\widehat{P}}(M|\mathcal{H}_t) = \widehat{M}_t \quad \forall t > 0 \quad (4.39)$$

where \widehat{Q} and \widehat{P} are the optional projections of P and Q onto \mathcal{H} .

For a local martingale, to be still such under a smaller filtration, it needs a reducing sequence of stopping times that is common for both filtrations. The extension for strict local martingales is mainly a technical detail. The main message is the same: if the two processes do not stop at common points and if these points are not qualitatively equals they cannot share the same properties. Comparing theorem (4.3.2) and (4.3.4) it is immediate to notice that the reason of the

⁶²Although the larger effects are usually manifested into the tails, the inability of capturing meaningful information Δ_t also appears into the central area of the EPK. The term pricing kernel puzzle has been originated to explain the non concavity of the functional into the central area of the graph.

instability of the projected processes lies on the reproducibility of the set of stopping times. Since under a sufficiently poorer information set $M_t \neq \widehat{M}_t$ and since the risk-neutral measure is assumed to be unbiased from an informative viewpoint⁶³ $\widehat{Q}_t = Q_t$, it follows naturally that the reason of the inequality is to be found into the projected objective measure $\widehat{P}_t \neq P_t$. What is determining for the change of nature of the pricing kernel is the difference in nullsets for the two measures. While under a theoretical set-up the risk neutral measure is absolutely continuous with respect to the physical one, the lack of crucial information in the latter produces a different view of the extreme events. Not sharing the same nullsets the two measures are then not absolutely continuous, a crucial requirement that impacts on the nature of the real-world projected pricing kernel.

It is a known fact that a strict local martingale is not a good candidate as a “potential” change of measure (Kreher and Nikeghbali (2011)). Even more importantly: from an economic viewpoint is when the nullsets are not common that the Dybvig theory can apply. Is therefore enough to find a non concavity in the EPK to set up his dominance strategy. Albeit true in principle, if the missing concavity is mainly due to a suboptimal use of the market information, the dominance of the contingent claim trading strategy will turn out to be only illusory hence only costly for whom will set it up. An augmentation of the filtration along the stopping times is needed to convert the strict local martingale into a valid martingale. The missing information under P is than what one needs to insert to complete the filtration. Or, more precisely, the missing risk component of the risk premium needs to be added so that the two measure can share again the same nullsets. The previously analyzed lower ability of channeling all the information from the stock market to the theory is here explained from the viewpoint of a suboptimal EPK.

To summarize in economic terms, the missing information (risk premium) not captured by a biased physical distribution leads to a sufficiently poorer filtration set, thus to missing stopping times that changes the nature of the EPK.

4.4 Empirical Application

As a final brick for our theoretical-empirical bridge, this section provides a graphical demonstration that the use of options market data, underpins and justifies what just defined theoretically. While the neoclassical literature requires the PK to be monotonically decreasing, in reality this is hardly achievable. Following Sala (2015)[143] here is shown how option data encodes crucial information that, if omitted, may bring to strong economical violations or, more precisely, to illusory puzzles

⁶³This sentence implicitly assumes the possibly non true condition of absence of market inconsistencies.

and illusory stochastic dominances. This point is analyzed comparing two informatively different EPKs. The former, the unconditional EPK, is not totally informative in the physical measure. The missing information of the real world measure is due to its suboptimal estimation which is based upon the use of stock market data only. The latter instead, the conditional one, uses both options and stock data for the estimate the physical measure. While of course both of them are still not totally unbiased with respect to the theory, the latter shows much smaller violations. The conditional one in fact, has an overall economically valid shape. What still makes it not perfect is the natural noise that arises from using real data. The link with the theory is immediate: the missing information at denominator - the unaccessible stopping times - are the cause of these violations thus showing how the econometric inability in using the market data may lead to illusory violations.

Figure (A.24), taken from [143], shows and compare two conditional and unconditional EPKs estimated from a random day, namely October 9, 2002. The left column shows the estimated densities, the right one the relative EPKs obtained taking the ratio of these same densities. The figure is also divided horizontally to show the entire spectrum of times-to-maturity that increases as it goes down starting from the shortest on top. The only difference among the two methodologies is that the physical measure of the unconditional EPK is estimated by fitting a GJR GARCH model onto 3500 daily stock returns only while the physical measure of the conditional EPK is made more informative by blending the same physical measure with the risk-neutral one, thus encoding the forward looking information of the option market into the measure⁶⁴. Both estimations are proposed with Gaussian and empirical innovations (FHS) thus providing a total of four different EPKs. For the densities, the dotted lines represent the risk-neutral and the unconditional physical measures while the continuous ones the revised conditional physical measure. For the EPKs the continuous (dotted) lines represent the conditional (unconditional) EPKs.

When the amount of data is high enough so that a meaningful statistical inference is possible, the conditional functionals are almost everywhere well behaved while the unconditional ones explode on both sides of the functional. As natural, the supports of both the distributions and the EPK get smaller as it moves toward shorter times-to-maturity. This is an expected result that is naturally related to degree of moneyness that gets smaller for short-living options. On the contrary, where the liquidity of the surface gets higher along the entire horizontal support, the information

⁶⁴The risk-neutral measure is extracted by minimizing the distance between real and simulated options market data.

extracted from the options' implied moments makes the stopping times accessible thus avoiding the economical inconsistencies of the exploding extremes. It is worth to emphasize that, to improve the readability of the graph, most of the unconditional EPK values have been cut as soon as too volatile. This happens as soon as the denominator of the function is not enough informative, hence with totally unaccessible stopping times.

To gain more insight from the analysis, the top panel of figure (A.26) shows another EPK estimation - here September 13, 2003 - but from a closer look, just focusing on a single time-to-maturity (94 days). To facilitate the analysis, the horizontal and vertical lines shows respectively the risk-neutral area and the current level of the market index. For this day and time-to-maturity is clearly visible from the functionals represented by the dotted lines how a suboptimal use of the information loads onto the left area of the EPK making the estimations very volatile. Due to the the missing information, the negligible or almost zero value of the real-world measure, the EPK represented by the blue dotted line explodes also for values that are not far out-of-the-money. While the use of empirical innovations alleviate the problem for the unconditional EPK, results are still puzzling and may be related with the apparent overpricing of put-options. Both conditional values have instead a more behaved monotonically decreasing behaviour. Bottom panel of figure (A.26) represents again the October 9, 2003 EPK but for a single time-to-maturity (346 days). As visible, the puzzling behaviour of the suboptimal EPK is on the right side of the functional thus showing the inability of extracting the information from the call options. Here it is clearly visible how the inaccessibility of physical measure is translated on the right side of the figure. While the partially conditional EPKs have exploding values thus producing a U shaped functional, the information extracted from the options produce an EPK that is in line with the neoclassical theory. The choice of these two figures is to show how the use of the options help in alleviating possible violations on both sides of the EPK

4.4.1 Extending the findings to the risk-neutral measure

We have shown how a suboptimal estimation of the information manifests in a smaller filtration set that affects strongly the validity of the measures that compose the EPK thus the EPK itself. As defined in proposition (4.3.1), the martingale property of the Radon-Nikodym derivative process is crucial for a correct pricing under the risk-neutral approach. It is in fact a necessary condition that makes sure that the risk-neutral measure has a total mass of one, allowing it to be a proper probability measure.

From this property it follows that if and only if the PK is a positive⁶⁵ martingale process, there exists a valid risk-neutral measure: the previously defined EMM. From the definition of the strict local martingale, it follows that if the PK is only locally a martingale but not a true martingale, it is not guaranteed that the related risk-neutral measure Q is a valid probability measure. Due to the central role of the PK and the measures that compose it, these different relations pose challenges in many areas of financial economics. As a demonstration, at the heart of the existence of financial bubbles and illusory arbitrages, there is the above cited relationship. In fact, it is when two investors are differently informed that one ends up valuing an investment as a martingale and one as a strict local martingale. The disalignment may lead to illusory mispricing, hence to non existing arbitrages and to bubbles.

Moreover, the entire risk neutral valuation theory assumes implicitly the existence of an efficient market. Either the incapability of proposing a fully informative measure or the possibility of finding alphas in the market would suggest to apply different pricing techniques than the risk neutral ones. An inefficient market in fact may lead to illusory arbitrages or bad hedging under the pricing theory (Jarrow (2013)[97]). It thus makes sense to test for the existence of bubbles before using these pricing techniques (i.e. following Jarrow, Kchia and Protter (2011)[95]). Whereas what it is showed empirically only partially explains this phenomenon, it surely gives an important intuition on this direction. What presented is indeed an empirical estimation that, although it substantially improves the degree of conditionality and richness of the information set, it is not immune to errors. Being estimated with a numerically intensive method and using noisy real data the model can only produce approximated final results. It is a classical result in numerical mathematics that algorithmic and data errors lie at the hearth of any numerical model that tries to describe a real-world phenomenon. Moreover, specifically for the estimation of the EPK there are different small sample bias (Leisen (2015)[109]) which may affect both the parametric and non-parametric estimation methodologies used in estimation. Therefore, we do not claim that the proposed conditional EPK is, from a theoretical point of view, a true martingale. Surely, results show that better estimation for a much larger fraction⁶⁶ than the proposed unconditional EPK. What is proposed is just a less biased EPK. As a main message, from the evaluation it follows that a better informed physical measure leads to an informatively less biased EPK.

Due to the tight interconnection among the measures and the PK, a proper estimation of the physical measure is of fundamental importance if one would extract the risk-neutral measure from

⁶⁵Positivity is only a requirement to be a valid PK under the FTAP.

⁶⁶This is true both horizontally for the domain of the gross returns and vertically for the domain of the EPK.

the estimated PK⁶⁷.

4.5 Fully and partially dynamic models, a comparison

If data would be independent and identically distributed (i.i.d.) there would not be any problem in comparing a conditional with an unconditional estimation since the two would be the same. In practice, finance data are highly non i.i.d. and conditionality is needed for a proper estimation. In this section we analyze the impact of comparing a conditional with an unconditional model for the basic pricing equation from consumption model. Given its centrality in asset pricing the impacts of the findings are wide and can affect different models i.e.: the mean-variance frontier, the CAPM, the APT and so on. Although of big interest we do not analyze these impact here and we leave it for future researches.

Let us recall the main finding of chapter (3): the PK is nothing but a generalization of the MRS which arises from the most basic consumer problem (CP). Modelling inter-temporally an investor by means of a utility function $U(\cdot)$ defined over current t and future $t+1$ scenarios s , the derivation of the investor's first order conditions leads to the basic consumption model. Let us recall the CP: an investor aims to maximize her welfare given her feasible budget set. Therefore, the resolution of the CP is the resolution of a constrained maximization which comprehends a conditional expectation under P . A wrong conditioning of this expectation with respect to a smaller information set leads to an overestimation of the MRS. In this chapter we quickly review the basic derivation of the investor's first order condition, then we modify it under a smaller information set. The result obtained is an indirect quantification of the information lost from the bad conditioning. What we analyze is then not a comparison of a dynamic versus a static model, but a comparison of a dynamic versus a partially dynamic model. While the former (of course) has a sharper impact, the latter is very often present and not considered in the finance literature. Cochrane (2001)[48] analyzes the former problem. He compares conditional and unconditional models. Among the others he proposes to use instrumental variables to move from a static to a dynamic model. At page 132 "Of course, another way to incorporate conditioning information is by constructing explicit parametric models of conditional distribution". Unfortunately it is known that a wrong parametrization could

⁶⁷The risk-neutral measure is biased if derived from a biased physical measure. In our case, the risk-neutral density is estimated using a cross section of market option data, hence independently from the risk-physical measure obtained from a time series of stock returns. As a consequence, unless the market data are mispriced, its estimate is unbiased.

lead to even more misleading results than using unconditionally a conditional model. What we propose in the next chapters (both theoretically and empirically), is a new non-parametric model to incorporate the missing information. Here we just consider the theoretical impact.

4.5.1 A conditional version of the basic consumption based model

We assume a generic utility function $U : \mathbb{R}^+ \rightarrow \mathbb{R}^+$ satisfies the properties listed in (3.1.3). The future welfare of a generic rational and small investor is defined as a two-period model:

$$U(c_t, c_{t+1}) = u(c_t) + \delta \mathbb{E}^P[u(c_{s_{t+1}})|I_t] \quad (4.40)$$

$$= u(c_t) + \delta \sum_s (u(c_{s_{t+1}})p_t(s_{t+1}|s_t)) \quad (4.41)$$

where: c represents consumption such that c_t is the today consumption and $c_{s_{t+1}}$ is the consumption for a given state s at the future time $t + 1$. The randomness of the future consumption is estimated by \mathbb{E}_t^P which is a subjective and conditional expectation. The same expectation can be translated as the sum, across states, of all feasible conditional future probabilities: $p_t(s_{t+1}|s_t)$. As required by the neoclassical theory, the investor's first order conditions are described in terms of conditional moments. These are visually represented by the subscript in the expectation \mathbb{E}_t but too often this condition is violated.

In this section we modify the definition of chapter (3) to underline the consequence of this omitted conditions. To emphasize, also visually, the conditioning of the expectation with respect to time t we omit the subscript t and we replace it by means of $(\cdot|I_t)$ where, under the neoclassical theory, I_t represents all the relevant and necessary information that affects the investor choice with respect to her welfare maximization.

Introducing ξ_t as the amount of asset the investor chooses to buy ⁶⁸ at time t , e as the agent's original wealth (endowment) and P_t as the price of the asset at time t with payoff $\psi_{s_{t+1}}$ ⁶⁹, then the two periods investor's optimization problem is:

$$\max_{\xi_t} \{u(c_t) + \delta \mathbb{E}^P[u(c_{s_{t+1}})|I_t]\} \quad (4.42)$$

subject to the budget constraint at time (t) and the Walrasian property at time ($t + 1$):

$$c_t = e_t - P_t \xi_t \quad (4.43)$$

$$c_{s_{t+1}} = e_{s_{t+1}} + \psi_{s_{t+1}} \xi_{s_{t+1}} \quad (4.44)$$

⁶⁸It is assumed that at each time t investors can buy/sell as much as they want.

⁶⁹The payoff is the price at time $t + 1$ plus any dividend, if exist.

Solving the optimization for P_t :

$$P_t = \mathbb{E}^P(P_t|I_t) \quad (4.45)$$

$$= \mathbb{E}^P \left[\underbrace{\delta \frac{u'(c_{s_{t+1}})}{u'(c_t)}}_{MRS} \psi_{s_{t+1}} | I_t \right] \quad (4.46)$$

$$= \mathbb{E}^P \left[\underbrace{\delta \frac{u'(c_{s_{t+1}})}{u'(c_t)}}_{MRS} \psi_{s_{t+1}} \right] \quad (4.47)$$

which leads to:

$$P_t = \mathbb{E}^P(P_t|I_t) \quad (4.48)$$

$$= \mathbb{E}^P[m_{s_{t+1}} \psi_{s_{t+1}} | I_t] \quad (4.49)$$

$$= \mathbb{E}^P[m_{s_{t+1}} \psi_{s_{t+1}}] \quad (4.50)$$

Equations (3.22) and (4.50) refers to a complete information set. Empirically, this theoretical requirement is very often violated. In most cases, the forward looking behaviours of the investor are not fully captured, thus producing a suboptimal information set I^{Sub} :

$$I_t^{\text{Sub}} < I_t \quad \forall t \in T \quad (4.51)$$

Under this information set, we have to set up a new optimization, which, as expectable, produces a sub-optimal MRS:

$$\widehat{P}_t = \mathbb{E}^{\widehat{P}}(P_t | I_t^{\text{Sub}}) \quad (4.52)$$

$$= \mathbb{E}^{\widehat{P}} \left[\underbrace{\delta \frac{u'(c_{s_{t+1}})}{u'(c_t)}}_{MRS} \psi_{s_{t+1}} | I_t^{\text{Sub}} \right] \quad (4.53)$$

$$= \mathbb{E}^{\widehat{P}} \left[\underbrace{\widehat{\delta} \frac{u'(\widehat{c}_{s_{t+1}})}{u'(\widehat{c}_t)}}_{\widehat{MRS}} \psi_{s_{t+1}} \right] \quad (4.54)$$

where we used the convention that:

$$\widehat{x} = \mathbb{E}^{\widehat{P}}[x | I^{\text{Sub}}] \quad (4.55)$$

Therefore, at each time $t \in T$ difference between the two prices is a value, say Δ , which represents indirectly the extra information missing into the non fully dynamic model:

$$P_t - \widehat{P}_t = \mathbb{E}^P(P_t|I_t) - \mathbb{E}^{\widehat{P}}(P_t|I_t^{\text{Sub}}) \quad (4.56)$$

$$= E^P \left[\underbrace{\delta \frac{u'(c_{s_{t+1}})}{u'(c_t)}}_{MRS} \psi_{s_{t+1}} | I_t \right] - E^{\widehat{P}} \left[\underbrace{\delta \frac{u'(c_{s_{t+1}})}{u'(c_t)}}_{MRS} \psi_{s_{t+1}} | I_t^{\text{Sub}} \right] \quad (4.57)$$

$$= E^P \left[\underbrace{\delta \frac{u'(c_{s_{t+1}})}{u'(c_t)}}_{MRS} \psi_{s_{t+1}} \right] - E^{\widehat{P}} \left[\underbrace{\widehat{\delta} \frac{u'(\widehat{c}_{s_{t+1}})}{u'(\widehat{c}_t)}}_{\widehat{MRS}} \psi_{s_{t+1}} \right] \quad (4.58)$$

$$= \Delta_t \quad (4.59)$$

Equivalently:

$$P_t - \widehat{P}_t = \mathbb{E}^P(P_t|I_t) - \mathbb{E}^{\widehat{P}}(P_t|I_t^{\text{Sub}}) \quad (4.60)$$

$$= \mathbb{E}^P[m_{s_{t+1}} \psi_{s_{t+1}} | I_t] - \mathbb{E}^{\widehat{P}}[m_{s_{t+1}} \psi_{s_{t+1}} | I_t^{\text{Sub}}] \quad (4.61)$$

$$= \mathbb{E}^P[m_{s_{t+1}} \psi_{s_{t+1}}] - \mathbb{E}^{\widehat{P}}[\widehat{m}_{s_{t+1}} \psi_{s_{t+1}}] \quad (4.62)$$

$$= \Delta_t \quad (4.63)$$

A good financial modeller should then minimize Δ_t as much as possible. It turns out that to minimize Δ_t is equal to optimize (maximize) the information set as much as possible. As presented in the next chapters, Δ_t is what we will reduce through the use of other external sources of information (i.e.: option data).

A small remark for completeness. We usually assume that the set with more information is the more informative one, such that $P_t - \widehat{P}_t > 0$. As a remote possibility, it might turn out that the forward looking behaviour of the investors (usually captured by means of the intrinsic volatility of the option prices) might produce biased future expectations. In such a case, the smaller information set would turn out to be the less biased. We still consider it a possible but very remote case.

4.6 From consumption to portfolio optimization problems, the conditional case

Although from an economic prospective consumption-based models would be the most comprehensive answer to all asset pricing questions, they are not fully reliable. At the origin of their biases there are the high difficulties concerning a proper estimation of the investors' consumption. To overcome the issue, a natural alternative approach to the classical Lucas asset pricing model is the following: given an economy where one can observe the prices of the available assets and try to model the subjective distribution of their final payoffs, what is the optimal portfolio for a small and rational investor⁷⁰?. Although the problem is very well-known in literature, not much emphasis has been put on the frequent biases that can arise whenever the investors' subjective beliefs that compose her distribution are suboptimally estimated.

Extending Øksendal (2005)[125] we answer to the problem starting from a simple economy described by a Levy-Itô mixed model and we analyze how the obtained outputs may impact on the total profitability of the investors.

The first part of the chapter presents the model. Then we apply the presented model to maximize the terminal wealth of a small and rational investor with logarithmic or power utility function showing the existence of an information premium that is then confirmed non-parametrically by means of the cross entropy. Finally, after an examination of the impacts on the optimal bounds for the PK, we make use of the Girsanov theorem to extend the impact of the findings to the risk-neutral measure passing from the PK itself.

4.6.1 Presenting the problem

To answer to the above problem we model a simple economy made of two assets and we study the impact of a smaller information set onto a rational investor that, endowed with a positive initial capital, wants to maximize her final welfare choosing among the set of admissible portfolios:

$$\Lambda(x_t) = \sup_{\xi_t \in \mathcal{A}_{x_t}} E^P[U(X^{\xi_t}(T))] \quad (4.64)$$

Under this framework, the conditionality of the expectation in (4.64) is represented by the set of feasible portfolios \mathcal{A}_{x_t} where \mathcal{A} represents the set of admissible portfolios and its subscript \mathcal{I}_t

⁷⁰For small investor we define an unsophisticated investor that cannot affect the market prices with her trading. Rationality is defined under the neo-classical theory as a risk averse investor that always prefer more to less.

restricts this set to the information available at time t . We will compute and compare the same optimization problem conditional to the complete and a suboptimal information set. Given a fixed time window, $[0, T]$, and for each $t \in T$ the optimal portfolio weight the investor, conditional to the information set (if it exists), is represented by ξ_t , so that $\Lambda_t < \infty$ is the optimal value of the problem. Both random variables are then \mathcal{I}_t -adapted stochastic processes.

Equation (4.64) is nothing but a consumer problem (CP) where the optimization is with respect to some financial asset class. As a consequence x_t is a financial instrument that proxies consumption: i.e. wealth. The resolution of the optimization problem is linked to the arbitrary choice of a regular utility function $U(\cdot) : [0, \infty] \rightarrow [-\infty, \infty]$ ⁷¹. For simplicity we start with a generic logarithmic utility function⁷²:

$$U(x_t) = \ln x_t \quad x_t > 0 \quad (4.65)$$

The resolution of the asset pricing problem (4.64) leads to the optimal expected logarithmic utility of the investor terminal wealth $X^{\xi_T}(T)$.

The entire optimization is performed under the physical measure, P , and conditional to the filtration set of the investor, \mathcal{I}_t . This justifies our partial information approach. For the problem it assumed a fixed time frame $0 \leq t \leq T \leq \infty$ such that, for each time $t \geq 0$:

$$\mathcal{I}_t = \mathcal{H}_t \subset \mathcal{F}_t \subset \mathcal{F} \quad (4.66)$$

It follows that, depending on the filtration in use, the degree of adeptness (measurability) of the stochastic process may change thus impacting in various form on the final outcomes of the optimization. As a technical remark: both filtrations lie in filtered probability spaces:

$$(\Omega, (\mathcal{H}_t)_{0 \leq t \leq T}, P, \mathbb{H}) \subset (\Omega, (\mathcal{F}_t)_{0 \leq t \leq T}, P, \mathbb{F}) \quad (4.67)$$

and satisfy the usual hypotheses.

The goal of the presented framework is to model the economy of a single investor with a suboptimal information set, where the stochastic processes that affect the optimal portfolio choice are \mathcal{H}_t and not \mathcal{F}_t -adapted, thus reflecting the poorer decisional power of the investor due to the lack of forward looking information.

⁷¹A utility function is assumed to be regular if it is differentiable, concave and non-negative.

⁷²With no loss of generality our results apply to different types of utility functions. Our main goal is to provide an intuitive explanation of the problem more than just a set of mathematical solutions of the optimization. Therefore, we keep the model one-dimensional and we only study the cases of log and power utility functions. Although much less explicit, similar results can be achieved with other utility functions and in a multidimensional framework.

4.6.2 The Lévy-Itô model and the portfolio optimization problem

Probably as a direct consequence of the much larger theoretical literature relative to the enlargement of filtration with respect to the one relative to the shrinkage of filtration, the same degree of richness is reflected on their different applications, i.e.: portfolio optimization. Equation (4.64) mirrors the well-known stochastic control problems related to the insider information (i.e. see Biagini and Øksendal (2005)[24]). While the insider information cases are characterized by an informed trader that has a larger information set with respect to the one of the “honest” trader, here things go on the opposite directions. In both cases extra assumptions and computational tools are required. Differently than the insider case, the literature for these problems is far more scarcer.

Following [125] our starting point is a Lévy-Itô market model composed by two assets: one risky and one risk-free. Given this simple economy⁷³, the investor implements her portfolio through a dynamic trading of the two assets. The choice of modelling the stochastic part of the risky investment by means of a Brownian motion (Itô process) and a pure jump process (Lévy process) instead of just using the classical diffusion-Itô model is justified by the higher descriptive power of the former. For modelling details we refer to Barndorff-Nielsen (1998)[14] and Cont and Tankov (2004) [152].

More in detail, the two assets that compose our economy are:

- A risksless asset, whose rate of return is allowed to fluctuate but that is otherwise risk-free, which is represented by a risk-free bond P_0 , whose unit price $P_0(t)$ at time t is:

$$dP_0(t) = r(t)P_0(t)dt \quad 0 \leq t \leq T \quad (4.68)$$

$$P_0(0) = 1 \quad (4.69)$$

where $r(t) = r(t, \omega) > 0$ is \mathcal{I}_t -adapted and represents the time invariant risk-free rate in the market⁷⁴.

- A risky asset, represented by a stock P driven by a one dimensional Brownian motion and a

⁷³Extending the problem to a model with 1 risk-free asset and n , $1 < n < \infty$ risky assets is surely more realistic but the obtained results would be much less immediate to interpret. As a main difficulty there is the delicate calibration of the possible correlation among the assets.

⁷⁴Equation (4.68) could be further refined by using a time varying risk-free rate as proposed by i.e.: Korn and Kraft (2001)[108]. The usual precision/tractability trade-off arises. To guarantee a higher tractability in the model we leave it deterministic.

pure jump process (Lévy-Itô process), whose unit price $P(t)$ at time t is:

$$dP(t) = P(t^-) \left[\mu(t)dt + \sigma(t)dB(t) + \int_{\mathbb{R}} \theta(t, z)\tilde{N}(dt, dz) \right] \quad (4.70)$$

$$P(0) > 0 \quad 0 \leq t \leq T \quad (4.71)$$

Technical assumptions (T.A.):

For each $t \in [0, T]$, $\omega \in \Omega$ and $z \in \mathbb{R} \setminus \{0\}$ we assume that the parameters of the continuous part of the process $\mu(t, \omega)$, $\sigma(t, \omega)$ and $B(t, \omega)$ satisfy the following technical assumptions (T.A.):

T.A.1 \mathcal{I}_t -progressively measurable, hence time dependent and non-anticipating

T.A.2 bounded on $[0, T] \times \Omega$

T.A.3 parameters $\mu(t) = \mu(t, \omega)$ and $\sigma(t) = \sigma(t, \omega)$ represent respectively the investors' expected returns form $P(t)$ and the volatility of $P(t)$

T.A.4 $B(t) = B(t, \omega)$ is an \mathcal{I}_t -adapted one dimension Brownian Motion

While parameters governing the jump part of the process are such that:

T.A.5 $\theta(t, z, \omega)$ is \mathcal{I}_t -adapted hence time dependent, and non-anticipating

T.A.6 $\tilde{N}(dt, dz) = N(dt, dz) - \nu_{\mathcal{I}_t}(dz)dt$ is the \mathcal{I}_t -compensated Poisson random measure of $\eta(t) = \eta(t, \omega) : [0, T] \times \Omega \rightarrow \mathbb{R}$ where:

$$\eta(t) = \int_0^t \int_{\mathbb{R}} \theta(t, z)\tilde{N}(dt, dz) \quad (4.72)$$

T.A.7 to prevent the process to be ≤ 0 , we set: $\theta(t, z) > -1$ for a.a. t, z with respect to $dt \times \nu_{\mathcal{I}_t}(dz)$

T.A.8 $\mathbb{E}^P[\eta^2(t)] < \infty$ for all $t \geq 0$

Further details on the main characteristics that govern the jumps part of (4.70) are in appendix (A.4). We remind to [152] for more information about Lévy processes in finance.

The application of the Itô formula for semimartingales plays a key role for the resolution of our problem. It turns out that also its application is fully dependent to a proper assessment of the filtration set. Let us quickly recall it:

Theorem 4.6.1. *If $A(t)$ is an Itô-Lévy process:*

$$dA(t) = at + bB(t) + \int_{\mathbb{R}} \alpha(t, z) \tilde{N}(dt, dz) \quad (4.73)$$

such that:

$$\int_0^t |a(s)| ds + \int_0^t b^2(s) ds + \int_0^t \int_{\mathbb{R}} \alpha^2(s, z) \nu(dz) ds < \infty \quad (4.74)$$

and $D(t) = f(t, A(t))$, where $f : [0, T] \times \mathbb{R}$ and $f \in \mathcal{C}^{1,2}$, then, by the Itô formula for semimartingales:

$$\begin{aligned} dD(t) &= \left\{ \frac{df}{dt}(t, A(t)) + \frac{df}{da}(t, A(t))a(t) + \frac{1}{2} \frac{d^2f}{da^2}(t, A(t))b^2(t) \right\} dt \\ &\quad + \frac{df}{da}(t, A(t))b(t)dB(t) \\ &\quad + \int_{\mathbb{R}} \left\{ f(t, A(t) + \alpha(t, z)) - f(t, A(t)) \right\} \tilde{N}(dt, dz) \\ &\quad + \int_{\mathbb{R}} \left\{ f(t, A(t) + \alpha(t, z)) - f(t, A(t)) - \frac{df}{da}(t, A(t))\alpha(a, z) \right\} \nu(dz) dt \end{aligned} \quad (4.75)$$

the process $D(t)$ is an Itô-Lévy process as well.

Let us apply theorem (4.6.1) to our portfolio problem. If the presented $T.A.$ are satisfied and if the t and z dependent parameters of equations (4.68) and (4.70) are \mathcal{I}_t -adapted such that:

$$\mathbb{E}^P \left\{ \int_0^T \left[|r(s)| + |\mu(s)| + \sigma^2(s) + \int_{\mathbb{R}} \theta^2(s, z) \nu_{\mathcal{I}_t}(dz) \right] ds \right\} < \infty \quad (4.76)$$

then, by the Itô formula for semimartingales, the solution of (4.70) is:

$$\begin{aligned} P(t) &= P(0) \exp \left\{ \int_t^T \left[\mu(s) - \frac{1}{2} \sigma^2(s) - \int_{\mathbb{R}} (\theta(s, z) - \ln(1 + \theta(s, z))) \nu_{\mathcal{I}_t}(dz) \right] ds \right. \\ &\quad \left. + \int_t^T \sigma(s) dB(s) + \int_t^T \int_{\mathbb{R}} \ln(1 + \theta(s, z)) \tilde{N}(dt, dz) \right\}; \quad 0 \leq t \leq T \end{aligned} \quad (4.77)$$

Now, suppose that $\xi(t) = \xi(t, \omega) : [0, T] \times \Omega \rightarrow \mathbb{R}$, $\omega \in \Omega$ is an \mathcal{I}_t -measurable stochastic process representing the fraction of wealth $X(t)$ invested the investor in the risky asset and $(1 - \xi(t))$ is invested in the risk-free asset, then the evolution of the total wealth process, $X(t) = X^{\xi_t}(t)$ ⁷⁵, of the investor is:

$$\begin{aligned} dX(t) &= (1 - \xi(t))X(t)r(t)dt \\ &\quad + \xi(t)X(t^-) \left\{ \mu(t)dt + \sigma(t)dB(t) + \int_{\mathbb{R}} \theta(t, z) \tilde{N}(dt, dz) \right\} \end{aligned} \quad (4.78)$$

$$X(0) = x > 0 \quad (\text{Initial capital}) \quad (4.79)$$

⁷⁵The parameter ξ is time-varying and we represent its time dependency equivalently with ξ_t or $\xi(t)$.

or, collecting terms:

$$dX(t) = X(t^-) \left\{ [r(t) + (\mu(t) - r(t))\xi(t)]dt + \sigma(t)\xi(t)dB(t) + \xi(t) \int_{\mathbb{R}} \theta(t, z)\tilde{N}(dt, dz) \right\}; \quad 0 \leq t \leq T \quad (4.80)$$

$$X(0) = x > 0 \quad (\text{Initial capital}) \quad (4.81)$$

A key element for the analysis of the analysis is how to define an admissible portfolio under the different filtrations in use.

Definition 4.6.1. *Given a small and rational investor, a portfolio process $\xi(t)$ is assumed to be \mathcal{I}_t -admissible for each $t \in T$ if:*

- *is \mathcal{I}_t -adapted for each t , where $0 < t < T < \infty$*
- $\mathbb{E}^P \left\{ \int_0^T \left[|\mu(t) - r(t)| \cdot |\xi(t)| + \sigma^2(t)\xi^2(t) + \int_{\mathbb{R}} \xi^2(t)\theta^2(t, z)\nu_{\mathcal{I}}(dz) \right] dt \right\} < \infty$
- $\xi(t)\theta(t, z) \geq -1$ a.s. for $dt \times \nu_{\mathcal{I}_t} dz$ for a.a. t and z

Given this framework we analyze the stochastic control problem of a small investor whose goal is to maximize $\mathbb{E}^P(U^{\xi_t}(T))$ over a finite time window $[0, T]$ and over the class of all possible time (t) admissible portfolios $\mathcal{A}_{\mathcal{I}_t}(t)$. Formally:

Proposition 4.6.1. *We solve a finite horizon $T < \infty$ stochastic control problem of a rational and small investor endowed with a positive initial capital $X(0) = x > 0$ and described by a generic utility function $U(\cdot)$, here a logarithmic utility function, whose goal is to maximize her expected utility from terminal wealth, i.e.: $\mathbb{E}^P[U(\cdot)(T)]$ by investing continuously in a risky, P , and in a risk-free asset, P_0 .*

The optimal value of the problem, denoted by, $\Lambda_{\mathcal{I}_t}^{Log}$, is valid only if, for each $t \in [0, T]$ and $\omega \in \Omega$, there exists an optimal portfolio process, $\xi(t, \omega) = \xi_{\mathcal{I}_t}^(t, \omega)$, which belongs to the set of admissible portfolios, $\mathcal{A}_{\mathcal{I}_t}$, s.t.:*

$$\Lambda_{\mathcal{I}_t}^{Log}(x_t) = \sup_{\xi_t \in \mathcal{A}_{\mathcal{I}_t}} \mathbb{E}^P[\ln X^{\xi_t}(T)] = \mathbb{E}^P[\ln X^{\xi_t^*}(T)] \quad (4.82)$$

where the value function of the problem is assumed to be:

$$\Lambda_{\mathcal{I}_t}^{Log}(x_t) < \infty \quad \forall x_t \in [0, \infty) \quad (4.83)$$

Given proposition (4.6.1) it emerges clearly that the solution of the problem is strongly related to the admissibility of the portfolio with respect to the information set, \mathcal{I}_t . To better underline its importance, the same portfolio problem will be analyzed under two different information sets:

1. Theoretical case: $\mathcal{I}_t = \mathcal{F}_t$: the information set is complete
2. Real-world case: $\mathcal{I}_t = \mathcal{H}_t \subset \mathcal{F}_t$: some relevant information is missing, hence cannot be reflected in the final asset price

The former case pertains to an investor which is able to set up an asset pricing model that fully capture *all* past, present and future relevant pricing information, hence also and above all her forward looking beliefs with respect to the future outcome. The latter instead pertains to a more realistic suboptimal case. Illusory unexisting arbitrages may be the naive consequence of an investor not being fully aware to pertain to the former or the latter group.

While is at least since Merton (1969)[119] and Samuelson (1969)[146] that the stochastic control problem for $(\mathcal{F}_t)_{t>0} = \mathbb{F}$ -adapted portfolios is a well-known problem in literature (a good review of the subject is, among the others, Cvitanic and Karatzas (1992) [51]), the one for an Itô-Lévy market model with a suboptimal information set is not. We extend the literature answering to this problem and analyzing its effect in asset pricing under different viewpoints.

4.6.3 The theoretical optimal choice

In this chapter we solve the portfolio optimization problem for an investor with logarithmic utility function under the theoretical case: when the filtration set is the complete one:

$$\text{Theoretical Case : } \mathcal{I}_t = \mathcal{F}_t \tag{4.84}$$

Assuming for all $t \in T$, that $\xi_t \in \mathcal{A}_{\mathcal{I}_t}$ and applying the Itô theorem for semimartingales to (4.80) the admissible investor's terminal wealth is:

$$\begin{aligned} X^{\xi_{t \in \mathcal{F}_t}}(T) = & x_t \exp \left\{ \int_t^T \left[r(s) + (\mu(s) - r(s))\xi_{\in \mathcal{F}_s}(s) - \frac{1}{2}\sigma^2(s)\xi_{\in \mathcal{F}_s}^2(s) \right] ds \right. \\ & + \int_t^T \xi_{\in \mathcal{F}_s}(s)\sigma(s)dB(s) \\ & + \int_t^T \int_{\mathbb{R}} \ln(1 + \xi_{\in \mathcal{F}_s}(s)\theta(s, z))\tilde{N}(ds, dz) \\ & \left. - \int_t^T \int_{\mathbb{R}} (\xi_{\in \mathcal{F}_s}(s)\theta(s, z) - \ln(1 + \xi_{\in \mathcal{F}_s}(s)\theta(s, z)))\nu_{\mathcal{F}_s}(dz, ds) \right\} \end{aligned} \tag{4.85}$$

Given the evolution of the total wealth of the investor, and assuming that:

$$\mathbb{E}^P[\ln X^{\xi_{t \in \mathcal{F}_t}}(T)] < \infty \tag{4.86}$$

the expected value of the problem is:

$$\mathbb{E}^P[\ln X^{\xi_{\in \mathcal{F}_t}}(T)] - \ln(x_t) = \mathbb{E}^P \left\{ \int_t^T \left[r(s) + (\mu(s) - r(s))\xi_{\in \mathcal{F}_s}(s) - \frac{1}{2}\sigma^2(s)\xi_{\in \mathcal{F}_s}^2(s) - \int_{\mathbb{R}} (\xi_{\in \mathcal{F}_s}(s)\theta(s, z) - \ln(1 + \xi_{\in \mathcal{F}_s}(s)\theta(s, z)))\nu_{\mathcal{F}_s}(dz) \right] ds \right\} \quad (4.87)$$

Fixing s and ω , the objective function to maximize is:

$$h(\xi_{\in \mathcal{F}_s}(s)) = (\mu(s) - r(s))\xi_{\in \mathcal{F}_s}(s) - \frac{1}{2}\sigma^2(s)\xi_{\in \mathcal{F}_s}^2(s) - \int_{\mathbb{R}} (\xi_{\in \mathcal{F}_s}(s)\theta(s, z) - \ln(1 + \xi_{\in \mathcal{F}_s}(s)\theta(s, z)))\nu_{\mathcal{F}_s}(dz) \quad (4.88)$$

where h is a positive and concave function. Taking the first order condition with respect to $\xi_{\in \mathcal{F}_s}(s)$ and equating the result to zero we obtain the solution $\xi_{\in \mathcal{F}_s}(s) = \xi_{\in \mathcal{F}_s}(s, \omega)$:

$$0 = h'(\xi_{\in \mathcal{F}_s}(s)) = \mu(s) - r(s) - \sigma^2(s)\xi_{\in \mathcal{F}_s}(s) - \int_{\mathbb{R}} \left(\theta(s, z) - \frac{\theta(s, z)}{1 + \xi_{\in \mathcal{F}_s}(s)\theta(s, z)} \right) \nu_{\mathcal{F}_s}(dz) \quad (4.89)$$

Collecting terms, the extra return from the portfolio is:

$$\mu(s) - r(s) = \sigma^2(s)\xi_{\in \mathcal{F}_s}(s) + \int_{\mathbb{R}} \frac{\xi_{\in \mathcal{F}_s}(s)\theta^2(s, z)}{1 + \xi_{\in \mathcal{F}_s}(s)\theta(s, z)}\nu_{\mathcal{F}_s}(dz) \quad (4.90)$$

which, for the the case of no jumps⁷⁶, $\theta(s, z) = 0$:

$$\xi_{\in \mathcal{F}_s}(s) = \frac{\mu(s) - r(s)}{\sigma^2(s)} \quad (4.91)$$

whose validity depends on $\sigma(s) \neq 0$ a.s. for a.a. (s, ω) , $s \in [0, T]$ and $\omega \in \Omega$. $\xi_{\in \mathcal{F}_s}(s)$ is the optimal portfolio process $\xi_{\in \mathcal{F}_s}^*(s)$ only if is in the set of admissible portfolios given the information set:

$$\xi \in \mathcal{A}_{\mathcal{F}_s} \rightarrow \xi_{\in \mathcal{F}_s}(s) = \xi_{\in \mathcal{F}_s}^*(s) \quad \text{for all } s \in [0, T] \quad (4.92)$$

It follows that the optimal portfolio value for the case of a theoretical (or full) information set \mathcal{F}_t is:

$$\Lambda_{\mathcal{F}_t}^{\text{Log}} = \mathbb{E}^P[\ln X^{\xi_{\in \mathcal{F}_t}^*}(T)] = \ln x_t + \mathbb{E}^P \left[\int_t^T \left(r(s) + \frac{(\mu(s) - r(s))^2}{2\sigma^2(s)} \right) ds \right] \quad (4.93)$$

With no loss of generality we can set the investor initial capital $X(0) = 1$ s.t.:

$$\Lambda_{\mathcal{F}_t}^{\text{Log}} = \mathbb{E}^P[\ln X^{\xi_{\in \mathcal{F}_t}^*}(T)] = \mathbb{E}^P \left[\int_t^T \left(r(s) + \frac{(\mu(s) - r(s))^2}{2\sigma^2(s)} \right) ds \right] \quad (4.94)$$

In conclusion: for a finite-time complete market model with one risk-less and one risky asset, the maximal expected logarithmic utility of the terminal wealth for a small and rational investor with a complete set of information is the integrated sum of the risk-free rate and a fraction of the Sharpe ratio squared.

⁷⁶An alternative and more rigorous solution that considers jumps can be found in Platen and Heath (2006)[132].

4.6.4 The real-world optimal choice

In this section we analyze again the same problem but for the viewpoint of an econometrician:

$$\text{Real-world case : } \mathcal{I}_t = \mathcal{H}_t \subset \mathcal{F}_t \quad (4.95)$$

We thus focus our attention, both theoretically and conceptually, on how the missing information of the filtration set propagates and affect onto the final profit of the investor.

Except for $\xi_{t \in \mathcal{H}_t}$ and $\nu_{\mathcal{H}_t}(dz, dt)$, which are affected by the scarcer informative content given by the suboptimal filtration set, the starting point for the restricted case is the same as the complete one:

$$\begin{aligned} X^{\xi_{t \in \mathcal{H}_t}}(T) = & x_t \exp \left\{ \int_t^T \left[r(s) + (\mu(s) - r(s))\xi_{\in \mathcal{H}_s}(s) - \frac{1}{2}\sigma^2(s)\xi_{\in \mathcal{H}_s}^2(s) \right] ds \right. \\ & + \int_t^T \xi_{\in \mathcal{H}_s}(s)\sigma(s)dB(s) \\ & + \int_t^T \int_{\mathbb{R}} \ln(1 + \xi_{\in \mathcal{H}_s}(s)\theta(s, z))\tilde{N}(ds, dz) \\ & \left. - \int_t^T \int_{\mathbb{R}} (\xi_{\in \mathcal{H}_s}(s)\theta(s, z) - \ln(1 + \xi_{\in \mathcal{H}_s}(s)\theta(s, z)))\nu_{\mathcal{H}_s}(dz, ds) \right\} \end{aligned} \quad (4.96)$$

Given the evolution of the total wealth of the investor and assuming that:

$$\mathbb{E}^{\hat{P}}[\ln X^{\xi_{t \in \mathcal{H}_t}}(T)] < \infty \quad (4.97)$$

the expected value of the problem is:

$$\begin{aligned} \mathbb{E}^{\hat{P}}[\ln X^{\xi_{t \in \mathcal{H}_t}}(T)] - \ln(x_t) = & \mathbb{E}^{\hat{P}} \left\{ \int_t^T \left[r(s) + (\mu(s) - r(s))\xi_{\in \mathcal{H}_s}(s) - \frac{1}{2}\sigma^2(s)\xi_{\in \mathcal{H}_s}^2(s) \right. \right. \\ & \left. \left. - \int_{\mathbb{R}} (\xi_{\in \mathcal{H}_s}(s)\theta(s, z) - \ln(1 + \xi_{\in \mathcal{H}_s}(s)\theta(s, z)))\nu_{\mathcal{H}_s}(dz) \right] ds \right\} \end{aligned} \quad (4.98)$$

To account for the smaller amount of information for the optimization of the restricted optimal portfolio $\xi_t \in \mathcal{A}_{\mathcal{H}_t}$, we need to insert, for all $t \in T$, an extra conditioning of the expectation with respect to \mathcal{H}_t :

$$\begin{aligned} \mathbb{E}^{\hat{P}}[\ln X^{\xi_{t \in \mathcal{H}_t}}(T)] - \ln(x_t) = & \mathbb{E}^{\hat{P}} \left\{ \int_t^T \mathbb{E}^P \left[r(s) + (\mu(s) - r(s))\xi_{\in \mathcal{H}_s}(s) - \frac{1}{2}\sigma^2(s)\xi_{\in \mathcal{H}_s}^2(s) \right. \right. \\ & \left. \left. - \int_{\mathbb{R}} (\xi_{\in \mathcal{H}_s}(s)\theta(s, z) - \ln(1 + \xi_{\in \mathcal{H}_s}(s)\theta(s, z)))\nu_{\mathcal{H}_s}(dz) | \mathcal{H}_s \right] ds \right\} \end{aligned} \quad (4.99)$$

or, applying the same convention of before:

$$\widehat{x}(t) = \mathbb{E}^{\widehat{P}}[x(t)|\mathcal{H}_t] \quad (4.100)$$

then:

$$\begin{aligned} &= \mathbb{E}^{\widehat{P}} \left\{ \int_t^T \left[r(s) + (\widehat{\mu}(s) - r(s))\xi_{\in\mathcal{H}_s}(s) - \frac{1}{2}\widehat{\sigma}^2(s)\xi_{\in\mathcal{H}_s}^2(s) \right. \right. \\ &\quad \left. \left. - \int_{\mathbb{R}} (\xi_{\in\mathcal{H}_s}(s)\widehat{\theta}(s, z) - \mathbb{E}^P(\ln(1 + \xi_{\in\mathcal{H}_s}(s)\theta(s, z))|\mathcal{H}_s))\nu_{\mathcal{H}_s}(dz) \right] ds \right\} \end{aligned} \quad (4.101)$$

where $(\widehat{\cdot})$ identifies the value of the parameters under the \mathcal{H}_t restriction.

Dealing with the subjective beliefs that affect the investor's investment decisions, the consequences of a suboptimal filtration set have an impact only on the risky assets of the investor's portfolio. Although allowed to fluctuate, the risk-free asset is in fact assumed to be free of any subjective and objective risks thus not affecting the overall investor's subjective behaviour with respect to risky decision to undertake.

This can be easily "demonstrated" starting from the definition of the classical physical pricing equation:

$$f_{t,T}(B_{t,T}) = \mathbb{E}^P(m_{t,T} \cdot B_{t,T} | \mathcal{F}_t) \quad (4.102)$$

where $f_{t,T}(B_{t,T})$ represents the time t pricing functional of a risk-free bond, B , which matures at time T and $m_{t,T}$ its relative PK. By construction, the role of the PK in asset pricing is to adjust the future payoff accounting for all possible risky future outcomes that can affect the underline. A risk-free product is, by definition, without risk⁷⁷. Not having to consider any risk or time preference of the investor, its expected final payoff is at any point in time equal to one, no matter what happens in the market. Given a finite time window, $t = 0, \dots, T < \infty$, this is translated in a PK:

$$\mathbb{E}^P(m_{t,T} | \mathcal{F}_0) = 1 \quad \text{for all } t \in T \quad (4.103)$$

As a consequence, at any Δt , the risk-free bond is immune to any risk and equation (4.102) can be iterated-back in time such that its time 0 price is:

$$f_{0,T}(B_{0,T}) = \mathbb{E}^P(B_{0,T} | \mathcal{F}_0) \quad (4.104)$$

⁷⁷This is not entirely true in reality since all assets may be affected by some risk. Nevertheless, although no products may have an a.s. probability of being totally risk-free, some of them have negligible risk potential so that they can easily be assumed as risk-free assets.

Given the role of the PK, the absence of risk for the underlying in question, and differently than (4.100), equation (4.104) is not affected by any possible missing forward looking information:

$$\mathbb{E}^P(m_{t,T}|\mathcal{F}_0) = \mathbb{E}^{\hat{P}}(m_{t,T}|\mathcal{H}_0) = 1 \quad \text{for all } t \in T \quad (4.105)$$

It follows that the riskless bond has the same value, no matter under which filtration:

$$f_{0,T}(B_{0,T}) = \mathbb{E}^P(B_{0,T}|\mathcal{F}_0) \quad (4.106)$$

$$= \mathbb{E}^{\hat{P}}(B_{0,T}|\mathcal{H}_0) \quad (4.107)$$

It follows naturally that we don't need to put any hat on any of the risk-less parameters present in our portfolio problems i.e.: $r(t)$.

As for the complete case, we fix s, ω such that the suboptimal objective function is:

$$\begin{aligned} \widehat{h}(\xi_{\in \mathcal{H}_s}(s)) = \\ (\widehat{\mu}(s) - r(s))\xi_{\in \mathcal{H}_s}(s) - \frac{1}{2}\widehat{\sigma}^2(s)\xi_{\in \mathcal{H}_s}^2(s) - \int_{\mathbb{R}} \left[\xi_{\in \mathcal{H}_s}(s)\widehat{\theta}(s, z) - \mathbb{E}^{\hat{P}}(\ln(1 + \xi_{\in \mathcal{H}_s}(s)\theta(s, z))|\mathcal{H}_s) \right] \nu_{\mathcal{H}_s}(dz) \end{aligned} \quad (4.108)$$

which is again positive and concave hence solvable:

$$0 = \widehat{h}'(\xi_{\in \mathcal{H}_s}(s)) = \widehat{\mu}(s) - r(s) - \widehat{\sigma}^2(s)\xi_{\in \mathcal{H}_s}(s) - \int_{\mathbb{R}} \left[\widehat{\theta}(s, z) - \mathbb{E}^{\hat{P}}\left(\frac{\theta(s, z)}{1 + \xi_{\in \mathcal{H}_s}(s)\theta(s, z)}|\mathcal{H}_s\right) \right] \nu_{\mathcal{H}_s}(dz) \quad (4.109)$$

It follows that the excess return under the suboptimal filtration set is:

$$\widehat{\mu}(s) - r(s) = \widehat{\sigma}^2(s)\xi_{\in \mathcal{H}_s}(s) + \int_{\mathbb{R}} \left[\widehat{\theta}(s, z) - \mathbb{E}^{\hat{P}}\left(\frac{\theta(s, z)}{1 + \xi_{\in \mathcal{H}_s}(s)\theta(s, z)}|\mathcal{H}_s\right) \right] \nu_{\mathcal{H}_s}(dz) \quad (4.110)$$

Remark 4.6.1. Equation (4.110) is valid only if $\xi_{\in \mathcal{H}_s}(s)\theta(s, z) \geq -1$ and if the a.s. uniform integrability with respect to $\nu_{\mathcal{H}_s}(dz)$ of:

$$\mathbb{E}^{\hat{P}}\left(\frac{\xi_{\in \mathcal{H}_s}(s)\theta^2(s, z)}{1 + \theta(s, z)}|\mathcal{H}_s\right) \quad (4.111)$$

applies for almost all s, ω .

Once more, for each $s \in [0, T]$, equation (4.111) is a consequence of:

$$\widehat{\theta}(s, z) = \mathbb{E}^{\hat{P}}(\theta(s, z)|\mathcal{H}_s) \quad (4.112)$$

In case of no jumps $\theta(s, z) = 0$:

$$\xi_{\in \mathcal{H}_s}(s) = \frac{\widehat{\mu}(s) - r(s)}{\widehat{\sigma}^2(s)} \quad (4.113)$$

where the above equation is valid only if $\widehat{\sigma}(s) \neq 0$ a.s. for a.a. (s, ω) , $s \in [0, T]$ and $\omega \in \Omega$.

Once more, $\xi_{\in \mathcal{H}_s}(s)$ is the optimal portfolio process $\xi_{\in \mathcal{H}_s}^*(s)$ only if is in the set of admissible portfolios given the information set:

$$\xi \in \mathcal{A}_{\mathcal{H}_s} \rightarrow \xi_{\in \mathcal{H}_s}(s) = \xi_{\in \mathcal{H}_s}^*(s) \quad \text{for all } s \in [0, T] \quad (4.114)$$

It follows that the optimal restricted portfolio value is:

$$\Lambda_{\mathcal{H}_t}^{\text{Log}} = \mathbb{E}^{\widehat{P}}[\ln X^{\xi_{\in \mathcal{H}_t}^*}(T)] = \ln x_t + \mathbb{E}^{\widehat{P}} \left[\int_t^T \left(r(s) + \frac{(\widehat{\mu}(s) - r(s))^2}{2\widehat{\sigma}^2(s)} \right) ds \right] \quad (4.115)$$

With no loss of generality we can set the initial capital of the investor to $X(0) = 1$ such that:

$$\Lambda_{\mathcal{H}_t}^{\text{Log}} = \mathbb{E}^{\widehat{P}}[\ln X^{\xi_{\in \mathcal{H}_t}^*}(T)] = \mathbb{E}^{\widehat{P}} \left[\int_t^T \left(r(s) + \frac{(\widehat{\mu}(s) - r(s))^2}{2\widehat{\sigma}^2(s)} \right) ds \right] \quad (4.116)$$

Now, combining (4.94) and (4.116) it follows that the use of a suboptimal information set leads to:

$$\Lambda_{\mathcal{F}_t}^{\text{Log}} - \Lambda_{\mathcal{H}_t}^{\text{Log}} = \frac{1}{2} \int_t^T \mathbb{E}^P \left[\frac{(\mu(s) - r(s))^2}{\sigma^2(s)} \right] ds - \frac{1}{2} \int_t^T \mathbb{E}^{\widehat{P}} \left[\frac{(\widehat{\mu}(s) - r(s))^2}{\widehat{\sigma}^2(s)} \right] ds \quad (4.117)$$

$$= \Delta_t^{\text{Log}} \quad (4.118)$$

where Δ_t^{Log} represents the time t information premium.

Following the neoclassical literature, the optimal (or theoretical) case assumes the investor to be a natural and fully rational optimizer capable to obtain the highest possible reward for unit of risk given the market scenario. On the contrary, the more realistic real-world case assumes and implies some deficiencies in optimization. It follows naturally that the above obtained information premium:

$$\Delta_t > 0 \quad (4.119)$$

Its connection with the PK bounds, which I will analyze in more detail in the next sections, justifies, reinforces and makes this statement even stronger. The empirical results of chapter 9 completes the connection between PK, information and portfolio performances. At the same time, the amount of

premium cannot be determined ex ante and is highly dependent to the market scenario (i.e.: level of volatility in the market).

Empirically, to have at each point in time a *full* information set as required by the neoclassical theory is usually a mere illusion. Therefore, the goal of a good financial modeller should be, for each $t \in T$, to minimize Δ_t as much as possible

$$\min \Delta_t \Rightarrow \max \mathcal{I}_t \quad (4.120)$$

It follows immediately that to minimize Δ_t means to collect and model in the best way possible all the information that is relevant for pricing the assets. A proper estimation of the real-world probabilities is then of key importance for many day-by-day operations (i.e.: trading, risk management, asset management).

4.6.5 The power utility case

A common drawback that pertains to the literature of the stochastic optimization problems (and not only) is its high reliance on the standardization of the parameters that govern most of models. In our case the proposed model is highly dependent to the choice of the utility function used to describe the investor decisions. Given their widespread use in literature and as an alternative to broaden our analysis, we propose a power utility function as a possible extension of the logarithmic utility function. Being aware that no investor can be properly and fully described by a parametric utility function, the proposed alternative is not a solution of the utility-problem but just a possible alternative. It follows that all results must then be considered as approximations.

The problem remains the same:

Proposition 4.6.2. *We want to solve a finite horizon $T < \infty$ stochastic control problem of a rational and small investor endowed with a positive initial capital $X(0) = x > 0$ and described by a generic utility function $U(\cdot)$, here a power utility function, whose goal is to maximize her expected subjective utility from terminal wealth, i.e.: $\mathbb{E}^P[U(\cdot)(T)]$ by investing continuously in a risky, P , and in a risk-free asset, P_0 .*

The optimal value of the problem, denoted by, $\Lambda_{\mathcal{I}_t}^{Pow}$, is valid only if, for each $t \in [0, T]$ and $\omega \in \Omega$, there exists an optimal portfolio process, $\xi(t, \omega) = \xi_{\mathcal{I}_t}^(t, \omega)$, which belongs to the set of admissible portfolios, $\mathcal{A}_{\mathcal{I}_t}$, s.t.:*

$$\Lambda_{\mathcal{I}_t}^{Pow}(x) = \sup_{\xi \in \mathcal{A}_{\mathcal{I}_t}} \mathbb{E}^P \left[\frac{1}{1-\lambda} X^{1-\lambda, \xi_t}(T) \right] = \mathbb{E}^P \left[\frac{1}{1-\lambda} X^{1-\lambda, \xi_t^*}(T) \right] \quad (4.121)$$

where the the value function of the problem is assumed to be:

$$\Lambda_{\mathcal{I}_t}^{Pow}(x) < \infty \quad \forall x_t \in [0, \infty) \quad (4.122)$$

As stated in proposition (4.6.2) and differently than (4.6.1) we model the investor by means of a Constant Relative Risk Aversion (CRRA) utility function:

$$U(x_t) = \frac{1}{1-\lambda} x_t^{1-\lambda} \quad x > 0 \quad \lambda \neq 1, \lambda > 0 \quad (4.123)$$

where $\lambda \in \mathbb{R}^+$ is a constant and measures the degree of relative risk aversion implicit in the utility function.

Given the new setup we follow the same procedure as before to solve the stochastic optimization problem under the two scenarios.

For a full information set: $\mathcal{I}_t = \mathcal{F}_t$ and with $\theta(s, z) = 0$ ⁷⁸:

$$\xi_{\mathcal{F}_s}(s) = \frac{1}{\lambda} \frac{\mu(s) - r(s)}{\sigma^2(s)} \quad (4.124)$$

where $\xi_{\in \mathcal{F}_s}(s)$ is the optimal portfolio process $\xi_{\in \mathcal{F}_s}^*(s)$ only if is in the set of admissible portfolios given the information set:

$$\xi \in \mathcal{A}_{\mathcal{F}_s} \rightarrow \xi_{\in \mathcal{F}_s}(s) = \xi_{\in \mathcal{F}_s}^*(s) \quad \text{for all } s \in [0, T] \quad (4.125)$$

For (4.124), to be valid, we need $\lambda \neq 0$ and $\sigma(s) \neq 0$ a.s. for a.a. (s, ω) , $s \in [0, T]$ and $\omega \in \Omega$ to hold.

Setting the investor initial capital to $X(0) = 1$:

$$\Lambda_{\mathcal{F}_t}^{Pow} = \mathbb{E}^P \left[\frac{1}{1-\lambda} X^{1-\lambda, \xi_{\in \mathcal{F}_t}^*}(T) \right] = \mathbb{E}^P \left[\int_t^T \left(r(s) + \frac{1}{\lambda} \frac{(\mu(s) - r(s))^2}{2\sigma^2(s)} \right) ds \right] \quad (4.126)$$

In conclusion, for a finite-time complete market with one risk-less and and one risky asset, the maximal expected power utility of the terminal wealth for a small and rational investor with a complete set of information, is the integrated sum of the risk-free rate and a fraction of the Sharpe ratio squared normalized by λ .

The same logic applies for the partial information set $\mathcal{I}_t = \mathcal{H}_t \subset \mathcal{F}_t$. In case of no jumps $\theta(s, z) = 0$:

$$\xi_{\in \mathcal{H}_s}(s) = \frac{1}{\widehat{\lambda}} \frac{\widehat{\mu}(s) - r(s)}{\widehat{\sigma}^2(s)} \quad (4.127)$$

⁷⁸Once more, Platen and Heath (2006)[132] provides the solution with jumps.

where λ , the indicator of the subjective degree of risk aversion of an investor, is fully impacted by the coarser filtration set so that:

$$\widehat{\lambda} = \mathbb{E}^{\widehat{P}}[\lambda | \mathcal{H}_s] \quad (4.128)$$

is a valid equivalence.

$\widetilde{\xi}_{\in \mathcal{H}_s}(s)$ is the optimal portfolio process $\xi_{\in \mathcal{H}_s}^*(s)$ only if is in the set of admissible portfolios given the information set:

$$\xi \in \mathcal{A}_{\mathcal{H}_s} \rightarrow \xi_{\in \mathcal{H}_s}(s) = \xi_{\in \mathcal{H}_s}^*(s) \quad \text{for all } s \in [0, T] \quad (4.129)$$

Also for the restricted Sharpe ratio, to be valid, we need $\widehat{\lambda} \neq 0$ and $\widehat{\sigma}(s) \neq 0$ a.s. for a.a. (s, ω) , $s \in [0, T]$ and $\omega \in \Omega$ to hold.

Setting the initial capital of the investor to $X(0) = 1$ the optimal restricted portfolio value is

$$\Lambda_{\mathcal{H}_t}^{\text{Pow}} = \mathbb{E}^{\widehat{P}} \left[\frac{1}{1-\lambda} X^{1-\lambda, \xi_{\in \mathcal{H}_t}^*}(T) \right] = \mathbb{E}^{\widehat{P}} \left[\int_t^T \left(r(s) + \frac{1}{\widehat{\lambda}} \frac{(\widehat{\mu}(s) - r(s))^2}{2\widehat{\sigma}^2(s)} \right) ds \right] \quad (4.130)$$

Now, combining (4.126) and (4.130) it follows that the difference in filtration is equal to:

$$\Lambda_{\mathcal{F}_t}^{\text{Pow}} - \Lambda_{\mathcal{H}_t}^{\text{Pow}} = \frac{1}{2} \int_t^T \mathbb{E}^P \left[\frac{1}{\lambda} \frac{(\mu(s) - r(s))^2}{\sigma^2(s)} \right] ds - \frac{1}{2} \int_t^T \mathbb{E}^{\widehat{P}} \left[\frac{1}{\widehat{\lambda}} \frac{(\widehat{\mu}(s) - r(s))^2}{\widehat{\sigma}^2(s)} \right] ds \quad (4.131)$$

$$= \Delta_t^{\text{Pow}} \quad (4.132)$$

which, as $\lambda \rightarrow 1$ ⁷⁹:

$$\Lambda_{\mathcal{F}_t}^{\text{Log}} - \Lambda_{\mathcal{H}_t}^{\text{Log}} = \frac{1}{2} \int_t^T \mathbb{E}^P \left[\frac{(\mu(s) - r(s))^2}{\sigma^2(s)} \right] ds - \frac{1}{2} \int_t^T \mathbb{E}^{\widehat{P}} \left[\frac{(\widehat{\mu}(s) - r(s))^2}{\widehat{\sigma}^2(s)} \right] ds \quad (4.133)$$

$$= \Delta_t^{\text{Log}} \quad (4.134)$$

4.6.6 Connection with the PK

There is a well known and strong interconnection between optimal portfolios and the PK. The latter is in fact bounded by the Sharpe ratio and vice-versa. The simplicity and generality of the rules that govern these bounds justify their widespread use in many aspects of financial economics (see Bekaert and Liu (2001)[22] for a review.). These bounds, also known as the Hansen-Jagganathan bounds (1991)[78], pose upper and lower limits on both random variables. Henrotte (2002)[83] investigates the tight but non-trivial relationship between the PK, its variance and the optimal portfolio choice. He defines that:

⁷⁹Which implies that also $\widehat{\lambda} \rightarrow 1$.

Definition 4.6.2. *The square of the Sharpe ratio of every portfolio is smaller than the variance of every normalized PK.*

From the definition, two are the main consequences in asset pricing:

- The square of the Sharpe ratio of any portfolio bounds the variance of any normalized PK:

$$\text{SR}^2 = \text{PK lower bound}$$

- No Sharpe ratios squared can be greater than the variance of the normalized PK:

$$\text{PK} = \text{SR}^2 \text{ upper bound}$$

where the variance is normalized through the mean of the PK itself and SR^2 states for Sharpe Ratio squared.

In this section I extend and generalize the above definition for the case of a suboptimal PK. Due to qualitative and quantitative importance of the missing information from the suboptimal filtration set it follows from equations (4.117) that, for each $t \in T$, if \mathcal{F}_t and \mathcal{H}_t differs significantly then:

$$\Delta_t^{\text{Log}} > 0 \tag{4.135}$$

which implies an a.s. inequality for the Sharpe ratios squared of the two portfolios:

$$\frac{(\mu(t) - r(t))^2}{\sigma^2(t)} > \frac{(\hat{\mu}(t) - r(t))^2}{\hat{\sigma}^2(t)} \tag{4.136}$$

All else equal, a better informed investor can aim to higher returns and lower volatility, which justify also conceptually the above inequality. These results show clearly that a suboptimal information set impacts directly on the total profitability of the investor's portfolio and also indirectly on the quality of their bounds.

Different papers refine and extend the Hansen-Jagannathan bounds in several directions⁸⁰. One of the main difficulties, common to many papers, is the time-varying estimation of the elements that compose the PK. Some authors propose unconditional PKs thus lowering the effectiveness of the findings. A time independent PK would be in fact useless for many day-by-day operations (i.e.: asset and risk management). Working on the insights of Gallant, Hansen and Tauchen (1990)[75] a more recent paper of Bekaert and Liu (2004)[22] extends the theory on the optimal bounds putting emphasis on the optimal use of the conditioning of the information. As a main result they show

⁸⁰See Ferson and Siegel (2002) [71] and references therein.

how, given some technical conditions⁸¹ the best bounds are the ones that maximizes the squared Sharpe Ratio:

$$\text{S.R.}(t)_{\text{Max}}^2 := \text{Best Bounds}(t) \leq \text{PK lower bound}(t) \quad (4.137)$$

Applying the theorem to our findings emerges how a more informed investors can not only benefit from superior returns but also from sharper, thus better, PK bounds. Given this result and from definition (4.6.2) it follows that:

$$\frac{\text{Var PK}_t(\mathcal{F})}{\text{Mean PK}_t(\mathcal{F})} > \frac{\text{Var PK}_t(\mathcal{H})}{\text{Mean PK}_t(\mathcal{H})} \quad (4.138)$$

4.6.7 Expressing the information premium as the Kullback-Leibler divergence

In this section we show how, through the PK, a suboptimal filtration set may propagates onto the risk-neutral pricing. Results can be so summarized:

- The information premium is nothing but the difference between the optimal and the suboptimal Kullback-Leibler divergences
- A suboptimal information set may affect the risk-neutral pricing by means of the restricted market price of risk that enters into the Girsanov theorem

Let us start defining the time-dependent theoretical and suboptimal market prices of risk⁸² as:

$$\text{Optimal} = \theta_t^{\in \mathcal{F}} = \theta_t = \frac{\mu(t) - r(t)}{\sigma(t)} \quad (4.139)$$

$$\text{Suboptimal} = \theta_t^{\in \mathcal{H}} = \hat{\theta}_t = \frac{\hat{\mu}(t) - r(t)}{\hat{\sigma}(t)} \quad (4.140)$$

Then, if θ_t and $\hat{\theta}_t$ are locally square integrable and if:

$$\mathbb{E}^P \left[\exp \left(\frac{1}{2} \int_0^t \theta_s^2 ds \right) | \mathcal{F}_t \right] < \infty \quad (4.141)$$

$$\mathbb{E}^P \left[\exp \left(\frac{1}{2} \int_0^t \hat{\theta}_s^2 ds \right) | \mathcal{H}_t \right] < \infty \quad (4.142)$$

⁸¹See [22], section 1.4 pag 345 and 346 for the main theorem and its proof.

⁸²With no loss of generality, the market price of risk of the previous sections, being scaled by its variance, is a proportional version of the theoretical one.

then, for all $t \in [0, T]$ and $x_t \in \mathcal{F}_t$, $x_t \in \mathcal{H}_t$:

$$Q_t(X) = \mathbb{E}^P[M(t)\mathbf{1}_{X_t}|\mathcal{F}_t] \quad (4.143)$$

$$\widehat{Q}_t(X) = \mathbb{E}^{\widehat{P}}[\widehat{M}(t)\mathbf{1}_{X_t}|\mathcal{H}_t] \quad (4.144)$$

are the optimal and suboptimal probability measures defined respectively on \mathcal{F}_t and \mathcal{H}_t where:

$$M(t) = \exp\left(\int_0^t \theta_s dW_s + \frac{1}{2} \int_0^t \theta_s^2 ds\right) \quad (4.145)$$

$$\widehat{M}(t) = \exp\left(\int_0^t \widehat{\theta}_s d\widehat{W}_s + \frac{1}{2} \int_0^t \widehat{\theta}_s^2 ds\right) \quad (4.146)$$

are the respective optimal and suboptimal pricing kernels. For both pricing kernels, W_t and \widehat{W}_t are the respective optimal and suboptimal Brownian Motion defined on (Ω, \mathcal{F}, P) and (Ω, \mathcal{H}, P) so that θ_t and $\widehat{\theta}_t$ are \mathcal{F}_t - and \mathcal{H}_t -adapted.

It follows that:

$$\mathbb{E}^P(M_t|\mathcal{F}_t) = 1 \quad (4.147)$$

such that M_t is a positive true martingale⁸³. As formerly demonstrated (section 4.3.2), the same may not be true under the suboptimal scenario. From the previous sections and justified by the rational behaviours of the investors under the neoclassical theory, it follows that the optimal quantities are always larger with respect to the suboptimal ones.

Now, let's recall the information premium:

$$\Delta_t^{\text{Log}} = \frac{1}{2} \int_t^T E^P \left[\frac{(\mu(s) - r(s))^2}{\sigma^2(s)} - \frac{(\widehat{\mu}(s) - r(s))^2}{\widehat{\sigma}^2(s)} \right] ds \quad (4.149)$$

Since the theory in object applies independently to the utility function used, we visually omit it. Assuming that the usual technical assumption holds (square local integrability and Novikov condi-

⁸³The Brownian motion under the new measure is:

$$W_t^Q = W_t - \int_0^t \theta_s ds \quad (4.148)$$

tion) the Kullback-Leibler divergence among the optimal densities is:

$$D_t^{KL}(P(t)|Q(t)) = \mathbb{E}^P \left(\log \frac{dP(t)}{dQ(t)} \right) \quad (4.150)$$

$$= \mathbb{E}^P \left(\log \frac{1}{M(t)} \right) \quad (4.151)$$

$$= \mathbb{E}^P \left(\int_0^t \theta dW_s + \frac{1}{2} \int_0^t \theta_s^2 ds \right) \quad (4.152)$$

$$= \frac{1}{2} \int_0^t \theta_s^2 \quad (4.153)$$

$$= \theta_t \quad (4.154)$$

The same hold for the suboptimal case:

$$\widehat{D}_t^{KL}(\widehat{P}(t)|Q(t)) = \mathbb{E}^{\widehat{P}} \left(\frac{d\widehat{Q}(t)}{d\widehat{P}(t)} \right) \quad (4.155)$$

$$= \mathbb{E}^{\widehat{P}} \left(\log \frac{1}{\widehat{M}(t)} \right) \quad (4.156)$$

$$= \mathbb{E}^{\widehat{P}} \left(\int_0^t \widehat{\theta} d\widehat{W}_s + \frac{1}{2} \int_0^t \widehat{\theta}_s^2 ds \right) \quad (4.157)$$

$$= \frac{1}{2} \int_0^t \widehat{\theta}_s^2 \quad (4.158)$$

$$= \widehat{\theta}_t \quad (4.159)$$

where:

$$D_t^{KL}(P(t)|Q(t)) = \begin{cases} < \infty, & \text{If } Q_t \ll P_t, \\ \infty, & \text{otherwise} \end{cases} \quad (4.160)$$

Condition (4.160) determines the highly remote but theoretically possible extreme case of an unbounded value due to the lack of absolute continuity of the measures⁸⁴.

Now, taking the difference among the two distances:

$$D_t^{KL}(P(t)|Q(t)) - \widehat{D}_t^{KL}(\widehat{P}(t)|Q(t)) = \theta_t - \widehat{\theta}_t \quad (4.161)$$

$$= \Delta_t \quad (4.162)$$

$$= \mathcal{F}_t - \mathcal{H}_t \quad (4.163)$$

⁸⁴Under some conditions, i.e.: the existence of a measure under which both P_t and Q_t are absolutely continuous, it is possible to have a finite value even if Q_t is not absolutely continuous with respect to P_t as well as we can have non finite values if the two measures are mutually absolutely continuous.

In finance the EMM has the advantage of being fully neutral and unaffected by subjective beliefs. Among the others, one of the main feature is that it prevents the problem of picking a parametric utility function to describe the fully non-parametric investors behavior. In this chapter we showed how, in presence of information premiums, both the risk-neutral and the real-world measure and the relative pricing kernel can be strongly affected. Overall, the asset pricing bias can be so summarized:

Missing information: $\Delta_t \Rightarrow$ Suboptimal filtration set: $\mathcal{H}_t \subset \mathcal{F}_t \Rightarrow$ Biased risk-physical measure:
 $\widehat{P}_t \neq P_t \Rightarrow$ Biased market price of risk: $\widehat{\theta}_t \neq \theta_t \Rightarrow$ Biased PK: $\widehat{M}_t \neq M_t \Rightarrow$ Biased risk-neutral
 measure: $\widehat{Q}_t \neq Q_t$
 \downarrow
 Mispricing: $X_t \neq \widehat{X}_t$

4.7 Future ideas

In this last section i show two possible theoretical extension that I plan to do as future works. I only sketch the idea without going in any detail, neither conceptual nor technical.

Optimal and suboptimal Riesz representation theorems, a comparison

Firstly I plan to analyze how the Riesz representation theorem may be affected by a suboptimal use of the information. Harrison and Kreps propose a time dependent conditional model by extending the time-independent original findings of Hansen and Richard. To do that they propose a new pricing model which uses all the necessary and relevant information at time t . Therefore, the passage from an unconditional to a conditional model is based upon the strong assumption of knowing, on each day, this information. Technically, the new functional is a pricing function which maps a time T payoff with a time t price. Both quantities are random variables. Given the above assumption and knowing that very often it is violated, the conditioning with respect to the time T payoff is still correct, the one with the time t price is instead too big (i.e.: if one relies only on historical data). I then will propose and compare a new mapping function from the time T payoff to a more conservative $t - \Delta_t$, where the delta represents as usual the information lost form the suboptimal use of the information.

Modelling and comparing optimal and suboptimal information processes

As a second theoretical idea I plan to model the stochastic evolution of the information from a theoretical viewpoint (complete) and from a real world viewpoint (suboptimal).

Assuming a generic form of the information to process:

$$\psi_{t,T} = \sigma_t X_T + \beta_{t,T}$$

where X_T are the info provided by the factors that compose the price (I assume only one factor). Is the factor which contains the true market info “market signal”.

The term σ represents the “information flow rate” for the factor X_T , the higher σ the quicker the info are revealed. The term $\beta_{t,T}$ is a Brownian Bridge which aim to represents the “market noise”. The idea of the Brownian Bridge is that we know the value of the process at time 0 and T but there are random movements in between.

This can represents the info provided by a conditional pricing kernel.

With a non or partially conditional pricing kernel, instead, we might assume the presence of a

Brownian motion, where the information is known at time 0, but not at time T .

The usual connection between Brownian motion and Brownian Bridge would suit perfectly with respect to a complete and an incomplete information. Moreover, given the nature of the underlines, I also plan to apply a Stochastic expansion (Platen) to model even better the two information processes.

CHAPTER 5

The Methodology - part I

This chapter describes the detailed estimation methodology employed in this thesis to estimate independently the two measures that govern the PK. The first part of the estimation model follows BEM (2008)[15] which, although econometrically advanced, it still only produces partially conditional results. The next section will go over this problem. The reason behind the extension and comparison of the BEM paper is demanding: extending an econometrically already advanced paper makes my exercise more challenging, hence more valuable for the literature. Starting from a more basic technology (i.e.: a basic kernel on a rolling window) would make my proposal less valuable. BEM 2008 already tries to propose a time varying and forward looking real world measure. While the former goal is achieved by reconstructing the daily stochastic volatility through a GARCH model, the latter is at best only partially achieved through an ad hoc use of the equity risk premium. The main limitation of their approach lies in the type of assets used in estimation: only a time series of stock return which, as anticipated in the previous chapters, makes impossible to infer naturally the risk physical density. After presenting the model, I analyze and discuss the main limitations of the original paper, above all from the point of view of the conditioning of the information. I conclude the chapter presenting how I propose to go over these limitations with the new estimation methodology. A full description of the new methodology is then carefully explained in the next chapter.

5.1 GJR GARCH-FHS and Monte Carlo simulation

The estimation of the fully-conditional EPK is based upon three steps. First I estimate, individually, the objective and the risk-neutral measures using stock and options data. Then, after the estimation of the adjusted risk-neutral measure, I make the still partially-conditional physical measure fully-conditional through the Dirichlet process. Taking the present value of the ratio of the now fully-conditional measures completes the estimation.

Following much of the literature, the physical density estimation of the first step, is based upon the strong assumption that historical data are correctly priced and fully informative. I relax it by means of the second step.

The intuition behind the risk-neutral density estimation is the following: at same way in which option prices are used to extract the otherwise unobservable implied volatilities, here I use options data to extract the otherwise unobservable investors preferences⁸⁵. To do it I reconstruct the stochastic volatility by means of an asymmetric GJR GARCH - FHS model. The main problem lies in the incompleteness⁸⁶ of the market which is implicitly assumed by a stochastic volatility. Knowing that a primitive asset can always be created by means of options, I complete the market by simulation but starting from real market data. Once that all states are priced, I can use them to estimate the relative pricing densities (Ross (1976)[139]). The idea is similar in spirit to Breeden and Litzenberger (1978)[34] although their model requires a high market densities for all states which is empirically implausible, even for the more liquid assets. I achieve it by simulation using real data.

5.1.1 The estimation of the physical parameters

Due to its documented flexibility and goodness of fit in estimating financial data in general (Christoffersen and Jacobs (2004)[46]) and the S&P 500 in particular, I reconstruct the stochastic volatility by means of an asymmetric GJosten, Jagannathan and Runkle (1993)[76] (GJR) GARCH (1,1) model. After testing the presence of GARCH effects in the time series in consideration (figures (A.6), (A.7), (A.8)), I describe the daily index dynamics under the objective distribution, p_t , through a GJR GARCH (1,1) model fitted onto the historical daily returns of the S&P 500, going back 3500 daily

⁸⁵The common approach, both for the industry and the academia, is to model parametrically the investor's preferences and then use the model for pricing. Unfortunately, investors preferences are highly non-parametric as well as unobservable.

⁸⁶Duan (1995)[55] proposes a direct change of variable by means of the Girsanov theorem but he does not allow for the market completeness as required by the theorem.

observations⁸⁷:

$$\log \frac{S_t}{S_{t-1}} = r_t = \mu + \epsilon_t \quad (5.1)$$

$$\epsilon_t = \sqrt{\sigma_t^2} z_t \quad (5.2)$$

$$z_t | \mathcal{F}_{t-1} \sim f(0, 1) \quad (5.3)$$

$$\sigma_t^2 = \omega + \beta \sigma_{t-1}^2 + \alpha \epsilon_{t-1}^2 + \gamma \epsilon_{t-1}^2 \mathbb{1}_{t-1} \quad (5.4)$$

$$\mathbb{1}_{t-1} = \begin{cases} 1, & \text{if } \epsilon_{t-1} < 0, \\ 0 & \text{if } \epsilon_{t-1} \geq 0. \end{cases} \quad (5.5)$$

where: S_t represents the time t S&P 500 price, μ the constant drift term⁸⁸ and ϵ_t the index innovation. For each time t the parameter ϵ_t is decomposed into σ_t : the daily market volatility and z_t , the standardized historical innovation.

Among the others, one of the main feature of the GJR GARCH is its ability in modelling skewness: the stronger impact caused by negative news. When $\gamma > 0$, the model accounts for the leverage effect⁸⁹: given the same magnitude of the shocks bad news, $\epsilon_{t-1} < 0$, raise future volatility more than good news, $\epsilon_{t-1} > 0$.

I refer to Appendix (A.5) for the minor technical assumptions behind the stationarity and non-negativity of the volatility estimation from the GJR GARCH model.

The scaled return innovation, z_t , is sampled from its empirical density function f , which is obtained by dividing each estimated return innovations, $\hat{\epsilon}_t$, by its estimate of the conditional volatility,

⁸⁷The analysis spans from January 1, 2002 for two years. Going back for 3500 days I stop right before the 1987 crash. More on this in section (9.1.4).

⁸⁸Although all estimated factors are time varying, to assume a constant value for μ in a small period of time has negligible effects on the final estimation. The same would not be true for the variance. In fact, for a small enough Δt :

$$\sigma_t^2 = c_t^2 \Delta t = O(\Delta t)$$

while for the mean:

$$\mu^2 = c2\Delta t^2 = O(\Delta t^2)$$

In any case, is better to consider (5.1) as an approximation.

⁸⁹Although still under debate (Figlewsky and Wang (2000)[72]), following much of the literature and after the 1976 paper of Black[26], I refer to the leverage effect as the asymmetric reaction of volatility to bad and good news, hence to positive or negative return innovations. Economically: a negative return implies a drop in the equity firm's value, thus increasing its leverage which, in turn, leads to a higher equity-return volatility.

$\sqrt{\hat{\sigma}_t^2}$ so that the standardized residual return is:

$$\hat{z}_t = \frac{\hat{\epsilon}_t}{\sqrt{\hat{\sigma}_t^2}} \quad (5.6)$$

This set of estimated scaled innovations gives an empirical density function that incorporates excess skewness, kurtosis, and other extreme return behaviours that cannot be captured under a normal density estimation. This approach, due to Barone-Adesi et al. (1999)[16], is called filtered historical simulation (henceforth: FHS). The essence of FHS is in some sense comparable to the bootstrapping. By applying a valid econometric model to historical prices such that it can filter out some of the well know characteristics of financial data like the fat tails, the leverage effect and the clustering⁹⁰, the remaining residuals thus form an empirical innovation of the assets. The resulting empirical innovations is a set of scaled innovations which capture possible non-normality presents in market data and can then be used for option prices simulation with the Monte Carlo machinery.

Table A.5 summarizes the set of estimated daily physical parameters $\theta_t = f(\omega, \alpha, \beta, \gamma)$ obtained by maximizing the pseudo maximum likelihood (PML), under the nominal, not necessarily true, assumption of normal innovations, following Bollerslev and Wooldridge (1992)[32]. This technique provides consistent parameter estimates even when the true innovation density is not normal, e.g., White (1982)[155], and Gouriéroux et al. (1984)[77]. Rosenberg and Engle (2002)[64] use the same approach to estimate the objective distribution of S&P 500 returns in their analyses.

5.1.2 The estimation of the risk-neutral parameters

For a given date (t), a GJR GARCH model (5.1) - (5.5) is calibrated to the cross section of out-of-the-money call and put options on the S&P 500, to capture the index dynamic under the risk neutral probability density function q .

For a given set of risk neutral parameters $\tilde{\theta} = f(\tilde{\omega}, \tilde{\alpha}, \tilde{\beta}, \tilde{\gamma})$, a return path is simulated by drawing an estimated past innovation, say, \hat{z}_1 , updating the conditional variance $\hat{\sigma}_{t+1}$, drawing a second innovation \hat{z}_2 , updating the conditional variance $\hat{\sigma}_{t+2}$, and so on up to $t + \tau$:

$$\frac{S_T}{S_t} = \exp(\tau\mu^* + \sum_{t=1}^{\tau} \sigma_{t+1}^2 \hat{z}_t) \quad (5.7)$$

where, following the FTAP, the risk neutral drift equals the risk free rate:

$$\mu^* = \text{R-N Drift} = \mathbb{E}^Q \left(\frac{S_t}{S_{t-1}} \middle| \mathcal{F}_{t-1} \right) = e^{r_t} \quad (5.8)$$

⁹⁰The non-normality of financial data is a well known fact, at least from Fama (1965)[68] and Mandelbrot (1966)[118]. The other features as well have been deeply analyzed and tested both theoretically and empirically.

The τ periods gross return, $\frac{S_{t+\tau}}{S_t}$, is simulated by means of a Monte Carlo simulation optimized by the Empirical Martingale Simulation (EMS) method of Duan and Simonato (1998)[56]. At each time I simulate $L = 20,000$ return paths from t to $t + \tau$.

The time t FHS - GJR GARCH call option price, with strike price K , and time to maturity τ , is given by the average of simulated payoffs, discounted for the daily risk-free rate. Put prices are computed similarly⁹¹.

A numerical search in the space of the risk neutral GJR GARCH parameters $\tilde{\theta} = f(\tilde{\omega}, \tilde{\alpha}, \tilde{\beta}, \tilde{\gamma})$ (table A.6) leads to a simulation that changes the simulated return paths, so as to best fit the cross-section of option prices available at date t , minimizing the mean squared pricing error:

$$\min_{(\tilde{\omega}, \tilde{\alpha}, \tilde{\beta}, \tilde{\gamma})} \sum_K \sum_{\tau} [|\text{Mkt P.}(K, \tau) - \text{Model P.}(K, \tau)|]^2 \quad (5.9)$$

where for each time t the above optimization represents the difference between the GJR-GARCH-FHS option pricing (Model P.) and the actual market values (Mkt P.) estimated for each vector of strikes K and time-to-maturities τ . The calibration is achieved when, varying the risk neutral GJR GARCH parameters $\tilde{\theta}$, the reduction in the mean square pricing error is negligible or below a given threshold.

For the minimization I use the MatLab built-in function `fminsearch` which performs the Neadler-Mead simplex method. This is a derivative-free unconstrained nonlinear optimization method which, starting at an initial settled value, finds the minimum value of a non-constrained multivariable function exploiting the simplex method. The Neadler-Mead algorithm has the advantage of being derivative-free but this at the cost of being a time expensive numerically intensive method. As alternative I propose a faster quasi-Newton algorithm (function `fminunc` in MatLab). Although faster Quasi-Newton methods may provide less precise results. By properly setting the initial conditions I am able to obtain qualitatively the same results in much less time. See table A.6 and A.7 for a comparison of the results. Being our methodology to calibrate the risk neutral parameters an intensive and time consuming method, to propose a time saving algorithm that produces same results is not of negligible importance.

From the above estimation it follows that both p_t and q_t follows a stochastic process of the same family but with different parameters. The asymmetry parameter γ , as well as the empirical innovations are of key importance to better capture the moments of the option surface.

⁹¹Each call payoff is computed as $\max(S_t^{Sim.} - K, 0)$ where $S_t^{Sim.}$ is the MC simulated underline price. By the same token each put payoff is defined as $\max(K - S_t^{Sim.}, 0)$

5.2 The main limitation: the incompleteness of the physical measure

As already anticipated theoretically in the previous chapters, the main limitation of the presented approach, that my thesis plans to address, is that BEM assumes that all the information about the objective probabilities of future returns, p , can be extrapolated from the historical record. This assumption may lead to large errors, that can be reduced if significant information on the physical distribution of security returns is available from other sources. In particular, option prices, used to estimate the risk neutral density, q , must reflect all the information publicly available to investors. Following the Bayesian approach, will take this information into account by centering on q our prior distribution for p (see chapter 5 for details).

Following Barone-Adesi, Mancini and Shefrin (2013)[18], I argue that the risk-neutral density from option prices may provide most of the missing information, once a risk premium is included to make it consistent with the objective measure. The above adjustment however implies that state prices are function of index return only. In the presence of other priced risk, such as the variance risk premium discussed by Heston and Nandi (2000)[87], or higher premium like the skewness or kurtosis risk premium it is interesting to investigate whether better objective estimates may be obtained rescaling the risk-neutral distribution to remove the volatility premium. However, recent results in Chorro et al. (2012)[43] suggest that the improvement induced by the variance risk premium, should be marginal. That should hold even more if parameters are re-calibrated.

The risk neutral density is then divided by the density of the revised (following the Bayesian non-parametric paradigm, as detailed later) physical probability, p^\dagger , to compute the pricing kernel. I do not plan to modify the numerator, that is the risk neutral density. In fact that would imply inefficiencies in the option market, which are not object of this thesis but can be an interesting investigation as future work.

CHAPTER 6

The Bayesian approach

The key concept at the core of my intuition for the modified physical measure is the idea of making it more complete by mixing the information extractable from the joint use of stock and option market data. Moreover, knowing that neither the PK nor real-world densities can be defined parametrically I aim to model both quantities without posing any parametric constraints but the ones required by the FTAP. Starting from these two stringent assumptions: the mixing property and the non parametric form of the result, I show in this chapter why the best tool to model the real-world distribution by means of options and stock data is through a Bayesian non-parametric model, namely the Dirichlet process (henceforth DP), which falls into the broader category of the Bayesian non-parametric distributions.

After showing why the non-parametric Bayesian approach suits well for my problem I introduce formally the main features of the DP focusing the attention on the precision parameter and its asymptotical behaviour. The precision parameter will be the key parameter for the final estimation of the measure. Building a bridge between statistics and economics, I then show how I use the DP to produce a time-varying conditional physical measure. As far as I know, the DP has never been used in literature for the estimation of the real-world measure. I close the chapter proposing possible empirical and theoretical Bayesian extensions that could provide even better final results than the one proposed in these thesis. Most of these models will be analyzed in future works.

6.1 Why a Bayesian non-parametric proposal?

As extensively defined in the previous chapter, the PK is the present value of a ratio of conditional densities, defined over two different spaces, the risk neutral one for the numerator and the physical one for the denominator (i.e.: equation (2.1)). Needless to say a correct procedure for the determination of the two conditional measures is of key importance for a correct pricing and estimation of the financial asset which the PK is referred to.

Under the most general definition, the problem of density estimation requires to properly define a coherent distribution that fits well the random sample of interest which, in my case, is the time series of returns of the S&P 500 index and its related European options (SPX).

All these data are highly non-normal, and to impose a parametric approach (a model indexed by a finite-dimensional parameter) for their estimation may lead to severe biases. Section 2.2.1 lists some of the main theoretical advantages of using a non-parametric approach for problems of density estimation. Empirically, there are many papers which show how different non-parametric approaches may lead to a better asset pricing, risk management and toward the resolution of some puzzles present in the literature i.e.: BEM[15] and Aït-Sahalia et al. (1998 and 2003)[3][2].

From a mere statistical point of view, a wrong use of a parametric models with a fixed and finite number of parameters can lead to over/under-fitting of data when there is a mismatch of complexity of the model (usually a bad parametrization) and the amount of data available.

To select the right model is often a difficult task, here it is where the Bayesian approach is of big help. Informally, a nonparametric Bayesian prior is a distribution over models such that the complexity of the model is unbounded, so random as well. Modelling using a Bayesian non-parametric approach the under-fitting problem is mitigated by choosing a functional form that is unbounded on the complexity; at the same time, the over-fitting is mitigated by the computation (approximation) of the full posterior over parameters.

Encouraged by these theoretical facts and the subsequent empirical applications, I adopt a fully non-parametric methodology to estimate a more informative real-world measure. As far as my problem is concerned, having some extra information to exploit and wanting to avoid any constraints the final functional form of both the PK and the real-world measure, a Bayesian non-parametric density estimation suits perfectly for this particular estimation problem.

As a consequence I estimate the conditional physical measure by means of a non-parametric Bayesian model. Following the Bayesian paradigm, a non-parametric density estimation is conducted by placing a prior probability model on the unknown density, where the parameters of this density are in some infinite-dimensional function space. Assume to have a set of observations (x_1, x_2, \dots, x_n) with an unknown distribution, A , I wish to infer given some observed data:

$$x_i \sim A \quad \text{for } i = 1, \dots, n$$

To solve this problem I place a prior over A and then I compute the posterior over A given the set of observed data. If the prior is chosen such that it lies wrongly in a constrained parametric family, then over/under-fitting problems automatically arise. On the contrary, a non-parametric approach places a prior over distributions with the widest possible support, usually the space of all distributions. Given the wideness of the space, I have to be wary of the tractability of the posterior which could not be obvious with extreme distributions. Much of the problem is by now solvable by the use of the Monte Carlo Markov chain (MCMC) techniques⁹². Given the problem and our goal a good choice for such a prior is the Dirichlet Process (DP, Ferguson, (1973)[69]). The process is fully characterized by two parameters α and C : $B \sim \text{DP}(\alpha, C)$ and has a support over discrete with a countable infinite number of point masses (more details in the next section).

The absence of dimensional constraints is a key point that allow me to have more freedom in the estimation of the functional form of the real-world measure thus leading to less biased final PK estimations. This goes against a huge amount of literature that models parametrically the fully non parametric PK.

In a nutshell, I adopt a Bayesian non-parametric approach to avoid both critical dependence on parametric assumptions and to exploit, trough the prior, the information in the risk neutral density, in order to improve the estimate of both the objective probability and, as a consequence, of the PK.

6.2 The Dirichlet Process

From the family of the Bayesian non-parametric distribution models the DP is a discrete random distribution introduced by Ferguson (1973)[69], as a direct consequence of the Kolomogorov con-

⁹²As for most of the Bayesian techniques, also the use of the DP, although very appealing in principle, has been historically halted by storage and computational intractability problems. As a consequence of the improved computational powers of computers many Bayesian techniques are now living a second childhood. The same of course applies to the here used DP.

sistency theory. The DP is in fact a stochastic process (a random distribution) whose samples are probability measures (generically referred as Random Probability Measures (RPM)) with probability one and such that its marginal with respect to any finite partition has Dirichlet distribution. By the definition of stochastic processes⁹³ the probability measures upon which lies the distribution of the DP has to be on some probability space, Ω , and such that the draws from the DP are fully random distributions. Given the Kolomogorov consistency theorem⁹⁴ for a random distribution, (B) , to be distributed as a DP, its marginals have to be Dirichlet distributed as well.

Following the Kolomogorov consistency theory⁹⁵, a compact formal definition of the DP is as follows: let (Ω, \mathcal{F}) be a finite, additive, non-negative and non-null measurable space, where Ω is a set and \mathcal{F} a σ -field on the set and let μ be a measure (unnormalized density) on this space. A random probability measure P^μ on (Ω, \mathcal{F}) is a DP with parameter μ if: whenever $\{D_1, D_2, \dots, D_r\}$ is a measurable partition of Ω (i.e., each $\mu(D_r) > 0$ for all r ⁹⁶), then the joint distribution of random probabilities is distributed as:

$$(P^\mu(D_1), P^\mu(D_2), \dots, P^\mu(D_r)) \sim \text{Dir}(\mu(D_1), \mu(D_2), \dots, \mu(D_r)) \quad (6.1)$$

where Dir denotes the standard Dirichlet distribution. Informally, P^μ is a DP if it behaves as if it were a Dirichlet distribution on any finite partition of the original space. Reducing the compactness of the definition it is possible to write $\mu = \alpha C$ as the product of the two parameters where $\alpha \in \mathbb{R}^+$ and $\alpha = \int_{\Omega} d\mu$ while the latter describes a density and is such that $C = \mu/\alpha$ and $C : \mathcal{F} \rightarrow (\mathcal{F} \rightarrow [0, 1])$. Then:

$$B : \mathcal{F}' \rightarrow (\mathcal{F} \rightarrow [0, 1]) \quad (6.2)$$

is a DP: $B \sim \text{DP}(\alpha, C)$ ⁹⁷. Focusing on the parameters α is a measure of precision (also known as concentration parameter) while C is a centering parameter (also known as base measure); both

⁹³Stochastic processes can be defined as distributions over function spaces with sample paths being random functions drawn from the distribution.

⁹⁴Given a probability space (Ω', \mathcal{F}') , is enough to provide a consistent definition on the marginals of the stochastic process itself to have a valid stochastic process. For the case of the DP, the marginals are Dirichlet distributions. See Appendix (A.6) for proof of the consistency of the marginals.

⁹⁵The same result can also be achieved using the stick breaking approach which is has the advantage of requiring less strict regularity assumptions.

⁹⁶This translates to: D_r are measurable disjoint events: $D_r \in \mathcal{F}$ and their union equal Ω .

⁹⁷By the Kolomogorov consistency theory, there exists a probability measure \mathbb{P} on $([0, 1]^{\mathcal{F}}, B^{\mathcal{F}})$ yielding these distributions where $([0, 1]^{\mathcal{F}}$ is the space of all functions from \mathcal{F} into $[0, 1]$ and $B^{\mathcal{F}}$ is the σ -field generated by the field of cylinder sets.

play an intuitive and key role in the definition of the process and its moments. The base parameter is in fact the mean of the DP while the precision parameter is the inverse of the variance. For any measurable set $D \subset \Omega$ (See Appendix (A.7)):

$$\mathbb{E}(B(D)) = C(D) \quad (6.3)$$

$$\mathbb{V}(B(D)) = \frac{C(D)(1 - C(D))}{(\alpha + 1)} \quad (6.4)$$

Measuring for how much one believes on the prior parameter is immediate to see that the variance is inversely proportional to the precision parameter: the smaller its value and the less the DP concentrates around the base. The concentration parameter is also called the strength parameter, referring to the strength of the prior when using the DP as a nonparametric prior over distributions in a Bayesian nonparametric model, and the mass parameter, as this prior strength can be measured in units of sample size (or mass) of observations.

6.2.1 Posterior distribution

By the definition of the DP the distribution $B \sim \text{DP}(\alpha, C)$ is a random distribution so that is possible to draw samples directly from it. Let $\rho_1, \rho_2, \dots, \rho_n$ be the sequence of independent draws from the random distribution B . It is immediate to see that $\rho_i \subset \Omega \quad \forall i = 1, \dots, n$.

Let D_1, D_2, \dots, D_r be a finite measurable partition of Ω and $n_k = \#\{i : \rho_i \in D_k\}$ be the number of observed values in D_k . By (6.1) the conjugacy between the Dirichlet and the multinomial distributions are still Dirichlet distributed:

$$((B(D_1), \dots, B(D_r)) | \Omega_1, \dots, \Omega_n) \sim \text{Dir}(\alpha C(D_1) + n_1, \dots, \alpha C(D_r) + n_r) \quad (6.5)$$

Being (6.5) valid for all finite measurable partitions, by the Kolomogorov consistency theory, the posterior distribution over B is a DP.

The relation between (6.1) and (6.5) shows that the DP has a conjugate family of priors over distributions that is closed under posterior updates given observations. After some algebra (Appendix (A.8)):

$$B | \Omega_1, \Omega_2, \dots, \Omega_n \sim \text{DP}\left(\alpha + n, \frac{\alpha}{\alpha + n} C + \frac{n}{\alpha + n} \frac{\sum_{i=1}^n \delta_{\rho_i}}{n}\right) \quad (6.6)$$

where δ_i is a point mass located at ρ_i and $n_k = \sum_{i=1}^n \delta_i(A_k)$.

Therefore the posterior distribution, once updated for the observations, is still a DP with an updated precision parameter: $\alpha + n$ and a base equal to a weighted average between the prior and the empirical distribution and where the average tilts toward one of the two depending on the weight given to the prior through α .

6.2.2 The DP for the physical estimation

Let $P_{t,T}$ and $Q_{t,T}$ be the probability measures associated with $p_{t,T}$ and $q_{t,T}$ respectively. My aim is to estimate the unknown density $p_{t,T}$, or, equivalently, the corresponding probability distribution, $P_{t,T}$. To avoid economical inconsistencies (i.e.: arbitrages) I center my DP prior on the risk-neutral measure $Q_{t,T}$, adjusted by a risk premium (more on this in section 7.1.1). This risk adjusted risk-neutral measure is here represented by $Q_{t,T}^*$. The DP has a conjugate family of priors over distributions that is closed under posterior updates given observations. Therefore, the DP posterior mean, after collecting for the observations, is still a DP:

$$P_{t,T}^\dagger = \frac{\alpha^*}{(\alpha^* + n)} C + \frac{1}{(\alpha^* + n)} \sum_{i=1}^n \delta(S_i) \quad (6.7)$$

$$= \frac{\alpha_{t,T}^*}{(\alpha_{t,T}^* + 1)} Q_{t,T}^* + \frac{1}{(\alpha_{t,T}^* + 1)} P_{t-\Delta t, T} \quad (6.8)$$

where: $\delta(S_i)$ is the point mass distribution at the return S_i (where S_i is the set of observations), and the precision parameter, $\alpha_{t,T}^*$ ⁹⁸, being calibrated daily, is time-varying.

Equation (6.8) is equivalent to (6.7) where, the base parameter, C , is the modified risk-neutral distribution $Q_{t,T}^*$ obtained by simulation and $\sum_{i=1}^n \delta(S_i) = P_{t-\Delta t, T}$. Given that both $P_{t-\Delta t, T}$ and $Q_{t,T}^*$ are empirical measures obtained by simulation there are no homogeneity problems in setting $P_{t-\Delta t, T} = \sum_{i=1}^n \delta(S_i)$ where T changes accordingly to what I need to simulate. To improve the readability of the problem, and with no loss of generality I normalize $n = 1$ (more details in section (7.1.1)).

The density $p_{t,T}^\dagger$, corresponding to the average of the posterior distribution, $P_{t,T}^\dagger$, is used in the denominator of the EPK giving rise to a revised version of it, $M_{t,T}^\dagger$. In other words, the proposed EPK, $M_{t,T}^\dagger$, is now the discounted Radon-Nikodym derivative of $Q_{t,T}$ relative to $P_{t,T}^\dagger$.

Note that, even though the DP is an a.s. discrete random probability measure, by smoothing the histograms of the simulated distributions with a kernel I obtain an a.s. continuous distributions. In general, by convolving B with kernels so that the resulting random distribution has a smooth density, I mean that: $B \sim \text{DP}(\alpha, C)$ and $f(x|\Omega)$ is a family of densities (kernels) indexed by Ω so that I have a non parametric density of x by:

$$p(x) = \int f(x|\Omega) B(\Omega) d(\Omega) \quad (6.9)$$

which is close to a natural extension of the DP: the mixture models.

⁹⁸To avoid confusion with respect to the ARCH parameter α , I identify the precision parameter used for the density estimation by $\alpha_{t,t+\tau}^* = \alpha_{t,T}^*$.

6.2.3 The precision parameter

Acting directly on the weight given to the prior and indirectly into the variance of the process, the crucial value of the DP lies in the precision parameter $\alpha_{t,T}^*$. For the estimation of the physical measure, the value of $\alpha_{t,T}^*$ tells us how much importance I want to give to the modified risk-neutral measure, $q_{t,T}^*$.

- When $\alpha_{t,T}^* = n$, equal weight is given to both measures
- Values of $\alpha_{t,T}^*$ increasingly smaller than n amount to give less and less weight to the prior opinion:

$$\lim_{\alpha_{t,T} \rightarrow +0} p_{t,T}^\dagger = p_{t-\Delta_t, T} \quad \text{s.t.} \quad M_{t,T}^\dagger = PV \left(\frac{q_{t,T}}{p_{t-\Delta_t, T}} \right) = M_{t-\Delta_t, T} \quad (6.10)$$

Asymptotically the conditional physical measure is equal to the original objective measure and the relative EPK turns the original one. Throughout the paper the subscript $t-\Delta_t, T$ represents, for each time t , the fraction of missing forward looking information.

- On the other hand, larger and larger values of $\alpha_{t,T}^*$ amount to give increasing importance to the information contained in the option prices:

$$\lim_{\alpha_{t,T} \rightarrow +\infty} p_{t,T}^\dagger = q_{t,T}^\dagger \quad \text{s.t.} \quad M_{t,T}^\dagger = PV \left(\frac{q_{t,T}}{q_{t,T}^\dagger} \right) \quad (6.11)$$

such that the obtained physical measure is now only composed by the scaled risk-neutral measure

Calibrating $\alpha_{t,T}^*$ to estimate a fully-conditional EPK, the precision parameter provides an indirect proxy with respect to the degree of missing conditionality provided by the partially-conditional objective measure $p_{t-\Delta_t, T}$. The larger the correction the higher the amount of missing information Δ_t . Or, from the opposite view, the higher the amount of sentiment in the market. By the same token, and due to the strong tightness between the measures and the EPK, the precision parameters has high explanatory power with respect to the non observable risk premium. Empirically, it emerges that for the EPK the best value for $\alpha_{t,t+\tau}^*$ is decreasing in options maturity. Increasing the precision parameter as the the option's time-to-maturity, τ , gets smaller means to give more weight on the forward looking beliefs extracted from the close-to-expiration options. The shorter the period, the more the information provided are reliable and less subject to changes and the more it makes sense to exploit the options information for the EPK . On the contrary, the larger the time-to-maturity

and the lower is the reliability of the forward looking information. As a consequence I give less and less importance to the prior information by reducing $\alpha_{t,t+\tau}^*$ correspondingly. The flexibility of the precision parameter allows us to adapt the proposed model to different market scenarios.

6.3 Possible extensions

To avoid the almost sure discreteness of the DP, which is inappropriate in many applications, DP mixtures (MDP) can be considered (see Escobar, (1988)[66], MacEachern (1994)[117], Escobar and West (1995)[67]. In MDP, the random probability measure is represented through an additional convolution: $P(x) \sim \int f(x|\Omega)H(\Omega)$, $H \sim DP(G, \alpha)$, where f is some kernel indexed by parameter Ω . When using a Gaussian kernel and mixing over both the mean and the variance parameter, the resulting density estimates resemble traditional kernel density estimates.

Due to the nature of the problem, other more flexible priors can be used. Among them the Pitman-Yor process (1992 and 1997)[129][131], which generalizes the DP having an additional “discount parameter” that allows to model power-law tails rather than only exponential tails (as for the DP). Completely Random Measures (Kingman (1967)[106], Kingman (1993)[107]) and Dependent Completely Random Measures (Lijoi et al. forthcoming[112]) are more general models that are useful when the observations are only partially exchangeable⁹⁹.

For priors more general and flexible than the DP one needs to resort to efficient MCMC algorithms in order to estimate and sample from the resulting posterior distributions (whose average will substitute the original p). Is it possible to use the seminal algorithms of Escobar (1994)[66] and Escobar and West (1995)[67] for the DP and the subsequent related literature that gave rise to the so called “marginal” sampling schemes, as well as “conditional” methods that rely on the stick-breaking representation of the DP (see the original paper by Ishwaran and James (2001)[90], all the way to the more recent ones by Walker, (2007)[153]; Papaspilopoulos and Roberts (2008)[127]; and Kalli et al. (2011)[104]).

From a statistical point of view is possible to read the proposed modification of the real-world measure and the subsequent PK as the construction of a different importance distribution used for pricing derivatives (see Asmussen and Glynn (2007)[8]and references therein for a review of importance sampling techniques): instead of the traditional p , the proposed importance distribution, p^\dagger ,

⁹⁹See Müller and Quintana (2004)[123] for a review of Bayesian non-parametric density estimation.

give rise to importance weights ($\frac{q}{p^*}$ i.e. the new PK) that are more stable and thus give rise to a variance reduction of the resulting importance sampling estimators of derivatives.

As described in section (2.2.1), parametric assumptions are difficult to justify for financial data and the misspecification of the model can lead to a misspecification of the PK. This can be extremely costly for an option seller, who would not be able to hedge his position well enough. Even by just assuming a more realistic stochastic process for the underlying asset, it becomes impossible to find a closed form solution for the PK. Furthermore, as empirically shown in Campbell, Lo, & MacKinlay (1997)[38], the parametric specifications of asset prices available in the literature have many pitfalls. For these reasons, it is also possible to “increase” the level of the non-parametrization with a Bayesian fully non-parametric estimator of the PK based on the Bernstein Polynomials (BP) framework from Wang and Ghosh (2012)[154]. The only information that still required to be used are basic properties of the problem to analyze. For option pricing purposes these comes from the no-arbitrage assumption and are: monotonicity and convexity of the call pricing function.

Although challenging, the BP framework allows for the implementation of the no-arbitrage restrictions directly in the estimation process, without artificially enforcing them after an unconstrained estimation. In this sense it is comparable to BEM. The no-arbitrage restrictions are vital for the whole option pricing theory, so I would want to make their preservation a top priority. Moreover, Carnicer and Pena (1993)[39] show that no other polynomial can beat the BP in terms of preserving the shape form of the approximated function.

CHAPTER 7

Making the Physical Measure Forward Looking

In this chapter I show how to make the still partially conditional physical measure estimated with the model presented in chapter 5 fully conditional. As anticipated in previous sections is econometrically not possible to make a measure forward looking, hence complete and conditional, by means of forward looking inputs only. The novelty of the proposed solution is then a byproduct of the type of inputs used, both options and stock rather than just the latter, and a novel econometric model that makes use of a Bayesian application. In this chapter in fact I see how to join the Dirichlet Process (DP) presented in chapter 6 to the econometric model presented in chapter 5. As often happens in financial literature, a non trivial task to perform is how to properly calibrate the risk premium in the model.

After a quick review of the unconditional methodology, I focus the attention on how to calibrate the different risk premiums needed for the estimation of the modified physical measure, hence of the modified pricing kernel. I close the chapter presenting theoretically how the intensive numerical nature of the econometric model behind the PK estimation can be a problem in presence of measures misestimations.

7.1 The Methodology - part II

Using the GJR GARCH physical parameters $\theta = f(\omega, \alpha, \beta, \gamma)$ calibrated from the time series of log returns and the risk neutral parameters $\tilde{\theta} = f(\tilde{\omega}, \tilde{\alpha}, \tilde{\beta}, \tilde{\gamma})$ calibrated from the cross section of OTM option prices, I simulate, for each day t of the time series, a stochastic volatility which is then used to as a data generating process (DGP) to simulate n log-prices through Monte Carlo.

On the simulated time series of log-prices I perform a non-parametric kernel density estimation on m^{100} equally spaced points to extract the densities $q_{t,T}$, $q_{t,T}^*$ and $p_{t,T}$. The measures $q_{t,T}^*$ and $p_{t,T}$ are then combined through the Dirichlet machinery to estimate $p_{t,T}^\dagger$. Then the modified pricing kernel $M_{t,T}^\dagger$ is estimated by taking the present value of $q_{t,T}$ over the modified physical measure I estimate .

7.1.1 The equity risk premium adjustments

Through the DP, the second step of the estimation makes the physical measure fully-conditional. If data would be independent and identically distributed (i.i.d.) there would not be any problem in comparing a conditional with an unconditional estimation since the two would be the same. In practice, finance data are highly non i.i.d. and conditionality is crucial for a proper estimation. The problem is very well-known in the finance literature. Proposing a model that uses instrumental variables to move from a static to a dynamic model Cochrane (2001) [48] analyzes the problem comparing a conditional with an fully unconditional model. From page 132 of [48] “Of course, another way to incorporate conditioning information is by constructing explicit parametric models of conditional distribution”. Unfortunately it is known that modelling a non-parametric functional parametrically leads to a.s. biased results. Differently with respect to the generic approach, I compare a fully with a partially dynamic model, where the former is obtained by means of a semi-parametric model.

For the estimation of the revised physical distribution, $p_{t,T}^\dagger$, three ingredients are needed. The “classical” physical, $p_{t-\Delta,t,T}$, and risk-neutral, $q_{t,T}$, measures and the modified risk-neutral measure, $q_{t,T}^*$. Not to violate the FTAP, a proper estimation of the former and the last, requires the calibration of a risk premium. The objective measure directly depends on the expected return, which is empirically recognized to be hard to estimate. In line with the neoclassical literature, I follow Merton (1980)[121] and I scale the drift used to simulate the log-prices by a fixed value of 8% or 4%¹⁰¹. Although violations are usually the product of different flaws, a miscalibrated risk premium can be the reason of the non-monotonicity of the EPK in the central area of the functional. Both the estimation of the risk premium and the precision parameter are non trivial tasks. Since their “role” is to model non-parametric random human behaviors, both parameters turn out to be highly stochastic. I propose to solve the issues empirically for $\alpha_{t,T}^*$, linking its value to the liquidity of

¹⁰⁰For robustness I estimate the density with a different number of points so that m that ranges from 100 to 10.000.

¹⁰¹While the former is the original value proposed by Merton, I also use half of it. The 8% is in fact based upon the returns in the US during a period of economic growth which might be too high given the period of my exercise.

the option market and calibrating it up to the best EPK, and theoretically for the risk premium. A deeper analysis for the estimation of these two quantities and their interconnections is left for future researches.

Representing the fixed daily percentage risk premium with, ϕ , and the daily percentage dividend yield with, d_t ¹⁰², the three different drifts used to simulate the daily asset prices are:

- Physical measure (p_t):

$$\text{Drift}_{p_t} = \frac{1}{365} \left(\sum_{t=1}^T r_t - d_t + \phi \right) \quad (7.1)$$

- Risk-neutral measure (q_t):

$$\text{Drift}_{q_t} = \frac{1}{365} \left(\sum_{t=1}^T r_t - d_t \right) \quad (7.2)$$

- Modified risk-neutral measure (q_t^*):

$$\text{Drift}_{q_t^*} = \frac{1}{365} \left(\sum_{t=1}^T r_t - d_t + \phi \right) \quad (7.3)$$

where $\frac{1}{365} \sum_{t=1}^T r_t$ represents the daily risk free rate¹⁰³.

Using (7.1) - (7.3) as input for the drift, the estimated θ_t for the objective and the FHS (or Gaussian) $\tilde{\theta}_t$ for the two risk-neutral distributions as inputs for the daily volatilities I obtain, as output, n simulated log-prices, where n are the number of simulations. Although for the entire estimation I set $n = 50.000$ the value can be flexibly changed depending on the different market scenarios and the computing power. For $i = n = n$. sim, the simulated log-price S_i is defined as:

$$\hat{S}_{i,t} = S_t \exp \left(\text{Drift} - \frac{\hat{\sigma}_{t,\text{sim}}^2}{2} \right) dt + \hat{\sigma}_{t,\text{sim}}^2 dW_t \quad (7.4)$$

where S_t represents the S&P 500 price at a given day, the Drift input changes dependently on what I need to simulate, $\hat{\sigma}_{t,\text{sim}}^2$ represents the simulated volatility obtained through the GJR GARCH parameters, dt is fixed at one day and dW_t represents the canonical Brownian motion. Depending on the exercise, I model dW_t by means of the FHS or the normal distribution.

Using a Monte Carlo pricing model, it is well-known that, due to the exponential semi-martingale structure of the simulated sample path (7.4), I need to correct it to prevent propagation errors that

¹⁰²Differently than the dividend yield and the risk free rate, the risk premium, being fixed, is assumed to be time-independent.

¹⁰³For each time t , the daily risk free rates, r_t , are obtained by interpolation

would lead to arbitrage violations. Using the Empirical Martingale Simulation (EMS) of Duan and Simonato (1998)[56] I optimize the Monte Carlo simulation:

$$S_{i,t}^{\text{corr.}} = S_t e^{r_t \cdot T} \frac{1}{\frac{1}{n} \sum_{i=1}^n \hat{S}_{i,t}} \quad (7.5)$$

where: $\hat{S}_{i,t}$ is the i^{th} simulated asset price at time t prior to the optimization adjustment and the maturity days are the common expiration days of the options. Equation (7.5) is made of a temporary created asset price at time t and the arithmetic average of it. I apply the correction to each iteration and time-to-maturity:

$$S_{i,t} = \hat{S}_{i,t} \cdot S_{i,t}^{\text{corr.}} \quad (7.6)$$

The effect of the empirical martingale adjustment is that the simulated sample (7.6) does not violate the martingale property thus producing rational option pricing bounds.

To extract the probability density functions from the estimated log-prices, I perform a non-parametric kernel density estimation evaluated at n equidistant points¹⁰⁴ with a Gaussian kernel and an optimal bandwidth computed following the Silverman rule of thumb¹⁰⁵. I repeat the same procedure for the three different measures and both for Gaussian and FHS innovations thus obtaining six different densities: three with the FHS innovations $p_{t-\Delta_t,T}^{\text{FHS}}, q_{t,T}^{\text{FHS}}, q_{t,T}^{*,\text{FHS}}$ and three with the Gaussian ones: $p_{t-\Delta_t,T}^{\text{Gauss}}, q_{t,T}^{\text{Gauss}}, q_{t,T}^{*,\text{Gauss}}$.

As of now, the objective densities are still biased. I solve the issue mixing the estimated measures through the DP. Taking the present value of the pricing measure over the revised physical measure complete the estimation.

As a consequence I pass from a partially-conditional EPK:

$$M_{t-\Delta_t,T} = e^{-r_t \cdot \text{Maturities}} \left(\frac{\text{Risk-neutral}_{t,T}}{\text{Risk physical}_{t-\Delta_t,T}} \right) \quad (7.8)$$

$$= e^{-r_t \cdot \tau} \left(\frac{q_{t,T}}{p_{t-\Delta_t,T}} \right) \quad (7.9)$$

$$= \text{PV} \left(\frac{q_{t,T}}{p_{t-\Delta_t,T}} \right) \quad (7.10)$$

¹⁰⁴Given the number of simulations of the experiment, I obtain the best estimates setting $n = 1000$.

¹⁰⁵The classical trade-off between the smoothness of the functional (bias) and jaggedness of the estimate (variance) arises here. Dealing with univariate data and using a Gaussian kernel the optimal bandwidth is:

$$h = \left(\frac{4\sigma^5}{3n} \right)^{1/5} \approx 1.06\sigma n^{-1/5} \quad (7.7)$$

where σ is the standard deviation of the distribution.

to a fully-conditional EPK:

$$M_{t,T}^\dagger = e^{-r_t \cdot \text{Maturities}} \left(\frac{\text{Risk-neutral}_{t,T}}{\frac{\alpha_{t,T}^*}{\alpha_{t,T+1}^*} \cdot \text{Modified risk-neutral}_{t,T} + \frac{1}{\alpha_{t,T+1}^*} \cdot \text{Risk physical}_{t-\Delta_t,T}} \right) \quad (7.11)$$

$$= e^{-r_t \cdot \tau} \left(\frac{q_{t,T}}{\frac{\alpha_{t,T}^*}{\alpha_{t,T+1}^*} \cdot q_{t,T}^* + \frac{1}{\alpha_{t,T+1}^*} \cdot p_{t-\Delta_t,T}} \right) \quad (7.12)$$

$$= \text{PV} \left(\frac{q_{t,T}}{p_{t,T}^\dagger} \right) \quad (7.13)$$

Proposing a numerically intensive estimation method it is not convenient to set the parameter n of the posterior means of the DP equal to the real number of simulated observations and then adequate $\alpha_{t,T}^*$. To prevent the use of non intuitive large numbers, which would make my analysis hard to follow, and with no loss of generality, I set $n = 1$. It follows that a value of $\alpha_{t,T}^* > 1$ would give more weight to the prior information, vice-versa for $\alpha_{t,T}^* < 1$.

The model behind the EPK (7.13) presets two interesting features: high flexibility, provided by the manipulation of the parameter $\alpha_{t,T}^*$ which can be easily adapted to different market conditions, and high generality since the model is fully independent of the estimation methods used to obtain the different measures.

7.2 A frequentist justification of the physical measure correction

The estimation technique presented in this thesis makes high use of numerically intensive simulation methods. In this subsection I insert a word of cautions toward possible statistical problems one can encounter estimating the EPK by means of a numerically intensive procedure.

Needless to say estimating a functional with biased inputs leads to a.s unexpected results. This is further amplified if the functional to estimate is a ratio and uses a.s. noisy real market data. It is in fact well-known (Jobson and Korkie (1980)[101] and Jobson and Korkie (1981)[102]) that ratios are highly unstable operators which may be biased even when their inputs are not and that market data, \hat{x} , are not frictionless with an a.s. mean zero noise: $x \neq \hat{x} = x + y$ where y , $\mathbb{E}^P(y) = 0$ is the possible noise. The effect of these biases may get even larger if the estimation is achieved as a result of a numerically intensive method. I show this point by means of a Taylor expansion adapted to a numerical experiment. Results are in line with the small sample bias of Leisen (2015)[109]

who warns for possible statistical explanations of the EPK non-monotonicity. The frequency of the estimation and the size of the sample used in estimation may increase/decrease these problems. As a consequence, no result can be seen as fully correct but only as an approximation.

As a largest picture, given a domain $D \in \mathbb{R}^+$, the main goal is to evaluate a bivariate function $f(x, y)$ where: $(x, y) \subseteq \mathbb{R}^2$, $D \subseteq \mathbb{R}^2$ and $z \subseteq \mathbb{R}$ such that:

$$f : D \subseteq \mathbb{R}^2 \rightarrow \mathbb{R}, (x, y) \rightarrow z = f(x, y) \quad (7.14)$$

Given the presented problem, the function, f , evaluates the ratio of two measures: q/p so that $z = \text{EPK}^{106}$.

More precisely I want to evaluate the expected value of this ratio under the physical measure:

$$\frac{q}{p} = \frac{\mathbb{E}^P(q)}{\mathbb{E}^P(p)} \quad (7.15)$$

Evaluating z by means of a numerical method which makes large use of different simulations, the obtained results are a.s. approximations. As a consequence, the simulated function is a.s. a biased estimator:

$$\frac{\mathbb{E}^P(q)}{\mathbb{E}^P(p)} \leq \mathbb{E}^P\left(\frac{q'}{p'}\right) = \mathbb{E}^P\left(\frac{q + \Delta q}{p + \Delta p}\right) \quad (7.16)$$

where the term on the right hand side, represents the expected value under the simulation method so that $\Delta p(q)$ represents the computational noise factor for $p(q)$.

To better explain how I approach the problem, a short excursus of numerical mathematics is needed¹⁰⁷. Whenever one deals with a generic computational problem where the goal is to find (x) such that:

$$f(x, d) = 0 \quad (7.17)$$

through a numerical method that approximates (\cdot) the above problem by a sequence of approximated models of the form:

$$\hat{f}_n(\hat{x}_n, \hat{d}_n) = 0 \quad n \geq 1 \quad (7.18)$$

different types of errors can arise. In the above equations $f : \mathbb{R} \rightarrow \mathbb{R}$ is the functional form that relates (x) and (d) , where (d) represents the dataset. Errors arise a.s. because equation (7.17) models a real-world problem which is solved by means of a computational model (7.18).

¹⁰⁶To lighten the notation and with no loss of generality, in this section I omit visually the time dependence t, T .

¹⁰⁷For a comprehensive analysis of the subject see, through the others, Quarteroni et al. (2007)[137].

Assuming (7.17) to be a Ill posed problem, the total error (e_T) from the approximation of the numerical (\hat{x}_n) and the real value (x) is the sum of mathematical (e_m) and computational (e_c) errors:

$$|x - \hat{x}_n| = e_T \quad (7.19)$$

$$= e_m + e_c \quad (7.20)$$

$$= (|x - f(x)|) + (|x - \hat{f}_n(\hat{x}_n)|) \quad (7.21)$$

$$= (e_{algo} + e_{dt}) + (e_{ds} + e_r + e_a) \quad (7.22)$$

From (7.22), the mathematical error (e_m) can be decomposed as the the sum of the mathematical model error (e_{algo}): how Ill the algorithm chosen to model (7.18) describes the real problem, and the data error (e_{dt}): how Ill the dataset (\hat{d}_n) used describes the real dataset. A further decomposition of e_m would show that the type of algorithm chosen to compute (7.17) has no impact on e_{dt} . Given the framework of my problem: $e_m \geq 0$.

The computational (or machine) error (e_c) is instead formed by: (e_{ds}) the numerical discretization error or truncation error that fills the gap between the analytical and the approximated solution, (e_r) the machine roundoff error which, using floating-point arithmetics, inevitably arises during the concrete resolution of (7.18) and (e_a) the error introduced by the numerical algorithm. In general and also in my case $e_c > 0$. It follows that, a.s. $e_T > 0$.

Back to main case, I define $e_T = \Delta p(q)$ as the sum of all possible errors for the estimation of the physical (risk-neutral) measure. Of the two measures, I assume the risk-neutral one to be unbiased. Supported by the empirical literature, the higher reliability of the risk-neutral measure with respect to the physical one can be also implicitly deduced from one of the most common PK violation. By a flex in the central area of the functional some papers¹⁰⁸ show the existence of a risk seeking behavior in the area of zero or nearly zero returns. This is by far the area with the highest amount of option prices available. Liquidity and mispricing are not an issue here. It follows that the central area is where the risk-neutral measure is at its highest precision thus moving on the risk physical measure the responsibility of the violation. This empirical fact, summed with the superior structure of the option data in reflecting the investors information conditionally support strongly my assumption.

¹⁰⁸From JackIrth (2000) and Engle and Rosenberg (2002)[64], the term ‘‘PK puzzles’’, in principle, has been coined to represents exactly this bias in the central part of the distribution.

It follows a.s. that $\Delta p > 0$ while $\Delta q \approx 0$ so that $p' = p + \Delta p$ while $q' = q$; which leads to:

$$\mathbb{E}^P \left(\frac{q'}{p'} \right) = \mathbb{E}^P \left(\frac{q + \Delta q}{p + \Delta p} \right) = \mathbb{E}^P \left(\frac{q}{p + \Delta p} \right) = q \mathbb{E}^P \left(\frac{1}{p + \Delta p} \right) \quad (7.23)$$

Assuming $(p + \Delta p) \in C^n$ for $n \geq 2$ and $\neq 0$, I approximate the expectation of the ratio through a second order Taylor series expansion (see appendix:(A.2)):

$$\mathbb{E} \left(\frac{q'}{p'} \right) = \mathbb{E}^P \left(\frac{q}{p + \Delta p} \right) \quad (7.24)$$

$$= \frac{q}{p} + f'_p \Delta p + \frac{1}{2} f''_{pp} \Delta p^2 + O(n^{-1}) \quad (7.25)$$

$$= \frac{q}{p} - \frac{q}{p^2} \Delta p + \frac{1}{2} \frac{2q}{p^3} \Delta p^2 + O(n^{-1}) \quad (7.26)$$

$$\simeq \frac{q}{p} + \frac{q}{p^3} \text{Var}(\Delta p) \quad (7.27)$$

Where f'_p and f''_{pp} are the first and second derivative with respect to p and the noise has a.s. zero expectation so that: $\mathbb{E}^P(p + \Delta p) = p$.

By the same token, the variance of the ratio is:

$$\mathbb{V} \left(\frac{q'}{p'} \right) \simeq \frac{q^2}{p^2} \left(\frac{\text{Var}(q)}{q^2} \right) \quad (7.28)$$

From the expansion the inequality (Jensen's inequality) in (7.16) is demonstrated. The greater the variance of the physical error, the larger the distance between the numerical $\mathbb{E}^P(q/p + \Delta p)$ and the mathematical $\mathbb{E}^P(q)/\mathbb{E}^P(p)$ models used to estimate the EPK.

Exploiting the flexibility of the proposed model, I work on this inequality adjusting the denominator of the ratio up to finding the the best functional form: the one that gives an economically valid EPK. The correction is executed by mixing differently the two measures that compose the modified physical measure. Thanks to the higher informativeness of the model, I can reduce the effect of these errors by reducing the variability of the physical noise. If the dataset in use does not pose possible small sample issues, this is achieved exploiting other sources of information which, completing the measure, would make the whole ratio more stable, hence less biased. As it will be made clear in the empirical sections (chapter 9), this is particularly true at the extremes of the distribution.

CHAPTER 8

The Datasets

To test empirically the efficiency of the proposed physical measure I test my model using two years of market data (2002-2004). Following much of the literature, the data are properly filtered before being used. This creates a natural trade-off between the quantity and the quality of the data used. In these type of problems the filtering operations are of extreme importance for the final result of the problem. Imposing too tight constraints I would end up having a non liquid database, thus producing invalid and misleading outputs, This is especially true for those values that lie deeply into the tails of the distributions that could be very informative to explain possible puzzles that arise at the extremes of the EPK. On the contrary, not imposing any constraints, I would end up with possibly mispriced option prices which could lead to other biases. Stocks and options market data refer to the S&P 500 Index returns and the European options on the S&P 500 Index (henceforth, SPX) traded on the Chicago Board of Options Exchange (CBOE). To complete the picture, I use the daily time series of risk free rate and dividend yields to capture the time preference of the PK. All data are from OptionMetrics.

8.1 The S&P 500 Index Options - The SPX

The cross-section of European SPX option prices ranges from January 2, 2002 to December 29, 2004. I only use use daily closing prices of out-of-the-money (OTM) and deep-out-of-the-money (DOTM) European put and call options, duly filtered as explained below.

The reasons behind the choice of the dataset are manifold. First and foremost: one of the biggest problem for the EPK literature is to find a valid benchmark. Aside from the theoretical geometrical

shape superimposed by the neoclassical theory is difficult to imagine the right PK. This is true and probably even truer also for “famous” days (i.e.: the Lehman Brothers bankruptcy day). Starting from this issue I consider to use as a benchmark an econometrically advanced paper concerning the estimation of both the real world measure and the PK. Therefore, for consistency and robustness, I use the same dataset and filtering criteria as BEM. Second, the SPX is among the most actively traded option index market in the world; this assures richness of data and reduces liquidity problems¹⁰⁹. Third the SPX has no wild-card features which makes the overall evaluation process easier.

The SPX market contracts traded on the CBOE have strike price intervals of 5 and 25 points. The choice of using only OTM and DOTM options is to reduce possible liquidity problems. If the choice is mainly effective for put options, the same is not always true for call options. This creates an asymmetry problem which is reflected in the empirical results. It is an empirically fact that, starting from the October 1987 crash, the daily volumes of OTM puts is much larger than the relative ITM puts; this reflects the strong demand of portfolio managers of protective puts to hedge against possible negative movements. At the same, there is not such a strong demand for OTM and DOTM call options. The lower liquidity of call options is at the core of most of the problems that arise at the right tail of the pricing kernel. Despite that, I will show in the next chapter the high importance of exploiting properly the information embedded in the call options to fix the U-shaped EPK violation.

Option moneyness is a way to define how deeply an option is in or out of-the-money. There are many different ways to define it. For sake of simplicity I go for the most basic and used definition:

$$\text{Moneyness}_t = \frac{\text{Strike price}_t}{\text{Asset price}_t} \quad (8.1)$$

$$\varrho_t = \frac{K_{i,t}}{S_t} \quad (8.2)$$

where $K_{i,t}$ represents the i^{th} strike for a given day t . From it, the degree of moneyness for put options is:

- out-of-the-money (OTM) $\rightarrow 0.85 \leq \varrho < 1$
- deep out-of-the-money (DOTM) $\rightarrow \varrho < 0.85$

while for call options is:

¹⁰⁹This is the main reason why it has been used by several authors for many different empirical problems. As far as the pricing kernel is concerned, these index has been used often times (Äit-Sahalia and Lo (1998)[3], Chernov and Ghysels (2000)[42], Heston and Nandi (2000)[87], and Polson and Stroud (2003)[133].)

- out-of-the-money (OTM) $\rightarrow 1.00 \leq \varrho < 1.15$
- deep out-of-the-money (DOTM) $\rightarrow \varrho \geq 1.15$

This classification will be of key importance for the calibration of the precision parameter.

From the point of view of the time-to-expiration ($T = t + \tau$), the SPX expiration months are the three near-term and the three additional months from the March, June, September and December quarterly cycle. To filter out the data for our analysis I also divide the options by time to maturity in:

- Long maturity: $\tau > 160$ days
- Medium maturity: $60 \leq \tau \leq 160$ days
- Short maturity: $\tau < 60$ days

From the entire sample size, only Wednesday prices are considered. Wednesday is in fact the least likely day to be a holiday and the less likely to be affected by the day-of-the-week effect, which is usually more present at the extreme of the working days (Monday and Friday). When Wednesday happens to be a holiday (three cases), I take the next first-available working day¹¹⁰.

Options prices are intended as the average of the bid-ask prices while the implied volatility estimated by OptionMetrics has been replaced by the Black & Scholes implied volatility extracted in MatLab knowing all other required data¹¹¹.

To avoid further liquidity problems, extra filtering criteria for the option prices, expirations and volatility has been applied. Option prices with values lower than \$0.05 are discarded. The same holds for options with a too short ($\tau < 10$ days) or too long ($\tau > 360$ days) time-to-maturity and for Black and Scholes implied volatilities greater than 70%.

After filtering the entire dataset the sample size accounts for 29,211 observations which are statistically summarized in table A.1. The sample size account for an average of 186.1 option prices for

¹¹⁰After an extended holiday, data may be erroneously more volatile than usual. For example, if the stock exchange remains closed for three days in row, at the first opening day the volatility could be higher than the reality. Of course, the value would not be exactly equal to the sum of the possible realized volatilities of the previous three closing days, but still may show unexpected higher values, thus producing extreme unjustified outliers that need to be carefully examined and filtered out, if necessary.

¹¹¹Knowing and trusting on the underline market prices, the option market prices, the strike prices, the expiration periods, the risk-free rate and dividend yield downloaded from OptionMetrics, and by using the `blsimpv` function built-in MatLab, I extract the volatility surface formed by the Black & Scholes implied volatility.

each Wednesday with a minimum of 142 up to a maximum of 237 contracts (standard deviation of 22.3). Overall: put and call options are evenly represented: 50.7% and 49.3%. The range of prices goes from short maturity DOTM put (call) of \$0.77 (\$0.34) to \$38.80 (\$34.82) of long maturities OTM options. The volatility smile is also clearly present: as the time-to-maturity gets larger, the smirk gets flatter while the bid-ask spread becomes narrower. Short/medium and long term maturities are almost equally represented with: 33%/31%/36% while OTM put (call) options account for 25% and 27% respectively.

8.2 The S&P 500 Index

The time series of S&P 500 index prices ranges for a larger time period starting from January 2, 2002 and going back for 3500 observations up to the 1988. The time series, voluntarily, stops exactly after the 1987 crash.

The choice of using indexes with respect than single stocks is justified by the different empirical advantages of the former. From Browns & Gibbons (1985)[36], under some mild assumptions, consumption data can be correctly substituted by financial asset data. To accomplish this task, a good proxy that covers and represents as much of the entire economy as possible is needed. This translates into a broad and liquid enough financial instrument. Among the different financial products present on the market, the S&P 500 index and options on it, best fulfill these requirements. Another advantage of using indexes with respect than single stocks is the diversification power. Indexes are by definition a large portfolio of stocks. By construction are more diversified and less likely to be affected by heavy jumps thus making their modellization easier.

Table A.2 shows summary statistics of the S&P 500 index prices and log-returns for the two periods of interest. Different statistics of the table confirm the presence of the usual features that characterize financial data, among them the strong non-normality of the distribution emerges.

8.3 The risk-free rate and the dividend-yield

After downloading the time series of daily zero coupon from OptionMetrics, the curve of the daily risk-free interest rates is obtained by interpolation. For each time-to-maturity (τ), the previous (τ_-) and next (τ_+) time period zero-coupon whose values straddle the time τ are linearly interpolated. Finally, the time series of S&P 500 index dividend yield from January 2, 2002 to December 29, 2004 has been downloaded and left untouched. The option price literature treats dividends in several

different ways with greater or minor impact on the final result. In our case I use the time series of dividend to compute an ex-dividend spot index level.

CHAPTER 9

Empirical Results

In this chapter I present and discuss the empirical results obtained applying the methodologies explained in chapter 5, 6 and 7 on the databases presented in chapter 8. In addition to compare the results with the main empirical findings present in literature, I use the BEM's paper as a benchmark to analyze in which directions the machinery of the modified EPK can improve it.

As a main finding, following the proposed methodology I obtain monotonically decreasing EPKs, hence results in agreement with the neoclassical economic theory. Results hold during days of “normal” market returns and, even more interestingly, during periods of bearish/bullish and bullish/bearish market returns¹¹². In addition to explain some of the puzzles governing the option literature, the benefits given by the proposed model are: the high flexibility provided by the precision parameters, the full generality with respect to the inputs used, the wider range of supports within is possible to find economically valid results, the high frequency of results which goes up to each single day and time-to-maturity, a faster convergence in simulation and a higher robustness to numerical noise errors (Sala and Barone-Adesi (2015)[145]). In a nutshell, the proposed model reduces the informative gap between the theoretical requirements and the empirical findings. As a weakness, the model depends to the economically and econometrically consistent but still semi-arbitrary choice of the starting value of the precision parameter. Related to the paper but independently with respect to the mixing procedure, the proposed estimation method depends from the arbitrary choice of the risk premium used in simulation and, indirectly, from the degree of liquidity of the underlying.

¹¹²For “normal” days I mean either periods of low volatilities or with a low activity (liquidity) in the market.

The empirical analysis focuses on both the EPK and on the densities of it. For all empirical findings I study the economic implications of the obtained results.

It is of key importance to underline that, for all tests, I always use the same estimated risk-neutral measure while changing the physical one. This approach gives me stability in the methodology in addition to an empirical confirmation. Since under the same risk-neutral measure I pass from economically invalid to economically valid EPK, I claim that a possible reason of the EPK non-monotonicity is an econometric issue that lies at the heart of the objective density estimation.

9.1 The classical and the revised EPK, a comparison

Applying the methodologies presented in (4) and (6) for each Wednesday, t , from January 2, 2002 to December 29, 2004 and for each time-to-maturity, τ , I obtain 862 observations for both the fully-conditional EPKs ($M_{t,t+\tau}^\dagger$) and the partially-conditional EPKs ($M_{t-\Delta_t,t+\tau}$).

From a practical point of view, a proper estimation of the EPKs on a day-to-day basis has an impact on many operations, i.e.: asset pricing and risk management. Despite its crucial role and due to its nature, a proper daily estimation of the EPK is as much important as non trivial from an econometric viewpoint. It is common knowledge that estimating the ratio of two highly volatile random values may lead to very unstable results. The higher the frequency of the results required, the greater is the complexity in estimation. Differently than what has been proposed in the literature, my methodology, although numerically intensive, does not put any unreal stationary assumption neither on the measures nor on the EPK, allowing us to have time-dependent, hence reliable results, for the desired time period (from daily to weekly, monthly and/or yearly, depending on the needs). Exploiting the implied moments of the options surface, allow us to have an objective measure that is non-zero in those areas where the information provided by the time series of stock returns are, by structure, negligible or null. The presented model is then interesting both from an econometrically and from an economic prospective.

Not fully capturing sentiment and time dependence, most of models may still work decently during days with a low trading of OTM and DOTM options and poorly during volatile days. Most of results are mainly biased into the tails. Enlarging the time window of the estimation usually fixes (or better, hides) the problem. By averaging out all daily estimates of the EPKs it is possible to wash out possible extreme and meaningless data, thus obtaining an EPK that is a valid

approximation on a yearly basis. Aït Sahalia and Lo (1998)[3] and Jackwerth (2000)[93] are two examples of non-time-varying EPKs¹¹³. Resorting on kernel estimations they need of not implied nor required by the theory stationary assumption on the estimation of the objective measure. It turns out that what they propose are time-invariant EPKs averaged on a yearly basis. Not fully capturing sentiment and time dependence, some models may still work decently during days with a low trading of OTM and DOTM options and poorly during volatile days. Most of results are mainly biased into the tails. Enlarging the time window of the estimation usually fixes (or better, hides) the problem. By averaging out all daily estimates of the EPKs it is possible to wash out possible extreme and meaningless data, thus obtaining an EPK that is a valid approximation on a yearly basis. Using kernels over past returns not only leads to unconditional measure, it also limits the estimations at low frequencies. Second, it is common knowledge that estimating the ratio of two highly volatile random values may lead to very unstable results. The higher the frequency of the results required, the greater is the complexity in estimation. Differently than what has been proposed in literature, the presented methodology, although numerically intensive, doesn't need any unreal stationary assumption neither on the measures nor on the EPK to provide fully time varying results. Relying on more informative data, from which making inference is a natural task, and exploiting the high flexibility provided by the DP, the proposed model has no limitations on the time-frequencies thus making it interesting both from an econometric and economic viewpoint.

Given the importance of the estimation frequency and to show the robustness of the findings over different time periods, I divide the empirical results into daily and yearly estimates. As a matter of precision, both the estimated EPK cannot be considered as non-parametric methods but only as semi-parametric ones. In fact, even though all EPKs are by construction fully non-parametric since I do not put any constraints on the shape of the functionals aside from the non-arbitrage constraints, all estimated measures are still based upon fully parametric GJR GARCH estimations¹¹⁴. Much more important than the classification of the degree of parametrization of the model that I use, is the classification of the degree of homogeneity and informativeness of the model. With the proposed method I pass from a biased non-homogeneous definition of the EPK - where I compare a

¹¹³Or at best, time-varying on a yearly basis.

¹¹⁴In general also the majority of the so called non-parametric models cannot be considered as fully non-parametric. In fact, at the root of any empirical analysis one needs to impose at least some minimum regularity conditions to guarantee a valid statistical inference - thus reducing implicitly the class of models - or directly assume a class of model - explicit restriction. This is valid in general and is even truer whenever one deals with problems that involve the estimation of the EPK and (above all) the estimation of the risk-neutral measure.

conditional risk-neutral measure with a non-conditional physical measure - to an unbiased homogeneous one. This is true by construction since the risk-neutral measure, being extracted from option prices only is fully conditional to all the available information in the market. On the contrary, the physical measure, being based upon a stream of past historical returns cannot be fully conditional nor informative.

As a result, by empirically mixing the two measures, I have from one side an economically valid denominator, and from the other a conditional measure which can now be compared with the conditional risk-neutral measure, thus producing an homogeneous EPK. I thus obtain an economically and econometrically unbiased function.

9.1.1 Daily estimates

Figure (A.11) shows the estimation of a random day, $t = 61$, taken from the time series of Wednesdays; namely February 26, 2003. To provide a complete picture of the time horizon, the figure is divided vertically in different times-to-maturity (from the shortest maturity, the ones depicted on top, all the way down to the longest). For $t = 61$ the set of τ ranges from 24 to 297 days¹¹⁵. The figure is then divided vertically in two parts. The graphs of the left column show the estimated measures, respectively: the risk-neutral density, $q_{t,t+\tau}$ ¹¹⁶ dotted in red, the partially-conditional physical density, $p_{t-\Delta,t+\tau}$ dotted in green and the thicker continuous blue line for the fully-conditional risk physical density, $p_{t,t+\tau}^\dagger$. For all graphs the index price is struck at S_t , here equal to \$827,56. The graphs of the right column show, for the same time and for each time-to-maturity, the relative conditional EPKs $M_{t,t+\tau}^\dagger$. The green(red) line represents the estimates obtained with 50.000 simulations and FHS(Gaussian) errors.

The precision parameter, $\alpha_{t,t+\tau}^*$, is flexibly calibrated such that it best explains the economical conditions present on the market. For $t = 61$ it goes from 2 for the closest time-to-maturity and decreases to 0.8 for the longest τ (more on the calibration of $\alpha_{t,t+\tau}^*$ in the next section (9.1.2)).

The figure shows all pros and cons of the proposed model which I divide in: values of the functional, values of the horizontal support (domain of stock returns) and values of the vertical support (domain of the EPK). Although at different magnitudes, the entire sample presents results in line with figure (A.11).

¹¹⁵All times-to-maturity are filtered to be within 360 and over 10 days.

¹¹⁶Given the structure of the figures, which are divided horizontally for the different times-to-maturity, I prefer to use $_{t,t+\tau}$ to make the concept clearer and easier to follow. Anyhow, $_{t,t+\tau} =_{t,T}$ are equal and can be used interchangeably.

Starting from the functional, aside for possible numerical frictions, all conditional EPKs are almost everywhere monotonically decreasing. Second, most of sentiment lies into the tails of the distribution. As a consequence, the higher informativeness of the model allows us to extract these tails information from the OTM and DOTM options, hence to enlarge the horizontal support of the EPK. The capability of exploiting a larger support is one of the main feature of the proposed model. Relying only on stock data at the denominator, most of papers present in literature suffer from being informative only in a very narrow area, the one ATM. Appendix (A.9) shows a brief literature review of some of the most important EPK horizontal supports.

Not having the tails under control may lead to big problems. This is particularly true into the left region of the EPK (i.e.: Risk management). A larger support also increases the degree of precision and effectiveness of daily pricing and hedging operations. To enlarge the support is important as well as difficult for contract with very short time-to-maturity. The reason of the difficulty is a byproduct of possible low liquidity of DOTM options and mispricings. It is in fact known that the area of interest for a trader gets narrower as the expiration approaches. Moreover, tail contracts may have erratic behaviors due to mispricing from trading strategies related to contract expirations and to hedging strategies which requires to rollover the position into later expirations. Although for a narrower area, short time-to maturity data are not silent and, above all for the put options, may be very liquid, hence informative. For short-term operations it is thus of key importance to be able to extract as much valuable information as possible. However, this is impossible just using historical stock data. Although econometrically nontrivial, a scenario analysis without short term results would be economically incomplete thus much less usable.

In some of the figure of the paper I may leave part of the supports unbounded to show the higher explanatory power of a proper use of the information set. As a drawback, the plots can sometimes be noisy as the liquidity dries up. As a demonstration, in a day with a low liquidity of D/OTM call(put) options, the right(left) fraction of the graph may get unstable as I go into the right tail. This confirms the sensitivity of the model to options liquidity.

Third, results on the vertical support (domain of the EPK) are in line with the ones of a utility maximizer rational investor. As I will show in section (9.1.4) only FHS values produces non puzzling results.

To corroborate and emphasize the above points, I compare the fully and partially-conditional EPKs for another day of the time series. Figure (A.12) shows the January 16, 2002 EPKs where the partially-conditional FHS(Gaussian) are in black(blue). Dealing with options information,

the chosen day aims to represent the average properties of the entire data sample. With 36(33) DOTM(OTM) put options, out of a total of 69 puts traded and 40(53) DOTM(OTM) call out of 93 call options traded, the chosen day is in fact in line with the overall average liquidity of 52%(48%) for the DOTM(OTM) put and 54%(45%) for the DOTM(OTM) call of the entire sample. Very similar results are obtained for days with analogous characteristics i.e.: $t = 85/86/87/88/89/90$ thus confirming the robustness of the model. The graphs confirm the findings on the functional and on both the horizontal and vertical supports. Not sharing the exploding values at the extremes and the flex into the central area of the functional, the fully-conditional EPKs produces valid, bounded and stable results for both supports. On the contrary, the partially-conditional values becomes very unstable also for values close to S_t . Taken from a large viewpoint, it seems that the fully-conditional EPKs extracted with Gaussian and the FHS are similar. It turns out that this is not entirely true. I will show this point in the next sections along with the importance of using empirical innovations, above all during strongly non-Gaussian days¹¹⁷. As visible, while for the left side of the graphs there is a big difference between the two models, there is a common convergence on the call side. As I will show in the next sessions, things change dramatically when the majority of the information is embedded into the call options.

To summarize, the theoretical foundation behind my model leads to results in agreement with classical economic theory. Empirically the good quality of the results are a bivariate product of the forward looking information extrapolated on a daily basis from the implied moments through the GJR GARCH model and the empirical mixing of the measures through the DP.

9.1.2 Linking the precision parameter to the options market

Starting from the insights of the previous section, it is now easier to show how I calibrate the values of $\alpha_{t,t+\tau}^*$. The main goal of this paper is to test, through the calibration of the precision parameter of the DP, the ability of a fully-conditional EPK to lead to a monotonically decreasing, hence economically valid, function. More precisely, calibrating the precision parameter by “completion”, that is, using $\alpha_{t,t+\tau}^*$ as a proxy for the missing sentiment, I test the potentiality of extracting relevant information from different market data. Given the structure of my methodology, effects load primarily on the physical measure and then, by construction, on the entire EPK thus improving both quantities.

¹¹⁷Although the Gaussian fully-conditional EPKs produce good results for $t = 3$, it still shares the flex in central area for the shortest time-to-maturity. Feature not shared by the FHS estimation which, being able to better capture the information of the call options, produce fully monotonically decreasing EPKs.

Previous analysis have shown empirically how most, if not the entire amount of information, lies into the option market¹¹⁸. More precisely lies into the OTM and DOTM options. Dealing with an informational issue and solving it letting the data speak as much as possible, I link the value of the precision parameter to the relative proportion of DOTM and OTM options traded out of the total amount of options in circulation.

Table (A.1) shows summary statistics of the OTM and DOTM options of my dataset. As expectable, both for call and put, DOTM options have mainly long times-to-maturity, while the opposite holds for OTM options. Specifically for the period, my sample size shows a substantial amount of long-term DOTM call options. Option prices share the expected behavior with respect to moneyness and time-to-maturity and, from the implied volatilities surface extracted by inverting the Black & Scholes formula given the market prices, it emerges the so-called volatility smile. Graphically, the left column of figure (A.13) on the top panel, shows with a blue(red) line with a circle(star) at each edge, the daily percentage of DOTM call(put) options. As visible, while the amount of DOTM put options traded is quite always more stable and around 50%, the percentage of DOTM call options is extremely high in the first half, up to almost 80%, and then decreases strongly in the second part up to less than 20%, which is in line with the decreasing level of the volatility for that period. The left and right bottom figures show the time series of the total percentage of DOTM and OTM put and call options and the relative precision parameters. The former uses the scale on the left axis and the latter the right one.

To be able to extract as much information as possible and to reach an high degree of precision and flexibility in estimation, $\alpha_{t,t+\tau}^*$ is linked to the daily liquidity of the option market by a fixed percentage thus producing a time series of time-varying precision parameters. For each single day, the highest value of alpha is for the shortest time-to-maturity, $\alpha_{t,t+\tau_{\text{Min}}}^*$. The subsequent precision parameters have decreasing values as the single maturities get far in time. The highest value is then the starting point to determine all the others. To initialize the entire time series of $\alpha_{t,t+\tau_{\text{Min}}}^*$, I search for the day with highest amount of options traded - divided for moneyness - and I manually

¹¹⁸Other analysis in the next sessions will support and confirm this sentence.

calibrate its best alpha. Daily moneyness is defined as $\lambda_t = \frac{K_t}{S_{t,T}}$:

$$\lambda_t = \begin{cases} \text{DOTM Call} = \kappa_t(D)_C, & \text{if } \lambda_t > 1.15, \\ \text{OTM Call} = \kappa_t(O)_C, & \text{if } 1 < \lambda_t \leq 1.15, \\ \text{OTM Put} = \kappa_t(D)_P, & \text{if } 0.85 \leq \lambda_t < 1, \\ \text{DOTM Put} = \kappa_t(O)_P, & \text{if } \lambda_t < 0.85, \end{cases} \quad (9.1)$$

Since the horizontal domain of the EPK accounts for both put and call options and since, given the daily option surface, I want to analyze where is the highest amount of information available, I sum up the amount of OTM or DOTM put and call options for each day of the time series so that κ_t represents the fraction of OTM or DOTM options traded out of the total:

$$\kappa_t = \begin{cases} \kappa_t(\text{OTM}) = & (\kappa_t(O)_C + \kappa_t(O)_P) \\ \kappa_t(\text{DOTM}) = & (\kappa_t(D)_C + \kappa_t(D)_P) \end{cases} \quad (9.2)$$

The maximum value is taken over all days of the time series where, as usual, $t = 1, \dots, 157$. Therefore, to determine the time series of alphas I use:

$$\vartheta_t = \frac{\alpha_t}{\max_t \left(\frac{\kappa(\cdot)_t}{\text{Total}_t} \right)} \quad (9.3)$$

where ϑ_t represents the fixed percentage, Total_t is the total amount of traded options for a each day of the time series and $\kappa(\cdot)_t$ is the daily sum of options divided for moneyness.

For example, to link the precision parameter to the market of the DOTM options:

$$\vartheta_t = \frac{\alpha_t}{\max \left(\frac{\kappa_t}{\text{Total}_t} \right)} \quad (9.4)$$

$$0.0368 = \frac{2.5}{67.91} \quad (9.5)$$

where the value of 67.91% is achieved for $t = 41$. Once that the “starting point” is determined, here is equal to 2.5, the other values of $\alpha_{t,t+\tau}^*$ follow just by inverting (9.3).

Of the entire process, the only “arbitrary” choice is the value of 2.5. This is only partially arbitrary since, to determine it, I let again the data speak. It turns out that for the highest values of both OTM and DOTM options, the value of 2.5 is the one that empirically leads to the best shape¹¹⁹ given the data. To account for the decreasing reliability of the information provided by the options

¹¹⁹For the best shape I mean the one that produces a monotonically decreasing function, as required by the neoclassical theory.

as the time-to-maturity increases, for each time t , once settled the starting values, the value of $\alpha_{t,t+\tau}^*$ decreases monotonically as τ increases. For example, for the case of $t = 61$ presented above, the precision parameter goes from 2 to 0.8 with 6 times-to-maturity. It follows that the values of the precision parameter are not only dynamic on a day-to-day basis, but also with respect to the different times-to-maturity thus assuring an high degree of precision in estimation. As a consequence my time series of $\alpha_{t,t+\tau}^*$ is made of 862 values, one for each estimated EPK.

Top and bottom panel of the right column of figure (A.13), show the same analysis but for OTM options. While confirming the results relative to the behavior of put options¹²⁰, an opposite trend characterizes the behavior of the OTM call options. As a direct consequence of the high amount of DOTM call options traded in the first half, also the graph related to the total amount of DOTM and OTM options reflect this mirror behavior. Being the values of $\alpha_{t,t+\tau}^*$ linked to them by a fixed proportion, the time series of the precision parameter shows similar features. Table A.4 shows summary statistics of the different $\alpha_{t,t+\tau}^*$. While of course the maximum values are in line, means, medians and standard deviations reflect the behavior given by the underlying used. As a minor technical detail, I increase the readability of the analysis - with no loss of generality - by using rounded numbers for the starting values of $\alpha_{t,t+\tau_{\text{Min}}}^*$. Although both satisfactory, the method that provides better results is when $\alpha_{t,t+\tau}^*$ is linked to the percentage of DOTM options: another indirect confirmation of the importance of the information that lies into the deepest tails of the option surface.

9.1.3 Option moments and the conditional physical measure

The good results showed so far are even amplified if I analyze the single operations more closely. The theoretical backbone behind Harrison and Kreps (1979)[80] is the thigh interconnection between the EPK and its measures. This feature allows us to create an equivalent martingale measure (EMM) through the Radon-Nikodym derivative. From the FTAP, it follows that interconnection and equivalence are determinant properties to obtain a monotonically decreasing EPK. As a immediate consequence, a non puzzling EPK can only be obtained if its measures are in line. Equivalence that is lost, by construction, if the measures do not share the same degree of conditionality with respect to the information set. Although valid in general this is even truer into the tails: the areas of higher noise.

¹²⁰The overall percentage of OTM put options is, as expectable, always above 50% and even higher than the amount of DOTM put options.

In this section I study, graphically and numerically, the statistical properties of the conditional physical distribution and I compare them with the partially-conditional and the risk-neutral ones. As a main finding, passing from a partially to a fully-conditional measure, the obtained new measure, benefiting from the “extra” forward looking information extracted from the options surfaces, is more complete. The smaller gap between what is theoretically required and empirically delivered is translated, in statistical terms, in a new measure that is now closer with respect to the risk-neutral one¹²¹. This turns out to be particularly true with respect to the third and fourth moments of the distribution, which are of key importance for understanding the behavior of any finance phenomenon. As expected, I obtain better results with FHS innovations than with Gaussian ones. To show the higher explanatory power provided by the implied moments to the modified physical measure, table (A.9) reports summary statistics of the difference between the fully and partially-conditional physical measures with respect to the “benchmark” unbiased risk neutral one:

$$\Delta_{t,T}^{\text{Cond.}} = q_{t,T} - p_{t,T}^{\dagger} \quad \text{And} \quad \Delta_{t-\Delta_t,T}^{\text{Part. Cond.}} = q_{t,T} - p_{t-\Delta_t,T} \quad (9.6)$$

Results are on a yearly basis¹²² and duly divided in short ($\tau < 60$), medium ($60 \leq \tau \leq 180$) and long ($\tau > 180$) time-to-maturity. As a main finding, while the impact of the information on the first two moments is, as expected, negligible, things change for skewness and kurtosis. As a common result, the partially-conditional measure has a delta for higher moments that are, at least, 1.9 times bigger than the fully-conditional ones. Results decrease as the time-to-maturity increases which confirms the higher impact of sentiment for the short run. Over the sample in consideration, the magnitude of the results decreases for the end 2002 beginning 2003 and has big values for the mid and long 2003. Due to the constant up-trend of the S&P 500 and its volatility during the 2003, results confirm the higher ability of the model to capture the informative content given by the presence of call options. Although I only report the FHS results all analysis have also been performed using Gaussian innovations. Results confirm the FHS findings and amplify the Δ_t : a further confirmation toward the usefulness of the empirical innovations.

Given the importance for the day-to-day estimation of the EPK and to gain insights on behavior of the measures, figure (A.14) shows the time series of $\Delta_{t,t+\tau}$ estimated with (9.6) with τ =short-time. The continuous blue(dotted red) line represents the difference between the fully(partially) condi-

¹²¹Results confirm and extend broadly the findings of Linn et al. (2014)[113].

¹²²To prevent possible inconsistencies and following the same time frequency of the empirical models presented above, results are obtained taking the arithmetic mean of the measures computed on a weekly basis. As a result I have respectively 52/53 and 52 values for 2002/2003 and 2004. Daily results confirm the same findings found on a yearly basis and are available upon request.

tional objective measure and the risk-neutral one. As a main finding, the absolute distance between the fully-conditional and the risk-neutral measures is always smaller than the same distance with respect to the partially-conditional physical measures. The spread between the two lines represents the paucity of forward looking information which cannot be captured by only using historical data. Results can confirm my theoretical insights and show the higher stability provided by the conditional measures for all moments: a well seen feature, above all for practitioners. Although omitted visually, the daily analysis for the medium and long time-to-maturity confirm the statistical analysis presented above.

From the above results it is now clearer that, by mixing the measures, the information extrapolated from the options' moments complete the information set, thus producing closer densities. Dealing with leptokurtic finance data, the magnitude of the completeness gets higher and clearer into the extremes. Thus, after the analysis of the entire surface, I now put a focus on the tails of the same distributions.

Figures (A.15) and (A.16) are a focus on the whole spectrum of the left tails of the distributions¹²³.

The impact on the left tail of the conditional physical distribution is stronger as I go for the shortest times-to-maturity and for more volatile days. As visible, while the partially-conditional physical measure has a shape that is very distant from the (assumed) unbiased risk-neutral measure, the conditional one is much closer. This is true also where the shape of the tail is very erratic. As expected, changes on $\alpha_{t,T}^*$ has a much stronger impact on the tails than a change of the risk premium. By increasing the former, the distance between the conditional and the partially-conditional measure with respect to the risk-neutral increases more than changing the risk premium. The central columns show the same experiment with higher α_t^* , while the rightmost show the same tails with a lower risk premium.

For the day in consideration, due to the lower liquidity of DOTM call options, the same quality of results cannot be achieved for the right tails. Figure (A.17) and (A.18) show that, although the conditional physical measures provide better results, there are still many divergent values. Not much can be done where there are almost no data. The paucity of data is also confirmed by the shorter length of the tail: the longer the higher is the number of liquid DOTM call options¹²⁴. Also

¹²³For robustness, I repeated the same analysis with different kernels and different number of points, from 500 to 5000. Best results are obtained with a normal kernel. At the cost of a higher computation time and only a negligible refinement on the shape, an increasing number of points does not provide a significative improvement on the final findings. The proposed figures have been computed with a normal kernel and 5000 points. Other results are available upon request to the authors.

¹²⁴By construction, results improve for days with a dense amount of DOTM call options, i.e $t = 63$ (March, 12,

for the right tails, results improve as I increase the time-to-maturity thanks to the higher amount of long term DOTM call options and the values of $\alpha_{t,T}^*$.

To summarize the above results I show the time series of daily left distances. Given its importance in risk management I keep working on the left tails. Figure (A.19) shows the log distance between the risk-neutral and the other two densities taken at the first decile¹²⁵. The blue continuous (green dotted) line represents the distance with respect to the fully (partially) conditional measure. Confirming the above findings, the proposed technique produces more stable and less distant results. Given the role of the EPK, to propose the first decile is a compromise between usability and precision. The proposed model in fact allows us to extend the horizontal support. Showing the log distances of the first decile is thus a good way to show how the use of the options' information provides great stability up to the tails. The bottom figure of (A.19) proposes the same analysis but with a 95% percentile. Confirming the above results, these results may be of importance for risk management operations. As visible, the obtained results show the higher noise that is common as I go deeper into the tails. At different magnitudes, the same findings have been found for the right tails. Results are available upon request to the author.

To conclude the analysis on the distributions I graph the yearly estimates of the three measures. The fully-conditional physical measure, $p_{t,T}^{\dagger,year}$, is in red, the partially-conditional physical measure, $p_{t-\Delta t,T}^{year}$, in green and the risk-neutral measure $q_{t,T}^{*,year}$ is blue. Plots represent the average of the daily measures divided for years (2002 - 2003 - 2004) and times-to-maturity (short - medium - long). The strong departure from the normal distribution is clearly visible (and expected). Distributions are in fact irregular, skewed and leptokurtic. Focusing the attention on the tails two usual results hold: the risk-neutral distribution is always the higher, a common behavior in the options literature (i.e.: Christoffersen et al.[45]) and the conditional physical measure behaves better than the partially-conditional one. Representing the average values, the differences are smaller. These findings confirm that averaging out the extreme values by enlarging the time window, the distances from a misspecified model gets lower. For robustness the same experiment has been proposed with different $\alpha_{t,T}^*$ and different risk premium. Results are robust and available upon request.

2003).

¹²⁵A decile divides the area of interest in 10 parts. Analyzing the first decile is equal to set a percentile at 10%

9.1.4 Focus on a single time t and time-to-maturity τ

To increase the degree of precision for the daily operations I analyze the behavior of the EPKs more closely. To guarantee continuity with the previous analysis I use the same days, but differently than before I focus on the different times-to-maturity. Figure (A.21) shows, for $t = 90$, the single values of $\tau = 31$ and $\tau = 94$. The tails problem here is clearly visible. While the extra information provided by the options' implied moments allows the fully-conditional EPKs to explain a much larger fraction of the horizontal support, the partially-conditional EPKs explode quite soon. Once again, the further I go into the left tail, the lower the liquidity of the DOTM put options and the more noisy are also the fully-conditional EPKs. By the same token, due to the physiological lower or zero trading volume of short-term DOTM call options, there is no way to naturally reduce by the same magnitude the volatility of the right tail of the EPK. There are in fact too many zeros or almost zero values that bring the EPK to infinity. As a non-natural approach to obtain valid results at the extremes of the functional one could cut the support, model the extreme values of the tails with more extreme (but usually parametric) distributions or put strong modelling assumptions (such as the requirement of non-increasing values for the functional).

Two of the main factors needed for a good modelling of the distributions are: an appropriate definition of the conditional stochastic process to capture the randomness of the underlying and a proper definition of the innovations which impact strongly the shape of the tails. A lot has been said on the ability of the GJR GARCH model in fitting the S&P 500 (i.e.: Christoffersen P. and K. Jacobs (2004)[46]). Here I focus the attention on the latter feature. A correct modellization of the distributions governing the EPK may answer to important mispricing problems: i.e. the apparent over and under-pricing of put and call options. A recent strand of the literature focuses on the interaction between mispricing and tail risks. To better capture the tails behavior of the option surface Andersen et al. (2014)[5] propose a risk-neutral parametric model made of three factors. Focusing on the left tail and through a non-normal "tail factor", the model explains the apparent overpricing of the short term put options. Using different ingredients, namely: the γ parameter, the FHS for the innovations and the informative content provided by the options' implied moments, my model features a more flexible and complete "tail factor" and answers to mispricing of both put and call options. Starting from the put options: rational investors are unlikely to buy protection in presence of extremely high costs. It follows that too high values on the left portion of the EPK are hardly justified economically since would imply an over-expensive insurance. Even in presence

of very low values of the underlying, for which an investor would be prone to spend money to buy protection, an irrationally too high price for short (and long) term put options is translated, in economic terms, in an over-expensive insurance.

Starting from the estimated EPKs, a direct way to unveil the presence of possible mispricings is throughout the analysis of the expected return obtainable from investing in the primitive asset. The today expected rate of return over τ days is defined as¹²⁶:

$$v_{t,t+\tau} = \left(\frac{1}{M_{t,t+\tau}} \right) - 1 \quad (9.7)$$

It follows by structure that $v_{t,t+\tau}^\dagger$ and, $v_{t-\Delta t,t+\tau}$ estimated with both FHS and Gaussian innovations, mirror the good and bad features related to the single estimates of the EPKs. Depending on the types of options traded on the market mispricings appear strongly for both put and call.

Keep working on $t = 90$ figure (A.22) shows the expected rates of return over τ days for the entire set of times-to-maturity. The horizontal line, starting from 0, represents the area of no return. The vertical one the degree of moneyness. While the fully-conditional returns have a monotonically increasing shape, as required by the neoclassical literature, the partially-conditional ones exhibit a concavity in the central area and extreme values into the tails. The former violation is mainly due to the missing forward looking information provided by the options information. The latter is a byproduct of the lack of information and the underestimation of the innovations. Both produce irrational values for an utility maximizer investor i.e: an investor would expect to have a positive return from the investment (concavity) and not to pay an irrational cost for protection/investment (left/right tail). Results are strong and show a decreasing magnitudes as τ increases. As done for the EPKs, to corroborate the above points, I analyze the mispricing problem from a closer look. The figure for $t = 90$ and $\tau = 31$ shows clearly both divergences. Starting from values around $S_t = \$850$ which is not far from the S&P 500 spot price of $S_0 = \$1026$, the estimated partially-conditional EPK with Gaussian innovations is roughly equal to 3.5, thus implying a negative expected rate of return over 31 days of roughly -70%: a very high price to be paid to insure against a not very extreme outcome. For the same case but with FHS innovations, the expected rate of return is roughly -50%, which is smaller but still very high. For both the Gaussian and the FHS cases the overpricing gets even bigger up, to -100%, as I go further into the tail.

On the other hand, as a direct consequence of the better estimate of the inputs, both conditional $v_{t,t+\tau}^\dagger$ do not share these problems thus explaining the apparent overpricing of short-term put

¹²⁶See Appendix A.9.

options. The above violations are present and persistent for the whole sample in object.

Probably even more interesting, being less debated in literature, is the ability of the model to explain possible call options mispricing. While some mispricing are visible also for $t = 90$, the low liquidity of DOTM options for that day make them less evident. To better analyze the ability of the model to capture mispricing also on the right side, I pick a day with a significant amount of DOTM call options i.e.: $t = 63$ ¹²⁷. The high presence of long call options for this day is in line with a possible protection scheme of a risk-averse investor that shorted heavily the index given the long sequence of negative returns.

Figure (A.22) shows the strong departure of the partially-conditional expected returns from the monotonically increasing shape required by the neoclassical literature. Violations are strong and persistent for all times-to-maturity. Where the EPKs are U shaped, the relative expected returns show a strong underpricing of call options. Throughout the entire sample, for all days for which the liquidity of DOTM and OTM options is high enough, the partially-conditional EPK leads to a U-shaped functional, thus to apparent call option underpricing, while the fully-conditional EKP, capturing the information provided by the call options, have the right fraction which correctly decreases monotonically. Mathematically this is possible only if the denominator is not small or null.

Aside from the single examples proposed, I find that, using an information-miscalibrated model, the apparent mispricing of put and call options is strong and persistent. Results are in line with Babaogolu et al. (2014)[10]. While put options, having a payout on the downside are affected by the EPK puzzle only with respect to their apparent mispricing (which lead to too high values), the positive payouts of call options are helpful in explaining the U shaped EPK puzzle both in magnitude of the values and for the shape. While I confirm the need of a correct modellisation of the innovations for of a proper estimation of the high moments of the distributions, I find that, a non puzzling PK, is also fully determinant for a correct pricing model. Estimating the expected return from investing in a contingent claim my results show and confirm, from a different angle, the stochastically dominating portfolio theory of Dybvig (1988)[58] and Beare (2011)[20]. Letting the data speak, my model is as much sensitive as flexible to the degree and the direction of the liquidity of the market. As a consequence of its explanatory power, my “tail factor” not only confirms Andersen et al. (2014)[5] it also extends the analysis for the call options thus confirming Bakashi et al. (2009)[12]. “Ranking” my findings, mispricings are highest for the partially-conditional

¹²⁷For $t = 63$ (March 12, 2003) the total amount of traded options is 155 divided in: 47 put options of which 54% are DOTM and 105 call options of which 81%, are DOTM and the remaining OTM.

EPKs with Gaussian innovations, then for the FHS innovations while almost non existing for the conditional modified EPKs¹²⁸.

Flexibility at work: bullish and bearish estimates One of the major problem of using only backward data is the incapability of capturing possible future market changes. I analyze a scenario in which the recent history has been bearish for some time and then, for any reason, the trend changed suddenly. Empirically, using a non homogeneous estimation methods, the obtained EPKs are strongly U-shaped (days with majority of D-OTM call options) or, at best, non monotonically decreasing (days with majority of D-OTM put options).

The empirical confirmation that liquidity matters and that option surfaces embed powerful information to explain economic movements, are in line with the puzzling results of Jackwerth (2000) and the following papers. It is in fact not by coincidence that the EPK puzzle emerges from the S&P 500 right after the 1987 crash: when the liquidity of put options used as insurance against possible market crashes grew exponentially. What is missing in most of the post-crisis estimations is the extra information which, due the higher liquidity, is now strongly present at numerator but not captured at the denominator. As a consequence, the degree of homogeneity between the risk-neutral and the objective measures collapses thus producing non-homogeneous ratios, hence puzzling EPKs. The proposed model overcome this issue.

Given my sample, the lowest value touched by the S&P 500 is $S_{\text{Min}} = \$776.76$ in October 10, 2002 ($t = 41$). Starting from a few weeks before, with a value of \$949.36 (August 21, 2002), S_{Min} is the result of a stream of bearish price movements which is then followed by a sudden change that culminates with the index back to \$938.87 on November 27, 2002. S_{Min} , or $t = 41$ is not by chance one of the day of the sample with the highest amount of traded DOTM put and call options (68%). The high amount of sentiment in those months is manifested through a byproduct of high volatility (average Vix at 34.85%) and a sharp V movement. Can I extrapolate meaningful results from this richness of information? Even more, is the option market able to anticipate the change of trend in the underlying?

Figure (A.24) shows the $M_{t-\Delta t, t+\tau}$ and $M_{t, t+\tau}^\dagger$ estimated with a starting $\alpha_{t, t+38}^*$ of 2.5 that decreases in time and a R.P.=8%. While for the entire set of $M_{t, t+\tau}^\dagger$ all features required by a valid EPK holds, for $M_{t-\Delta t, t+\tau}$ two results are striking: the right portion of the graph is upward sloping, and the Gaussian estimations are extremely overpriced at both ends. Violations become even more apparent once I give a closer look at each time-to-maturity. To better support my results I plot a

¹²⁸The FHS, describing better the non-normality of the distributions, provides superior results.

vertical line which, starting from S_t , highlights the put and call area of the EPKs and a vertical line that identifies the risk-neutral area. Figure (A.25) shows how, for all τ s, the partially-conditional EPKs present an economically invalid strong U shape. Differently than the first stream of literature, which focuses on the apparent violations in the central area of the functional, this is what the recent papers call the new EPK puzzles.

Things change using a fully-conditional denominator¹²⁹. Aside from the usual physiological problem at the extremes of the functional, the conditional EPKs produce monotonically decreasing, thus economically valid estimates. It is not by chance that a missing convergence of the right part of the functional happens right before a strong up-movement. In such a case, the relevant forward looking information are present into the highly traded D/OTM call options. Capturing these information lead to monotonically decreasing EPKs.

In a nutshell: if under normal market conditions the partially-conditional EPK has, as a principal problem, a strong overpricing of put options, under strong market up/downturns - that implies a higher than average trading of both put and call options - things change dramatically. If in the former case the right side of the EPK correctly converges monotonically to low values, in the latter the classical methodology has problem of convergence that produces strongly U-shaped estimates. Highly volatile days may impact also on the estimation of the risk-neutral measure. As documented by Song and Xiu (2014)[149] the absence of a properly calibrated stochastic volatility may hit strongly the final values of the SPD¹³⁰. BEM shows the high correlation between the proposed GJR-GARCH-FHS estimates and the VIX thus confirming the ability of the model in explaining volatility and the higher moments of the distribution into the SPD.

Combining the above findings I claim that, due to the high randomness of the EPK, the unconditional and partially-conditional econometric models proposed in literature are simply not enough to capture all necessary information required to produce a valid EPK. Even if the last historic values would be informative of the actual market scenario, the magnitude of the last information is washed out by the long stream of data used (because needed) in estimation. This big mass, averaging out all values, would put same weights to past and present information make them all

¹²⁹Also in this case the model answers to the apparent mispricing of call options as for $t = 63$. Not by chance, also for $t = 63$, the S&P 500 value is at its lowest value right before an up movement, again a V-shape behavior. Opposite results are obtained from a bullish period followed by a downturn (i.e. for $t = 55$).

¹³⁰To confirm this fact the authors compare their SPD estimations with the ones of Ait Sahalia and Lo (1998)[3]. While the former condition their SPD for the stochastic volatility, the latter propose an unconditional estimation. For days with high/low VIX values the final values differ greatly while are similar in days with average volatility.

the same unimportant. If investors have expectations that are absent or divergent with the past, it becomes impossible to capture their heterogeneity of beliefs (findings are in line with Ziegler (2007)[159]). Although always valid, this fact emerges more strongly for days with high volatility. A naive example may confirm what just stated. The artificial Black & Scholes economy assumes constant first and second moments. Under this framework, the EPK holds with no needs of complex econometric adjustments. But an economy with these characteristics is unreal, since moments of the distributions are not constant in reality. This is true under normal market conditions and even truer under strong market up/downturns. Extremely puzzling results during days with high volatility and kurtosis and low skewness explains how, under normal market conditions, a partially-conditional physical measure can on average still produce valid EPKs.

Both the empirical and the naive examples presented show the importance of having a non trivial and highly flexible econometric model to account for all the publicly available information in the market. While always valid, this is especially true under very noisy market conditions. Results confirm and amplify the importance of the high moments of the distributions for a correct estimation of the EPK and are in line with Agarwal et al. (2008)[1], Bo-Young et al. (2013)[29] and Chabi-Yo (2012)[157]¹³¹.

The length of the rolling window: when more data doesn't imply more quality

Starting from the insights of the previous subsections, the short literature review about the physical measure estimation of the EPK and given the estimation techniques proposed in this paper, it's worth focusing on the trade-off between the degree of conditionality of the estimate and the quality of the information provided by the data. The trade-off arises naturally whenever one deals with intensive numerical methods and/or non-parametric models. Estimating the physical measure by means of historical data, the more I increase the time length of the rolling window the more I gain from the point of view of the degree of conditionality of the estimation but, at the same time, the more I lose from the point of view of the informative content of the recent most important information.

Statistically, given the structure of the estimation technique, it is only by increasing the rolling window that one can increase the degree of conditionality of the estimates. Even more: making use of a non-parametric kernel density estimation, it is well-known that the quality of the outputs is strongly related to the number of inputs used. At the same time though, the more one adds data,

¹³¹Whom find that market skewness and kurtosis are priced in the cross section of returns.

the less is the weight put on each single value and, consequently, the less is the informative content that can be extracted by “significant” data points.

On the contrary, since my model makes inference from a rich admixture of data, it becomes, by structure, less data intensive. As long as the estimated parameters are accurate, it’s natural to achieve a fully-conditional output using the same or even a shorter rolling window of data. Exploiting the informative content present in other assets it is thus possible to save time and increase the overall quality of the outputs.

To validate econometrically the choice of going back exactly 3500 observations, and to check for the effectiveness of the GJR GHARCH in accounting for the volatility clustering phenomenon typical in the financial datasets and its ability to remove the heteroscedasticity in daily returns, I perform two robustness checks on the model: namely the portmanteau test of Ljung and Box (1978)[114] and the Lagrange multiplier (LM) ARCH test (Engle 1982)[62].

For each Wednesday of the sample, I perform both tests on four different sets of physical residuals estimated with (5.1) - (5.5) and going back 2000/3500/5000 and 9818 observations. The choice of the length of these rolling windows is arbitrary and aim to check for the performance of a shorter period of observations - 2000 - and two longer ones.

Recall (5.1) the residuals of the mean equation: $\epsilon_t = r_t - \mu$. I square them ϵ_t^2 to check for conditional heteroscedasticity, aka ARCH effects. Two main tests are available: the Ljung-Box statistics Q_K to the squared residual series where the null hypothesis is that the first K lags of the ACF of the ϵ_t^2 series are zero.

For my specific case the Ljung-Box null-hypothesis tests if the first 21 correlations are equal to zero. The hypothesis is tested on the normal (figure (A.28)), squared (figure (A.29)) and absolute (figure (A.30)) returns of the S&P 500 and their residuals:

$$Q_K = n(n+2) \sum_{j=1}^{21} \frac{\rho_j^2}{n-j} \quad (9.8)$$

where ρ_j^2 is the sample autocorrelation at lag j , n is the sample size and I test for $K = 21$ lags. It is evident that having 3500 observations is enough to remove the autocorrelation in the residuals. Adding more observations is not necessary. Therefore, given my sample, a GJR GARCH with 3500 observations is a properly calibrated model to account for the volatility clustering phenomenon.

The LM test which is equivalent to the F statistics for testing $\alpha_i = 0$, for $i = \dots, K$) in the linear regression:

$$\epsilon_t^2 = \alpha_0 + \alpha_1 x \epsilon_{t-1}^2 + \dots + \alpha_K \epsilon_{t-K}^2 + a_t \quad t = K+1, \dots, T, \quad (9.9)$$

We perform a LM ARCH with ($K = 10$) legs where the null hypothesis $H_0 : \alpha_1 = \dots = \alpha_{10} = 0$ tests for the homoscedasticity of the sample.

Also in this case the effectiveness of the GJR GARCH model with 3500¹³² observations in removing the heteroscedasticity in returns is evident. In fact, while innovations are strongly homoscedastic, returns are not. Due to the high power of the model, to add more observations is not required since it cannot increase significantly the quality of the model. Therefore, as long as these tests produce valid results I can obtain conditional data with no need of going back for too long into the time series of the observations.

Yearly Estimates

To provide a graphical summary and to check for the validity of my results for the entire dataset, I enlarge the time window passing from a single day to an yearly estimation. Following Jackwerth (2000), I perform the estimation of $M_{t,T}^\dagger$ and $M_{t-\Delta_t,T}$ on a grid of 100 points of the gross return, S_T/S_t . The estimation is carried out for both put and call, for each day and time-to-maturity. For each year, the daily values are averaged across short, medium and long times-to-maturity. To make my results stronger, I perform the same experiment with different calibrations of $\alpha_{t,T}^*$, of the risk premium and changing the number of simulations. Figure (A.27) show both the conditional and the partially-conditional results duly divided in years (horizontally) and short, medium and long time-to-maturity (vertically). For both cases I cut the support as soon as the functional explodes, thus implying a negligible or null physical density. From the figure, two main findings arise, one for the horizontal support and one for the vertical one:

- Corroborating the daily results, the fully-conditional EPKs spread over a wider - if not the entire - horizontal support (domain of gross returns). Differently than the vertical one, the larger the support the better. Due to the overall larger amount of traded DOTM and OTM put options, the main enlargement of the support affects the area of negative returns: the one more sensitive and important for risk management operations. Confirming the strong informative power of the proposed model, major improvements are achieved for the short and medium times-to-maturity. While with the fully-conditional methodology I can quite always go up to 0 (exceptions are the 2003 and 2004 short times-to-maturity which starts from 0.3 and 0.4 respectively) with the partially-conditional methodology is never possible to go over 0.6 for the short time-to-maturity and over 0.4 for the medium time-to-maturity. Although with

¹³²Both tests have been proposed with different sets of observations, ranging from 1000 to over 9000, and for different legs. Results are omitted but available upon request.

an overall higher noise for the partially-conditional EPKs, the right sides of the functionals are similar for both methods and converge analogously. Unfortunately, this well behavior on the right side of the EPKs lacks of robustness at high frequency. As documented in section (9.1.4) there is no convergence on the call side of the EPKs for some particular days thus confirming how non puzzling low frequency averaged EPK may hide otherwise existing EPK puzzles on a day-by-day basis.

- Confirming the mispricing problems presented on a daily basis, the conditional EPKs live in a smaller vertical support (domain of the EPK) than the partially-conditional ones. Results are graphically evident and economically valid, above all for the upper-bound. While the conditional values never go over a value of 3.1 for $M_{t,T}^{\dagger Gauss.}$ and 2.5 for $M_{t,T}^{\dagger FHS}$, the partially-conditional $M_{t-\Delta_t,T}^{Gauss.}$ almost arrives to touch the value of 10 for the 2002 long time-to-maturity¹³³.

A final remark: for all estimates results may look slightly different if the two main random variables - the precision parameter and the risk premium - would be set at different values. While the literature about the risk premium is enormous (but still unanswered), to my knowledge nobody has never analyzed the problem of a proper calibration of the precision parameter to estimate a density. As anticipated, as a possible extension, I can still follow the Bayesian paradigm and use statistically more advanced and flexible priors and distributions or link the precision parameter to other markets. I keep the analysis of both issues for future researches. In all cases, although different calibration proposal can be taken into consideration, the main concept behind the proposed model is robust and untouched: I always obtain better results proposing a more informative fully time-varying EPK with respect to a partially-conditional EPK.

9.2 Robustness checks

After a quick presentation of the GJR GARCH parameters estimated from stock and options market data, in the next section I analyze how good is the GJR GARCH model in fitting the returns of my databases.

To test this feature I implement the same methodology presented in chapter 5 but fitted on a semi-synthetic database that, by construction, does not present any liquidity issue. Starting from

¹³³The domain of the EPK is even bigger than $[0,10]$. To propose readable plots, the value of 10 is the maximum I report. This is achieved considering only values < 10 to be valid EPKs. In reality, the real domain should be $[0, \infty)$. For the presented figure, the maximum value visible is 4.5.

market values, the new dataset is complete. Completeness is related to the degree of liquidity in the market. The new dataset in fact has a price for each value that goes from the highest to the lowest strike value, for each day of the sample. By removing the empirical issues that could be caused by market data, the quality of the obtained parameters tells us how good is the model in reconstructing the randomness of the market.

As empirical exercise I briefly comment the parameters estimated using the asymmetric GJR GARCH model presented in 5.1 fitted on the empirical dataset presented in 8.2.

Table A.5 shows summary statistics of the physical parameters while tables A.6 and A.7 show summary statistics for the pricing parameters estimated with the two different minimization procedures (Simplex and quasi-Newton) calibrated each Wednesday from January 2, 2002 to December 29, 2004 thus producing 157 observations. I only analyze the former table since results are comparable.

The tables present the first two moments of the four estimated parameters and their persistence, duly divided for time ¹³⁴ and probability measures. The parameters estimated under the pricing measure are further divided in FHS and Gaussian errors. The physical parameters are estimated using 3500 historical S&P 500 log-returns while the risk-neutral ones using the cross-section of OTM SPX option prices. For each Wednesday I have a set of pricing parameters and 3500 scaled innovations which are then used for the FHS parameters estimate.

All parameters but α are fairly stable over time. This finding, summed with the larger values of γ with respect to α confirms the presence of a strong leverage effect into the time series of interest ¹³⁵.

The physical parameters of β are always larger than the risk-neutral ones thus indicating a less pronounced presence of the volatility clustering effect in the risk-neutral world.

The most important empirical result for option pricing comes from the γ which is always larger under \mathbb{Q} than under \mathbb{P} . Even more, the risk-neutral Gaussian parameters are the highest, followed by the FHS and then by the physical one.

It is well known that a biased leverage parameter sensitively affects the option pricing. A too low (high) γ parameter is economically translated into an investor that expects to receive a higher (lower) compensation when the S&P 500 declines under the risk-neutral measure. If one would wrongly use the physical parameters for an exercise of option pricing would incur in serious pricing errors due to a strong volatility underestimation. The same is true for the use of Gaussian risk-

¹³⁴Each parameter has been estimated on a daily basis and then averaged yearly.

¹³⁵Negative asset returns raises volatility more than positive ones.

neutral parameters, where the volatility is likely to be overestimated.

9.2.1 GJR GARCH robustness check

The huge amount of different GARCH specifications gives us the possibility to properly use different methodologies to model the conditional volatility of the market.

Following much of the literature¹³⁶ I use the GJR GARCH model due to its ability in fitting the S&P 500 data and its flexibility in capturing the leverage effect. To corroborate these known good qualities, I perform a new test on the GJR GARCH model which confirms the above mentioned features.

Using a semi-synthetic database I first fit the GJR GARCH model as presented in chapter 4 using standard initial values for the unconstrained non-linear minimization.

Then, using the calibrated GJR GARCH parameters and the fitted option prices, I calibrate again the same GJR GARCH model¹³⁷.

Figure A.32 shows the “goodness of fit” of the first and second estimated parameters. I only perform the robustness check for the FHS innovations.

The blue line represents the first set of parameters estimation while the red one the second set of estimates. The analysis is regularly performed on a daily basis, thus producing 157 observations for each parameter.

Figure A.33 perfectly shows the numerical nature of the experiment. Due to possible errors that can normally arise performing any numerical simulation and, in combination to the high collinearity of the models, the figure shows that there are some physiological difference on a single daily basis (above all for γ , the more unstable of the parameters). Nevertheless table A.10 shows the pointwise numerical difference between the two estimations. The first (second) panel collects summary statistics of the first (second) estimations, while the third one is the pointwise difference between the two. As a main result, the difference is always negligible: I do not “lose” any important

¹³⁶Among the huge amount of papers that analyze the performances of the different GARCH specifications Rosenberg and Engle (2002)[64] find that the GJR GARCH model is the best one in fitting the S&P 500 returns, while Engle and Ng (1993)[63] and BEM document the overall good performance in modelling the news impact curve (NIC).

¹³⁷Due to the high complexity in estimating correctly the drift of the daily returns, I again follow Merton and I fix the values of μ equal to 4/6 or 8%. The three different specifications produce similar values thus enhancing the robustness of the model to possible changing values.

information during the numerical exercise of estimation, thus confirming the high quality of the GJR GARCH model in reconstructing the conditional volatility.

APPENDIX A

Appendix

A.1 Absolute Continuity and the Radon-Nikodym Derivative

If ζ and φ are two measures on a σ -algebra \mathcal{B} of subsets of X , then φ is absolutely continuous with respect to ζ if $\varphi(A) = 0$ for any $A \in \mathcal{B}$ such that $\zeta(A) = 0$. In symbol: $\varphi \ll \zeta$.

A stronger generalization of the above is: if the measure φ is a.s. finite, i.e. $\varphi(X) < \infty$, the property $\varphi \ll \zeta$ is equivalent to the following: for any $\alpha > 0$ there is a $\gamma > 0$ such that $\varphi(A) < \alpha$ for every A with $\zeta(A) < \gamma$ (this follows from the Radon-Nykodin theorem, see below, and the absolute continuity of the integral).

The generic definition can be extended to signed measures and to vector-valued measures φ . The vector-valued extension can be generalized to vector-valued ζ 's: in such a case the absolute continuity of φ with respect to ζ amounts to the requirement that $\varphi(A) = 0$ for any $A \in \mathcal{B}$ such that $|\zeta|(A) = 0$, where $|\zeta|$ is the total variation of ζ .

Assuming that ζ is σ -finite, the Radon-Nikodym theorem characterizes the absolute continuity of φ with respect to ζ with the existence of a nonnegative Borel function $f \in L^1(\varphi)$ such that $\varphi = f\zeta$, i.e.:

$$\varphi(A) = \int_A f d\zeta \quad \text{for every } A \in \mathcal{B}$$

A.2 Approximating to the mean and the variance of a ratio using Taylor

To compute the approximated mean and the variance of a ratio, I follow Arnold (1998)[151] and John Wiley & Sons (1980)[60].

My goal is to compute $\mathbb{E}(C)$ and $\mathbb{V}(C)$ where $C = c(A, B) = \frac{A}{B}$ and A and B are two random variables with B having either no mass at 0 or support $[0, \infty)$. To approximate the values I perform a Taylor expansion of $c(\cdot)$.

For a generic univariate function $f(x)$ that has n continuous derivatives on a neighborhood of θ , the Taylor expansion of $f(x)$ is:

$$f(x) \approx \sum_{n=0}^{\infty} \frac{f^n(\theta_x)}{n!} (x - \theta_x)^n \quad (\text{A.1})$$

By the same token, the Taylor expansion of a generic bivariate function $f(x, y)$ is defined as:

$$f(x, y) \approx \sum_{n_1=0}^{\infty} \sum_{n_2=0}^{\infty} \frac{(x - \theta_x)^{n_1} (y - \theta_y)^{n_2}}{n_1! n_2!} \frac{\partial^{n_1+n_2} f(x, y)}{\partial x^{n_1} \partial y^{n_2}} \quad (\text{A.2})$$

The double infinite series (A.2) that approximates $f(x, y)$ can also be written as a first order expansion around θ :

$$f(x, y) = f(\theta) + f'_x(\theta)(x - \theta_x) + f'_y(\theta)(y - \theta_y) + \text{remainder} \quad (\text{A.3})$$

where f'_x and f'_y are the first derivative with respect to x and y and the Lagrange remainder term is obtained applying the mean variance theorem to its continuous counterpart:

$$R_n(x, \psi) = f(x) - \sum_{k=0}^n f^k(\theta) \frac{(x - \theta_x)^k}{k!} \quad (\text{A.4})$$

$$= f^{n+1}(\psi) \frac{(x - \theta_x)^{n+1}}{(n+1)!} \quad (\text{A.5})$$

assuming $f^{n+1}(\psi)$ to be continuous on $[\theta; x]$.

If $\lim_{n \rightarrow \infty} R_n(x, \psi) = 0$, then I turn back to (A.2).

If $\theta = (\mathbb{E}(x), \mathbb{E}(y))$, the approximated expectation of the function is $\mathbb{E}[f(x, y)]$, so that its

expansion is:

$$\mathbb{E}[f(X, Y)] = f(\theta) + f'_x(\theta)(0) + f'_y(\theta)(0) + O(n^{-1}) \quad (\text{A.6})$$

$$\approx f(\mathbb{E}(X), \mathbb{E}(Y)) \quad (\text{A.7})$$

The bivariate second order Taylor expansion is:

$$f(x, y) = f(\theta) + f'_x(\theta)(x - \theta_x) + f'_y(\theta)(y - \theta_y) \quad (\text{A.8})$$

$$+ \frac{1}{2} \left\{ f''_{xx}(\theta)(x - \theta_x)^2 + 2f''_{xy}(\theta)(x - \theta_x)(y - \theta_y) + f''_{yy}(\theta)(y - \theta_y)^2 \right\} + \text{remainder} \quad (\text{A.9})$$

where f''_{xx} , f''_{yy} , f''_{xy} are the second derivative with respect to x and y and the reminder is defined as in (A.5). The expansion can be better approximated as:

$$\mathbb{E}[f(X, Y)] = f(\theta) + \frac{1}{2} \left\{ f''_{xx}(\theta)\text{Var}(X) + 2f''_{xy}(\theta)\text{Cov}(X, Y) + f''_{yy}(\theta)\text{Var}(Y) \right\} + O(n^{-1}) \quad (\text{A.10})$$

Applying the expansion to the random variables (A, B) :

$$c''_{AA} = 0, \quad c''_{AB} = -B^{-2}, \quad c''_{BB} = \frac{2A}{B^3}$$

So that, the approximated expectation of the function c is:

$$\mathbb{E}\left(\frac{A}{B}\right) = \mathbb{E}(c(A, B)) \quad (\text{A.11})$$

$$\approx \frac{\mathbb{E}(A)}{\mathbb{E}(B)} - \frac{\text{Cov}(A, B)}{\mathbb{E}(B^2)} + \frac{\text{Var}(B)\mathbb{E}(A)}{\mathbb{E}(B^3)} \quad (\text{A.12})$$

By the same token, the variance can be approximated by a first order Taylor expansion:

$$\mathbb{V}(f(X, Y)) = \mathbb{E} \left\{ [f(x, y) - \mathbb{E}(f(X, Y))]^2 \right\} \quad (\text{A.13})$$

$$\approx \mathbb{E} \left\{ [f(x, y) - f(\mathbb{E}(X)\mathbb{E}(Y))]^2 \right\} \quad (\text{A.14})$$

$$\begin{aligned} &= \mathbb{E} \left\{ [f(\theta) + f'_x(\theta)(x - \theta_x) + f'_y(\theta)(y - \theta_y) + O(n^{-1})] - \right. \\ &\quad \left. [f(\theta) + f'_x(\theta)(0) + f'_y(\theta)(0) + O(n^{-1})] \right\} \\ &= \mathbb{E} \left\{ [f'_x(\theta)(X - \theta_x) + f'_y(\theta)(Y - \theta_y)]^2 \right\} \\ &\approx f'_x(\theta)^2 \text{Var}(X) + 2f'_x(\theta)f'_y(\theta)\text{Cov}(X, Y) + f'_y(\theta)^2 \text{Var}(Y) \end{aligned} \quad (\text{A.15})$$

Following the same procedure:

$$c'_A = B^{-1}, \quad c'_B = -\frac{A}{B^2}$$

Such that:

$$\mathbb{V}\left(\frac{A}{B}\right) \approx \frac{1}{\mathbb{E}(B^2)} \text{Var}(A) - 2 \frac{\mathbb{E}(A)}{\mathbb{E}(B^3)} \text{Cov}(A, B) + \frac{\mathbb{E}(A^2)}{\mathbb{E}(B^4)} \text{Var}(B) \quad (\text{A.16})$$

$$= \frac{\mathbb{E}(A^2)}{\mathbb{E}(B^2)} \left[\frac{\text{Var}(A)}{\mathbb{E}(B^2)} - 2 \frac{\text{Cov}(A, B)}{\mathbb{E}(A)\mathbb{E}(B)} + \frac{\text{Var}(B)}{\mathbb{E}(B^2)} \right] \quad (\text{A.17})$$

A.3 The Riesz Representation Theorem

In this Appendix I prove existence and uniqueness of the Riesz Representation Theorem

Existence:

If $F = 0$ I can just take $k_f = 0$ and thereby have $F(x) = 0 = \langle x, 0 \rangle \forall x \in \mathcal{H}$, where \mathcal{H} is a Hilbert space.

Assume that F is not a zero functional. Let \mathcal{A} define the null space of a nonzero functional as:

$$\mathcal{A} = \{x \in \mathcal{H} : F(x) = 0\} \quad (\text{A.18})$$

which is the definition of the $\text{Ker}(F)$. Since $F \neq 0$, i.e. $\text{Ker}(F) \neq \mathcal{H}$, and since F is continuous it follows that $\text{Ker}(F)$ is a closed subspace of \mathcal{H} .

From the orthogonal decomposition theorem I can decompose the Hilbert space as:

$$\mathcal{H} = \mathcal{A} + \mathcal{A}^\perp \quad (\text{A.19})$$

or, equivalently:

$$\mathcal{H} = \text{Ker}(F) \oplus \text{Ker}(F)^\perp \quad (\text{A.20})$$

where $\mathcal{A}^\perp \neq \{0\}$ is the orthogonal complement of the null space. The same holds for $\text{Ker}(F)^\perp$

Let $b > 0 \in \mathcal{A}^\perp$ be a non-zero vector s.t. multiplying b by a scalar $F(b) = 1$.

It follows that any vector $x \in \mathcal{H}$ can be decomposed in its dual form as:

$$x = (x - F(x)b) + F(x)b \quad (\text{A.21})$$

Since $b \in \mathcal{A}^\perp$ it follows that $F(x)b \in \mathcal{A}^\perp$ and $(x - F(x)b) \in \mathcal{A}$.

This implies that:

$$b \cdot x = b \cdot (F(x)b) \quad (\text{A.22})$$

Setting the vector:

$$k_f = \frac{b}{(b \cdot b)} \quad (\text{A.23})$$

and plugging it into (A.22):

$$k_f \cdot x = \frac{F(x)(b \cdot b)}{(b \cdot b)} = F(x) \quad (\text{A.24})$$

Uniqueness:

Suppose that both $k_f, \widetilde{k}_f \in \mathcal{H}$ satisfy (3.6) such that for every $x \in \mathcal{H}$,

$$F(x) = \langle x, k_f \rangle = \langle x, \widetilde{k}_f \rangle \quad (\text{A.25})$$

or, equivalently that:

$$(k_f - \widetilde{k}_f) \cdot x = 0 \quad \forall x \in \mathcal{H} \quad (\text{A.26})$$

It follows that:

$$\langle x, k_f - \widetilde{k}_f \rangle = 0 \quad \text{for every } x \in \mathcal{H} \quad (\text{A.27})$$

such that, taking $x = k_f - \widetilde{k}_f$ I obtain $\|k_f - \widetilde{k}_f\|^2 = 0 \Rightarrow k_f = \widetilde{k}_f$.

Or, equivalently, that:

$$(k_f - \widetilde{k}_f) = 0 \quad \text{for every } x \in \mathcal{H} \quad (\text{A.28})$$

□

A.4 Properties of Lévy processes

Given a filtered probability space which satisfies the usual hypothesis $(\Omega, \mathcal{F}_t, \mathbb{F}, P)$ and a fixed time period $0 \leq t \leq T$ with $T < \infty$, a generic stochastic process $(X_t)_{t \geq 0}$ whose values are in \mathbb{R}^d and $X_0 = 0$ is a Lévy process:

$$X_t = X(t, \omega) : [0, \infty) \times \Omega \rightarrow \mathbb{R} \quad (\text{A.29})$$

if satisfies the following properties:

- Independent increments: for an increasing sequence of times $t_0, t_1, t_2, \dots, t_n$ with $t_0 < t_1 < t_2 < \dots < t_n$, the random variables are time independent: $X_{t_0}, X_{t_1} - X_{t_0}, \dots, X_{t_n} - X_{t_{n-1}}$.
- Stationary increments: the law of $(X_{t+h} - X_t)$ is independent to t
- $(X_t)_{t \geq 0}$ is a right continuous with left limits - cadlag - process s.t. the paths $t \rightarrow X_t$ are non-anticipating
- Stochastic continuity: $\forall \delta > 0, \lim_{h \rightarrow 0} \mathbb{P}(|X_{t+h} - X_t| \geq \delta) = 0$

where the last property assures that jumps occur at random times. If the probability of having a jump is not a.s. equal to 0, I would have the so called “calendar effect”.

Given a Lévy process and a measurable subset A , its measure ν on \mathbb{R}^d is called the Lévy measure $X : \nu(A)$ and is defined as:

$$\nu(A) = \mathbb{E}[\#\{t \in [0, 1] : \Delta X_t \neq 0, \Delta X_t \in A\}] \quad A \in \mathcal{B}(\mathbb{R}) \quad (\text{A.30})$$

The Lévy measure represents the expected number, per unit of time, of jumps whose size belongs to the measurable set $A \subset \mathbb{R}^d$.

Now, given a Lévy process $(X_t)_{t \geq 0}$ on \mathbb{R}^d and a Lévy measure ν , by the Lévy-Ito decomposition theorem any such a process can be decomposed as:

$$X_t = \alpha t + \beta B_t + X_t^l + \lim_{\epsilon \downarrow 0} \hat{X}_t^{\{\epsilon\}} \quad (\text{A.31})$$

where $\alpha, \beta \in \mathbb{R}$ are constants, B_t is an \mathcal{F}_t -adapted Brownian Motion. The first two terms on the right hand side compose the continuous Gaussian Lévy process.

The discontinuous part is composed by:

$$X_t^l = \int_{|x| \geq 1, s \in [0, t]} x J_X(ds \times dx) \quad (\text{A.32})$$

and (A.33)

$$\hat{X}_t^\epsilon = \int_{\epsilon \leq |x| < 1, s \in [0, t]} x \{J_X(ds \times dx) - \nu(ds)dx\} \quad (\text{A.34})$$

$$\equiv \int_{\epsilon \leq |x| < 1, s \in [0, t]} x \hat{J}_X(ds \times dx) \quad (\text{A.35})$$

where the last term converges a.s. and uniformly in $t \in T$. The former process is a compound Poisson process while the latter indicates the compensated version of the process¹³⁸. All terms in (A.31) are independent.

The process X_t is identified by means of the characteristic triplet (A, ν, γ) where A is the covariance matrix of the Brownian motion, ν is the Lévy measure and γ is the drift of the continuous Gaussian Lévy process.

¹³⁸The jump integral is X_t^ϵ replaced by its compensated version to avoid singularities and assure convergence.

A.5 Constraints for stationarity

In this appendix I present the parameter transformations required to correctly implement the unconstrained non-linear optimization presented in chapter 4.

The GJR GARCH model, to properly estimate positive and stationary values, needs of some technical assumptions.

To produce strictly positive estimates for the variances σ_t , it is necessary that:

- $\omega > 0$
- $\alpha \geq 0, \beta \geq 0$ and $\gamma \geq 0$

Whil the first assumption is strictly required, the second ones, if violated, only brings to different model generalizations.

To assure stationarity:

$$\alpha + \beta + \frac{\gamma}{2} < 1 \quad (\text{A.36})$$

where α and β need again to be positive.

Given the set of unrestricted parameters $\bar{\theta} = \{\bar{\omega}, \bar{\alpha}, \bar{\beta}, \bar{\gamma}\}$, a parameters restriction is achieved by:

$$\omega = e^{\bar{\omega}} \quad (\text{A.37})$$

$$\alpha = \frac{e^{\bar{\alpha}}}{1 + e^{\bar{\alpha}} + e^{\bar{\beta}} + \frac{1}{2}e^{\bar{\gamma}}} \quad (\text{A.38})$$

$$\beta = \frac{e^{\bar{\beta}}}{1 + e^{\bar{\alpha}} + e^{\bar{\beta}} + \frac{1}{2}e^{\bar{\gamma}}} \quad (\text{A.39})$$

$$\gamma = \frac{e^{\bar{\gamma}}}{1 + e^{\bar{\alpha}} + e^{\bar{\beta}} + \frac{1}{2}e^{\bar{\gamma}}} \quad (\text{A.40})$$

By substituting the relations in equations (A.37) to (A.40), one can easily verify whether the conditions of the above assumptions are properly met.

A.6 Consistency of DP marginals

Thanks to the Kolomogorov consistency theory, it is enough to provide a consistent definition on the marginals of the stochastic process to have a valid stochastic process. To prove consistency I make use of a Lemma (without proving it): **Lemma:** If $E_1, \dots, E_r \sim \Gamma(k_i, \theta) \equiv \text{Gamma}(k_i, \theta)$ are independent, with k_i representing the shape parameter and θ the scale parameter, then:

$$\left(\frac{E_1}{\sum_r E_r}, \dots, \frac{E_r}{\sum_r E_r} \right) \sim \text{Dir}(K_1, \dots, K_r) \quad (\text{A.41})$$

That leads to a Proposition: **Proposition:** Given (Ω, \mathcal{F}) , let (X_1, X_2, X_3) and (Y_1, Y_2) be two measurable partitions on the set Ω : If B is a DP s.t. $B \sim \text{DP}(\alpha, C)$, then:

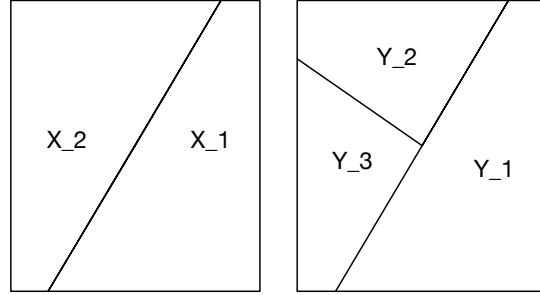


Figure A.1: Measurable partitions of Ω

$$(B(X_1), B(X_2))(Z_1, Z_2) \quad \text{and} \quad (B(Y_1), B(Y_2), B(Y_3))(H_1, H_2, H_3)$$

where, for both partitions, the second set of parenthesis represent the random variables relative to the partitions i.e.: $(Z_1 = B(Z), \text{etc.})$ I.e.: $X_1 = Y_1$ and (Y_2, Y_3) is a partition of (X_2) .

If the above is true, then there is an equality in distribution: $(Z_1, Z_2) \stackrel{\text{distr.}}{=} (H_1, H_2, H_3)$. **Proof:**

Let $E_i \sim \text{Gamma}(\alpha C(Y_i), \theta)$ be independent and θ be an arbitrary scale parameter $\theta > 0$ then,

applying the above lemma:

$$(H_1, H_2 + H_3) \stackrel{\text{distr.}}{=} \left(\frac{E_1}{\sum_r E_r}, \frac{E_2 + E_3}{\sum_r E_r} \right) \quad (\text{A.42})$$

$$\stackrel{\text{distr.}}{=} \left(\frac{E_1}{\sum_r E_r}, \frac{E^\dagger}{\sum_r E_r} \right) \quad (\text{A.43})$$

$$\stackrel{\text{distr.}}{=} (Z_1, Z_2) \quad (\text{A.44})$$

where, from standard properties of Gamma random variables:

$$E^\dagger \sim \text{Gamma}(C(Y_2) + C(Y_3), \theta) = \text{Gamma}(C(X_2), \theta)$$

Following the same argument, I can apply the above for any finite number of blocks.

The invariance under permutations applies obviously for this case and allows us to have a valid stochastic process.

A.7 Moments of DP

For the derivation of the moments of the DP I exploit the definition of Kolomogorov consistency applied to the partition (D, D^c) of Θ . Given a random distribution $B(D)$, $B \sim \text{DP}(\alpha^{DP}, C)^{139}$ and for $C \subset \Theta$:

$$(B(D), B(D^c)) \sim \text{Dir}(\alpha^{DP}C(D), \alpha^{DP}C(D^c)), \quad (\text{A.45})$$

By considering the partition (D, D^c) it is seen that $B(D)$ has a Beta distribution:

$$B(D) \sim \text{Beta}(\alpha, \beta) \quad (\text{A.46})$$

where α represents $(\alpha^{DP}C(D))$ and β represents $(\alpha^{DP}C(D^c))$.

The first and the second moment of the Beta distributions are:

$$\mathbb{E}[B(D)] = \frac{\alpha}{\alpha + \beta} \quad (\text{A.47})$$

and:

$$\mathbb{V}[B(D)] = \frac{\alpha\beta}{(\alpha + \beta)^2(1 + \alpha + \beta)} \quad (\text{A.48})$$

Applying the (A.47) and (A.48) I have that:

$$\mathbb{E}[B(D)] = \frac{\alpha}{\alpha + \beta} \quad (\text{A.49})$$

$$= \frac{\alpha^{DP}C(D)}{\alpha^{DP}C(A) + \alpha^{DP}C(D)^c} \quad (\text{A.50})$$

$$= C(D) \quad (\text{A.51})$$

which is nothing but the base, and:

$$\mathbb{V}[B(D)] = \frac{\alpha\beta}{(\alpha + \beta)^2(1 + \alpha + \beta)} \quad (\text{A.52})$$

$$= \frac{(\alpha^{DP})^2 C(D)(1 - C(D))}{(\alpha^{DP})^2 (\alpha^{DP} + 1)} \quad (\text{A.53})$$

$$= \frac{C(D)(1 - C(D))}{(\alpha^{DP} + 1)} \quad (\text{A.54})$$

which shows why α is the precision parameter.

¹³⁹For the derivation I make use of the Beta distribution with parameters α and β , to avoid confusion I call the parameter used for the DP as α^{DP} . This modification is valid only for the appendix

A.8 DP posterior moments

For a complete derivation of the posterior moments of a DP process I refer to Sethraman (1994, from pag. 646)[147]. Here I propose a sketch of the derivation of the updated moments after collecting one observation, the generalisation for n observations is immediate.

After collecting the first observation, the updated concentration parameter α^U ¹⁴⁰ is the result of taking the sum of the parameters of any finite dimensional Dirichlet distribution:

$$\alpha^U = \alpha + 1 \tag{A.55}$$

The one-time updated base measure, C^U comes from the conjugacy of the Dirichlet-multinomial distribution (6.5):

$$((B(D_1))|\Omega_1 \sim Dir(\alpha C(D_1) + n_1) \tag{A.56}$$

which is normalized through its updated centering parameter (A.55):

$$C^U = \frac{\alpha}{\alpha + 1}C + \frac{1}{\alpha + 1}n_1 \tag{A.57}$$

Generalizing (A.57) for n updates yields (6.6)

¹⁴⁰For a better reading, I add to the updated parameters the subscript U .

A.9 Derivation of the expected rate of return of investing in contingent claims

The time-state preference model of Arrow (1964)[6] and Debreu (1959)[7], introduce the Arrow-Debreu security: a very basic financial instruments also known as pure or primitive security. The product pays one unit of numeraire i.e.: a currency or a commodity, on one specific state of nature and zero elsewhere. Passing from discrete to continuous states Arrow-Debreu securities are defined by the state price density (SPD). Under the continuous framework the security pays one unit of numeraire x if the state falls between x and $x + dx$ and zero elsewhere.

To go long one primitive assets at time t costs: $q e^{-r\tau}$ where $\tau = T - t$ represents the time-to-maturity. From the above, the payoff at maturity is:

$$\text{Payoff} = \begin{cases} 1, & \text{with probability } p \\ 0, & \text{with probability } (1 - p). \end{cases}$$

It follows that the time t expected rate of return of return over τ days is:

$$\mathbb{E}(v_{t,\tau}) = \frac{p - q e^{-r\tau}}{q e^{-r\tau}} \tag{A.58}$$

$$= \left(\frac{1}{M_{t,\tau}} \right) - 1 \tag{A.59}$$

$$= \left(\frac{1}{\text{PK}} \right) - 1 \tag{A.60}$$

Bibliography

- [1] V. Agarwal, G. Bakshi, and J. Huij. Do higher-moment equity risks explain hedge fund returns? *Robert H. Smith School Research Paper No. RHS 06-153*. Available at SSRN: <http://ssrn.com/abstract=1108635>, Aug 2009.
- [2] Y. Äit-Sahalia and J. Duarte. Non-parametric option pricing under shape restrictions. *Journal of Econometrics*, 116:9–47, 2003.
- [3] Y. Äit-Sahalia and A. W. Lo. Non-parametric estimation of state-price densities implicit in financial asset prices. *The Journal of Finance*, 53(2):499–457, Apr 1998.
- [4] I. Anagnou, B. Mascia, H. Stewart, and T. Robert. The relation between implied and realised probability density functions. *Working paper, University of Technology, Vienna*, 2001.
- [5] T.G. Andersen, N. Fusari, and V. Todorov. The risk premia embedded in index options. *CREATES Research Papers*, 2014.
- [6] J.K Arrow. The role of securities in the optimal allocation of risk-bearing. *The Review of Economic Studies*, 31(2):91–96, Apr 1964.
- [7] J.K Arrow and G. Debreu. Existence of an equilibrium for a competitive economy. *Econometrica*, 22(3):265–290, Jul 1954.
- [8] S. Asmussen and P.W. Glynn. *Stochastic Simulation. Algorithms and Analysis*. Springer Science and Business Media LLC, New York, 2007.
- [9] Francesco Audrino, Robert Huitema, and Markus Ludwig. An empirical analysis of the ross recovery. Available at SSRN: <http://ssrn.com/abstract=2433170>, February 2015.
- [10] K. Babaoglu, P. Christoffersen, S. Heston, and K. Jacobs. Option valuation with volatility components, fat tails, and nonlinear pricing kernels. *Working Paper*, 2014.

- [11] B. Bahara. Implied risk-neutral probability density functions from options prices: Theory and application. *Bank of England ISSN 1368-5562*, 1997.
- [12] G. Bakshi, D. Madan, and G. Panayotov. Returns of claims on the upside and the viability of u-shaped pricing kernel. *Journal of Financial Economics*, 97(1):130–154, July 2010.
- [13] M. Barkhagen, J. Blomvall, and E. Platen. Recovering the real-world density and liquidity premia from option data. *Quantitative Finance*, February 2016.
- [14] O. Barndorff-Nielsen. A process of normal inverse gaussian type. *Finance and Stochastics*, 1:41–68, 1998.
- [15] G. Barone-Adesi, R. Engle, and L. Mancini. A garch option pricing model in incomplete markets. *Review of Financial Studies*, 21(1223-1258), 2008.
- [16] G. Barone-Adesi, K. Giannopoulos, and Les Vosper. Backtesting derivative portfolios with filtered historical simulation (fhs). *Journal of future market*, 8(1):31–58, March 1999.
- [17] G. Barone-Adesi, L. Mancini, and H. Shefrin. *Sentiment, Asset Prices and Systemic Risk in Handbook of Systemic Risk*. Cambridge University Press, Cambridge, 2011.
- [18] G. Barone-Adesi, L. Mancini, and H. Shefrin. A tale of two investors: Estimating optimism and overconfidence. *Working Paper, Swiss Finance Institute Research Paper No. 12-21*, page Available at SSRN: <http://dx.doi.org/10.2139/ssrn.2060983>, 2013.
- [19] D.S. Bates. Testing option pricing models. *NBER Working Paper N.5129*, 1995.
- [20] B.K. Beare. Measure preserving derivatives and the pricing kernel puzzle. *Journal of Mathematical Economics*, 47:689–697, 2011.
- [21] B.K. Beare and D.W.L. Schmidt. An empirical test of pricing kernel monotonicity. *Journal of Applied Econometrics*, pages 1099–1255, 2014.
- [22] G. Bekaert and J. Liu. Conditioning information and variance bounds on pricing kernels. *The Review of Financial Studies*, 17(2):339–378, 2004.
- [23] J. Berger. *Statistical Decision Theory and Bayesian Analysis*. New York: SpringerVerlag, 1993.
- [24] F. Biagini and B. Øksendal. A general stochastic calculus approach to insider trading. *Applied Mathematics and Optimization*, 52(2):167–181, 2005.

- [25] M. Birke and K.F. Pilz. Non-parametric option pricing with no-arbitrage constraints. *Journal of Financial Econometrics*, 7:53–76, 2009.
- [26] F. Black. Studies of stock price volatility changes. *Proceedings of the 1976 Meetings of the American Statistical Association*, pages 171–181, 1976.
- [27] F. Black and M. Scholes. The pricing of options and corporate liabilities. *The Journal of Political Economy*, 81(3):637–654, May - Jun 1973.
- [28] R.R. Bliss and N. Panigirtzoglou. Option-implied risk aversion estimates. *Journal of Finance*, 59:407–446, 2004.
- [29] C. Bo-Young, P. Christoffersen, and K. Jacobs. Market skewness risk and the cross-section of stock returns. *Journal of Financial Economics*, 107(1):46–68, Jan 2013.
- [30] T. Bollerslev, R.Y. Chou, and K.F. Kroner. Arch modeling in finance: A review of the theory and empirical evidence. *Journal of Econometrics*, 52(1-2):5–59, 1992.
- [31] T. Bollerslev and V. Todorov. Tails, fears and risk premia. *Journal of Finance*, 66:2165–2211, 2011.
- [32] T. Bollerslev and J.M. Wooldridge. Quasi-maximum likelihood estimation and inference in dynamic models with time-varying covariances. *Econometric Reviews*, 11(2):143–172, 1992.
- [33] J. Borovicka, L. Hansen, and J. Scheinkman. Misspecified recovery. *Journal of Finance (forthcoming)*, 2015.
- [34] D. Breeden and R. Litzenberger. Prices of state-contingent claims implicit in option prices. *The Journal of Business*, 51(4):621–651, October 1978.
- [35] P. Brémaud and M. Yor. Changes of filtrations and of probability measures. *Zeitschrift für Wahrscheinlichkeitstheorie und verwandte Gebiete*, 45:269–295, 1978.
- [36] D.P. Brown and M.R. Gibbons. A simple econometric approach for utilitybased asset pricing models. *Journal of Finance*, 40:359–381, 1985.
- [37] D.P. Brown and J. Jackwerth. The pricing kernel puzzle: Reconciling index option data and economic theory. *Working Paper*, 2004.
- [38] Y. Campbell, A.W.Lo, and A. MacKinlay. *The Econometrics of Financial Markets*, volume chapter 2, pp. 27-85. Princeton University Press, 1997.

- [39] J. Carnicer and J. Pena. Shape preserving representations and optimality of the bernstein basis. *Advances in Computational Mathematics*, 1:173–196, 1993.
- [40] P. Carr and J. Yu. Risk, return and the ross recovery. *The Journal of Derivatives*, 20(1):38–59, Fall 2012.
- [41] R. Chaudhuri and M. Schroder. Monotonicity of the stochastic discount factor and expected option returns. *Review of Financial Studies*, 28(5):1463–1505, 2015.
- [42] M. Chernov and E. Ghysels. A study towards a unified approach to the joint estimation of objective and risk neutral measures for the purpose of options valuation. *Journal of Financial Economics*, 56:407–458, 2000.
- [43] C. Chorro, D. Guegan, and F. Ielpo. Option pricing in garch-type models with generalized hyperbolic innovations. *Quantitative Finance*, 12(7):1079–1094, 2012.
- [44] P. Christoffersen, S.L. Heston, and K. Jacobs. The shape and term structure of the index option smirk: Why multifactor stochastic volatility models work so well. *Working paper*, 2009.
- [45] P. Christoffersen, S. Heston, and K. Jacobs. Capturing option anomalies with a variance-dependent pricing kernel. *Review of Financial Studies*, 26:1963–2006, 2013.
- [46] P. Christoffersen and K. Jacobs. Which garch model for option valuation? *Management Science*, 50(9):1024–1221, 2004.
- [47] P. Christoffersen, K. Jacobs, and C. Ornathanalai. Dynamic jump intensities and risk premiums: Evidence from s&p 500 returns and options. *Journal of Financial Economics*, 106:447–472, 2012.
- [48] J.H. Cochrane. *Asset Pricing*. Princeton University Press, 2001.
- [49] A. Mas Colell, M.D. Whinston, and J.R. Green. *Microeconomic Theory*. Oxford University Press, 1995.
- [50] G.M. Constantinides and S. Perrakis. Stochastic dominance bounds on derivatives prices in a multiperiod economy with proportional transaction costs. *Journal of Economic Dynamics and Control*, 26:1323–1352., 2002.

- [51] J. Cvitanic and I. Karatzas. Convex duality in constrained portfolio optimization. *Annals of Applied Probability*, 2:767–818, 1992.
- [52] F. Delbaen and W. Schachermayer. A general version of the fundamental theorem of asset pricing. *Mathematische Annalen*, 300:463–520, 1994.
- [53] F. Delbaen and W. Schachermayer. Fundamental theorem of asset pricing for unbounded stochastic processes. *Mathematische Annalen*, 334:215–250, 1998.
- [54] J.L. Doob. *Stochastic Processes*. Wiley, 1953.
- [55] J.C. Duan. The garch option pricing model. *Mathematical Finance*, 5:13–32, 1995.
- [56] J.C. Duan and J.G. Simonato. Empirical martingale simulation for asset prices. *Management Science* 44, 9:1218–1233, 1998.
- [57] B. Dumas, J. Fleming, and R.R. Whaley. Implied volatility functions: Empirical tests. *Journal of Finance*, LIII(6):2059–2106, Dec 1998.
- [58] P.H. Dybvig. Inefficient dynamic portfolio strategies or how to throw away a million dollars in the stock market. *The Review of Financial Studies*, 1(1):67–88, Spring 1988.
- [59] P.H. Dybvig and S. Ross. *Arbitrage, State Prices and Portfolio Theory*, volume 1 of *Handbook of the Economics and Finance*. Elsevier, 1 edition, 2003.
- [60] Elandt-Johnson and Johnson, editors. *Survival Models and Data Analysis*. John Wiley and Sons, 1980.
- [61] K.D. Elworthy, X.-M.Li, and M. Yor. The importance of strictly local martingales: Applications to radial ornstein-uhlenbeck processes. *Probability Theory and Related Fields*, 115:325–355, 1999.
- [62] R. Engle. Autoregressive conditional heteroscedasticity with estimates of the variance of united kingdom inflation. *Econometrica*, 50(4):987–1007, Jul 1982.
- [63] R. Engle and V.K. Ng. Measuring and testing the impact of news on volatility. *Journal of Finance*, 48(5):1749–1778, Dec 1993.
- [64] R. Engle and V. Rosenberg. Empirical pricing kernels. *Review of Financial Studies*, 64(341-372), 2002.

- [65] B. Eraker. Do stock prices and volatility jump? reconciling evidence from spot and option prices. *Journal of Finance*, 59(1367-1403), 2001.
- [66] M.D. Escobar. Estimating normal means with a dirichlet process prior. *Journal of the American Statistical Association*, 89(425):268–277, 1994.
- [67] M.D. Escobar and M. West. Bayesian density estimation and inference using mixtures. *Journal of the American Statistical Association*, pages 577–588, 1995.
- [68] E. Fama. The behavior of stock-market prices. *The Journal of Business*, 38(1):34–105, Jan 1965.
- [69] T. Ferguson. Bayesian analysis of some non-parametric problems. *Annals of Statistics*, 1(2):209–230, 1973.
- [70] D. Fernzhold and I. Karatzas. On optimal arbitrage. *Annals of Applied Probability*, 20:1179–1204, 2010.
- [71] W.E. Ferson and F. Siegel. Stochastic discount factor bounds with conditioning information. *Review of Financial Studies*, 15(2):567–595, 2003.
- [72] S. Figlewski and X. Wang. Is the leverage effect a leverage effect? *Technical report, NYU Stern School of Business and City University of Hong Kong.*, 2000.
- [73] H. Follmer and P. Protter. Local martingales and filtration shrinkage. *ESAIM Probability and Statistics*, 14:825–838, 2011.
- [74] G. Frahm. Pricing and valuation under the real-world measure. *International Journal of Theoretical and Applied Finance*, 19(1), February 2016.
- [75] R. Gallant, L. Hansen, and G. Tauchen. Using conditional moments of asset payoffs to infer the volatility of the intertemporal marginal rates of substitution. *Journal of Econometrics*, pages 141–179, 45.
- [76] L.R. Glosten, R. Jagannathan, and D.E. Runkle. On the relation between the expected value and the volatility of the nominal excess return on stocks. *The Journal of Finance*, 48(5):1779–1801, 1993.
- [77] Y. Goulbev, W. Härdle, and R. Timonfeev. Testing monotonicity of pricing kernels. *SFB 649 Working paper*, 2008.

- [78] L. Hansen and R. Jagannathan. Implications of security market data for models of dynamic economies. *Journal of Political Economy*, 99(2), 1991.
- [79] L. Hansen and S. Richard. The role of conditioning information in deducing testable restrictions implied by dynamic asset pricing models. *Econometrica*, 55:587–613, 1987.
- [80] W. Hardle and Z. Hlavka. Dynamics of state price densities. *Journal of Econometrics*, 150:1–15, 2009.
- [81] J. M. Harrison and D.M. Kreps. Martingales and arbitrage in multiperiod securities markets. *Journal of Economic Theory*, 20:381–408, 1979.
- [82] J. M. Harrison and S.R. Pliska. A stochastic calculus model of continuous trading: Complete markets. *Stochastic Processes and their Applications*, 15:313–316, 1983.
- [83] P. Henrotte. Pricing kernels and dynamic portfolios. *Pricing Kernels and Dynamic Portfolios*, Available at SSRN: <http://ssrn.com/abstract=302047>, August 2002.
- [84] T. Hens and C. Reichlin. Three solutions to the pricing kernel puzzle. *Working paper, University of Zurich*, 2011.
- [85] S. Heston. A closed-form solution for options with stochastic volatility, with applications to bond and currency options. *Review of Financial Studies*, 6:327–343, 1993.
- [86] S. Heston, M. Lowenstein, and G. Willard. Options and bubbles. *Review of Financial Studies*, 20:359–390, 2007.
- [87] S. Heston and S. Nandi. A closed-form garch option valuation model. *Review of Financial Studies*, 13(585-625), 2000.
- [88] W. Hildebrand and A. Kirman. *Equilibrium Analysis*. Amsterdam, North-Holland, 1988.
- [89] J.E. Jr. Ingersoll. *Theory of Financial Decision Making*. Yale Univeristy, 1986.
- [90] H. Ishwaran and L.F. James. Gibbs sampling methods for stick-breaking priors. *Journal of the American Statistical Association*, 96:161–173, 2001.
- [91] K. Itô and S. Watanabe. Transformation of markov processes by multiplicative functionals. *Annals de l'institut Fourier*, 15:15–30, 1965.

- [92] J. Jackwerth. Option-implied risk-neutral distributions and implied binomial trees: A literature review. *Journal of Derivatives*, 7(2):66–82, Winter 1999.
- [93] J. Jackwerth. Recovering risk aversion from option prices and realized returns. *Review of Financial Studies*, 13:433–451, 2000.
- [94] J. Jackwerth and M. Rubinstein. Recovering probability distributions from option prices. *The Journal of Finance*, 51(5):1611–1631, Dec 1996.
- [95] R. Jarrow. How to detect an asset bubble. *SIAM Journal of Mathematical Finance*, 2:839–865, 2011.
- [96] R. Jarrow. The third fundamental theorem of asset pricing. *Annals of Financial Economics*, 7(2), December 2012.
- [97] R. Jarrow. Option pricing and market efficiency. *The journal of portfolio management*, pages 88–94, Fall 13 2013.
- [98] R. Jarrow and M. Larsson. The meaning of market efficiency. *Mathematical Finance*, 22:1–30, 2012.
- [99] R. Jarrow, P. Protter, and A.D. Sezer. Information reduction via level crossings in a credit risk model. *Finance and Stochastics*, 11(2):195–212, April 2007.
- [100] R. Jarrow, P. Protter, and K. Shimbo. Asset price bubbles in incomplete markets. *Mathematical Finance*, 20:145–185, 2010.
- [101] J.D. Jobson and B. Korkie. Estimation for markowitz efficient portfolios. *Journal of the American Statistical Association*, 75(371):544–554, Sep 1980.
- [102] J.D. Jobson and B. Korkie. Performance hypothesis testing with the sharpe and treynor measures. *The Journal of Finance*, 36(4):889–908, Sep 1981.
- [103] G. Johnson and L. Helms. Class "d" supermartingales. *Bulletin of the American Mathematical Society*, 69(1):59–62, 1963.
- [104] M. Kalli, J. Griffin, and S. Walker. Slice sampling mixture models. *Statistics and Computing*, 21:93–105, 2011.
- [105] B. Kelly and H. Jiang. Tail risk and asset prices. *Review of Financial Studies*, 27(10):2841–2871, 2014.

- [106] J.F.C. Kingman. Completely random measures. *Pacific J. Math.*, 21(1):59–78, 1967.
- [107] J.F.C. Kingman. *Poisson Processes*. Oxford University Press, 1993.
- [108] R. Korn and H. Kraft. A stochastic control approach to portfolio problems with stochastic interest rates. *SIAM Journal on Control and Optimization*, 40(4):1250–1269, 2001.
- [109] D. Leisen. Small sample bias in pricing kernel estimations. *Available at SSRN: <http://ssrn.com/abstract=2409661>*, Jan 2015.
- [110] S.F. LeRoy, J. Werner, and S. Ross. *Principles of Financial Economics*. Cambridge University Press, 2000.
- [111] H. Levy. Upper and lower bounds of put and call option value: Stochastic dominance approach. *Journal of Finance*, 40:1197–1217, 1985.
- [112] A. Lijoi, B. Nipoti, and I. Prünster. Bayesian inference with dependent normalized completely random measures. *Bernoulli forthcoming*, 2011.
- [113] M. Linn, S. Shive, and T. Shumway. Pricing kernel monotonicity and conditional information (january 22, 2014). *Available at SSRN: <http://ssrn.com/abstract=2383527>*, Jan 2014.
- [114] G.M. Ljung and G.E.P. Box. On a measure of lack of fit in time series of models. *Biometrika*, 65:297–303, 1978.
- [115] E.R. Jr. Lucas. Asset prices in an exchange economy. *Econometrica*, 46(6):1429–1445, Nov 1978.
- [116] R.D. Luce and H. Raiffa. *Games and Decision: Introduction and Critical Survey*. Dover Books on Mathematics, 1957.
- [117] S.N. MacEachern and P. Müller. Estimating mixture of dirichlet process models. *Journal of Computational and Graphical Statistics*, 7(2):223–238, 1998.
- [118] B. Mandelbrot. Forecast of future prices, unbiased market, and "martingale" models. *Journal of Business*, 39(1):242–255, 1966.
- [119] R.C. Merton. Lifetime portfolio selection under uncertainty: the continuous time case. *Review of Economics and Statistics*, 51:247–257, 1969.

- [120] R.C. Merton. An intertemporal capital asset pricing model. *Econometrica*, 41(5):867–887, Sep 1973.
- [121] R.C. Merton. On estimating the expected return on the market. *Journal of Financial Economics*, 8:323–361, 1980.
- [122] P-A. Meyer. Decomposition of supermartingales: the uniqueness theorem. *Illinois Journal of Mathematics*, 7(1):1–17, 1963.
- [123] P. Müller and F.A. Quintana. Non-parametric bayesian data analysis. *Statistics Science*, pages 95–110, 2004.
- [124] O. Nikodym. Sur une generalisation des integrales de m. j. radon. *Fundamenta Mathematicae*, 15:131–179, 1930.
- [125] B. Øksendal. The value of information in stochastic control and finance. *Australian Economic Papers*, 44(4):352–364, December 2005.
- [126] J. Pan. The jump-risk premia implicit in options: Evidence from an integrated time-series study. *Journal of Financial Economics*, 63(1):2–50, Jan 2002.
- [127] O. Papaspiliopoulos and G. Roberts. Retrospective markov chain monte carlo methods for dirichlet process hierarchical models. *Biometrika*, 95:169–186, 2008.
- [128] S. Pastorello, E. Renault, and N. Touzi. Statistical inference for random-variance option pricing. *Journal of Business and Economic Statistics*, 18(3):358–367, 2000.
- [129] M. Perman, J. Pitman, and M. Yor. Size-biased sampling of poisson point processes and excursions. *Probability Theory and Related Fields*, 1992.
- [130] S. Perrakis and P.J.Ryan. Option pricing bounds in discrete time. *The Journal of Finance*, 39(2):519–525, Jun 1984.
- [131] J. Pitman and M. Yor. The two-parameter poisson-dirichlet distribution derived from a stable subordinator subordinator. *Annals of Probability*, 25(2):855–900, 1997.
- [132] E. Platen and D. Heath. *A Benchmark Approach to Quantitative Finance*. MR22653260. Springer Finance, Berlin, 2006.

- [133] N.G. Polson and J.R. Stroud. *Bayesian Inference for Derivative Prices*. J.M. Bernardo and M.J. Bayarri and J.O. Berger and A.P. Dawid and D. Heckerman and A.F.M. Smith and M. West, 2003.
- [134] J.W. Pratt. Risk aversion in the small and in the large. *Econometrica*, 32(1-2):122–136, 1964.
- [135] P. Protter. *Stochastic Integration and Differential Equations*. Version 2.1. Springer Verlag, Heidelberg, second edition edition, 2005.
- [136] P. Protter. Strict local martingales with jumps. *Stochastic Processes and their Applications*, 125(4):1352–1367, April 2015.
- [137] A. Quarteroni, R. Sacco, and F. Saleri. *Numerical Mathematics*, volume 37. Springer Verlag, 2007.
- [138] F. Riesz. Sur les operations fonctionnelles lineaires. *C.R. Acad. Sci*, 149:974–977., 1909.
- [139] S. Ross. Options and efficiency. *The Quarterly Journal of Economics*, 90:74–89, 1976.
- [140] S. Ross. The recovery theorem. *The Journal of Finance*, 70(2):615–648, April 2015.
- [141] M. Rubinstein. The valuation of uncertain income streams and the pricing of options. *The Bell Journal of Economics*, 7(2):407–425, Autumn 1976.
- [142] M. Rubinstein. The jump-risk premia implicit in options: Evidence from an integrated time-series study. *Journal of Finance*, 49(3):771–818, Jul 1994.
- [143] C. Sala. Does the pricing kernel anomaly reflect forward looking beliefs? *Working Paper*, 2015.
- [144] C. Sala and G. Barone-Adesi. Conditioning the information in portfolio optimization. *Working Paper*, Available at SSRN: <http://ssrn.com/abstract=2675090>, 2015.
- [145] C. Sala and G. Barone-Adesi. The pricing kernel density: the case of the information that did not bark. *Working Paper*, 2015.
- [146] P.A. Samuelson. Lifetime portfolio selection by dynamic stochastic programming. *Review of Economics and Statistics*, 51:239–246, 1969.
- [147] J. Sethraman. A constructive definition of dirichlet priors. *Statistica Sinica*, 4:639–650, 1994.

- [148] H. Shefrin. *A Behavioral Approach to Asset Pricing*. Elsevier Academic Press, Boston, second edition edition, 2008.
- [149] Z. Song and D. Xiu. A tale of two option markets: Pricing kernels and volatility risk. *Fama-Miller Working Paper. Chicago Booth reserach paper No. 12-10*, 2014.
- [150] C. Stricker. Quasimartingales, martingales locales, et filtrations naturelles. *Wahrscheinlichkeitstheorie verw. Gebiete*, 39:55–63, 1977.
- [151] Stuart and Ord., editors. *Kendall's Advanced Theory of Statistics*, volume 1. Arnold, 6 edition, 1998.
- [152] P. Tankov and R. Cont. *Financial Modelling with Jump Processes*. CRC Press, 2004.
- [153] S. Walker. Sampling the dirichlet mixture model with slices. *Comm. Statist. Sim. Comput*, 36:45–54, 2007.
- [154] J. Wang and S. Ghosh. Shape restricted non-parametric regression with bernstein polynomials. *Computational Statistics and Data Analysis*, 56:2729–2741, 2012.
- [155] H. White. Maximum likelihood estimation of misspecified models. *Econometrica*, 50(1):1–25, 1982.
- [156] A. Yatchew and W. Härdle. Non-parametric state price density estimation using constrained least squares and the bootstrap. *Journal of Econometrics*, 133(2):579–599, 2006.
- [157] C. Yo. Pricing kernels with stochastic skewness and volatility risk. *Management Science*, 58(3):624–640, March 2012.
- [158] C. Yo, E. Renault, and R. Garcia. State dependence can explain risk-aversion puzzle. *The review of Financial studies*, 21(2):973–1011, 2008.
- [159] A. Ziegler. Why does implied risk aversion smile? *Review of Financial Studies*, 20(3):859–904, 2007.

Moneyness		Maturity					
		$\tau < 60$		$60 \leq \tau \leq 160$		$\tau > 160$	
K/S		Mean	St.D.	Mean	St.D.	Mean	St.D.
< 0.85	Put Price (\$)	0.77	1.20	2.54	3.20	8.69	7.63
	σ^{BS}	38.97	9.59	33.36	7.75	28.37	5.21
	Bid-Ask (%)	0.99	0.69	0.69	0.68	0.27	0.36
	Observations (n.)	1798		2357		2845	
0.85 – 1.00	Put Price (\$)	8.43	7.60	19.12	11.63	34.80	15.67
	σ^{BS}	22.38	7.05	21.88	5.35	21.28	3.91
	Bid-Ask (%)	0.19	0.21	0.10	0.05	0.06	0.03
	Observations (n.)	3356		2136		2314	
1.00 – 1.15	Call Price (\$)	7.45	7.81	15.79	12.60	34.82	18.91
	σ^{BS}	17.55	5.80	17.31	5.02	17.61	4.00
	Bid-Ask (%)	0.34	0.46	0.17	0.23	0.07	0.04
	Observations (n.)	2955		2118		2247	
> 1.15	Call Price (\$)	0.34	0.43	0.85	1.61	3.96	5.60
	σ^{BS}	34.87	12.35	23.34	8.27	18.87	4.48
	Bid-Ask (%)	1.81	0.47	1.49	0.71	0.83	0.78
	Observations (n.)	1633		2288		3154	

Table A.1: This table reports Mean, Standard Deviation (St.D.) and Number of observations (n.) for each moneyness/maturity category of out-of-the-money (OTM) and deeply-out-of-the-money (DOTM) SPX options observed on each Wednesdays from January 2, 2002 until December 29, 2004.

Moneyness is defined as the ratio of the strike price over the spot price (K/S).

Maturity (τ) is measured in calendar days and divided in short/medium/long time-to-maturity.

$\sigma^{B.S.}$ represents the Black & Scholes implied volatility estimated using the market prices and inverting the B&S function.

Bid-ask spread (Bid-Ask) is calculated as:

$$100 \cdot \frac{(\text{Ask price} - \text{Bid price})}{\text{Market price}}$$

where the market price is assumed to be equal to the average of the bid and ask prices.

Filtering criteria: $10 \leq \tau \leq 360$ days, $\sigma^{B.S.} \leq 70\%$ and/or price $> \$0.05 \rightarrow$ Total Sample: 29,201 observations.

Statistics	Daily Prices		Daily log-returns	
	1988 - 2002	2002 - 2004	1988 - 2002	2002 - 2004
Mean	698.580	1029.453	4.29 e-04	7.70e-05
Median	474.655	1058.305	4.62e-04	4.63e-03
Std. Dev.	392.922	108.964	0.009	0.012
MAD	346.922	95.597	0.006	0.008
IQR	700.58	196.24	0.009	0.012
Minimum	250.830	776.76	-0.071	-0.042
Maximum	1527.46	1213.54	0.049	0.055
Skewness	0.696	-0.4241	-0.3226	0.250
Kurtosis	1.963	1.893	7.597	4.989
t (Mu)	105.235	259.42	2.609	0.176
t (Sk)	16.819	-4.754	-7.792	2.811
t (Ku)	-12.519	-6.202	55.529	11.179
JB	439.645	61.075	3144.20	132.87
Lillief.	0.216	0.141	0.0638	0.055

Table A.2: Summary statistics of the S&P 500 index daily prices and log-returns from January 2, 2002 going back 3500 observations (1988) - first and third column - and from January 2, 2002 until December 29, 2004 - second and fourth columns. Mean (μ) is the average mean. Std Dev. (σ) is the standard deviation. MAD is the mean absolute deviation. IQR is the inter-quantile range. Skewness (ι) and Kurtosis (κ) are the third and fourth moment respectively.

$$t(\mu) = \frac{\mu}{\sigma/\sqrt{T}}$$

$$t(\iota) = \frac{\iota}{\sqrt{6/T}}$$

$$t(\kappa) = \frac{(\kappa - 3)}{\sqrt{24/T}}$$

where T represents the time window. JB is the Jarque Bera test:

$$JB = T \cdot \left(\frac{\iota^2}{6} + \frac{(\kappa - 3)^2}{24} \right)$$

Lillief. is the Lilliefor normality test of the default null hypothesis that the sample comes from a distribution in the normal family, against the alternative that it does not come from a normal distribution. The test returns the logical value = 1 if it rejects the null hypothesis at the 5% significance level, and = 0 if it cannot.

EPK Horizontal Support - Literature Review

Authors and Year	Min	Max
Ziegler (2007)	0.70	1.50
Beare-Shmidt (2011)	0.86	1.09
Engle and Rosenberg (2002)	0.90	1.08
Jackwerth (2000)	0.90	1.07
Chabi Yo (2012)	0.90	1.07
Jackwerth and Brown (2012)	0.95	1.05

Table A.3: This table provides short summary statistics of the minimum and maximum extension of some of the estimated pricing kernels present in literature. All papers have been cited throughout the thesis. All kernels taken into consideration are projected onto market gross returns. It follows that 1 is the area of no return while values > 1 (< 1) represents the call(put) side of the empirical pricing kernel.

The precision parameter - Summay Statistics

	Mean	Median	St.D	Max. value	Min. Value
$\alpha_{t,T}^*$ DOTM	1.8250	1.7918	0.3127	2.5000	1.2580
$\alpha_{t,T}^*$ OTM	1.9155	1.9497	0.3220	2.5000	1.2190

Table A.4: Mean, Median, Standard Deviation (St.D.), Maximum and Minimum values of the daily $\alpha_{t,T}^*$ calibrated using DOTM (top) and OTM (bottom) options.

GJR (Physical P)	ω		α		β		γ		Persistence	
	Mean	St.D.	Mean	St.D.	Mean	St.D.	Mean	St.D.	Mean	St.D.
Year										
2002	1.48e-06	1.33e-07	7.53e-03	1.02e-03	0.927	0.005	0.100	0.009	0.984	0.01
2003	1.33e-06	2.67e-07	7.76e-03	2.21e-03	0.926	0.007	0.107	0.007	0.987	0.01
2004	1.05e-06	4.60e-08	7.29e-03	2.21e-03	0.930	0.002	0.106	0.003	0.990	0.00

Table A.5: Yearly Mean and Standard Deviation (St.D.) of the physical (P) GJR GARCH parameters $\theta_t = f(\omega, \alpha, \beta, \gamma)$ calibrated each Wednesdays - from January 2, 2002 to December 29, 2004 - on the S&P 500 Index returns using the Pseudo Maximum Likelihood (PML) approach and 3500 log-returns.
The GJR GARCH (1,1) model under the physical measure is:

$$\log \frac{S_t}{S_{t-1}} = \mu + \epsilon_t$$

$$\sigma_t = \omega + \beta\sigma_{t-1}^2 + \alpha\epsilon_{t-1}^2 + \mathbb{1}_{t-1}\gamma\epsilon_{t-1}^2$$

Persistence is defined as: $= \alpha + \beta + \gamma/2$

GJR (R.N. Q - FHS)		$\tilde{\omega}$		$\tilde{\alpha}$		$\tilde{\beta}$		$\tilde{\gamma}$		Persistence		
Year	Mean	St.D.	Mean	St.D.	Mean	St.D.	Mean	St.D.	Mean	St.D.	Mean	St.D.
2002	3.14E-06	2.74E-06	9.64E-04	9.05E-02	0.866	0.091	0.216	0.143	0.975	0.252		
2003	3.29E-06	4.04E-06	1.31E-03	1.30E-01	0.866	0.130	0.213	0.207	0.973	0.363		
2004	2.08E-06	2.96E-06	3.55E-03	1.74E-01	0.832	0.174	0.293	0.301	0.981	0.498		
GJR (R.N. Q - Gauss)		$\tilde{\omega}$		$\tilde{\alpha}$		$\tilde{\beta}$		$\tilde{\gamma}$		Persistence		
Year	Mean	St.D.	Mean	St.D.	Mean	St.D.	Mean	St.D.	Mean	St.D.	Mean	St.D.
2002	4.02E-06	5.49E-06	5.84E-03	1.19E-02	0.845	0.093	0.271	0.165	0.985	0.187		
2003	3.57E-06	6.23E-06	4.48E-03	8.95E-03	0.829	0.206	0.273	0.327	0.969	0.378		
2004	1.69E-06	1.43E-06	2.05E-03	4.95E-03	0.827	0.110	0.315	0.198	0.986	0.214		

Table A.6: Upper panel: Mean and Standard Deviation of the Risk Neutral (Q) GJR GARCH parameters $\tilde{\theta} = f(\tilde{\omega}, \tilde{\alpha}, \tilde{\beta}, \tilde{\gamma})$ calibrated each Wednesdays - from January 2, 2002 to December 29, 2004 - on the cross-section of out-of-the-money (OTM) SPX options with Filtered Historical Innovations (FHS). Optimization method: Simplex method which is an unconstrained derivative free optimization for non linear functions.
The GJR GARCH model under the risk neutral measure is:

$$\log \frac{S_t}{S_{t-1}} = \mu + \epsilon_t$$

$$\sigma_t = \tilde{\omega} + \tilde{\beta}\sigma_{t-1} + \tilde{\alpha}\epsilon_{t-1}^2 + \mathbf{1}_{t-1}\tilde{\gamma}\epsilon_{t-1}^2$$

Persistence is defined as: $= \tilde{\alpha} + \tilde{\beta} + \tilde{\gamma}/2$
Center panel: same as the first table but with Gaussian innovations.

GJR (R.N. Q - FHS)		$\tilde{\omega}$		$\tilde{\alpha}$		$\tilde{\beta}$		$\tilde{\gamma}$		Persistence		
Year	Mean	St.D.	Mean	St.D.	Mean	St.D.	Mean	St.D.	Mean	St.D.	Mean	St.D.
2002	3.14e-06	2.74e-06	9.68e-04	0.0049	0.866	0.090	0.216	0.143	0.975	0.167		
2003	3.73e-06	5.48e-06	0.0012	0.0053	0.852	0.171	0.234	0.275	0.971	0.315		
2004	2.18e-06	2.87e-06	0.0020	0.0096	0.824	0.183	0.311	0.335	0.981	0.360		
GJR (R.N. Q - Gauss)		$\tilde{\omega}$		$\tilde{\alpha}$		$\tilde{\beta}$		$\tilde{\gamma}$		Persistence		
Year	Mean	St.D.	Mean	St.D.	Mean	St.D.	Mean	St.D.	Mean	St.D.	Mean	St.D.
2002	3.77e-06	5.32e-06	0.0057	0.0119	0.798	0.218	0.364	0.431	0.986	0.445		
2003	3.30e-06	4.70e-06	0.0040	0.0085	0.825	0.200	0.306	0.366	0.983	0.391		
2004	2.64e-06	3.67e-06	0.0017	0.0053	0.766	0.242	0.442	0.468	0.989	0.482		

Table A.7: Upper panel: Yearly Mean and Standard Deviation (St.D.) of the risk-neutral (Q) GJR GARCH parameters $\tilde{\theta}_t = f(\tilde{\omega}, \tilde{\alpha}, \tilde{\beta}, \tilde{\gamma})$ calibrated each Wednesdays - from January 2, 2002 to December 29, 2004 - on the cross-section of out-of-the-money SPX options with Filtered Historical Innovations (FHS). Optimization method: Quasi-Newton. The GJR GARCH (1,1) model under the risk-neutral measure is:

$$\log \frac{S_t}{S_{t-1}} = \mu + \epsilon_t$$

$$\sigma_t = \tilde{\omega} + \tilde{\beta}\sigma_{t-1}^2 + \tilde{\alpha}\epsilon_{t-1}^2 + \mathbb{1}_{t-1}\tilde{\gamma}\epsilon_{t-1}^2$$

Persistence is defined as: $= \tilde{\alpha} + \tilde{\beta} + \tilde{\gamma}/2$

Bottom panel: same as the top table but with Gaussian innovations.

Modified risk physical measure (p^\dagger)											
Mean		Median		St.err		Skewness		Kurtosis			
Error	FHS	Gauss	FHS	Gauss	FHS	Gauss	FHS	Gauss	FHS	Gauss	
$\tau = 24$	8.6810e-04	7.8914e-04	8.0607e-06	6.6434e-06	0.0020	0.0021	2.4877	3.0070	8.0742	11.0486	
$\tau = 52$	5.5896e-04	5.5906e-04	4.5874e-06	5.4007e-06	0.0014	0.0014	2.6881	2.7742	9.1031	9.6495	
$\tau = 87$	4.8558e-04	4.8558e-04	8.3001e-06	6.4017e-06	0.0011	0.0011	2.4312	2.4200	7.7090	7.6260	
$\tau = 150$	4.4294e-04	4.4293e-04	1.5757e-05	1.2377e-05	8.6436e-04	8.6331e-04	2.0429	2.0095	5.5847	5.6840	
$\tau = 241$	4.3414e-04	4.3412e-04	4.0719e-05	3.1092e-05	7.0910e-04	7.1155e-04	1.6029	1.5836	4.1410	4.0511	
$\tau = 332$	3.7623e-04	3.7621e-04	3.5994e-05	3.2425e-05	5.9316e-04	6.0141e-04	1.5117	1.5155	3.8257	3.8233	
Risk physical measure (p)											
Mean		Median		St.err		Skewness		Kurtosis			
Error	FHS	Gauss	FHS	Gauss	FHS	Gauss	FHS	Gauss	FHS	Gauss	
$\tau = 24$	0.0011	7.8955e-04	1.5011e-05	2.9181e-17	0.0022	0.0019	2.1011	2.5886	6.0873	8.4899	
$\tau = 52$	5.5900e-04	5.5901e-04	2.2058e-06	5.2233e-08	0.0014	0.0013	2.6246	2.5479	8.7220	8.2856	
$\tau = 87$	4.8557e-04	4.8559e-04	3.8759e-06	4.6127e-07	0.0011	0.0011	2.4270	2.3321	7.662	7.1397	
$\tau = 150$	4.4294e-04	4.4294e-04	8.6973e-06	2.1987e-06	8.8879e-04	8.8193e-04	2.0919	2.0166	6.0564	5.6967	
$\tau = 241$	4.3413e-04	4.3414e-04	2.8141e-05	1.5717e-05	7.3851e-04	7.4107e-04	1.6750	1.6249	4.3839	4.1753	
$\tau = 332$	3.7623e-04	3.7623e-04	2.3760e-05	1.3435e-05	6.2005e-04	6.3022e-04	1.5916	1.5639	4.0828	3.9581	
Risk Neutral measure (q)											
Mean		Median		St.err		Skewness		Kurtosis			
Error	FHS	Gauss	FHS	Gauss	FHS	Gauss	FHS	Gauss	FHS	Gauss	
$\tau = 24$	7.8960e-04	7.890e-04	5.2204e-06	7.8056e-06	0.0020	0.0022	2.8543	3.1483	10.0977	11.9673	
$\tau = 52$	5.5907e-04	5.5885e-04	5.4593e-06	6.9992e-06	0.0014	0.0014	2.7294	2.8836	9.3531	10.3352	
$\tau = 87$	4.8559e-04	4.8558e-04	8.6024e-06	9.4021e-06	0.0011	0.0011	2.4547	2.4943	7.8444	8.0441	
$\tau = 150$	4.4294e-04	4.4293e-04	1.8151e-05	1.9505e-05	8.5813e-04	8.6213e-04	2.0402	2.0439	5.8439	5.8528	
$\tau = 241$	4.3414e-04	4.3411e-04	4.5697e-05	4.54404e-05	6.9959e-04	7.0303e-04	1.5830	1.5990	4.0766	4.1280	
$\tau = 332$	3.7623-e04	3.7619e-04	4.3054e-05	4.2090e-05	5.8264e-04	5.8959e-04	1.4800	1.5209	3.7249	3.8669	

Table A.8: Summary statistics of the Conditional Risk Physical, the Partially Conditional Risk Physical and the Risk Neutral distributions

Risk-neutral measure - Fully conditional risk physical measure: $\Delta_{t,T}^{\text{Cond.}} = q_{t,T} - p_{t,T}^{\dagger}$					
Year	τ	Δ_t Mean	Δ_t Var	Δ_t Skewness	Δ_t Kurtosis
2002	Short	0	0	0.0530	0.3765
	Medium	0	0	0.0427	0.2592
	Long	0	0	0.0295	0.1636
2003	Short	0	0	0.0081	0.0777
	Medium	0	0	-0.0050	0.0018
	Long	0	0	0.0033	0.0345
2004	Short	0	0	0.0943	0.6352
	Medium	0	0	0.1408	0.8751
	Long	0	0	0.1857	1.0230

Risk-neutral measure - Partially conditional risk physical measure: $\Delta_{t-\Delta_t,T}^{\text{Part. Cond.}} = q_{t,T} - p_{t-\Delta_t,T}$					
Year	τ	Δ_t Mean	Δ_t Var	Δ_t Skewness	Δ_t Kurtosis
2002	Short	0	0	0.1583	1.1086
	Medium	0	0	0.0822	0.4956
	Long	0	0	0.0227	0.1503
2003	Short	0	0	0.0078	0.1406
	Medium	0	0	-0.0377	-0.1361
	Long	0	0	-0.0319	-0.1095
2004	Short	0	0	0.2885	1.9063
	Medium	0	0	0.3303	2.0117
	Long	0	0	0.3571	1.9243

Table A.9: Yearly pointwise difference between $q_{t,T} - p_{t,T}^{\dagger}$ (top) and $q_{t,T} - p_{t-\Delta_t,T}$ (bottom) computed with FHS, R.P.=4% and $\alpha_t^* = 2.5$ and decreasing. The times-to-maturity are: Short for $\tau < 60$, Medium for $60 \leq \tau \leq 180$ and Long for $\tau > 180$. For readability purposes, values smaller than $ne-03$ are set equal to 0.

GJR (R.N. Q - FHS - FIRST)		ω		α		β		γ		Persistence		
Year	Mean	St.D.	Mean	St.D.	Mean	St.D.	Mean	St.D.	Mean	St.D.	Mean	St.D.
2002	1.3069E-06	5.0656E-07	0.1068	0.0262	0.8064	0.0449	0.1600	0.0520	0.9932	0.097024	0.9932	0.097024
2003	9.5987E-07	4.1221E-07	0.1018	0.0872	0.8154	0.1209	0.1582	0.1586	0.9963	0.287439	0.9963	0.287439
2004	7.0199E-07	2.5966E-07	0.0628	0.0247	0.8673	0.0793	0.1379	0.1577	0.9991	0.182875	0.9991	0.182875
GJR (R.N. Q - FHS - SECOND)		ω		α		β		γ		Persistence		
Year	Mean	St.D.	Mean	St.D.	Mean	St.D.	Mean	St.D.	Mean	St.D.	Mean	St.D.
2002	1.2638E-06	8.1499E-07	0.1087	0.0320	0.8049	0.0567	0.1598	0.0564	0.9935	0.116895	0.9935	0.116895
2003	8.1789E-07	4.1555E-07	0.1021	0.1152	0.8056	0.1602	0.1776	0.2497	0.9965	0.40021	0.9965	0.40021
2004	6.4254E-07	2.9647E-07	0.0639	0.0283	0.8726	0.0588	0.1254	0.0904	0.9992	0.132294	0.9992	0.132294
SECOND - FIRST		ω		α		β		γ		Persistence		
Year	Mean	St.D.	Mean	St.D.	Mean	St.D.	Mean	St.D.	Mean	St.D.	Mean	St.D.
2002	0.0000	0.0000	0.0019	0.0058	-0.0015	0.0119	-0.0002	0.0044	0.0003	-1.9871E-02	0.0003	-1.9871E-02
2003	0.0000	0.0000	0.0003	0.0279	-0.0098	0.0393	0.0194	0.0911	0.0002	-1.1277E-01	0.0002	-1.1277E-01
2004	0.0000	0.0000	0.0011	0.0036	0.0053	-0.0205	-0.0125	-0.0674	0.0001	5.0581E-02	0.0001	5.0581E-02

Table A.10: Upper panel: First risk neutral (Q) GJR GARCH parameters $\theta^* = f(\omega^*, \alpha^*, \beta^*, \gamma^*)$ calibrated each Wednesday - 157 observations - on the cross-section of semi-synthetic SPX options with Filtered Historical Innovations (FHS).
 Central panel: Second risk neutral Q pricing GJR GARCH parameters $\theta^* = f(\omega^*, \alpha^*, \beta^*, \gamma^*)$ calibrated each Wednesday - 157 observations - on the cross-section of semi-synthetic SPX options with Filtered Historical Innovations (FHS) using the upper panel parameters.
 Lower panel, pointwise difference between the second minus to first panel.
 All variables are defined as before.

Figures

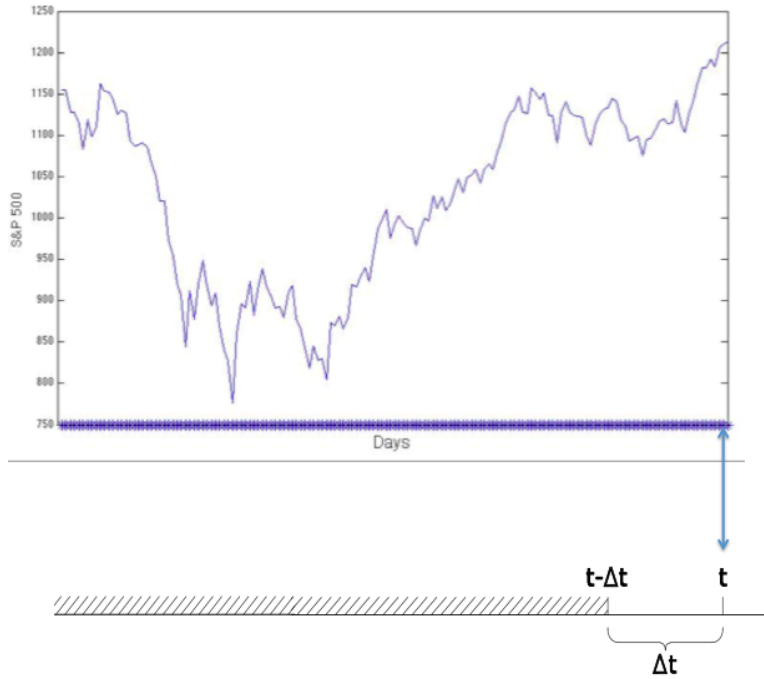


Figure A.2: Top: S&P 500 time series. The blue stars on the horizontal axis represent the set of daily observations. Below: *Suboptimal* S&P 500 time line. Starting from t and going back using the above daily observations, the resulting information set is, by construction, suboptimal, \mathcal{H}_t . What is missing are the investors *forward* looking beliefs, Δ_t .

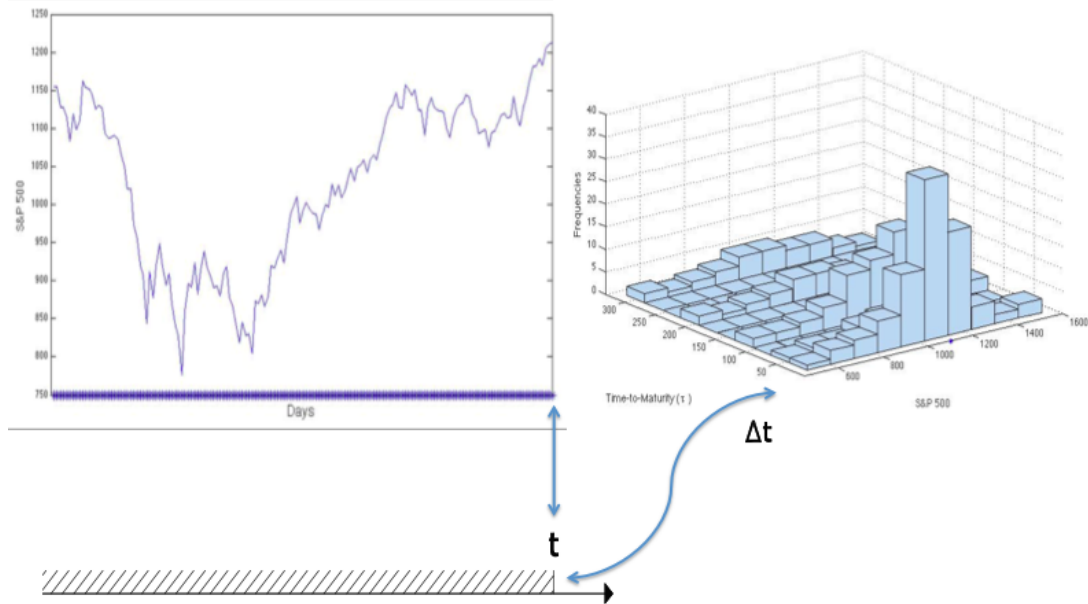


Figure A.3: Bottom: S&P 500 *time series* and *cross section* of option data. The option surface is represented by X axis: set of strike prices (the single blue star represent the t value of the S&P 500), Y axis: times-to-maturity τ ; Z axis: frequencies. Below: *Complete* S&P 500 time line. Completeness is achieved through the previously missing forward looking beliefs extracted from the time t cross-section of option data.

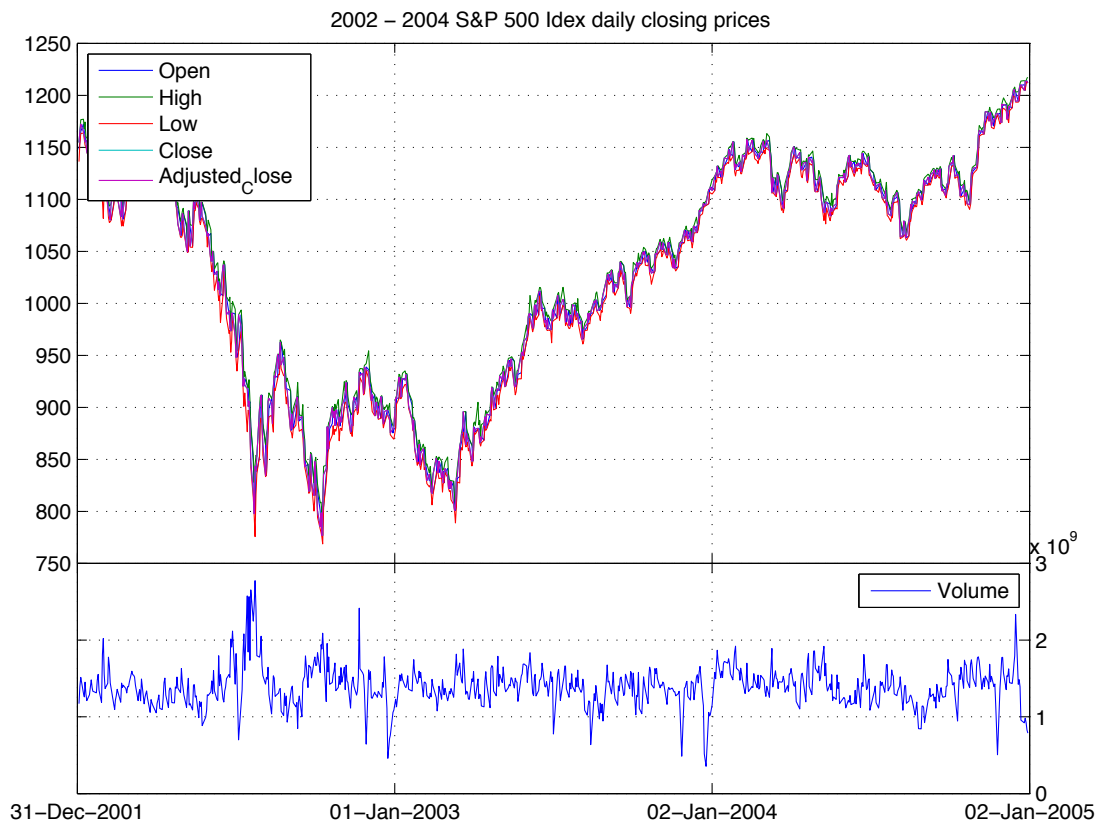
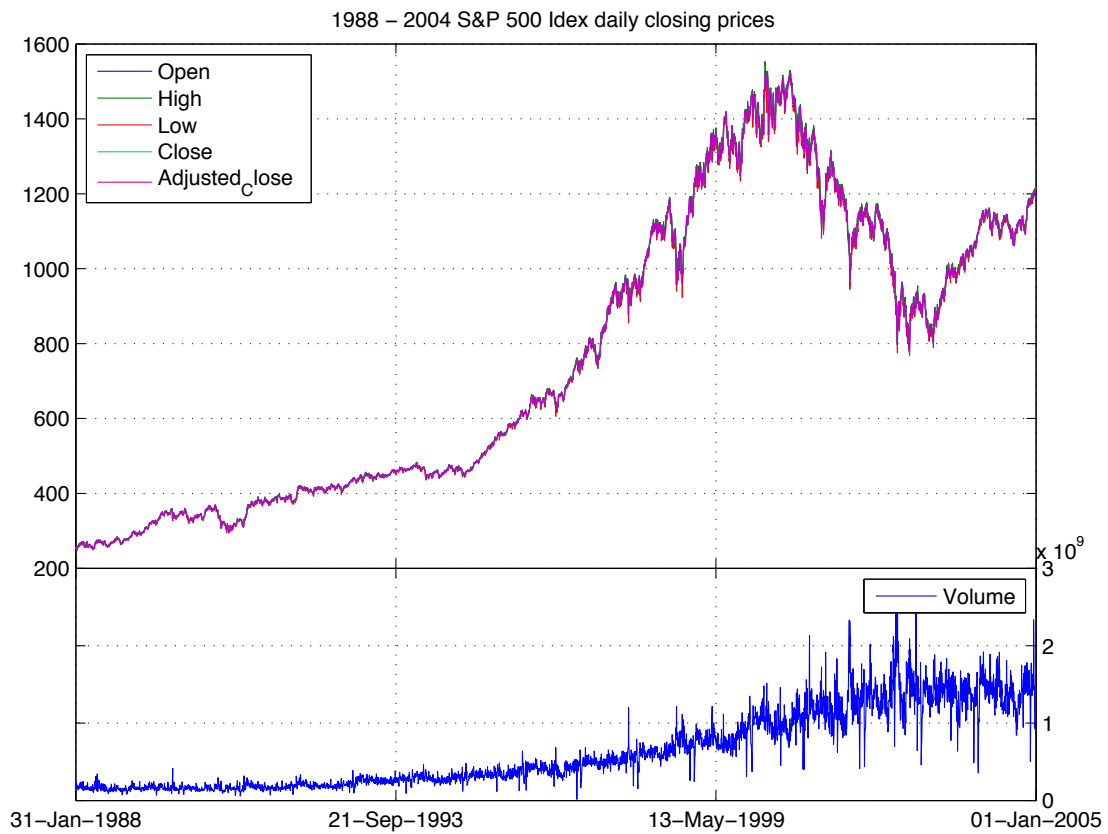


Figure A.4: Top: 1988 - 2002 time series of S&P 500 index daily closing prices (various indicator) and daily volume.
 Bottom: 2002 - 2004 time series of S&P 500 index daily closing prices (various indicator) and daily volume.

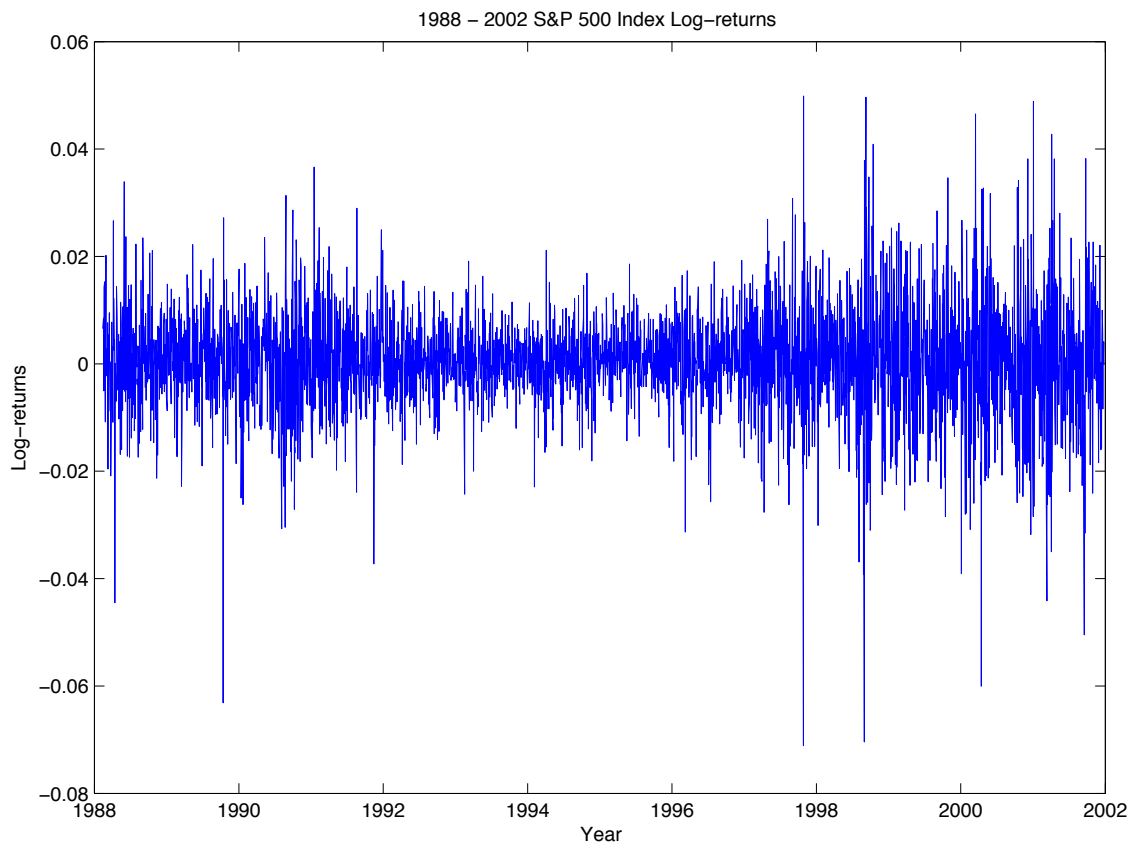
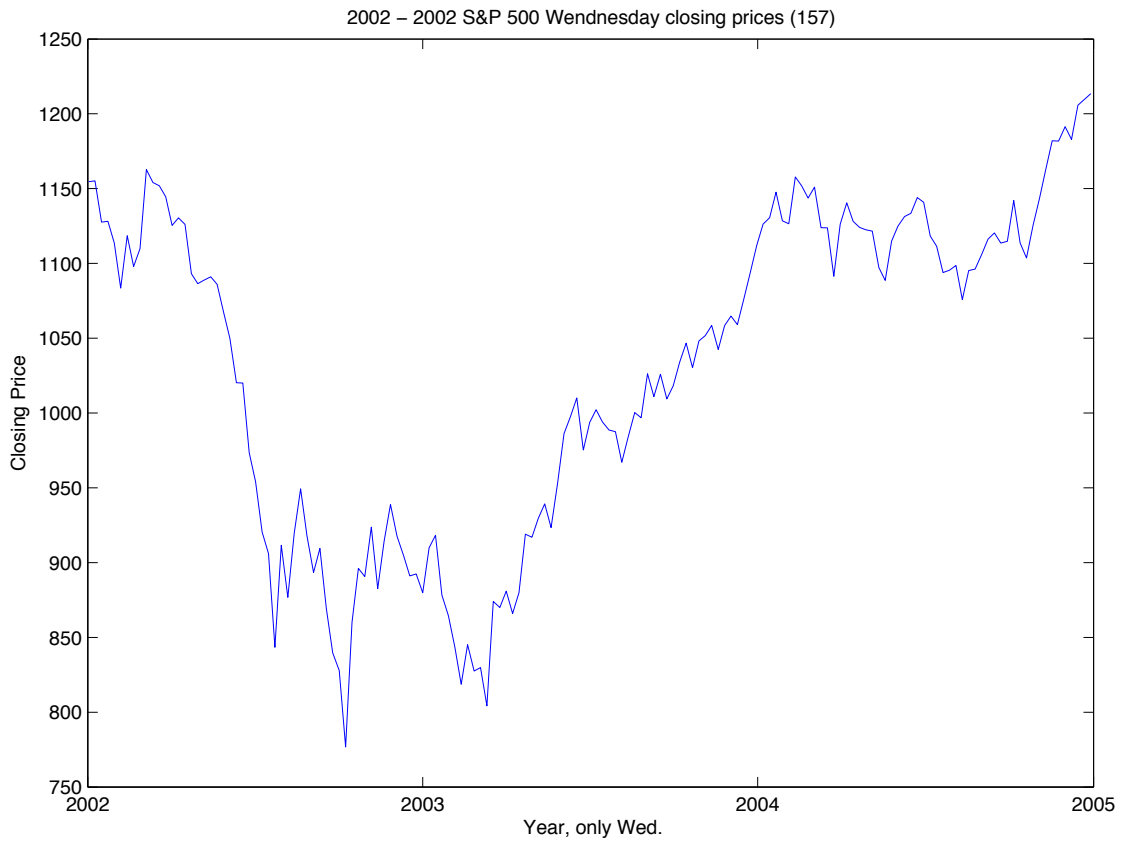


Figure A.5: Top: S&P 500 Index weekly closing prices (Wednesday only) from 02/01/2002 to 30/12/2004 - 157 daily observations. Bottom: 1988-2002 daily S&P 500 log-returns.

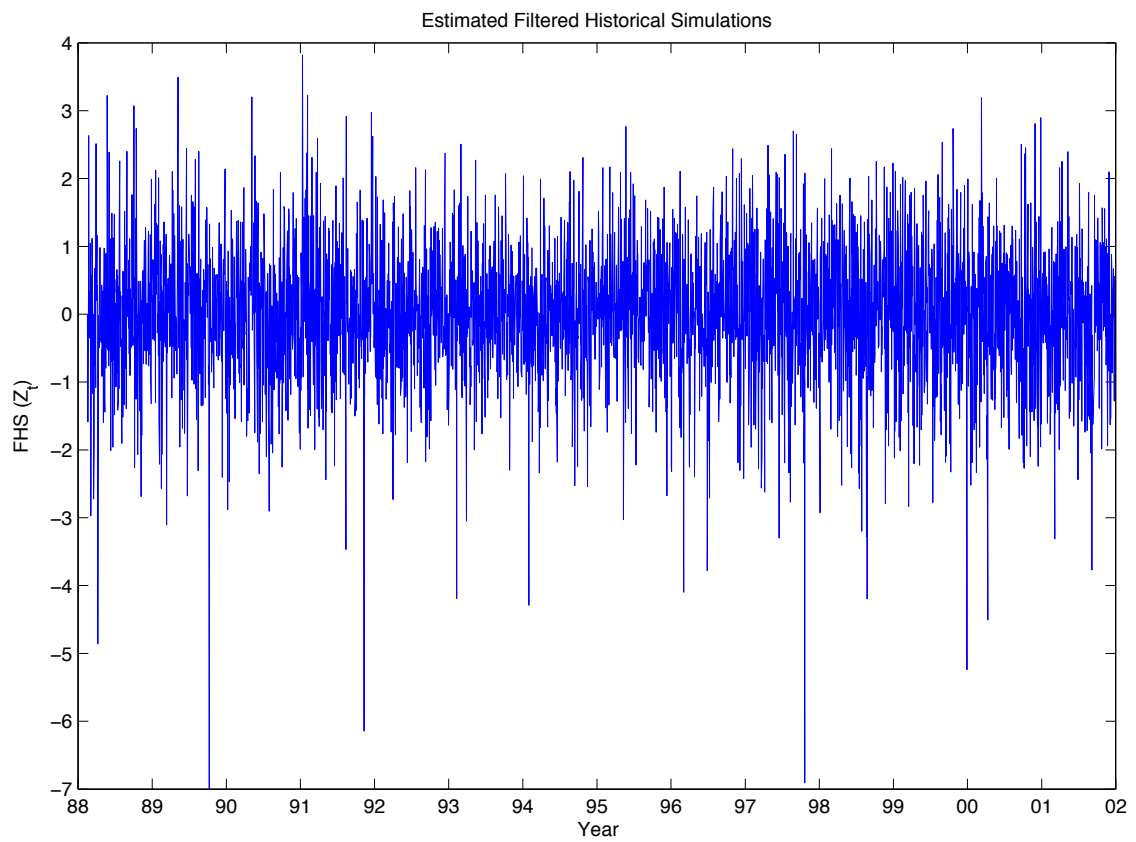
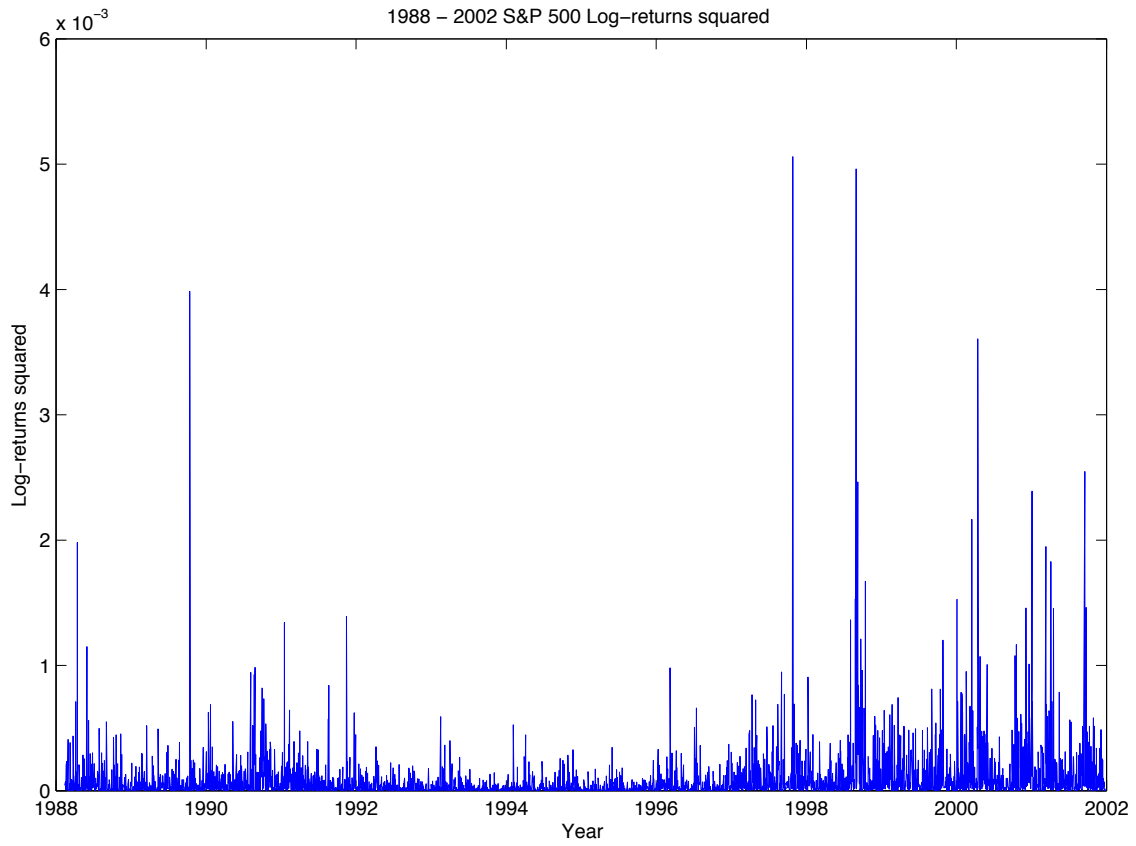


Figure A.6: Top: 1988-2002 daily S&P 500 squared log returns.
 Bottom: 1988-2002 time series of estimated (FHS) (z_t) from 02/01/2002 going back 3500 observations.

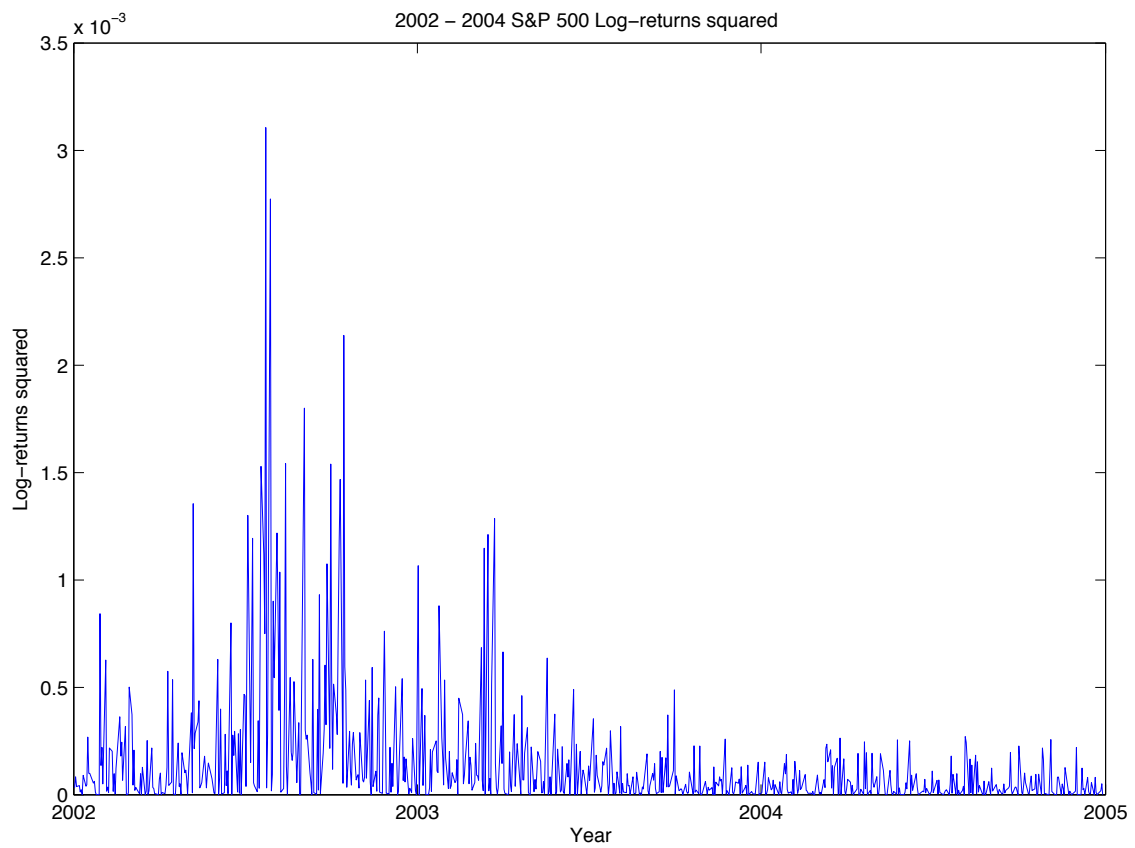
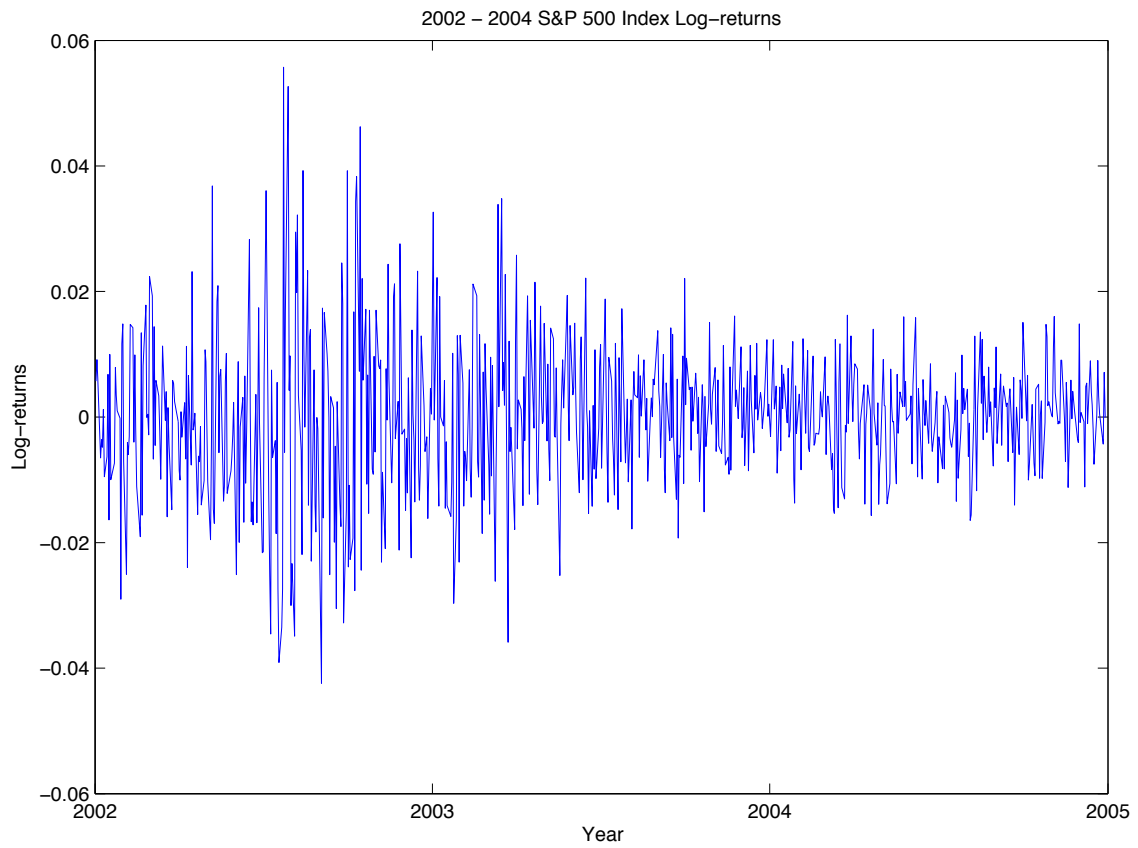
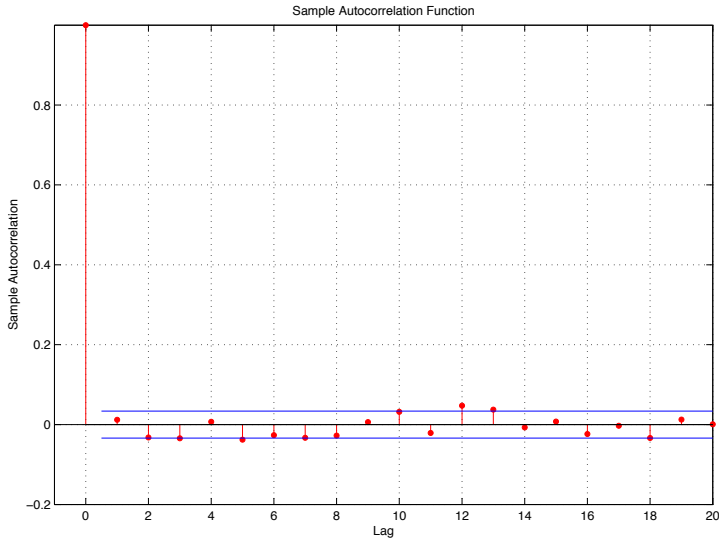
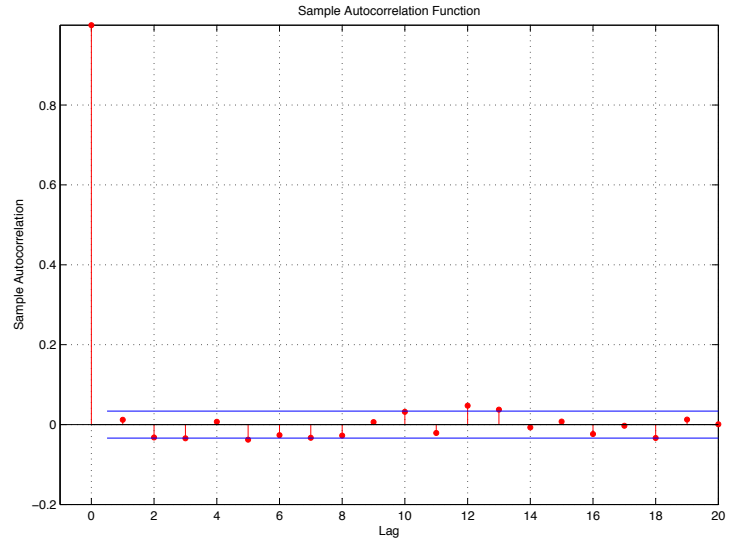


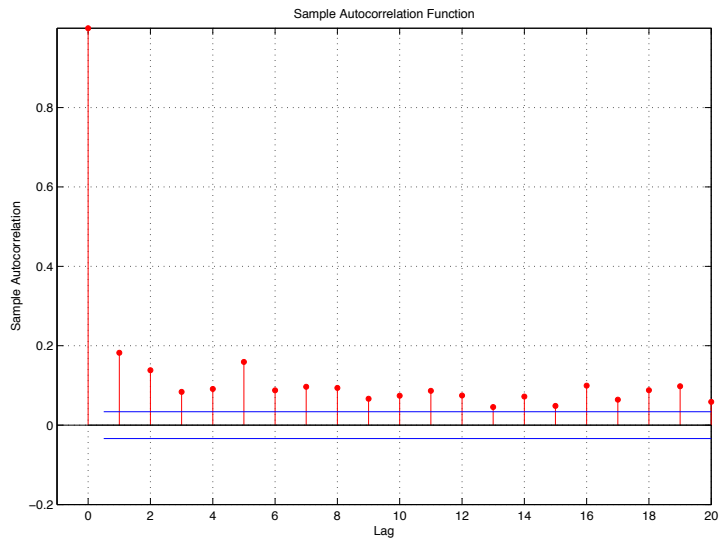
Figure A.7: Top: 2002-2004 daily S&P 500 log returns.
 Bottom: 2002-2004 daily S&P 500 squared log-returns.



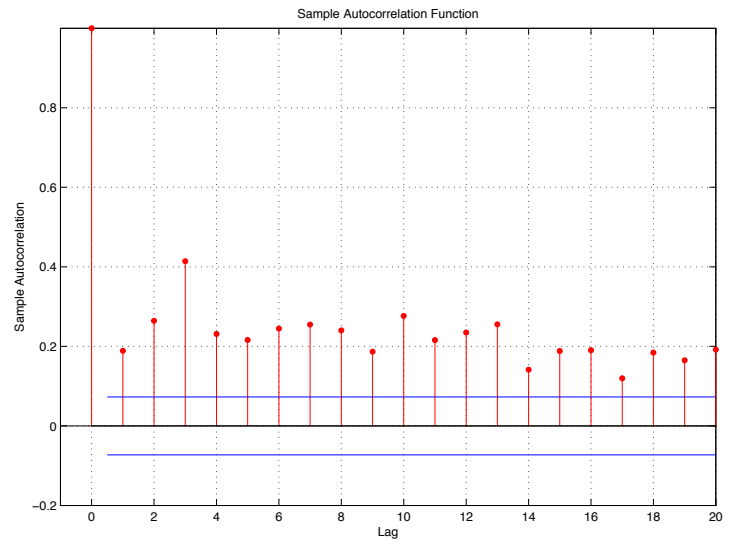
(a) ACF 1988-2002 log returns



(b) ACF 2002-2004 log returns



(c) ACF 1988-2002 squared log returns



(d) ACF 2002-2004 squared log returns

Figure A.8: Top: 1988-2002 and 2002-2004 ACF of daily S&P 500 log returns.
 Bottom: 1988-2002 and 2002-2004 ACF of daily S&P 500 squared log returns.

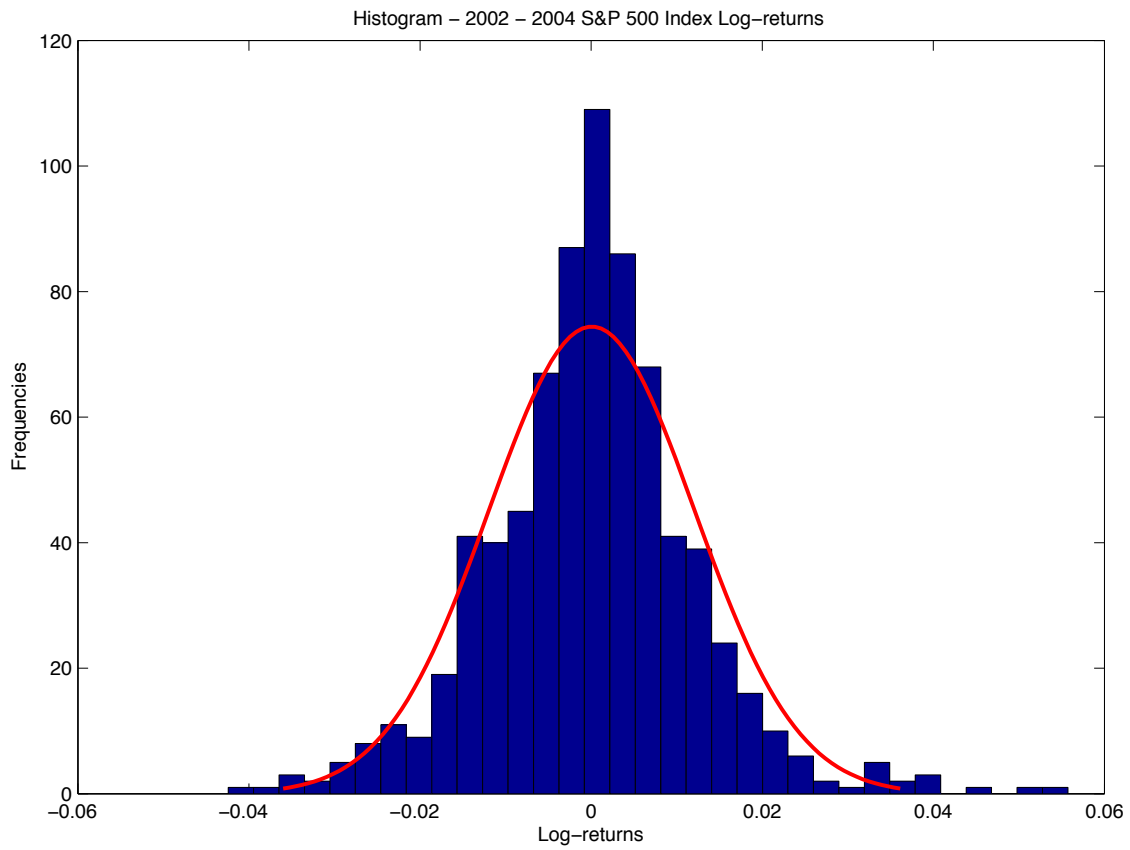
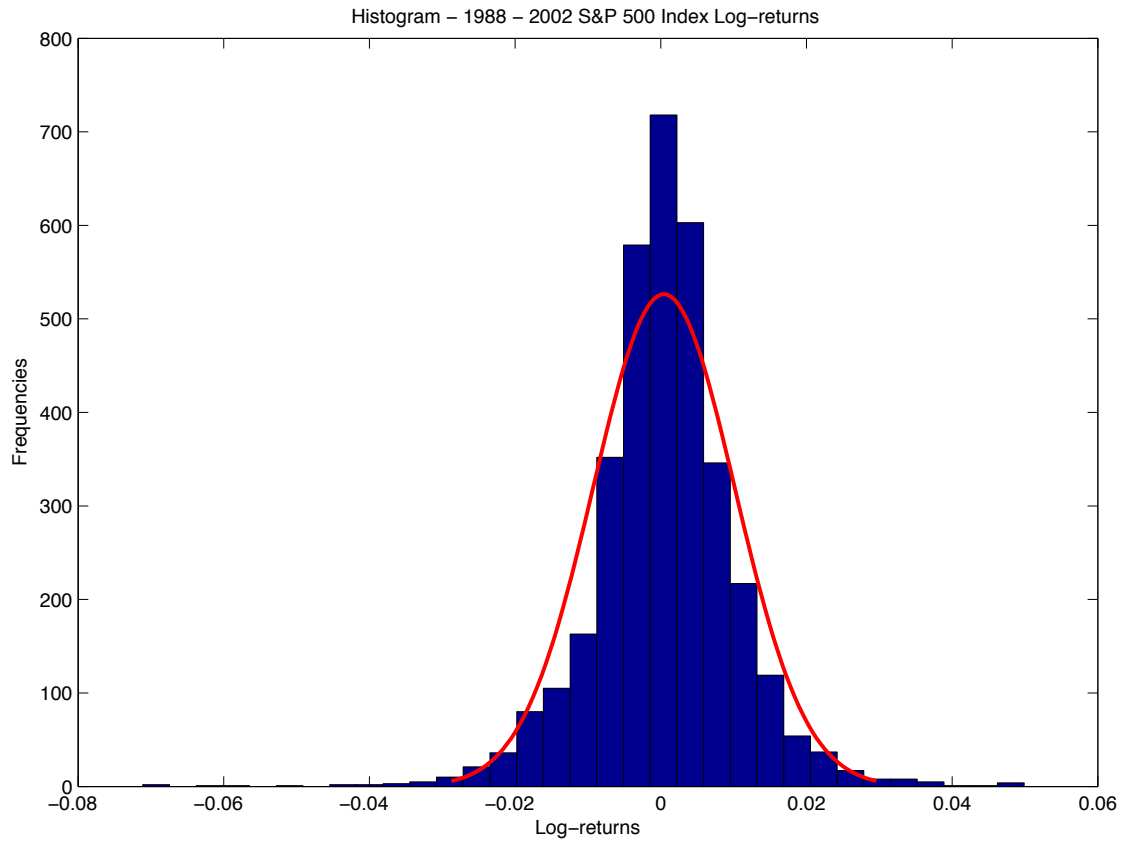
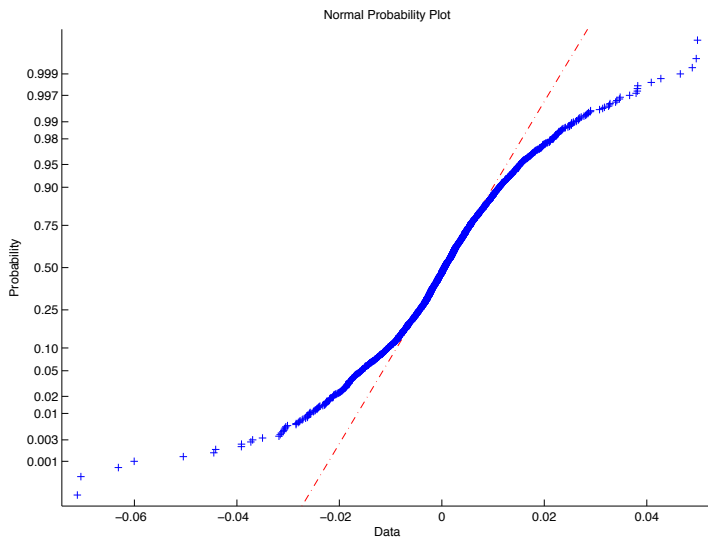
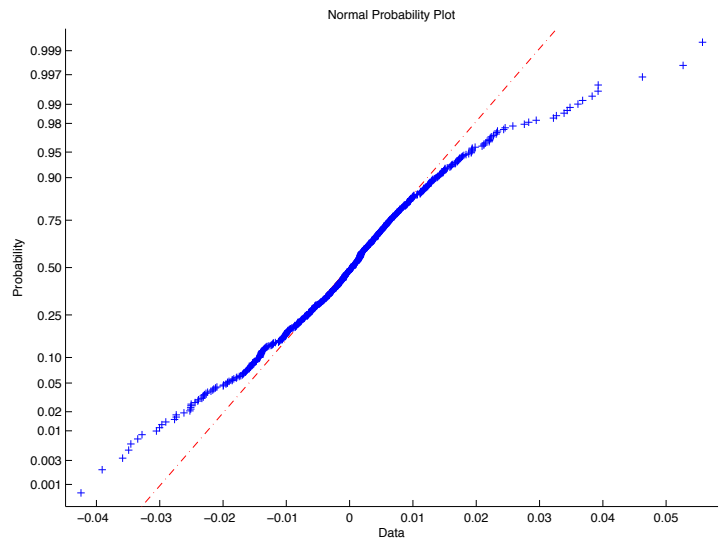


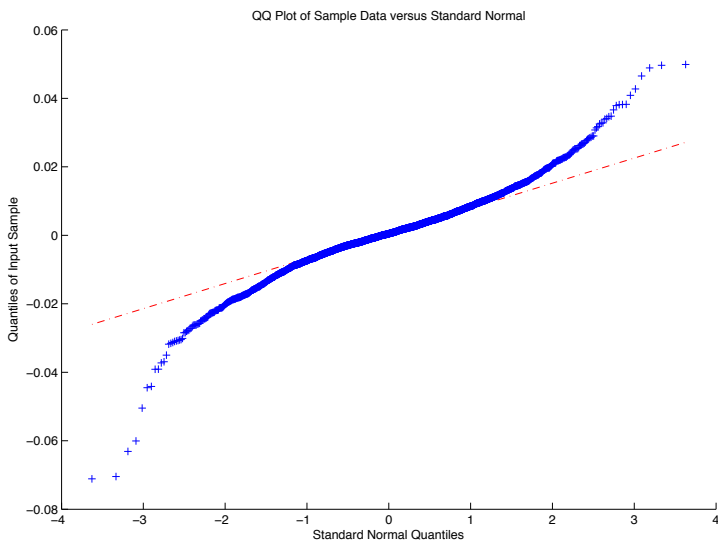
Figure A.9: Top: 1988-2002 Histogram of of daily S&P 500 log returns against normal distribution (red line).
 Bottom: 2002-2004 Histogram of of daily S&P 500 log returns against normal distribution (red line).



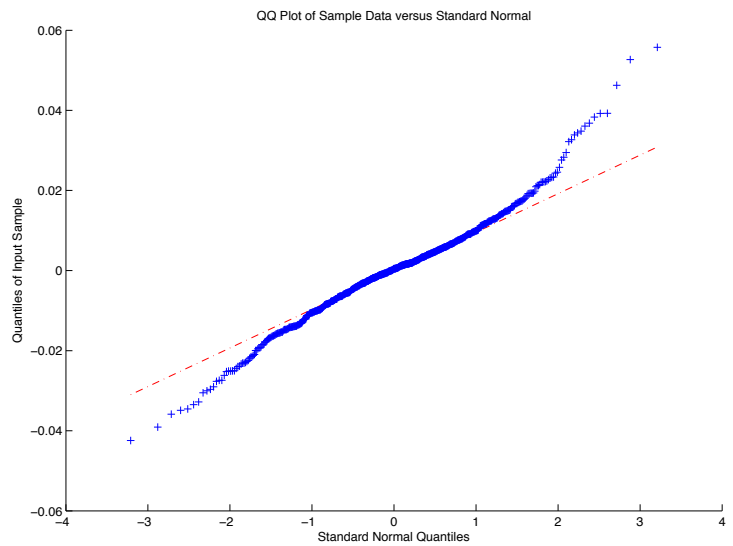
(a) 1988-2002 Normal plot



(b) 2002-2004 Normal plot



(c) 1988-2002 QQ plot



(d) 2002-2004 QQ plot

Figure A.10: Top: 1988-2002, 2002-2004 Normal plots of S&P 500 log returns.
Bottom: 1988-2002 and 2002-2004 QQ plots of S&P 500 log returns.

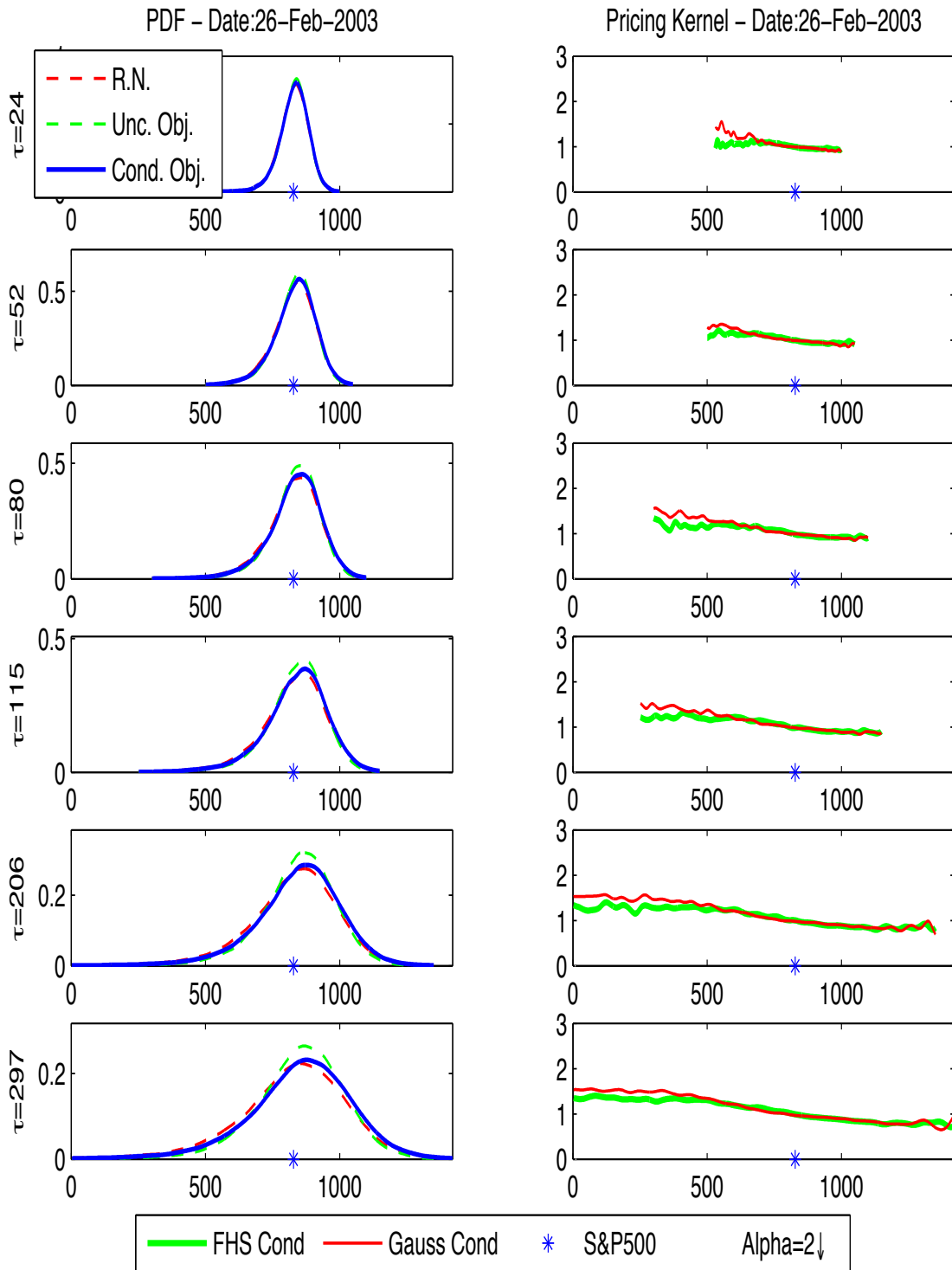


Figure A.11: Single day ($t = 61$) all range of times-to-maturity Probability Density Functions (left column) and conditional Empirical Pricing Kernels, $M_{t,t+\tau}^\dagger$ (right column), with starting $\alpha_{t,t+24}^* = 2$ and decreasing and R.P.= 4%.

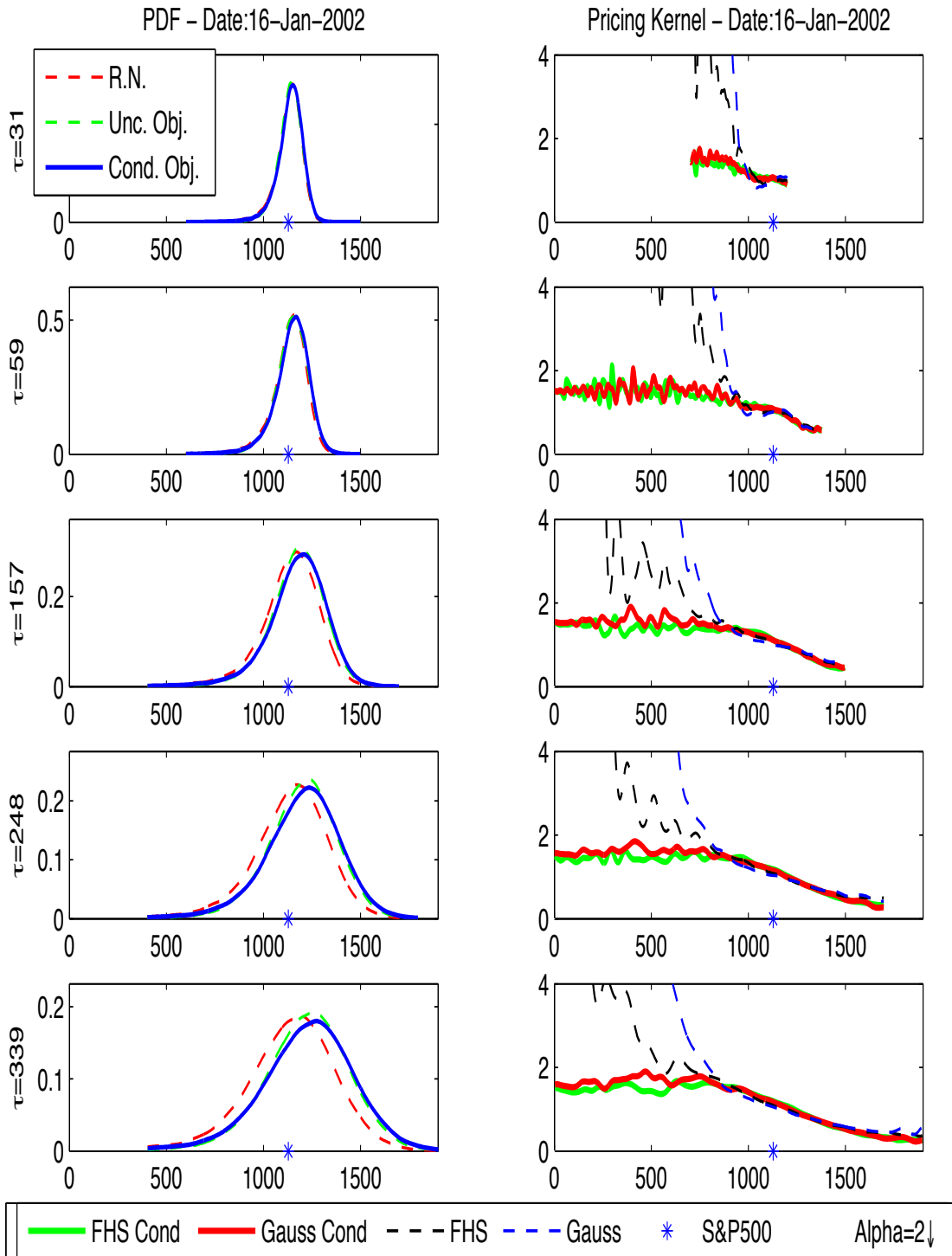


Figure A.12: Single day ($t = 3$) all range of times-to-maturity Probability Density Functions, PDFs, (left column), conditional ($M_{t,t+\tau}^\dagger$), and partially-conditional Empirical Pricing Kernels, EPKs ($M_{t-\Delta t,t+\tau}$), (right column) with starting $\alpha_{t,t+24}^* = 2$ and decreasing and R.P.= 8%. The fully (partially) conditional EPKs are represented with the continuous (dotted) line.

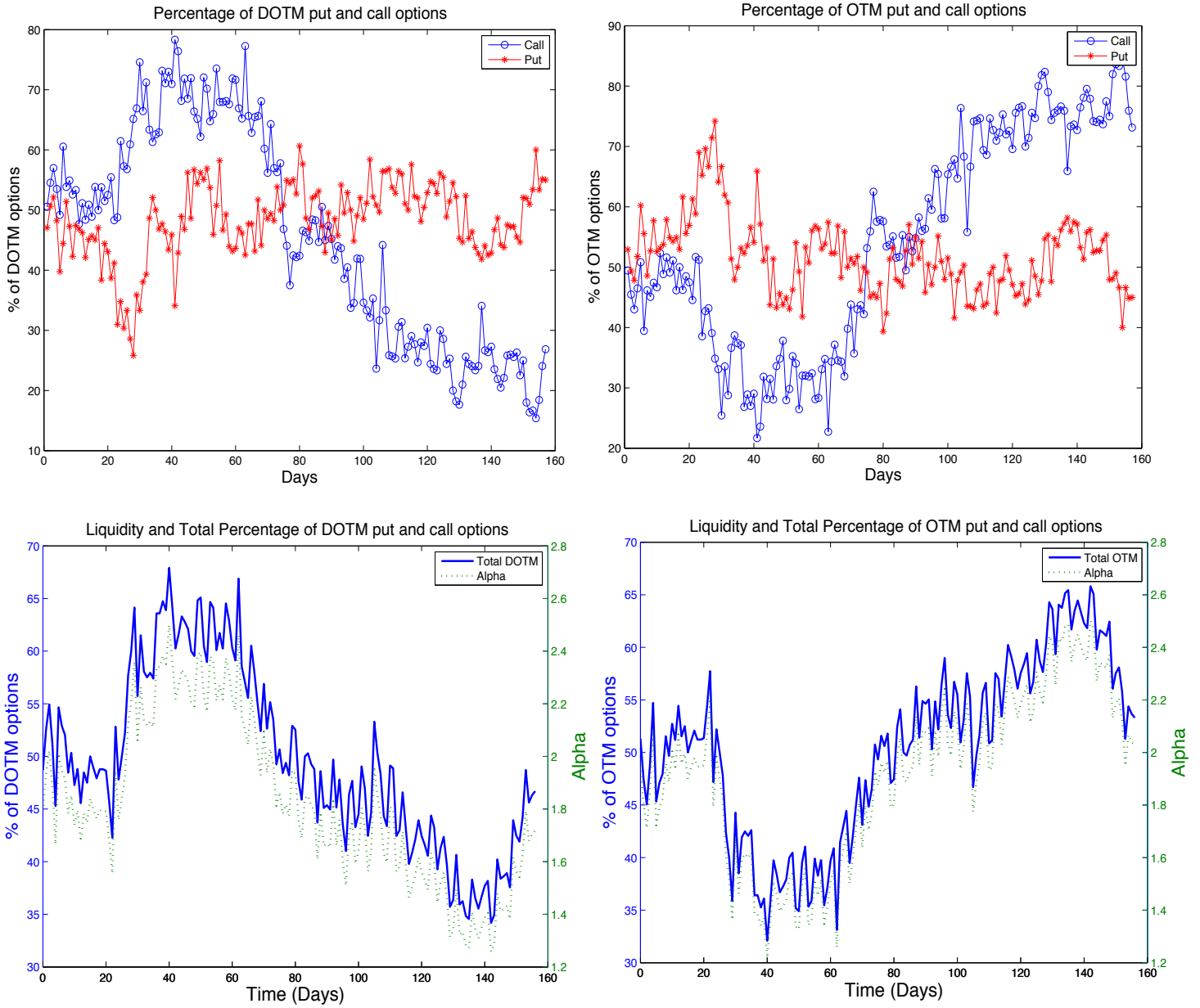


Figure A.13: Top: daily percentage of DOTM (left) and OTM (right) call and put options. Call(put) options are represented with a blue (red) line and with a circle (star) at each edge. Bottom: total fraction of DOTM (left) and OTM (right) options out of the total amount of liquidity:

$$\text{Daily OTM}_t = \frac{\kappa_t(\text{OTM})}{\text{Total}_t} = \frac{\kappa_t(O)_C + \kappa_t(O)_P}{\text{Total}_t} = \frac{\text{Daily OTM Calls}_t + \text{Daily OTM Puts}_t}{\text{Total Liquidity}_t}$$

where $\kappa_t(\cdot)$ represents the options moneyness. The same holds for DOTM options.

Linked to the option market there is the relative $\alpha_{t,T}^*$ computed inverting $\vartheta_t = \frac{\alpha_t}{\max_t \left(\frac{\kappa(\cdot)_t}{\text{Total}_t} \right)}$. The

former are represented by continuous blue line and use the left axis scale, the latter by the green dotted line and use the right scale.

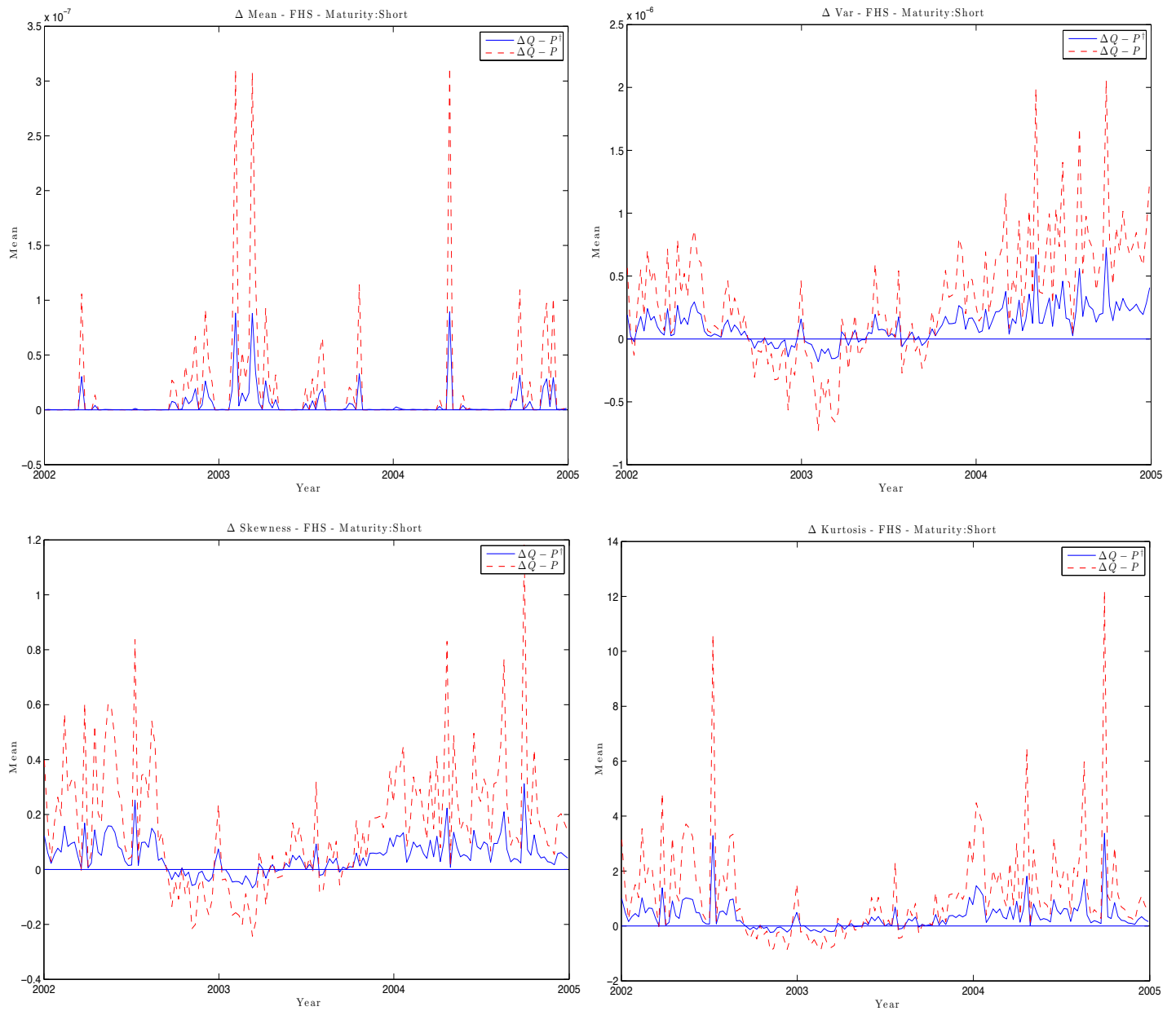


Figure A.14: Daily Δ_t moments between the fully and partially-conditional physical measure with respect to the risk-neutral measure for short time-to-maturity ($\tau < 60$) with $\alpha_{t,T}^* = 2.5$ and R.P.=4%. Top: Δ_t Mean and Δ_t Variance. Bottom: Δ_t Kurtosis and Δ_t Skewness.

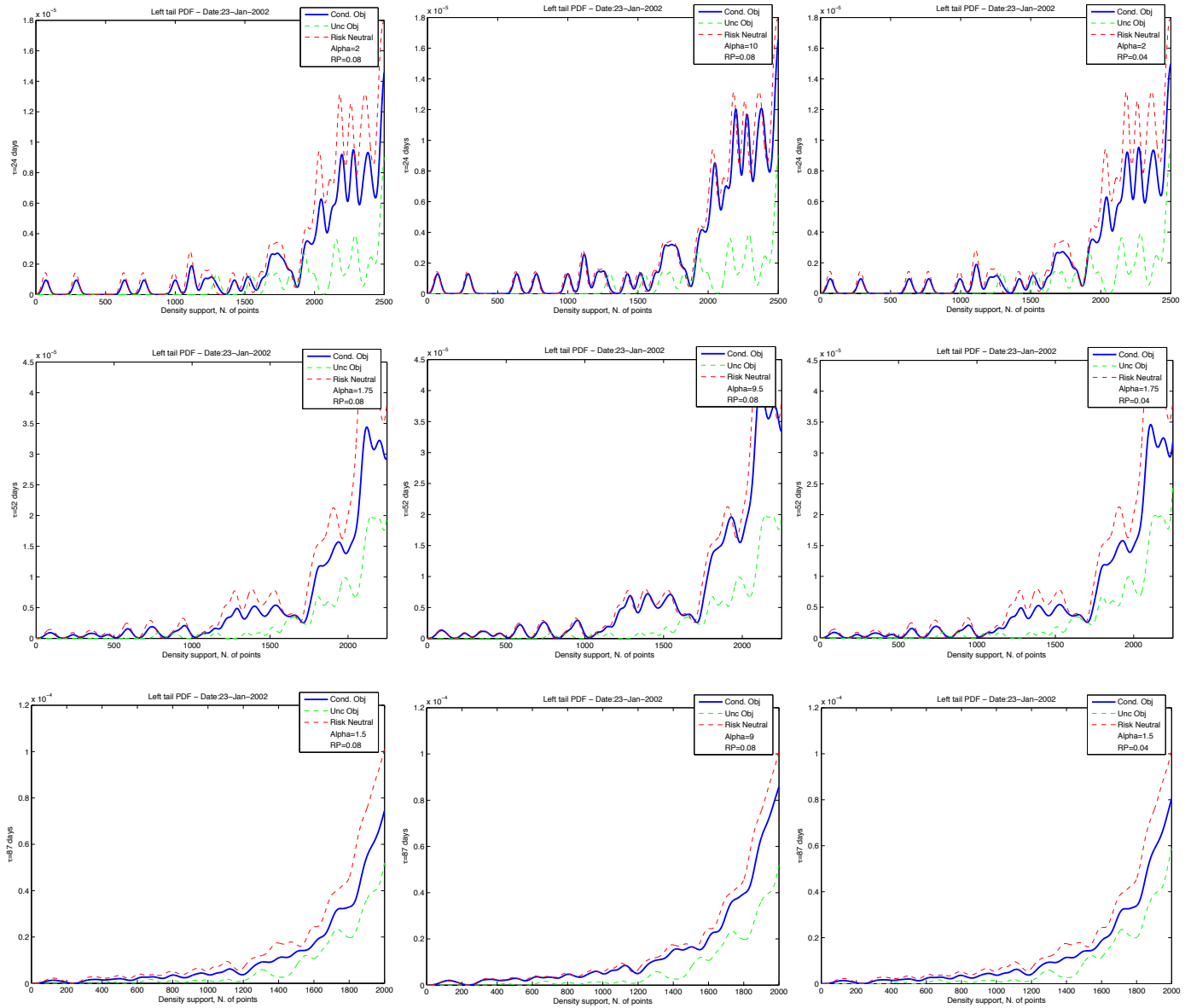


Figure A.15: Single day ($t = 4$), short and medium times-to-maturity ($\tau = 24/57/82$) focus on the left tails of the distributions: partially-conditional physical measure (blue), $p_{t-\Delta_t, t+\tau}$, fully conditional physical measure (black) $p_{t-\Delta_t, t+\tau}^\dagger$ and risk-neutral measure (red) $q_{t-\Delta_t, t+\tau}^*$. The first column is calculated with $\alpha_{t, t+24}^* = 2$ and decreasing and R.P.= 8%, the second column with $\alpha_{t, t+24}^* = 10$ and decreasing and R.P.= 8% and the third column with $\alpha_{t, t+24}^* = 2$ and decreasing and R.P.= 4%. All densities are computed non-parametrically using a normal kernel density estimation with 5000 points and optimal bandwidth (Silverman's rule of thumb).

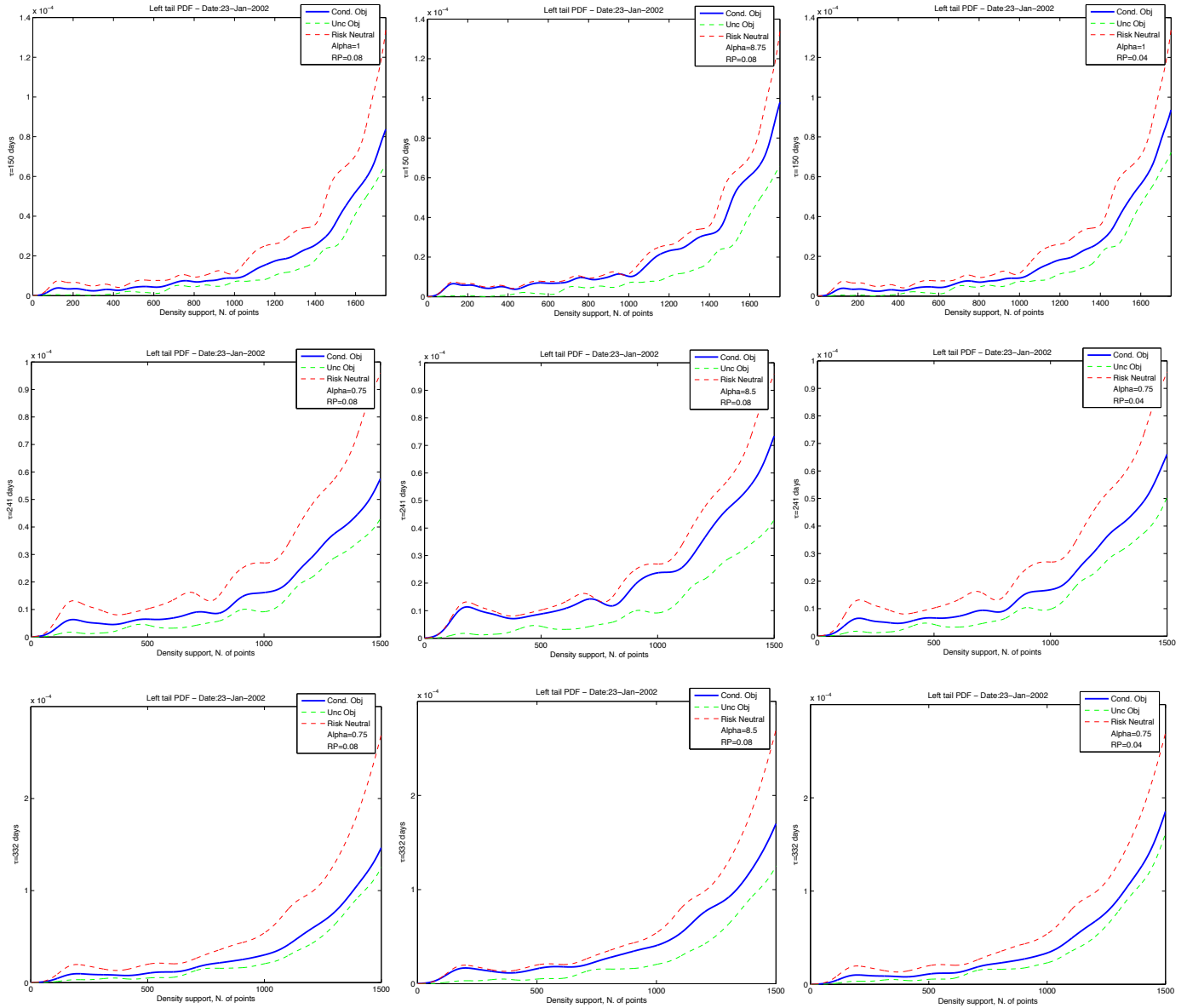


Figure A.16: Single day ($t = 4$), medium and long times-to-maturity ($\tau = 150/241/332$) focus on the left tails of the distributions: partially-conditional physical measure (blue), $p_{t-\Delta_t, t+\tau}$, fully conditional physical measure (black) $p_{t-\Delta_t, t+\tau}^\dagger$ and risk-neutral measure (red) $q_{t-\Delta_t, t+\tau}^*$. The first column is calculated with $\alpha_{t, t+150}^* = 1$ and decreasing and R.P.= 8%, the second column with $\alpha_{t, t+150}^* = 8.75$ and decreasing and R.P.= 8% and the third column with $\alpha_{t, t+150}^* = 1$ and decreasing and R.P.= 4%. All densities are computed non-parametrically using a normal kernel density estimation with 5000 points and optimal bandwidth (Silverman's rule of thumb).

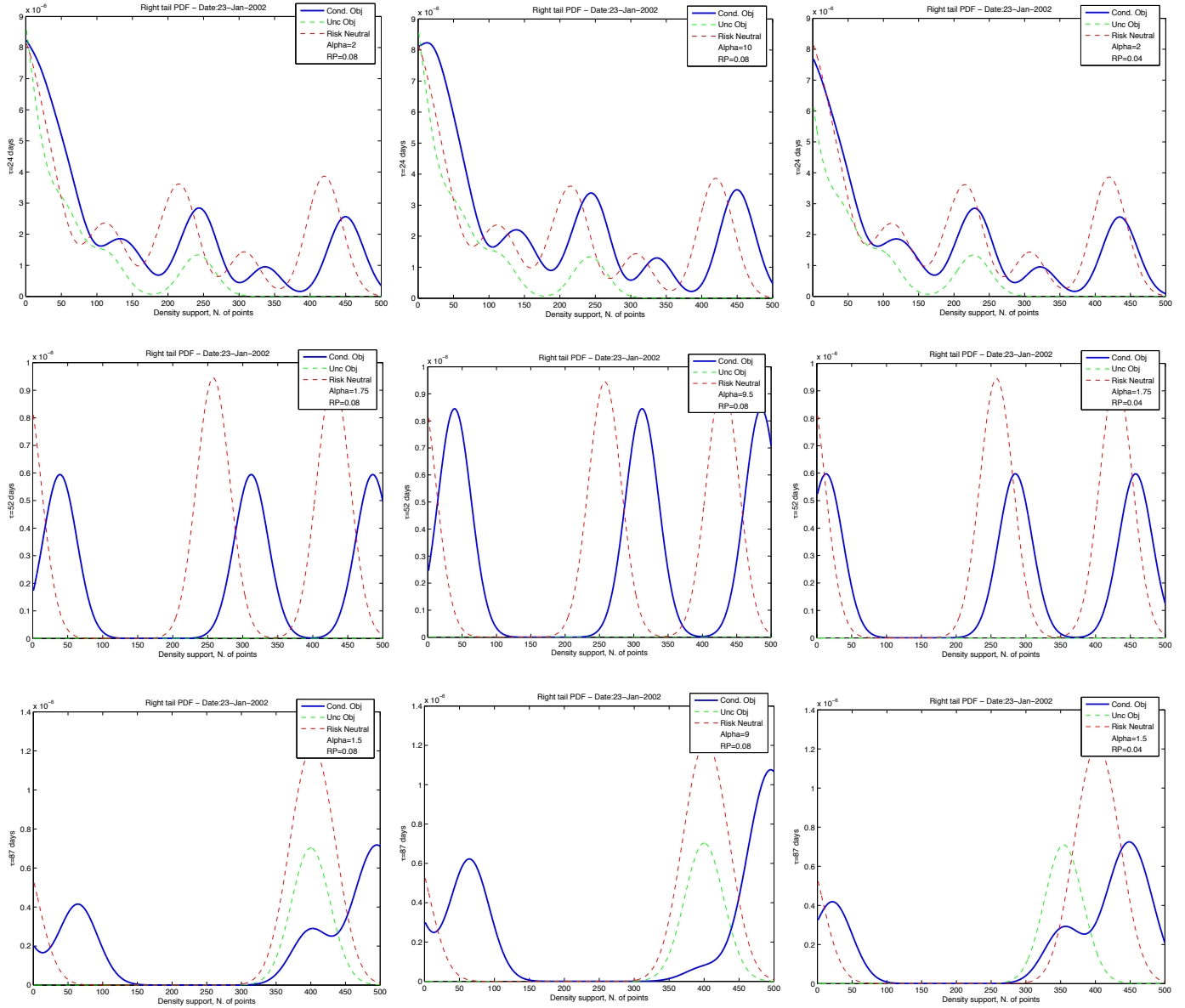


Figure A.17: Single day ($t = 4$), short and medium times-to-maturity ($\tau = 24/57/82$) focus on the right tails of the distributions: partially-conditional physical measure (blue), $p_{t-\Delta_t, t+\tau}$, fully conditional physical measure (black) $p_{t-\Delta_t, t+\tau}^\dagger$ and risk-neutral measure (red) $q_{t-\Delta_t, t+\tau}^*$. The first column is calculated with $\alpha_{t, t+24}^* = 2$ and decreasing and R.P. = 8%, the second column with $\alpha_{t, t+24}^* = 10$ and decreasing and R.P. = 8% and the third column with $\alpha_{t, t+24}^* = 2$ and decreasing and R.P. = 4%. All densities are computed non-parametrically using a normal kernel density estimation with 5000 points and optimal bandwidth (Silverman's rule of thumb).

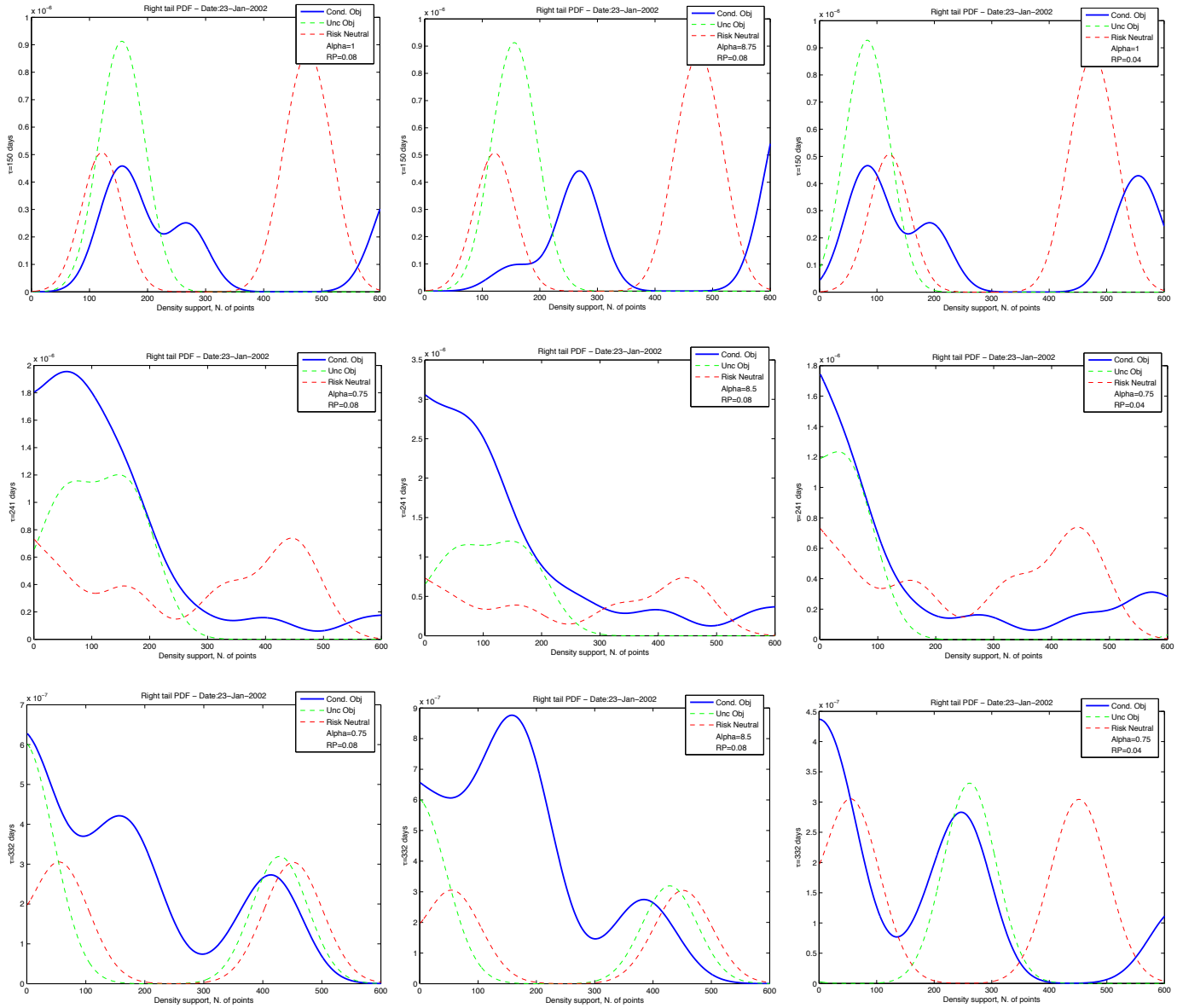


Figure A.18: Single day ($t = 4$), medium and long times-to-maturity ($\tau = 150/241/332$) focus on the right tails of the distributions: partially-conditional physical measure (blue), $p_{t-\Delta_t, t+\tau}$, fully conditional physical measure (black) $p_{t-\Delta_t, t+\tau}^\dagger$ and risk-neutral measure (red) $q_{t-\Delta_t, t+\tau}^*$. The first column is calculated with $\alpha_{t, t+150}^* = 1$ and decreasing and R.P.= 8%, the second column with $\alpha_{t, t+150}^* = 8.75$ and decreasing and R.P.= 8% and the third column with $\alpha_{t, t+150}^* = 1$ and decreasing and R.P.= 4%. All densities are computed non-parametrically using a normal kernel density estimation with 5000 points and optimal bandwidth (Silverman's rule of thumb).

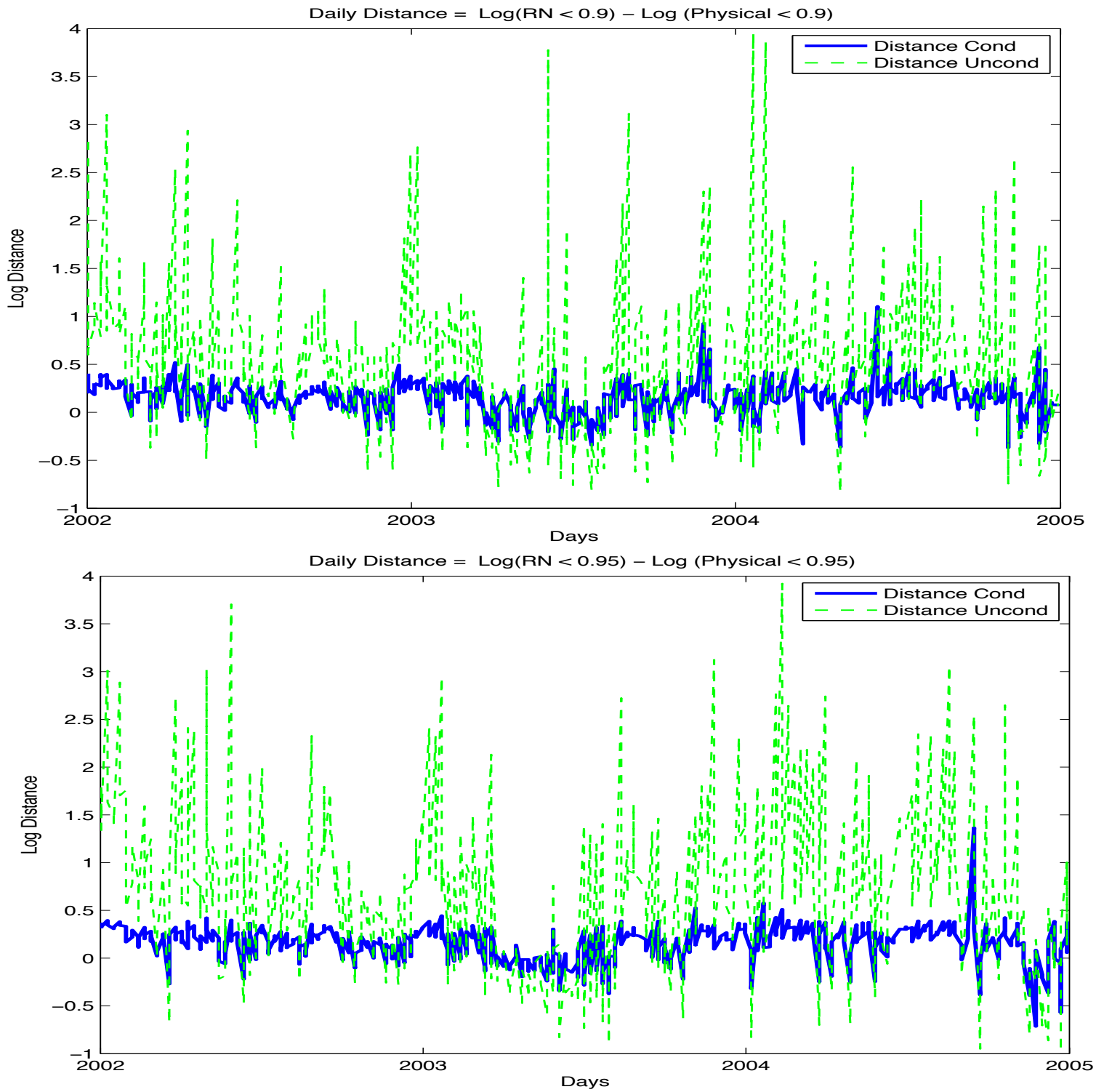


Figure A.19: Daily Log distance, $\log \Delta_t$, of fully and partially-conditional physical measures with respect to the risk-neutral measure for the first decile (Top) and for the first half decile (Bottom). The time series represent all the 862 daily estimations thus covering the entire time series for all times-to-maturity. All values are calculated with $\alpha_{t,T}^* = 2$ and R.P.=4%.

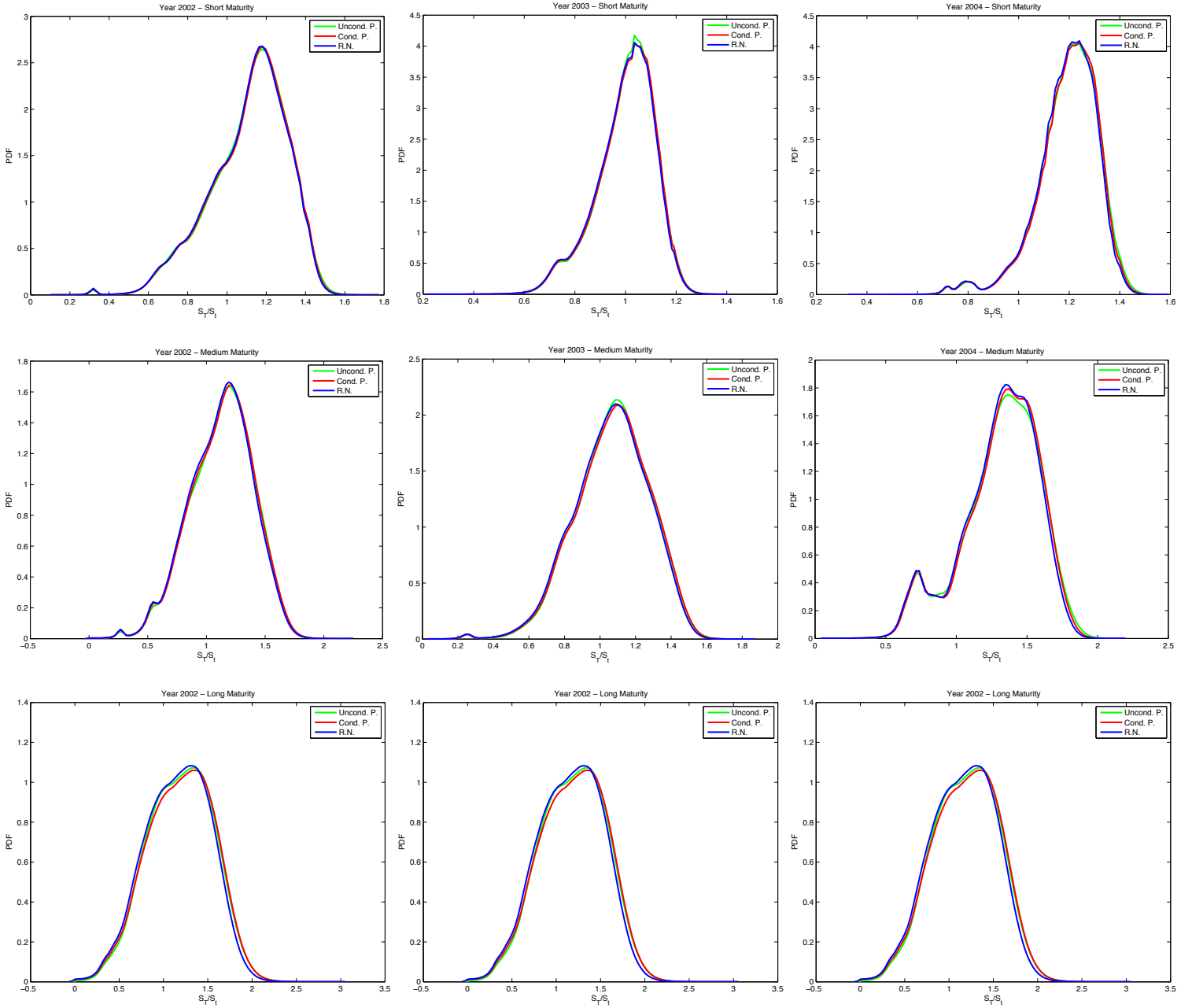


Figure A.20: 2002-2003-2004 yearly estimates of the FHS partially-conditional physical measure (green), $p_{t-\Delta_t, T}^{year}$, conditional physical measure (red) $p_{t, T}^{\dagger, Year}$ and risk-neutral measure (blue) $q_{t, T}^{*, year}$ with $\alpha_{t, T}^* = 2$ and R.P. = 4%. Graphs are divided horizontally in short-medium and long time-to-maturity and vertically for years.

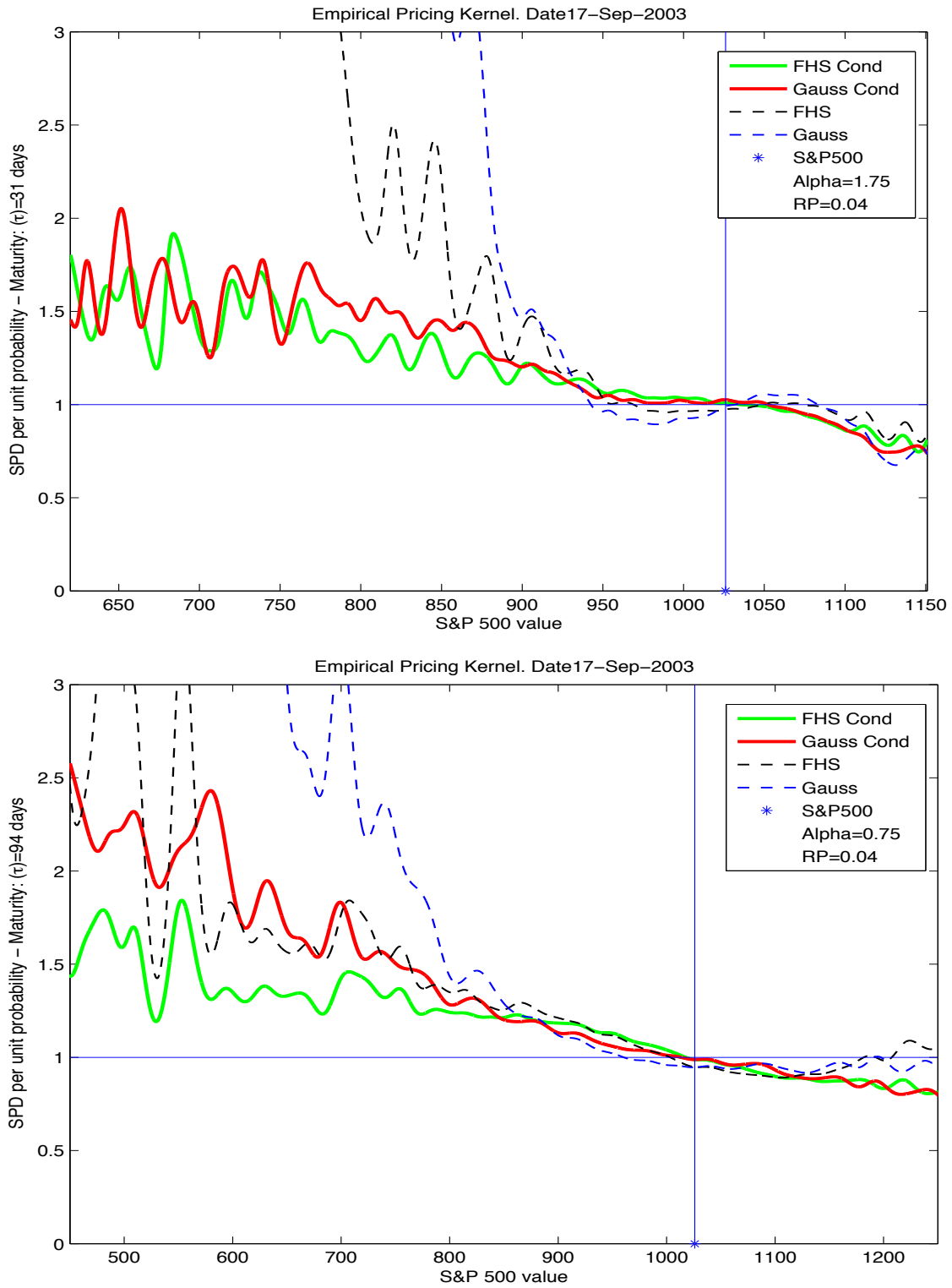


Figure A.21: Top: Single day ($t = 90$) single time-to-maturity ($\tau = 31$) conditional $M_{t,t+31}^\dagger$ and partially-conditional $M_{t-\Delta_t,t+31}$ Empirical Pricing Kernels, EPKs, with $\alpha_{t,t+31}^* = 1.75$ and R.P.=4%

Bottom:Single day ($t = 90$) single time-to-maturity ($\tau = 94$) conditional $M_{t,t+94}^\dagger$ and partially-conditional $M_{t-\Delta_t,t+94}$ Empirical Pricing Kernels, EPKs, with $\alpha_{t,t+94}^* = 0.75$ and R.P.=4%.

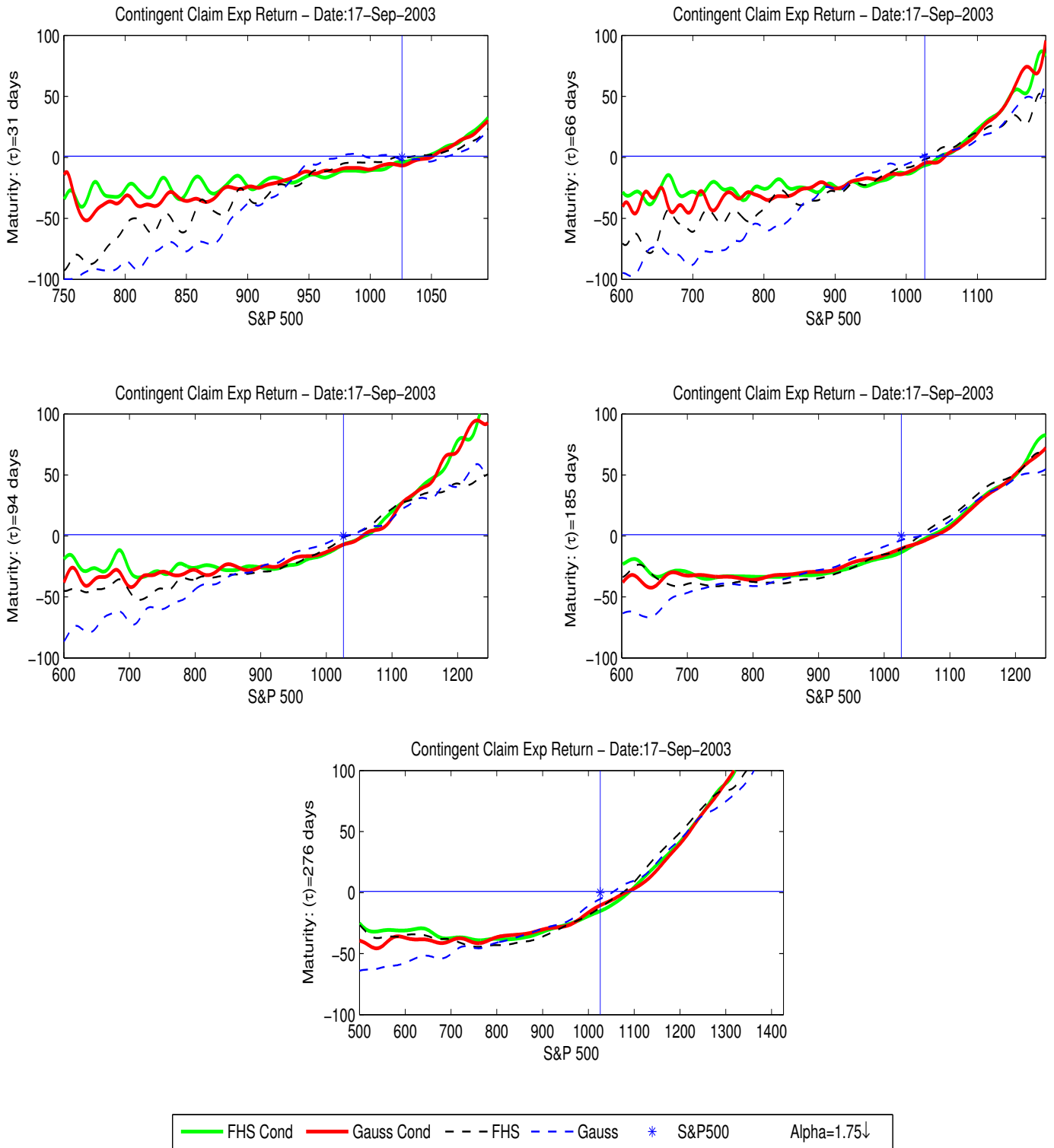


Figure A.22: Single day ($t = 90$) all range of times-to-maturity fully and partially-conditional expected return $v_{t,t+\tau}^\dagger$, $v_{t-\Delta t,t+\tau}$ from investing in a generic contingent claim estimated with $\alpha_{t,t+31}^* = 1.75$ and decreasing and R.P.=4%.

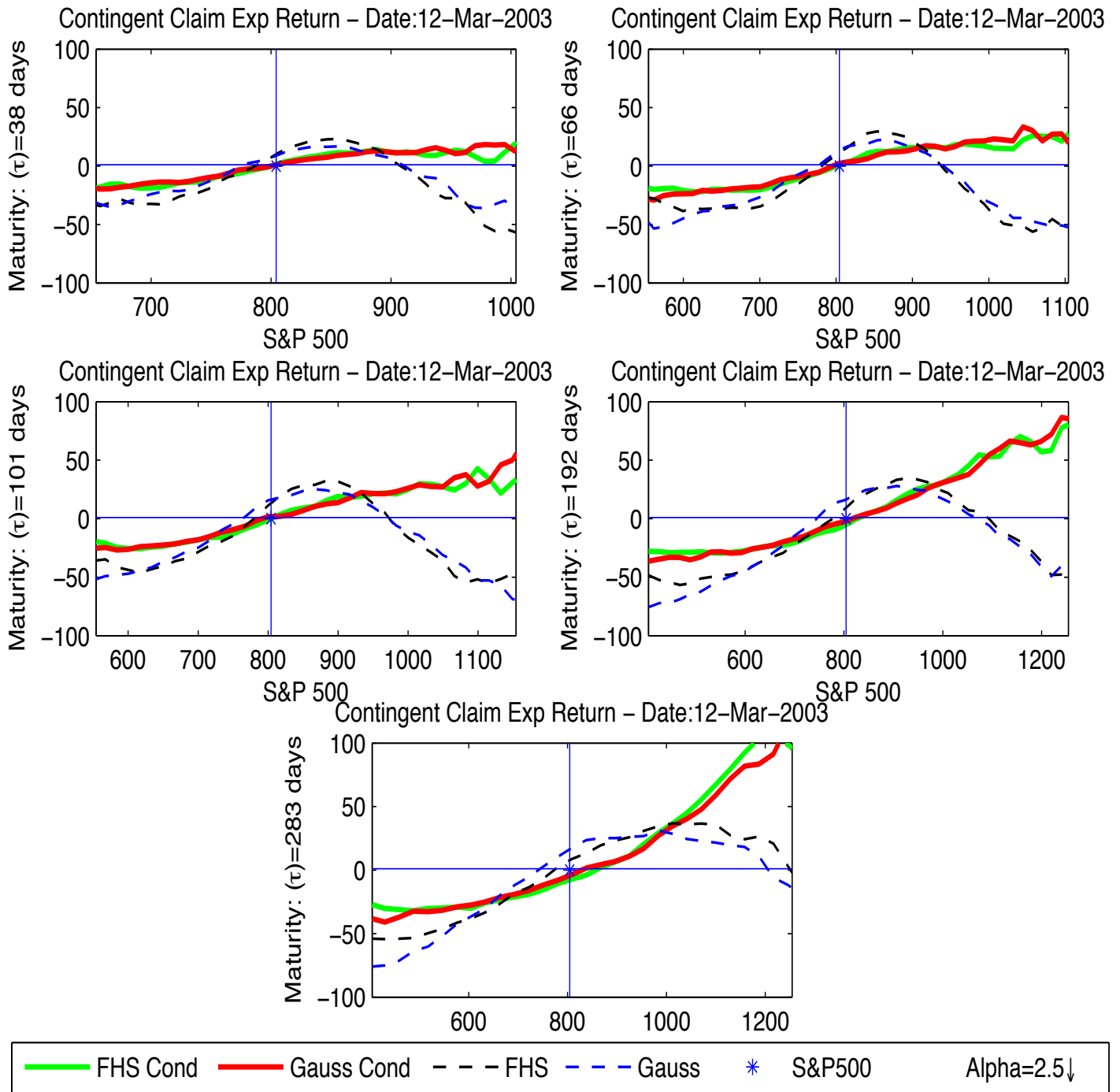


Figure A.23: Single day ($t = 63$) all range of times-to-maturity fully and partially-conditional expected return $v_{t,t+\tau}^\dagger$, $v_{t-\Delta t,t+\tau}$ from investing in a generic contingent claim estimated with $\alpha_{t,t+38}^* = 2.5$ and decreasing and R.P.=8%.

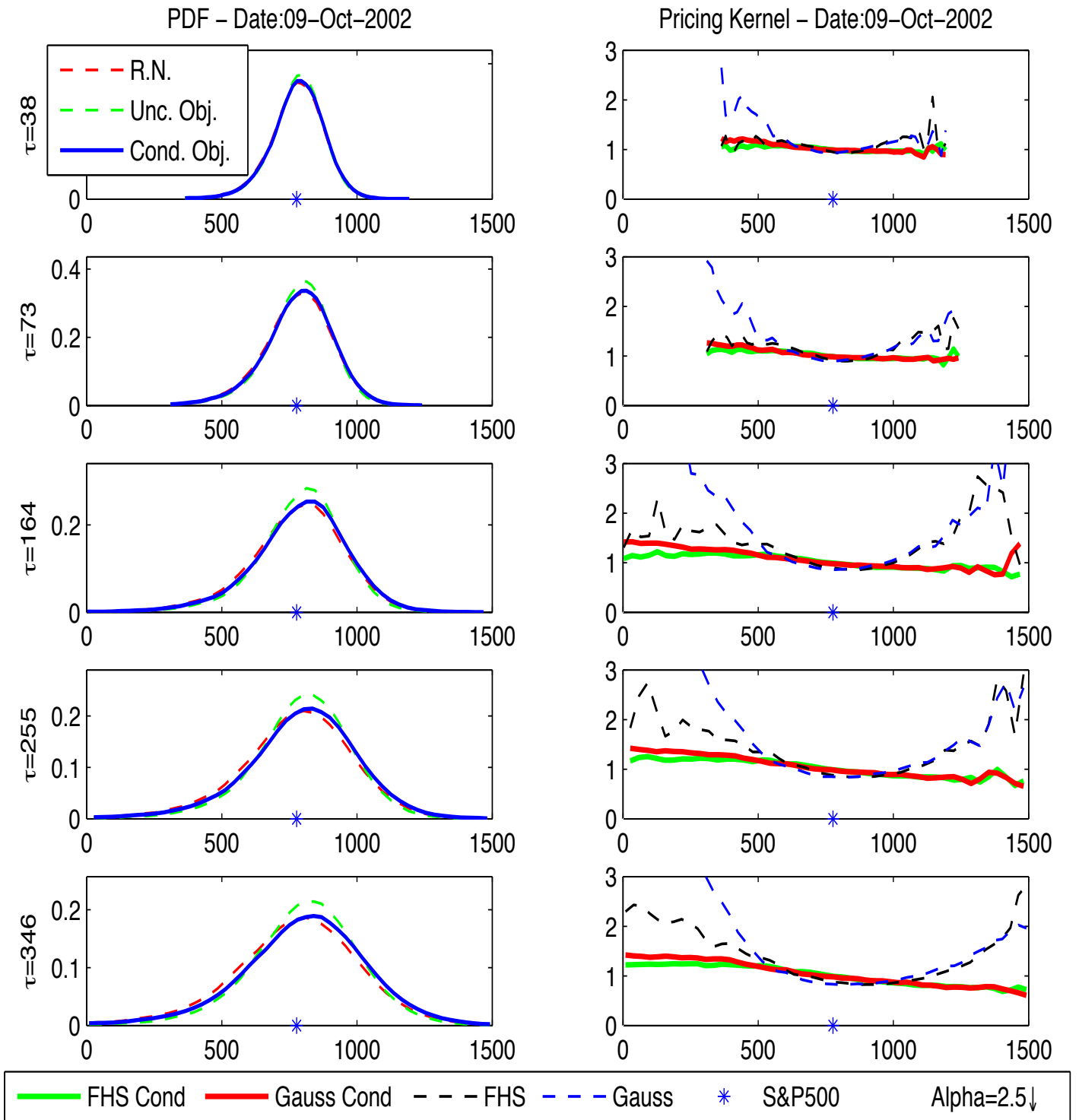


Figure A.24: Single day ($t = 41$) all range of times-to-maturity Probability Density Functions (left column), conditional and partially-conditional EPKs (right column) with $\alpha_{t,t+38}^* = 2.5$ and decreasing and R.P.= 4%

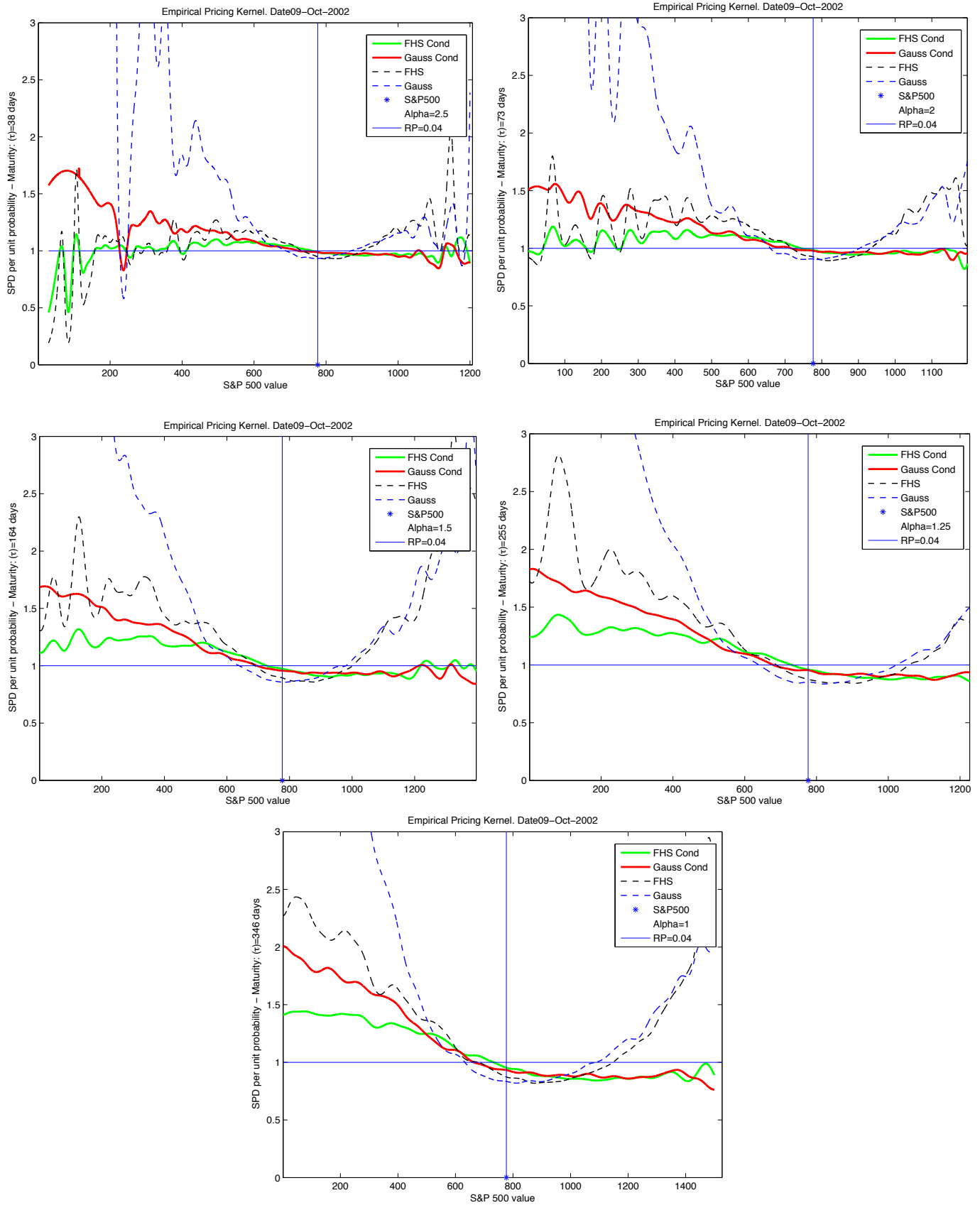


Figure A.25: Focus on the single day ($t = 41$) and all times-to-maturity ($\tau= 38/73/164/255/346$) conditional Empirical Pricing Kernels, EPKs, $M_{t,t+\tau}^\dagger$, and partially-conditional Empirical Pricing Kernels, EPKs, $M_{t-\Delta,t+\tau}^*$, with $\alpha_{t,t+38}^* = 2.5$ and decreasing and R.P.= 4%.

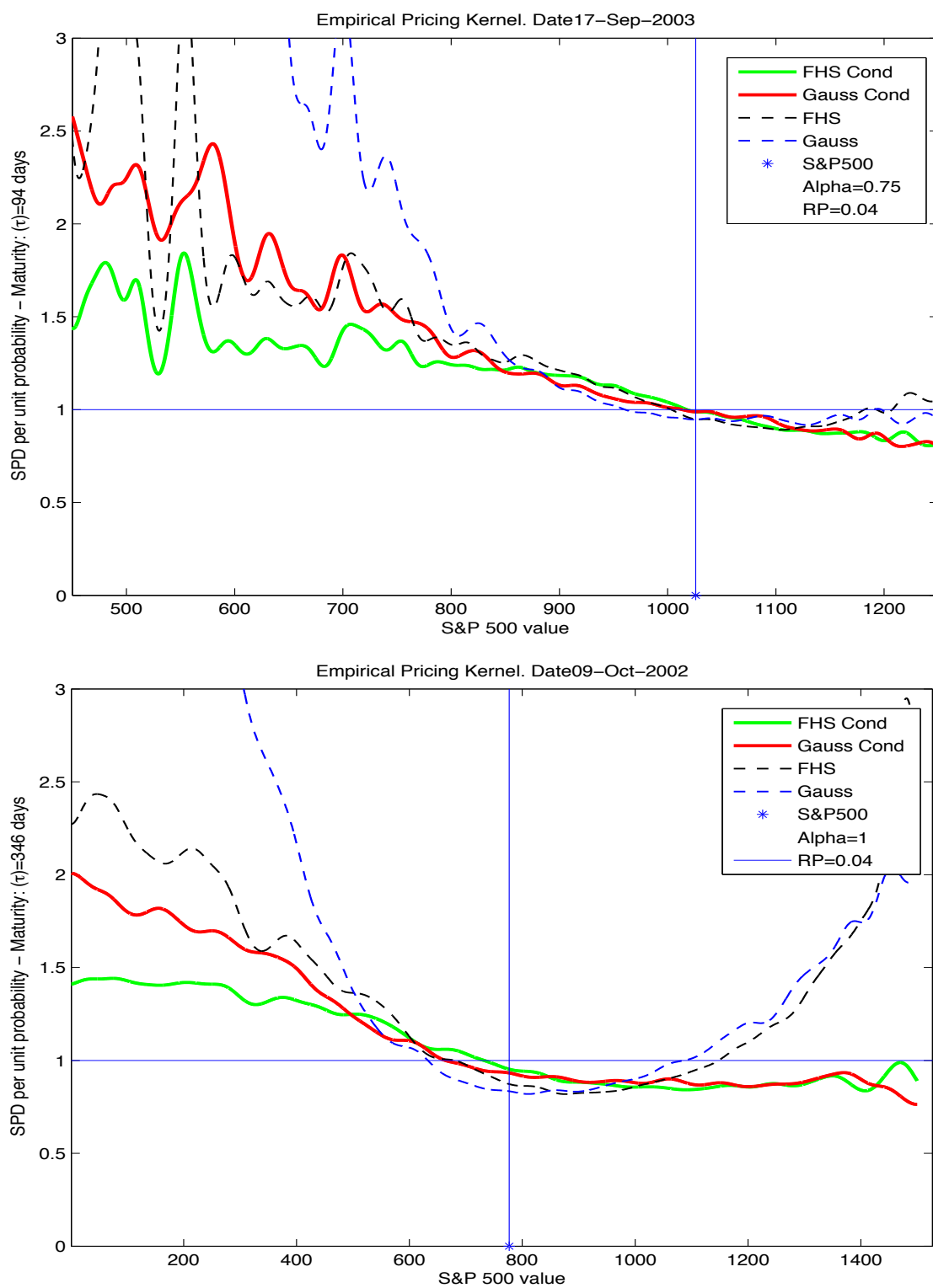


Figure A.26: Top: Single day (September 13, 2003) single time-to-maturity ($\tau = 94$) conditional, $M_{t,t+94}$, and unconditional, $\widehat{M}_{t-\Delta,t+94}$, EPKs with $\alpha_{t,t+94}^* = 1.5$. Bottom: Single day (October 9, 2002) and single time-to-maturity ($\tau = 346$) conditional EPKs, $M_{t,t+346}$, and unconditional EPKs, $\widehat{M}_{t-\Delta,t+346}$, with $\alpha_{t,t+346}^* = 1$.

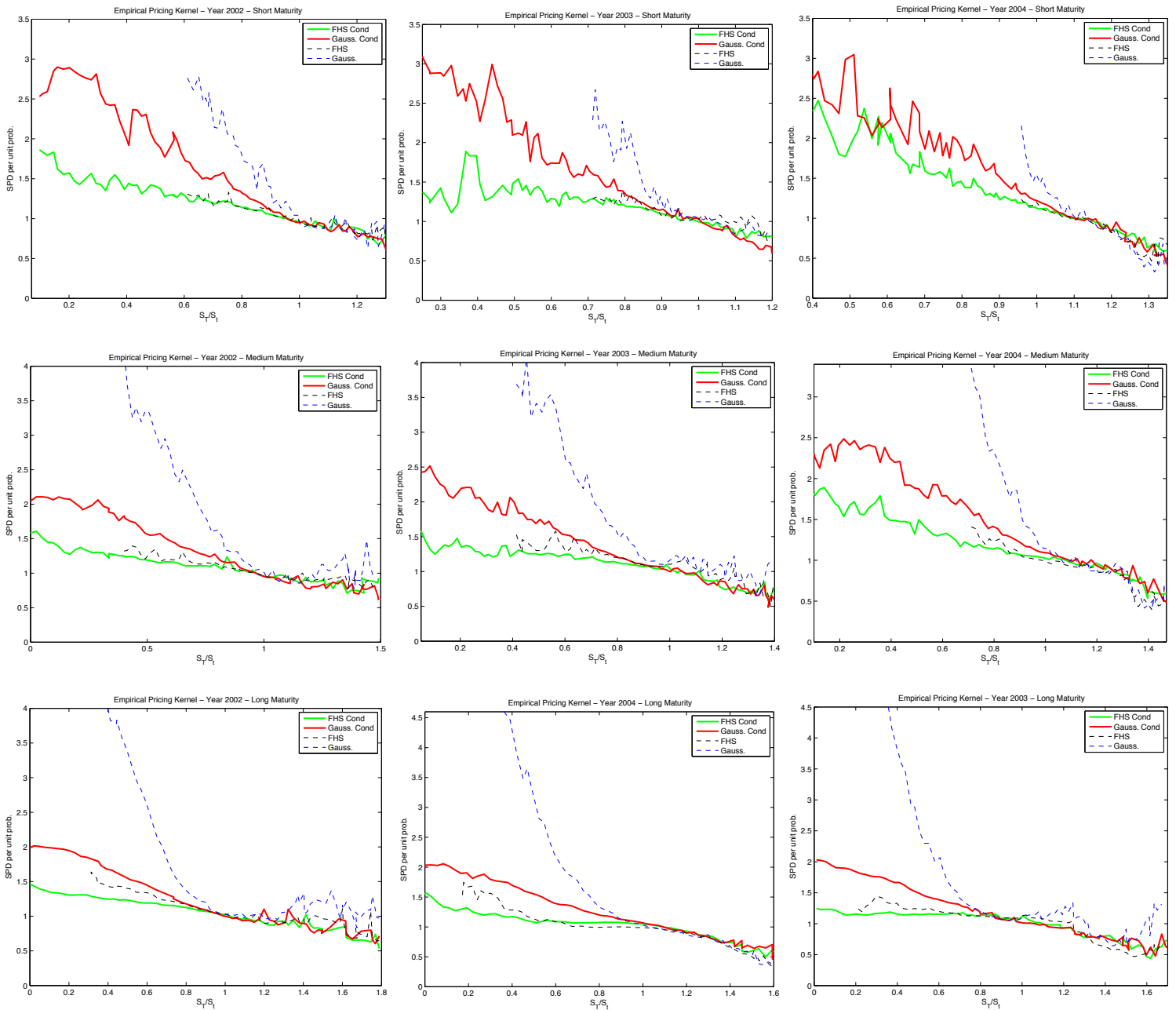
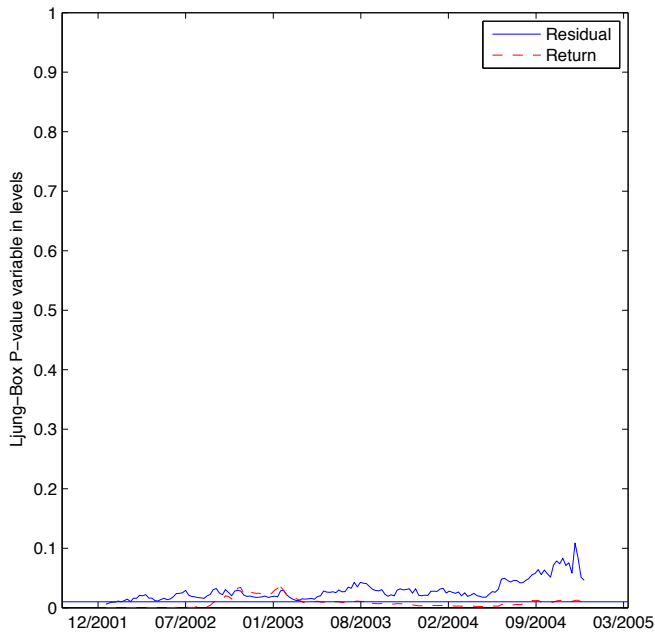
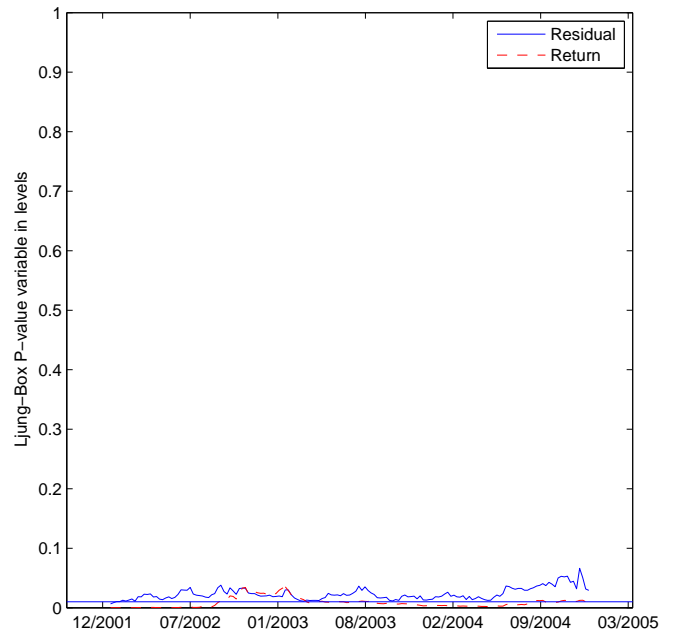


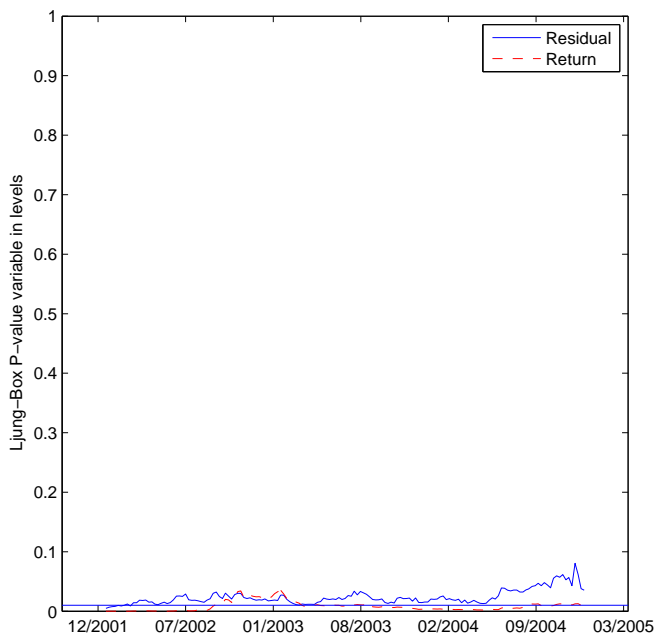
Figure A.27: 2002-2003-2004 yearly estimates of the FHS (green) and Gauss. (red) conditional $M_{t,T}^{\dagger,year}$ and FHS (black dotted) and Gauss. (blue dotted) conditional $M_{t-\Delta t,T}^{year}$ with $\alpha_{t,T}^* = 1$ and R.P.= 4%. Single graphs are divided horizontally in short-medium and long times-to-maturity and vertically for years.



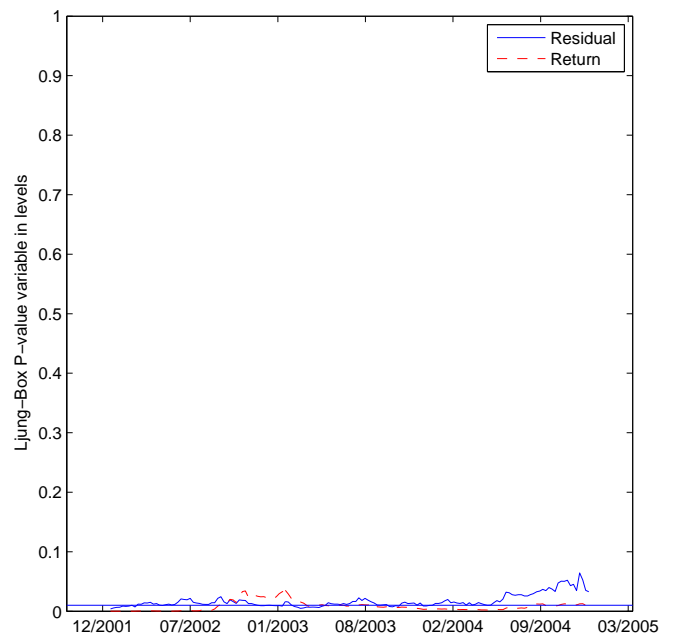
(a) 2000 observations



(b) 3500 observations

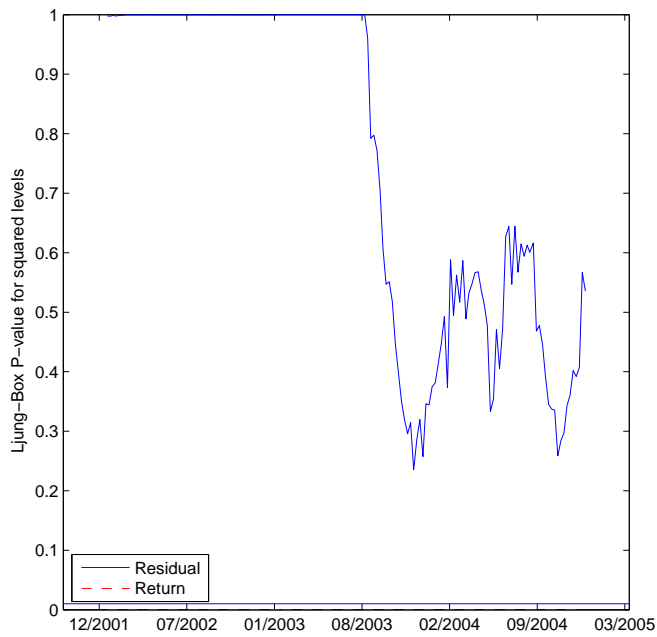


(c) 5000 observations

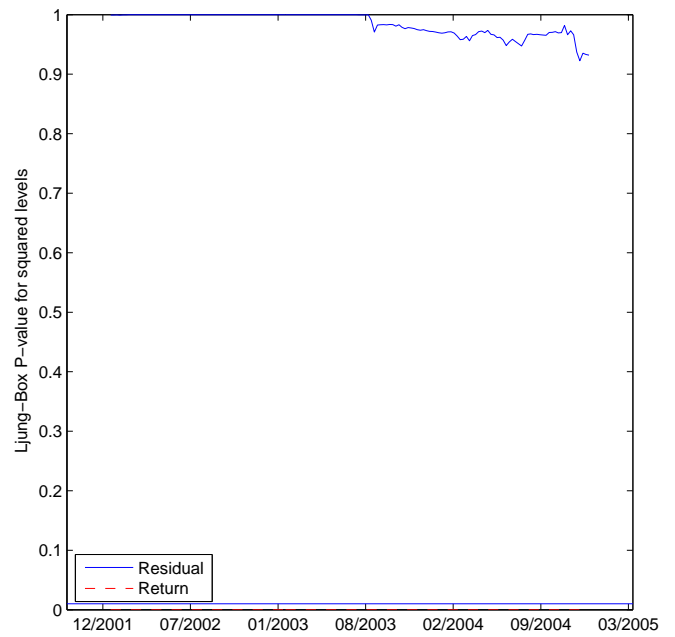


(d) 9818 observations

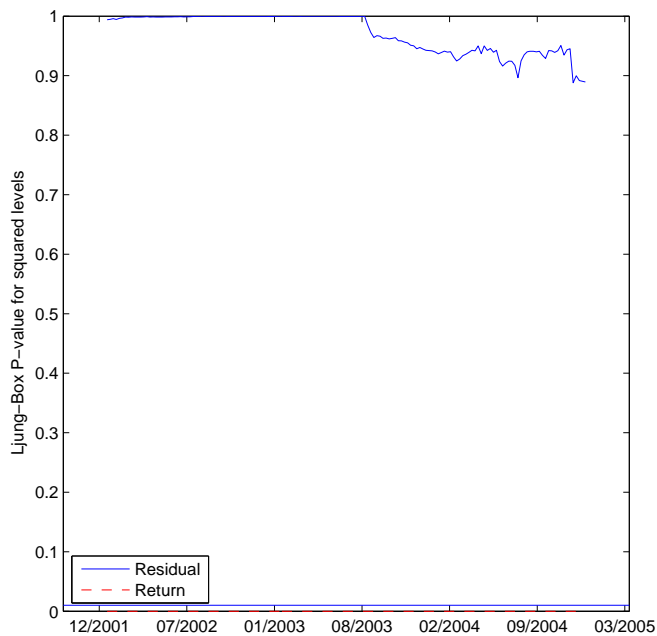
Figure A.28: Portmanteau test of Ljung and Box. The vertical line represents the p-value of the Ljung-Box test for 2000/3500/5000/9818 observations applied to daily S&P 500 returns (red dash line) and to the past squared standardized residuals (blue solid line) for each Wednesday of the time series (157 observations). The blue line on the p-value axe represents the 1% level.



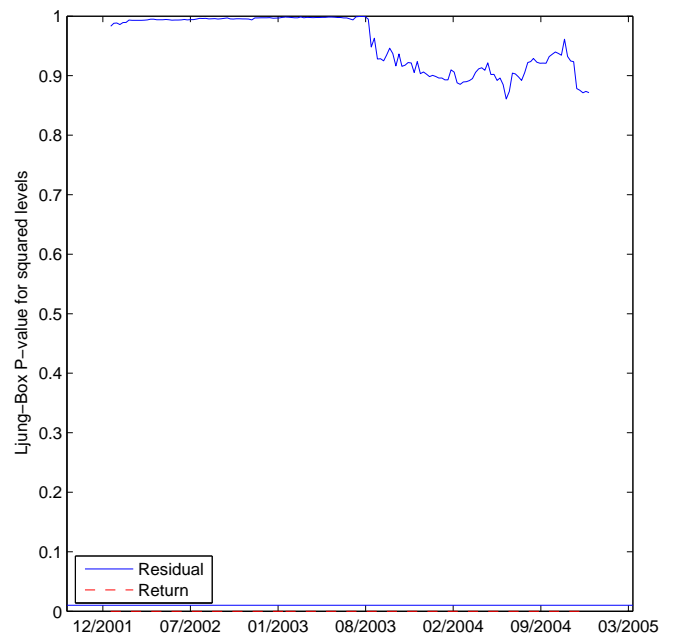
(a) 2000 observations



(b) 3500 observations

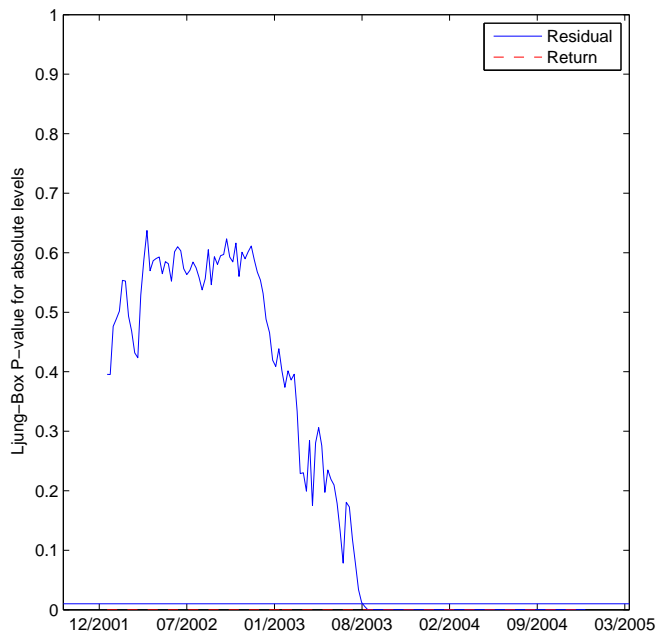


(c) 5000 observations

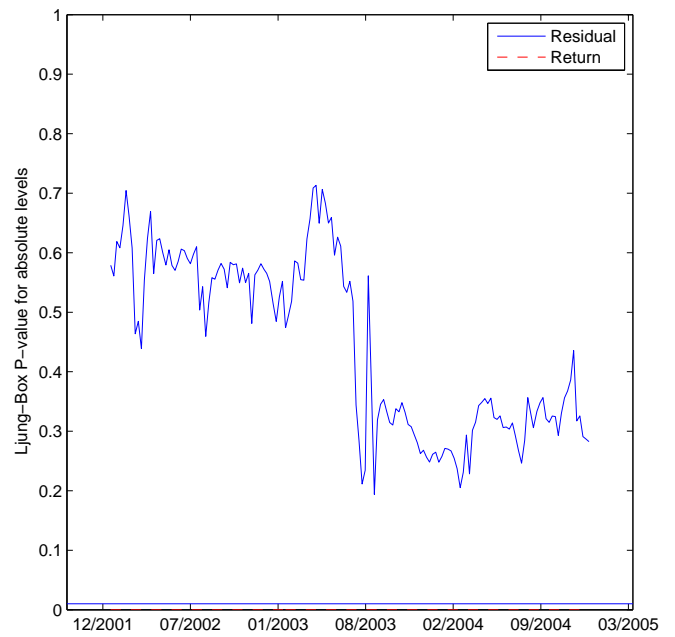


(d) 9818 observations

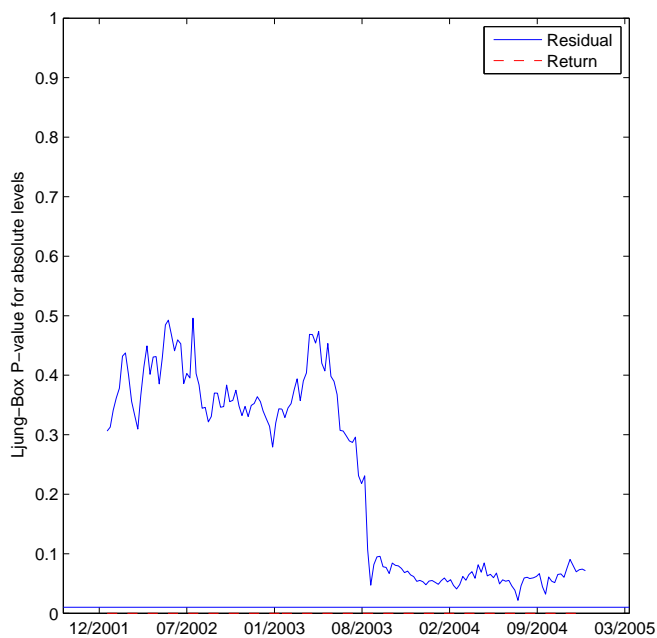
Figure A.29: Portmanteau test of Ljung and Box. The vertical line represents the p-value of the Ljung-Box test for 2000/3500/5000/9818 observations applied to daily S&P 500 squared returns (red dash line) and to the past standardized squared residuals (blue solid line) for each Wednesday of the time series (157 observations). The blue line on the p-value axe represents the 1% level.



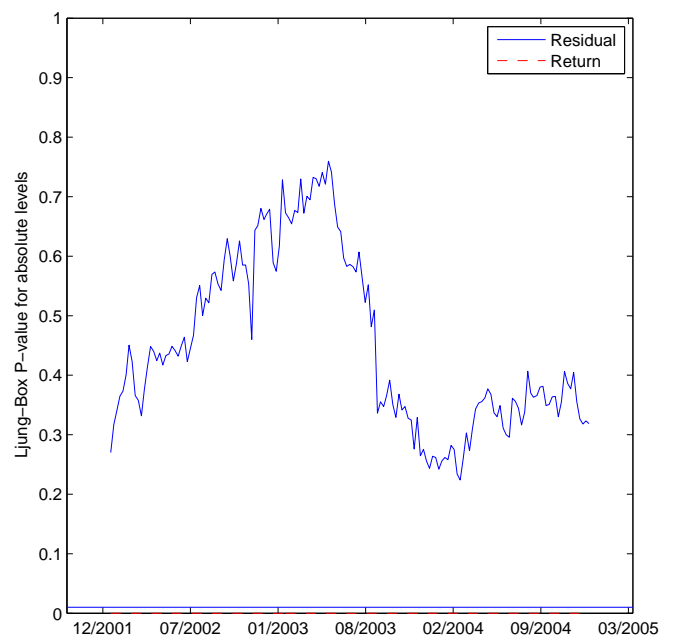
(a) 2000 observations



(b) 3500 observations

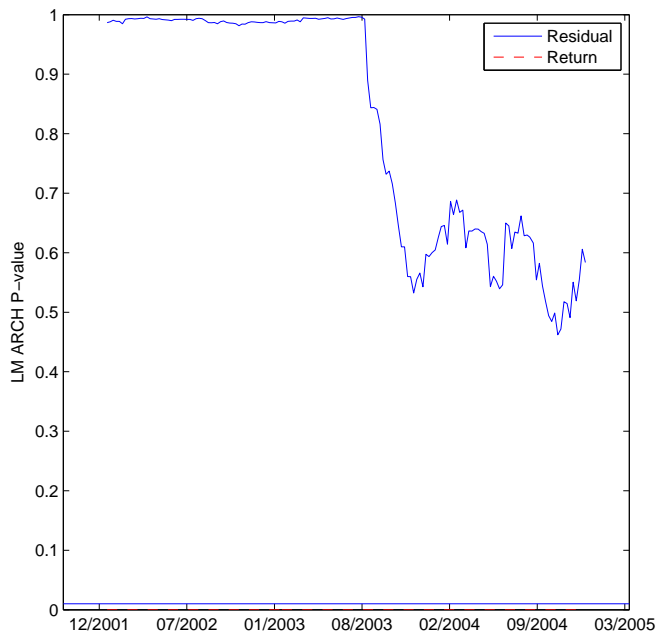


(c) 5000 observations

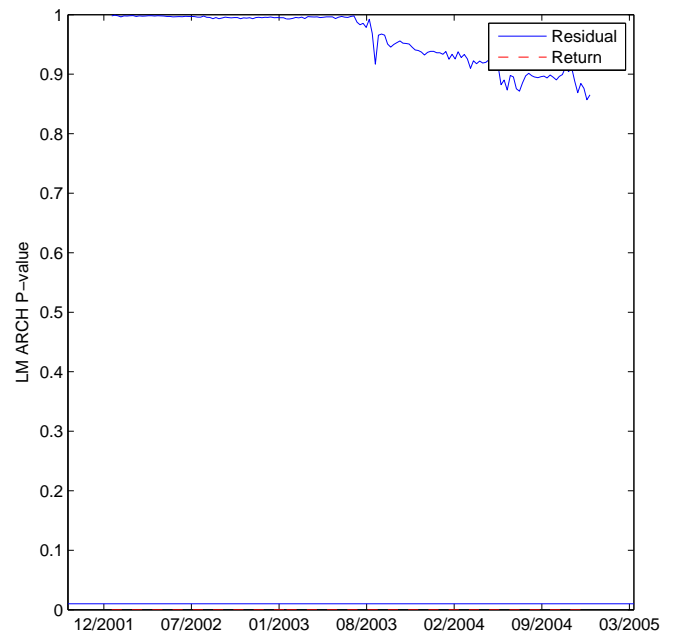


(d) 9818 observations

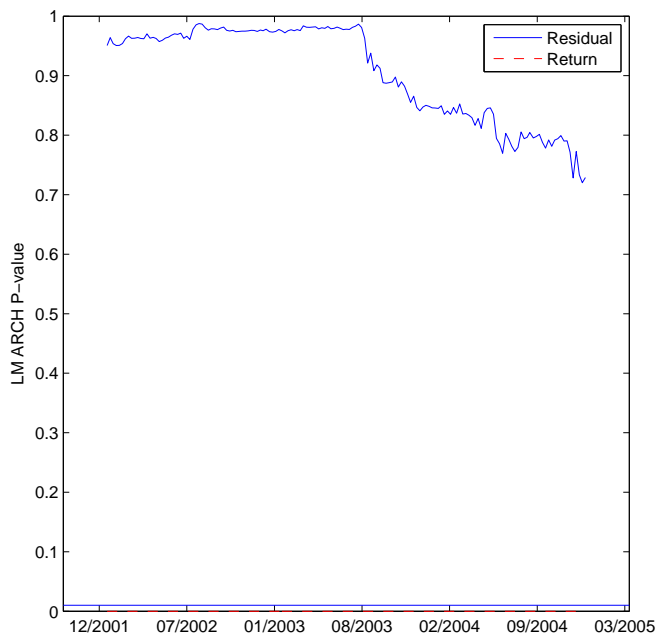
Figure A.30: Portmanteau test of Ljung and Box. The vertical line represents the p-value of the Ljung-Box test for 2000/3500/5000/9818 observations applied to daily S&P 500 absolute returns (red dash line) and to the past squared standardized residuals (blue solid line) for each Wednesday of the time series (157 observations). The blue line on the p-value axe represents the 1% level.



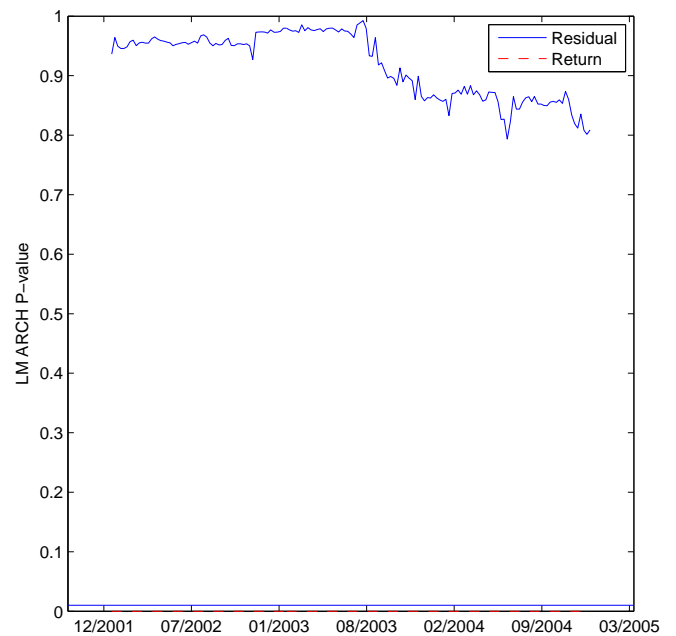
(a) 2000 observations



(b) 3500 observations

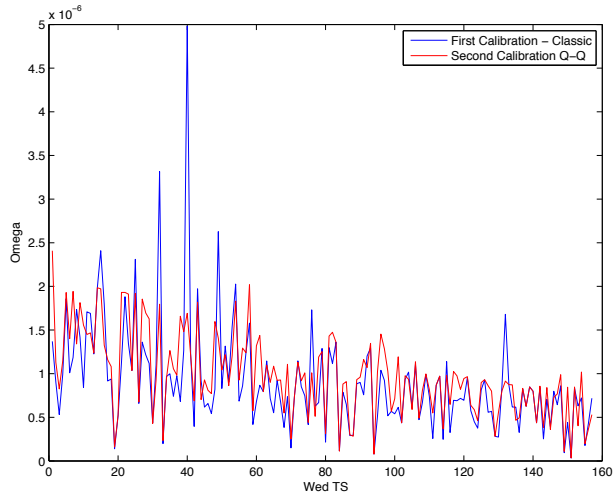


(c) 5000 observations

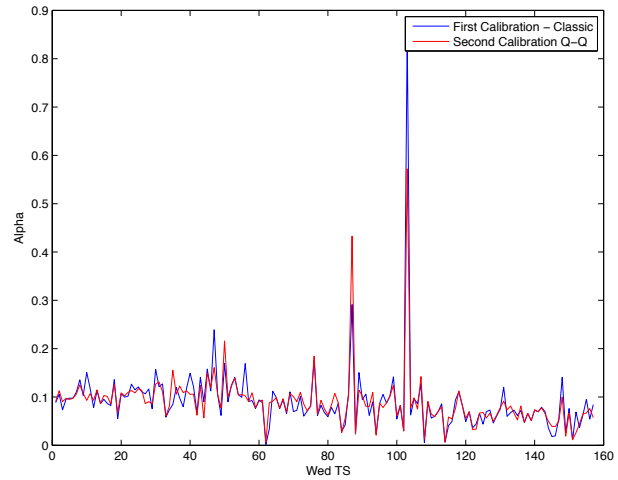


(d) 9818 observations

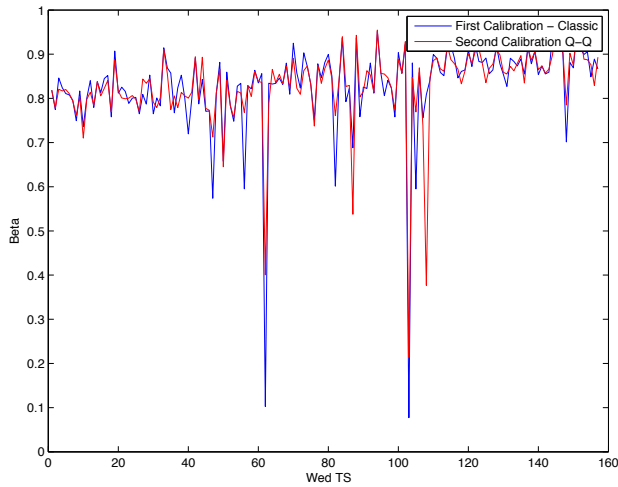
Figure A.31: Lagrange Multiplier (MA) ARCH test. The vertical line represents the p-value of a 10 legs LM ARCH test for 2000/3500/5000/9818 observations applied to daily S&P 500 returns (red dash line) and to the past standardized innovations (blue solid line) for each Wednesday of the time series (157 observations). The blue line on the p-value axe represents the 1% level.



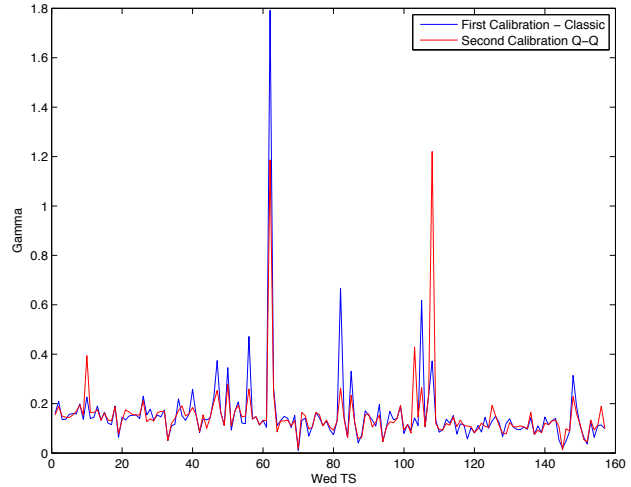
(a) omega



(b) alpha



(c) beta



(d) gamma

Figure A.32: First (blue line) and second (red line) set of parameters (θ) (in clockwise order: ω , α , β , γ) estimations obtained by fitting the semi-synthetic dataset onto the GJR GARCH model.

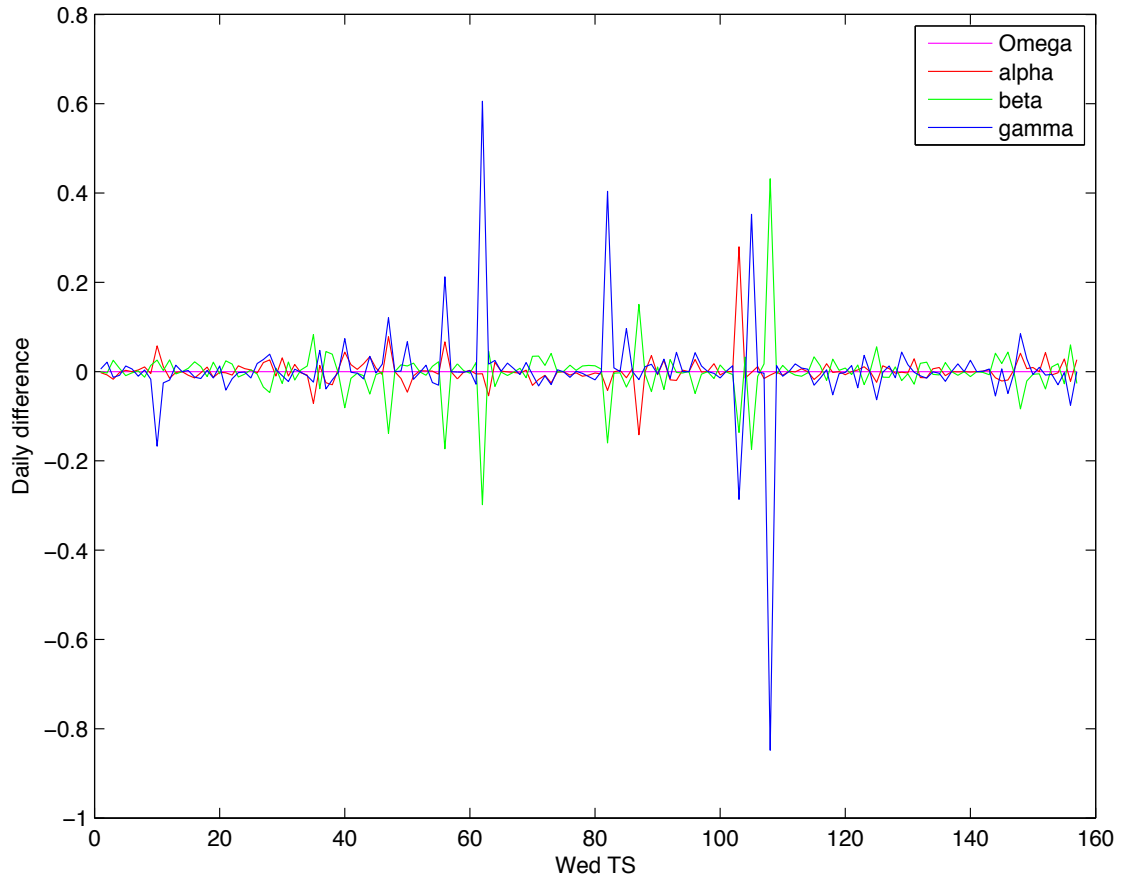


Figure A.33: Pointwise daily parameters (θ) difference between the second and first estimation.