The Pricing Kernel Anomaly: the Case
of the Information that did not Bark

Carlo Sala*  Giovanni Barone-Adesi†
Swiss Finance Institute  Swiss Finance Institute
and University of Lugano  and University of Lugano

September 30, 2015

---

*Carlo Sala, Institute of Finance, University of Lugano, Via G. Buffi 13, CH-6900 Lugano, Switzerland. E-mail: carlo.sala@usi.ch.
†Giovanni Barone-Adesi, Institute of Finance, University of Lugano, Via G. Buffi 13, CH-6900 Lugano, Switzerland. E-mail: giovanni.baroneadesi@usi.ch.
The Pricing Kernel Anomaly: the Case
of the Information that did not Bark

Abstract

This paper provides an empirical study on the conditioning of the information in the estimation of the pricing kernel. Although the neoclassical theory requires the pricing kernel to be a monotonically decreasing function, recent empirical studies found several violations: at the extremes as well as in the central part of the functional. Since Jackwerth (2000), this is known as the “pricing kernel puzzle”.

While the risk-neutral moments extracted from option surfaces are by construction forward looking, the ones obtainable from historical returns are only partially informative, thus suboptimal with respect to investors’ future beliefs.

This underestimation of the physical filtration produces a disalignment with respect to the full conditioning of the information set as required by the neoclassical theory. Empirically it turns out that most of papers present in literature are then affected by a non-homogeneity bias.

We propose a new flexible and highly informative non-parametric method to estimate a non-stationary and fully-conditional physical measure. Exploiting the informational content of the implied moments of option prices, the proposed measure embeds the missing forward looking information necessary to produce a time-varying and fully-conditional physical measure.

A natural approach to exploit simultaneously multiple data and provide statistical inference is the Dirichlet process. Using the precision parameter of the Dirichlet process as a proxy for the missing information, we calibrate it with respect to the daily liquidity of the options in the market. The obtained density is a mixture of the two measures and combines the prior forward looking information available from option data with the historical background provided by stock returns. The new measure is then used to investigate the pricing kernel monotonicity.

Keywords: Pricing Kernel, Measure Estimation, Radon-Nikodym Derivative, Monotonicity, Dirichlet process, Option Data
1 Introduction

The main goal of this paper is to propose a new flexible and generic non-parametric method for the estimation of the conditional physical measure which is then used to investigate the pricing kernel (henceforth: PK) monotonicity.

To do it, we 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 paper is that the two measures are estimated independently and 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 not considered in literature.

The proposed fully-conditional model, exploiting its high flexibility and being free of any constraints - but the ones required by the Fundamental Theorem of Asset Pricing (henceforth: FTAP) - improves the overall estimates of the empirical pricing kernel (henceforth: EPK\(^1\)) and answers to puzzles concerning the consistence of the distribution implicit in option prices and the time series properties of the underlying asset prices.

Our line of research is motivated by the recent empirical studies of the PK which, starting from Jackwerth 2000 \cite{32}, produced conflicting and often puzzling results.

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 and the EPK.

Mathematically, in a continuous world with no-arbitrage\(^2\), the time \(t\) PK, \((M_t)\), is defined as the Radon-Nikodym derivative of the risk-neutral measure with respect to the physical measure of security returns. 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, \(q_t(R)\), divided by the physical density, \(p_t(R)\):

\[
M_t = PV_t \left[ \frac{q_t(R)}{p_t(R)} \right] \tag{1}
\]

By theory, equation (1) must be fully-conditional with respect to the time \(t\) expectation. Being

---

\(^1\)The distinction between the pricing kernel (PK) and the empirical pricing kernel (EPK) may not be always so neat. From the point of view of the terminology, throughout the paper, we refer to the former for a more theoretically oriented analysis and to the latter when we go to the empirics.

\(^2\)Requiring the existence of no-arbitrage we are 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.
used for pricing purposes the PK must be a martingale and a strictly positive process: \( M_t > 0 \). By no-arbitrage constraints, the ratio (1) is related to the expected gross return, \( R_{t,T} \), from investing in simple state contingent claims:

\[
R_{t,T} = \frac{1}{M_{t,T} \cdot r_t T}
\]

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 ineffective hedging strategies. It follows that a proper modellisation of the PK can only be achieved by means of a time-varying stochastic model.

Although the neoclassical theory requires the PK to be a monotonically decreasing function, recent empirical studies found several violations in different areas of the functional: at the extremes as well as in the central portion. These violations lead to severe problems both under complete and incomplete market scenarios. We refer to the "PK puzzle" when the estimated PK is not general enough to properly explain the whole cross section of option data.

Evaluating (1) it is emerges that, overall, the estimation of a fully-conditional, hence fully informative, physical distribution is the most challenging task to perform. Due to the strong interconnection between the two densities and the EPK a biased denominator impacts strongly on the estimation of the EPK.

In this paper we solve this econometric issue proposing a new estimation technique that, exploiting a Bayesian non-parametric model for the estimation of an unknown measure (i.e.: the Dirichlet process (Ferguson 1973) [28]), combines the prior information available from different security markets and uses them to enrich the denominator of the PK, hence the PK overall.

Throughout the paper we explore and exploit three empirical facts, namely: the higher informative content provided by the joint use of stocks and options data with respect to just using raw historical stock returns, the econometrical difficulty encountered in the estimation of a fully-conditional and time-varying physical measure and the non-homogeneity bias which affects the two measures thus, by consequence, the PK.

First, options are, by structure, informative and forward looking financial assets. Since Chernov

\[3\]In this paper we refer primarily to index option prices but, with no loss of generality, the same conclusion can be extended to other financial asset classes.
and Ghysel (2000) [17], it is known that incorporating information provided by option prices assures an higher degree of precision in the estimation of the parameters than only using the information provided by stock returns. We exploit this fact proposing an empirical methodology that joins flexibly the risk-neutral and the physical measures. Tested on a daily basis and for different times-to-maturity, our model allows us to show how option prices mirror the investors’ risk preferences, errors and beliefs.

Second, the literature offers different parametric and non-parametric methodologies to estimate the physical distribution. Almost all of them are based upon a stream of backward looking stock returns, hence of past information. What is missing are the forward looking beliefs of the investor. These beliefs capture most of the risk appetite for the investment. The obtained measure is then non-conditional and only partially informative. Enlarging the time length of the estimation is not enough to obtain a measure that is conditional to the current time. Conditionality is of key importance with respect to the today volatility and higher moments. We estimate them by means of an asymmetric GJR GARCH - FHS model.

Last but not least, most of the literature related to the PK, the EPK and the PK puzzle compare a fully-conditional numerator extracted from options data with a partially-conditional denominator extracted from past stock returns. The obtained functional is thus non-homogenous with respect to the conditionality of the information. The annexed results are then spurious and misleading.

2 Intuition behind the model and preliminary results

The main limitation of the approaches present in literature is that they assume that all the information about the real-world probabilities of future returns, \( p_t \), 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.

Christoffersen at al.(2012) [20] and Bollerslev and Todorov(2011) [13] confirm that most of the risk and risk pricing information of the underlying asset can be extracted from derivatives product.

\[\text{Despite the overall documented higher quality of using the joint information and the difficulty in obtaining a sufficiently informative and "up to date" physical distribution, to our knowledge, still nobody has fully applied the proposed joint methodology for the estimation of the conditional physical measure.}\]

\[\text{Among the others, Jackwerth and Brown (2004) [15], Ziegler (2007) [58] and Beare (2011) [11] point out this issue but remain silent with respect to a possible resolution.}\]
Once adjusted for a risk premium\(^6\), 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, our approach is to solve the presented problem empirically and in a very natural way: letting the data speak as much as possible.

As a main novelty, in this paper, we propose a time-varying, flexible, non-parametric and fully-conditional physical measure that empirically adds part of the risk adjusted forward looking information embedded into the options moments into the unconditional historical denominator so that it can be homogeneously related with the fully-conditional risk-neutral measure thus producing an unbiased ratio.

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.

We 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 paper but can be an interesting investigation as a future work.

While since Ross (2015)[42] the use of the risk-neutral measure to extract the entire real world probability is gaining attention, the use of the risk-neutral density to improve the estimation of the physical density is not common in the finance literature. As Ross (2015)[42] shows, under particular assumptions\(^7\), it is possible to recover the entire physical density from the risk-neutral density but its mean. A complete recovery of the physical density becomes possible in a special case, namely when the state transition matrix has full rank, the diffusion is bounded and the utility function is state independent. Option prices and investors in practice are unlikely to satisfy these requirements. Carr and Yu (2012)[16] and Audrino et al. (2015)[6] propose alternative initial assumptions for the Ross recovery theorem. As main differences with respect to these papers we do not extract the entire physical measure from the risk-neutral one but we only use it to complete the objective.

As a main advantage, we do not need to put any structure on the functional form of the PK. 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

\(^6\)The 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)[31], 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)[18] suggest that the improvement induced by the variance risk premium, should be marginal. That should hold even more if parameters are re-calibrated.

\(^7\)The ones underpinning Girsanov’s theorem.
physical distribution. Our 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.

Our model is tested on US index and index options data. The obtained empirical results are robust and outperforms other methodologies. Being a data-driven methodology, interesting results are achieved in periods of strong market up/down turns, when the market movements produce high liquidity in the option market. Given our results, we claim that some of the puzzles present in literature are due to an econometric bias with respect to a bad modelling of the information. The 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 our research are likely to be relevant for risk management, regulation and financial policy. Asset management will also benefit from improved understanding of the PK.

3 Theoretical problem

Investors’ subjective beliefs are forward looking. An investor decides if and how to trade depending, among the others, on her personal beliefs. Since the investment horizon spans from the present into the future it is, by definition, uncertain. This uncertainty represent the degree of riskiness of an investment. Therefore, the valuation of any risky investment, has to take into account the forward nature of the subjective beliefs.

Counterintuitively with respect to their nature, investors beliefs are estimated with backward looking information (i.e.: Aït Sahalia and Lo (1998)[2], Jackwerth (2000)[32], Brown and Jackwerth (2001)[15], Rosenberg and Engle (2002)[27], Barone-Adesi et al.(2008)[9], Yatchew and Härdle (2006)[46]). Using these data, an important fraction of the investor’s risk and preferences are lost. As a consequence, a discrepancy between what is empirically obtainable and what is theoretically required by the neo-classical asset pricing literature arises. The larger the forward looking information bias, the larger is the subsequent mispricing.

From the point of view of the information filtration, this can be translated into a shrunk information
where the two information sets are increasing\textsuperscript{5} in time and contain all available and potentially usable information:

\[ \mathcal{H}_t = \{x_{-\infty}, \ldots, x_{t-1}, x_{t-\Delta_t}\} \quad \text{and} \quad \mathcal{F}_t = \{x_{-\infty}, \ldots, x_{t-1}, x_t\} \]

(4)

where \( \Delta_t \) represents the fraction of missing forward looking information that involves any risky decision to undertake from today \( t \) with respect to a future time \( t+\tau \). More on the theoretical effects of projecting the pricing kernel onto a coarser filtration set in Sala and Barone-Adesi (2015)\textsuperscript{[43]}.

4 The Empirical Pricing Kernel (EPK)

Playing 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 we present our model for the estimation of the EPK, a more rigorous definition of information filtration is needed. Unless differently stated, we assume a fixed and finite planning horizon \( t \in T \), where \( T < \infty \). We define the EPK on a rich enough filtered probability space \((\Omega, \mathcal{F}, P, \mathbb{F})\), where, for \( 0 \leq t \leq T \), the filtration \( \mathbb{F} = (\mathcal{F}_t)_{t \in T} \) satisfies the usual hypothesis.

A filtration is nothing but an increasing family of \( \sigma \)-algebras \( \{\mathcal{F}_t : t \in T\} \). It follows that:

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

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

The usual hypothesis are satisfied whenever the filtered probability space is complete and right continuous. Completeness of \((\mathcal{F}_t)_{t \in T}\) is achieved if the probability space is \( P \) complete and \( \mathcal{F}_0 \)

\textsuperscript{8}Two 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.

7
contains all the $P$-null sets of $\mathcal{F}$. Right continuity is defined as:

$$\mathcal{F}_t = \mathcal{F}^+_t = \bigcap_{u > t} \mathcal{F}_u \quad \text{for all } t \geq 0 \quad (5)$$

A correct modelling of the information flow is crucial for a proper measurability of random processes. A stochastic process $\{Z_t : t \in T\}$ is said to be adapted to a filtration $\{\mathcal{F}_t : t \in T\}$ if, for each $t \in T$, the process is measurable with respect to $\mathcal{F}_t$.

**Definition 4.1.** Defined in a continuous world with no-arbitrage, 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:

$$M_{t,T} = \mathbb{E}^P(M_{t,T}|\mathcal{F}_t) = e^{-r_t(T-t)} \frac{q_{t,T}(S_T|S_t)}{p_{t,T}(S_T|S_t)} \bigg| \mathcal{F}_t > 0 \quad (6)$$

where the conditional expectation is taken under the time $t$ physical measure $P$, hence conditioned to all subjective information up to time $t$, $q_{t,T}$ represents the conditional risk-neutral or state price density (SPD), $p_{t,T}$ the conditional historical density, $r_t$ the continuously compounded daily risk-free rate and $S_t$ is a proxy for the market portfolio.

In our case $S_t$ represents the S&P 500 index. Therefore the EPK is a projected PK where the projection is onto the extended positive real line occupied by all possible values taken by the index. Subscript $t,T$, which throughout the paper applies to all inputs of the PK and EPK, 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 homogenous PK. To lighten the notations we assume $\tau = (T - t)$ as fixed and equal to one year.

Function (6) can be estimated in different ways, depending on how is defined and how the ingredients that compose it are obtained. In all cases, given the role of the EPK, which represents a highly non-parametric phenomenon like the investors behaviour, it is always recommended to put as much less structure to the functional form as possible. Our 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.


5 A frequentist justification of the physical measure correction

Before we move to the main theorems, in this subsection we insert a word of cautions toward the possible statistical problems one can encounter during the estimation of the EPK through 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)\[33\] and Jobson and Korkie (1981)\[34\]) 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. We show this point by means of a Taylor expansion adapted to a numerical experiment. Our results are in line with the small sample bias of Leisen (2015)\[35\] who warns for possible statistical explanations of the PK 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 \(D \in \mathbb{R}^+\), our 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:

\[f : D \subseteq \mathbb{R}^2 \to \mathbb{R}, (x,y) \to z = f(x,y)\] (7)

In our case, the function, \(f\), evaluates the ratio of two measures: \(q/p\) so that \(z = \text{EPK}^P\).

More precisely we want to evaluate the expected value of this ratio under the physical measure:

\[\frac{q}{p} = \frac{\mathbb{E}^P(q)}{\mathbb{E}^P(p)}\] (8)

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:

\[\frac{\mathbb{E}^P(q)}{\mathbb{E}^P(p)} \leq \mathbb{E}^P \left( \frac{q + \Delta q}{p + \Delta p} \right)\] (9)

\({}^9\)To lighten the notation and with no loss of generality, in this section we omit visually the time dependence \(t,T\).
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 we approach the problem, a short excursion of numerical mathematics is needed. Whenever one deals with a generic computational problem where the goal is to find ($x$) such that:

$$f(x,d) = 0$$

through a numerical method that approximates ($\hat{\cdot}$) the above problem by a sequence of approximated models of the form:

$$\hat{f}_n(\hat{x}_n, \hat{d}_n) = 0 \quad n \geq 1$$

different types of errors can arise. In the above equations $f : \mathbb{R} \to \mathbb{R}$ is the functional form that relates ($x$) and ($d$), where ($d$) represents the dataset. Errors arise because equation (10) models a real world problem which is solved by means of a computational model (11).

Assuming (10) to be a well 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:

$$|x - \hat{x}_n| = e_T$$

$$= e_m + e_c$$

$$= (|x - f(x)|) + (|x - \hat{f}_n(\hat{x}_n)|)$$

$$= (e_{algo} + e_{dt}) + (e_{ds} + e_r + e_a)$$

From (15), the mathematical error ($e_m$) can be decomposed as the sum of the mathematical model error ($e_{algo}$): how well the algorithm chosen to model (11) describes the real problem, and the data error ($e_{dt}$): how well 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 (10) has no impact on $e_{dt}$.

For the resolution of our 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

\[\text{For a comprehensive analysis of the subject see, through the others, Quarteroni et al. (2007)[40].}\]
concrete resolution of (11) and (eₐ) the error introduced by the numerical algorithm. In general and also in our case eₐ > 0. It follows that, a.s. eᵣ > 0.

Back to our case, we define eᵣ = Δp(q) as the sum of all possible errors for the estimation of the physical(risk-neutral) measure. Of the two measures, we 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 puzzles”. By a flex in the central area of the functional some papers[11] show the existence of a risk seeking behaviour 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.

For (q) to be unbiased, it follows a.s. that Δp > 0 while Δq ≈ 0 so that (q') = (p + Δp) while (q') = (q); which leads to:

\[ \mathbb{E}^P \left( \frac{q'}{p'} \right) = \mathbb{E}^P \left( \frac{q + Δq}{p + Δp} \right) = \mathbb{E}^P \left( \frac{q}{p + Δp} \right) = q\mathbb{E}^P \left( \frac{1}{p + Δp} \right) \] (16)

Assuming (p + Δp) ∈ Cⁿ for n ≥ 2 and ≠ 0, we approximate the expectation of the ratio through a second order Taylor series expansion (see appendix:[A]):

\[ \mathbb{E} \left( \frac{q'}{p'} \right) = \mathbb{E}^P \left( \frac{q}{p + Δp} \right) = \frac{q}{p} + f'_p Δp + \frac{1}{2} f''_{pp} Δp^2 + O(n⁻¹) \] (17)
\[ \approx \frac{q}{p} + \frac{q}{p^2} Var(Δp) \] (18)

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: \( \mathbb{E}^P(p + Δp) = p \).

By the same token, the variance of the ratio is:

\[ \text{Var} \left( \frac{q'}{p'} \right) \approx \frac{q^2}{p^2} \left( \frac{\text{Var}(q)}{q^2} \right) \] (21)

From Jackwerth (2000) and Engle and Rosenberg (2002)[27], the term "PK puzzles", in principle, has been coined to represents exactly this bias in the central part of the distribution.
From the expansion the inequality (Jensen’s inequality) in (9) is demonstrated. The greater the variance of the physical error, the larger the distance between the numerical $E^p(q/p + \Delta p)$ and the mathematical $E^p(q/E^p(p)$ models used to estimate the EPK.

Due to the flexibility of our proposed model, we work on this inequality initially by trial and error, adjusting the denominator of the ratio up to finding 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\footnote{The entire methodology is fully explained in section (7.1).}. Due to the flexibility of our model, we 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 can be achieved by exploiting other sources of information which, completing the measure, would make the whole ratio more stable, hence more reliable. As we will see in the empirical section\footnote{The 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 may be highly non-parametric as well as unobservable.}, this is particularly true at the extremes of the distribution.

6 Estimating the measures

Our estimation of the fully-conditional EPK is based upon three steps. In the first we estimate, individually, the objective and the risk-neutral measures. Then, after the estimation of the scaled risk-neutral measure, we calibrate the fully-conditional physical measure through the Dirichlet process. Taking the present value of the ratio of the two fully-conditional measures completes the estimation.

The physical density estimation in the first step is based upon the strong assumption that historical data are correctly priced and fully informative. We 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 we use options data to extract the otherwise unobservable investors preferences\footnote{Duan (1995)\cite{22} proposes a direct change of variable by means of the Girsanov theorem but he does not allow for the market completeness required by the theorem.}. To do it we reconstruct the stochastic volatility by means of an asymmetric GJR GARCH - FHS model (details follow). The main problem lies in the incompleteness\footnote{Duan (1995)\cite{22}} of the market which is implicitly assumed by a stochastic volatility.

Knowing that a primitive asset can always be created by means of options, we complete the market by simulation. Once that all states are priced, we can use them to estimate the relative pricing densities (Ross (1976)\cite{41}). The idea is similar in spirit to Breeden and Litzenberger (1978)\cite{14}.
Their model requires a high market densities for all states which is empirically implausible, even for the more liquid assets. We achieve it by simulation using real data.

The first step is based upon Barone-Adesi, Engle and Mancini (2008)\(^9\) (henceforth: BEM). We use it as a benchmark for the empirical analysis to show the difference between a partially-conditional and a fully-conditional model. The estimation period spans from January 2002 until December 2004. All data are from OptionMetrics. Following the literature and using BEM as a benchmark the filtering methodology of the dataset is the same.

For each Wednesday \(t\) in our sample, we estimate two asymmetric Glosten, Jagannathan and Runkle (1993)\(^{29}\) (GJR) GARCH models. To describe the daily index dynamics under the objective distribution, \(p_t\), a GJR GARCH (1,1) model is fitted to historical daily returns of the S&P 500, going back 3500 daily observations\(^{15}\):

\[
\log \frac{S_t}{S_{t-1}} = \tau_t = \mu + \epsilon_t \tag{22}
\]

\[
\epsilon_t = \sqrt{\sigma_t^2} z_t \tag{23}
\]

\[
z_t | F_{t-1} \sim f(0, 1) \tag{24}
\]

\[
\sigma_t^2 = \omega + \beta \sigma_{t-1}^2 + \alpha \epsilon_{t-1}^2 + \gamma \epsilon_{t-1}^2 \mathbb{1}_{t-1} \tag{25}
\]

\[
\mathbb{1}_{t-1} = \begin{cases} 
1, & \text{if } \epsilon_{t-1} < 0, \\
0, & \text{if } \epsilon_{t-1} \geq 0. 
\end{cases} \tag{26}
\]

where \(\mu\) represents the constant drift term\(^16\) and (26) accounts for the leverage effect. Table (1) summarizes the set of estimated daily physical parameters \(\theta_t = f(\omega, \alpha, \beta, \gamma)\) obtained via Gaussian Pseudo Maximum Likelihood (PML).

To capture the index dynamic under the risk-neutral distribution, \(q_t = \tilde{\theta}_t\), another GJR GARCH (1,1) model is calibrated to the cross section of out-of-the-money (OTM) SPX option prices. The \(^{15}\)Our analysis spans from January 1, 2002 for two years. Going back for 3500 days we stop right before the 1987 crash. More on this in section (9.4.1).

\(^{16}\)Although all estimated factors are time-varying, to assume a constant value for \(\mu\) 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 \(\Delta t\) small enough:

\[
\sigma_t^2 = c_t^2 \Delta t = O(\Delta t)
\]

while for the mean:

\[
\mu^2 = c^2 \Delta t^2 = O(\Delta t^2)
\]

Although negligible, values from \(^{29}\) must be considered as approximations.
calibration is achieved via non-linear least squares, i.e. minimizing the sum of squared pricing errors with respect to the GARCH parameters:

$$\min_{(\tilde{\omega}, \tilde{\alpha}, \tilde{\beta}, \tilde{\gamma})} \sum_K \sum_\tau \left[ \left| \text{Market Prices}(K, \tau) - \text{Model Prices}(K, \tau) \right| \right]^2$$  (27)

where the model prices are obtained by simulation. Innovations are modelled using the filtered historical simulations (FHS) of Barone-Adesi, Giannopulos and Los Vosper (1999)[10] and the Gaussian distribution. See table (2) for the risk-neutral parameters. 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 $\gamma$, as well as the empirical innovation are of key importance to better capture the moments of the option surface.

7 Why a Bayesian non-parametric proposal?

To estimate the conditional physical measure, we make use 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 $(x_1, x_2, \ldots, x_n)$ with an unknown distribution, $A$, we wish to infer given some observed data:

$$x_i \sim A \quad \text{for} \quad i = 1, \ldots, n$$

To solve this problem we place a prior over $A$ and then we compute the posterior over $A$ given the set of observed data. If we chose the prior 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, we 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 MCMC techniques[17]. A typical choice for such a prior is the Dirichlet Process, DP, (Ferguson, (1973)[28]). The DP has support over discrete with a countable infinite number of point masses and is fully characterized by two parameters, $\alpha$ and $C$, so that: $A \sim \text{DP}$

\[^{17}\text{As for most of the Bayesian techniques, also the use of the DP, although very appealing given its features, was halted by storage and computational intractability problems. As a consequence of the improved computational powers of computers many Bayesian techniques are now living a second childhood.}\]
Parameter $\alpha$, a scalar defined over the positive real line $\mathbb{R}^+$, is a measure of precision which indicates how much weight is placed on the prior. Indeed, is the inverse of the DP variance: $\nabla(A) = V(A)(1 - V(A))/(1 + \alpha)$. The base parameter, $C$, is a centering parameter as well as the mean of the DP. For any measurable set $A \subset \Omega$ it follows that $E(A) = C(A)$. The absence of dimensional constraints, which assures more freedom in the estimation of the functional form of the EPK, turns out to be one of the key feature to obtain better and more realistic final estimations.

In a nutshell, adopting a Bayesian non-parametric approach allows us to avoid both critical dependence on parametric assumptions and to exploit, through 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 EPK in general.

7.1 The DP for the conditional physical estimation

Let $P_{t,T}$ and $Q_{t,T}$ be the probability measures associated with $p_{t,T}$ and $q_{t,T}$ respectively. Our 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) we center our DP prior on the risk-neutral measure $Q_{t,T}$, adjusted by a risk premium (more on this in section 8). We represent this risk adjusted risk-neutral measure by $Q_{t,T}^\ast$.

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:

\[
P_{t,T}^1 = \frac{\alpha^*}{(\alpha^* + n)} C + \frac{1}{(\alpha^* + n)} \sum_{i=1}^{n} \delta(S_i) \]

\[
= \frac{\alpha_{t,T}^*}{(\alpha_{t,T}^* + 1)} Q_{t,T}^* + \frac{1}{(\alpha_{t,T}^* + 1)} P_{t-\Delta_t,T} \]

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, being calibrated daily, is time-varying. Its updated version becomes

\[^{13}\text{Given its characteristics } \alpha \text{ takes different names in literature i.e.: the centering or the strength parameters. Throughout the paper we call it as the precision parameter.}\]

\[^{19}\text{Where } \Omega \text{ represents a probability space.}\]
Equation (29) is equivalent to (28) 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 we need to simulate. Moreover, to improve the readability of the problem, we can, with no loss of generality, normalize $n = 1$ (more details in section (8)).

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 semi-parametric 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 we obtain, with no loss of generality, an a.s. continuous distributions.

### 7.2 The precision parameter

Acting directly onto 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 we 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; in the limit:

$$
\lim_{\alpha_{t,T} \to +0} p_{t,T}^{\dagger} = p_{t-\Delta_t,T} \quad \text{s.t.} \quad M_{t,T}^{\dagger} = PV \left( \frac{q_{t,T}}{p_{t-\Delta_t,T}} \right) = M_{t-\Delta_t,T} (30)
$$

Asymptotically the conditional physical measure is equal to the original objective measure and the relative EPK turns the original one. For all partially-conditional measures the subscript $t-\Delta_t,T$ represents, at 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

\[ \text{To avoid confusion with respect to the ARCH parameter } \alpha, \text{ we identify the precision parameter used for the density estimation by } \alpha_{t,t+\tau} = \alpha_{t,T}^*. \]
the information contained in the option prices; in the limit:

\[
\lim_{\alpha_{t,T} \to +\infty} p_{t,T}^{\dagger} = q_{t,T}^{\dagger} \quad \text{s.t.} \quad M_{t,T}^{\dagger} = PV\left(\frac{q_{t,T}}{q_{t,T}^{\dagger}}\right)
\]

(31)

the conditional physical measure is only composed by the scaled risk-neutral measure.

Using \(\alpha_{t,T}^{*}\) to regulate the shape of the EPK, the precision parameter provides and indirect proxy with respect to the degree of unconditionality provided by the partially-conditional objective measure \(p_{t-\Delta t,T}\). The larger the correction the higher the amount of missing information \(\Delta_{t}\). 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, \(\tau\), gets smaller means to give more weight on the forward looking beliefs extracted from the options surface. 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 further we go in time and the less is the reliability of forward looking information so that we give less weight by reducing \(\alpha_{t,t+\tau}^{*}\). The flexibility of the precision parameter allows us to adapt the proposed model to different market scenarios.

Following the Bayesian paradigm and to further improve the performance, it is possible to place a higher level in the modelling hierarchy eliciting a prior also on the \(\alpha_{t,T}^{*}\) parameter, following West (1995)\[45\]. It can also be adopted a more flexible distribution, i.e. the Pitman-Yor(1997)\[39\] process which, through the discount parameters, allows for a better modelization of the tails of the distribution.

8 The conditional physical measure estimation and the EPK

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 needed for a proper estimation. The problem is well-known in the finance literature. Cochrane (2001)\[21\] analyzes the problem comparing a conditional with an unconditional model. Among the others, he proposes to use instrumental variables to move
from a static to a dynamic model. From page 132 of [21] "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, we compare a fully with a partially dynamic model, where the former is obtained by means of a non-parametric model.

For the estimation of the fully-conditional 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, \( q_{t,T}^\ast \). 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, we follow Merton (1980) [38] and we scale the drift used to simulate the log-prices by a fixed value of 8% or 4%[21]. Although violations are usually the product of different flaws, a miscalibrated risk premium can be the reason of the nonmonotonicity of the EPK in the central area of the functional. The estimation of the risk premium and the precision parameter are both non trivial tasks. Since their "role" is to model non-parametric random human behaviours, both parameters turn out to be highly stochastic. We propose to solve the issues empirically for \( \alpha_{t,T}^\ast \), linking its value to the liquidity of the options in the 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, \( \phi \), and the daily percentage dividend yield with, \( d_t \)[22], the three different drifts used to simulate the daily asset prices are:

- **Physical measure (\( p_t \)):**
  \[
  \text{Drift}_{p_t} = \frac{1}{365} \left( \sum_{t=1}^{T} r_t - d_t + \phi \right) \tag{32}
  \]

- **Risk-neutral measure (\( q_t \)):**
  \[
  \text{Drift}_{q_t} = \frac{1}{365} \left( \sum_{t=1}^{T} r_t - d_t \right) \tag{33}
  \]

---

[21] While the former is the original value proposed by Merton, we 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 our exercise.

[22] Differently than the dividend yield and the risk free rate, the risk premium, being fixed, is assumed to be time-independent.
• Modified risk-neutral measure \( (q_t^*) \):

\[
\text{Drift}_{q_t^*} = \frac{1}{365} \left( \sum_{t=1}^{T} r_t - d_t + \phi \right)
\]  

(34)

where \( \frac{1}{365} \sum_{t=1}^{T} r_t \) represents the daily risk free rate\(^\text{23}\).

Using (32) - (34) as input for the drift, the estimated \( \theta_t \) for the objective and the FHS (or Gaussian) \( \tilde{\theta}_t \) for the two risk-neutral distributions as inputs for the daily volatilities we obtain, as output, \( n \) simulated log-prices, where \( n \) are the number of simulations. Although for the entire estimation we 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 \( S_t \) is defined as:

\[
\hat{S}_{i,t} = S_t \exp \left( \text{Drift} - \frac{\hat{\sigma}_{t,\text{sim}}^2}{2} \right) dt + \hat{\sigma}_{t,\text{sim}}^2 dW_t
\]  

(35)

where \( S_t \) represents the S&P500 price at a given day, the Drift input changes dependently on what we 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, we model \( dW_t \) by means of the FHS or the normal distribution.

Using a MC pricing model, it is well-known that, due to the exponential semi-martingale structure of the simulated sample path (35), we need to correct it to prevent propagation errors that would lead to arbitrage violations. Using the Empirical Martingale Simulation (EMS) of Duan and Simonato (1998)\(^\text{23}\) we optimize the Monte Carlo simulation:

\[
S_{t,t}^{\text{corr.}} = S_t e^{r_t \cdot T} \frac{1}{n} \sum_{i=1}^{n} \hat{S}_{i,t}
\]  

(36)

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 (36) is made of a temporary created asset price at time \( t \) and the discounted arithmetic average of it. We apply the correction to each iteration and time-to-maturity:

\[
S_{t,t} = \hat{S}_{i,t} \cdot S_{t,t}^{\text{corr.}}
\]  

(37)

\(^\text{23}\)For each time \( t \), the daily risk free rates, \( r_t \), are obtained by interpolation
The effect of the empirical martingale adjustment is that the simulated sample does not violate the martingale property thus producing rational option pricing bounds.

To extract the probability density functions from the estimated log-prices, we perform a non-parametric kernel density estimation evaluated at \( n \) equidistant points with a Gaussian kernel and an optimal bandwidth computed following the Silverman rule of thumb. We 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 \( q_{t,T}^{FHS}, q_{t,T}^{*,FHS} \) and three with the Gaussian one: \( q_{t,T}^{Gauss}, q_{t,T}^{*,Gauss} \).

As of now, the objective densities are still biased. We solve the issue mixing the estimated measures. Taking the present value of the pricing measure over the conditional physical measure complete the estimation.

Mixing the measures we pass from a partially-conditional EPK:

\[
M_{t-,\Delta_t,T} = e^{-r_t \cdot \text{Maturities}} \left( \frac{\text{Risk-neutral}_t,T}{\text{Risk physical}_t,\Delta_t,T} \right) \quad (39)
\]

\[
= e^{-r_t \cdot T} \left( \frac{q_{t,T}}{p_{t-\Delta_t,T}} \right) \quad (40)
\]

\[
= \text{PV} \left( \frac{q_{t,T}}{p_{t-\Delta_t,T}} \right) \quad (41)
\]

to a fully-conditional EPK:

\[
M_{t,T}^\dagger = e^{-r_t \cdot \text{Maturities}} \left( \frac{\text{Risk-neutral}_t,T}{\frac{\alpha_t,T}{\alpha_t,T+1} \cdot \text{Modified risk-neutral}_t,T + \frac{1}{1+\alpha_t,T} \cdot \text{Risk physical}_t,\Delta_t,T} \right) \quad (42)
\]

\[
= e^{-r_t \cdot T} \left( \frac{q_{t,T}}{\frac{\alpha_t,T}{\alpha_t,T+1} \cdot q_{t,T}^* + \frac{1}{1+\alpha_t,T} \cdot p_{t-\Delta_t,T}} \right) \quad (43)
\]

\[
= \text{PV} \left( \frac{q_{t,T}^*}{p_{t,T}^*} \right) \quad (44)
\]
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 our analysis hard to follow, and with no loss of generality, we normalize to $n = 1$ the total number of observations. It follows that a value of $\alpha_{t,T}^* > 1$ would give more weight to the prior information, viceversa for $\alpha_{t,T}^* < 1$.

The model behind the EPK presents 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.

9 Empirical Results

As a main finding, the proposed Bayesian non-parametric methodology produces 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 returns\(^{27}\). 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 high robustness to numerical noise errors. All in all, 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 precision parameter. Related to the paper but independently with respect to the mixing procedure, the proposed estimation method depends from the fully arbitrary choice of the risk premium used in simulation and, indirectly, from the degree of liquidity of the underlying.

Following the above methodology for each Wednesday, $(t)$, of the January 2, 2002 to December 29, 2004 time period and for each time-to-maturity, $\tau$, we obtain 862 observations for the fully-conditional EPKs $(M_{t,t+\tau}^\dagger)$ and the partially-conditional EPKs $(M_{t-\Delta t,t+\tau})$.

\(^{27}\) For "normal" periods we mean either periods of low volatilities or with a low activity (liquidity) in the market.
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, the daily estimation of the EPK is as much important as non trivial from an econometric viewpoint. The estimate of 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, our 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. As a consequence, the model is interesting both from and econometrically and from an economic prospective. To highlight this feature and to show the robustness of the findings over different time periods, we divide the empirical results into daily and yearly estimates.

Ait Sähalia and Lo (1998)\(^2\) and Jackwerth (2000)\(^32\) are two examples of non-time-varying EPKs\(^28\). Putting a not implied nor required by the theory stationary assumption on the estimation of the objective measure, their results are time-invariant. It turns out that what they propose are EPKs averaged on a yearly basis. 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.

\section{Daily estimates}

Figure \((2)\) 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 horizontally in different times-to-maturity, \(\tau\). For \(t = 61\) the set of \(\tau\) ranges from 24 to 297 days\(^29\). The figure is then divided vertically in two parts. The graphs of the left column show

\(^{28}\)Or at least, time-varying on a yearly basis.

\(^{29}\)All times-to-maturity, \(\tau s\), are filtered such that are within 365 and over 10 days.
the estimated measures, respectively: the risk-neutral density, $q_{t,t+\tau}$ 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, $p_{t,t+\tau}^\dagger$. Due to the risk premium, the mean of the risk-neutral
measure is correctly shifted on the left with respect to the other two measures. 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, $t$, and for each time-to-maturity, $\tau$, 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, $\alpha^*_{t,T}$, is flexibly calibrated such that it best explains the economical
conditions present on the market. For $t = 61$ it goes from 1.75 for the closest time-to-maturity and
decreases to 0.75 for the longest $\tau$. More on the calibration of $\alpha^*_{t,T}$ in the next section (9.2).
The figure shows all pros and cons of the proposed model which, to simplify the exposition, we
divide in: values of the functional, values of the horizontal support (domain of stock return) and
values of the vertical support (domain of the EPK). Although at different magnitudes the entire
sample in analysis presents results in line with figure (2).
First of all, aside for possible numerical frictions, all conditional EPKs are almost everywhere
monotonically decreasing. Secondly, 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
possibility of exploiting a larger support is one of the main feature of the proposed model. Most of
papers present in literature suffer from being informative only in a very narrow area, the one ATM.
Above all for risk management, not having the tails under control may lead to big problems. A
larger support also increases the degree of precision and effectiveness of daily pricing and hedging
operations. This is particularly true into the left region of the EPK. 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 possible 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 behaviours 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, data for the short time-to-maturity are not
silent and, above all for the put options, are very liquid, hence informative. It is thus possible

\[30^\text{Given the structure of the figures, which are divided horizontally for the different times-to-maturity, we prefer to use}
\]
\[t,t+\tau \text{ to make the concept clearer and easier to follow. Anyhow, } t,t+\tau = t,T \text{ are equal and can be used interchangeably.}\]
and of key importance to be able to extract as much valuable information as possible. However, this would be impossible just using historical stock data, as a consequence, most of papers avoid this problem eliminating these maturities, thus providing incomplete results. In this paper we leave most of supports unbounded to show the higher explanatory power of a proper use of the information set. As a drawback, the plots can sometimes be very noisy as the liquidity dries up. As a demonstration, in a day with a low liquidity of D/OTM call options, the right fraction of the graph may get unstable as we go into the tail. This confirms the sensitivity of the model to options liquidity.

Finally, results on the vertical support (domain of the EPK) are in line with the ones of a utility maximizer rational investor. As we will see in section 9.4 only FHS values produces non puzzling results.

As a visual confirmation of the above points, we take another random day from the entire time series: January 23, 2002 and we compare the fully with the partially-conditional EPKs. Figure 3 shows both the fully and the partially-conditional EPKs with the partially-conditional FHS(Gaussian) in black(blue). From a liquidity viewpoint, for \( t = 4 \) we have 41 DOTM puts, out of a total of 85 puts traded, hence a 48.24% which is in line with an average of 49% for the entire sample. For the call we have 61/114, thus 53% of DOTM against a 51% for the entire sample. Very similar graphs are obtained for days with similar characteristics i.e.: \( t = 85 \) which has respectively 52/48% for DOTM put/call 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 function, the fully-conditional EPKs, being more informative all over the functional, produces valid, bounded and stable results for the entire functionals. On the other side, the partially-conditional values becomes very unstable also for values very close to \( S_t \). Taken from a larger 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. We will see this in the next sections along with the importance of using empirical innovations, above all during strongly non-Gaussian days.\(^{31}\)

To summarize, the theoretical foundation behind our 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

\(^{31}\)Although the Gaussian fully-conditional EPKs produce good results for \( t = 4 \), 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 sentiment non-parametrically into the call options, produce fully monotonically decreasing EPKs.
GJR GARCH model using FHS or Gaussian innovations and the empirical mixing of the measures through the Dirichlet Process.

9.2 Linking the precision parameter to the options market

Starting from the insights in the preceding section, it is now easier to show how we calibrate the value of $\alpha_{t,T}^\star$. As a main message, this paper, through the calibration of the precision parameter of the DP, tests the potentiality that a fully-conditional EPK lead to a monotonically decreasing function. More precisely, calibrating the precision parameter by "completion", that is, using $\alpha_{t,T}^\star$ as a proxy for the missing sentiment, we test the potentiality of extracting the information embedded in these stochastic values. Given the structure of our empirical methodology, the analysis is first performed on the physical measure and then, by construction, on the entire EPK thus improving both values.

In the previous analysis we have shown empirically the theoretical fact that most, if not the entire amount of information, lies into the option market\footnote{Other analysis in the next sessions will support and confirm this sentence.}. More precisely lies into the OTM and DOTM options. Since our empirical approach is to let the data speak as much as possible, we 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 (3) shows summary statistics of the OTM and DOTM options of our dataset. As expectable, both for call and put, DOTM options have mainly long times-to-maturity, while the opposite holds for OTM options. As we will show also graphically, our sample size shows a substantial amount of long term DOTM call options. Option prices share the expected behaviour 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 (5), 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 larger than 50%, the percentage of DOTM call options is extremely high in the first half, up to 100%, and then decreases strongly in the second part up to less than 20% which is in line with the level of the volatility for that period. The left bottom panel shows the time series of the total percentage of DOTM put and call options (left axis) and the relative values of $\alpha_{t,T}^\star$ (right axes).

To be able to extract as much information as possible and to reach an high degree of precision and
flexibility in estimation, $\alpha_{t,T}^*$ is linked to the daily liquidity of the option market by a fixed percentage thus producing a time series of time-varying precision parameters. The value of $\alpha_{t,t+\tau_{\text{Min}}}^*$ is determined directly for its highest value and then indirectly by inversion. The highest value is then the starting point to determine all the others. Given our sample we fix it equal to 2.5:

$$\vartheta = \frac{2.5}{\max_t \left( \frac{\kappa_t}{\text{Total}_t} \right)}$$

where $\vartheta$ represents the obtained fixed percentage, $\text{Total}_t$ is the total amount of traded options for each day of the time series and $\kappa_t = \frac{K_t}{S_{t,T}}$ takes into account the different moneyness:

$$\kappa_t = \begin{cases} 
\text{DOTM Call}, & \text{if } \kappa_t > 1.15, \\
\text{OTM Call}, & \text{if } 1 < \kappa_t < 1.15, \\
\text{OTM Put}, & \text{if } 0.85 < \kappa_t < 1, \\
\text{DOTM Put}, & \text{if } \kappa_t < 0.85,
\end{cases}$$

where $\tau_{\text{Min}}$ represents the shortest maturity which is the ones with the highest amount of information usable.

Since the horizontal domain of the EPK accounts for both put and call options and since, given the daily option surface, we want to analyze where is the highest amount of information available, for each day of the time series we sum up the amount of OTM or DOTM put and call options so that $\kappa_t$ represents the fraction of OTM or DOTM options traded out of the total. The maximum value is taken over all days of the time series where, as usual, $t = 1, \ldots, 157$.

For example, to link the precision parameter to the market of the DOTM call options:

$$\vartheta = \frac{2.5}{\max \left( \frac{\kappa_t}{\text{Total}_t} \right)}$$

$$0.0291 = \frac{2.5}{85.98}$$

where the value of 85.98% is achieved for $t = 63$. Once that the "starting point" is determined, the other values of $\alpha_t^*$ follow just by inverting (45).

Of the entire process, the only "arbitrary" choice is the value of 2.5. This is only partially arbitrary since, to determine it, we let again the data speak. It turns out that for the highest values of both $\tau_{\text{Min}}$, the EPK is more informative.

---

33 Where $\tau_{\text{Min}}$ represents the shortest maturity which is the ones with the highest amount of information usable.
OTM and DOTM options, the value of 2.5 is the one that empirically leads to the best shape given the data. To account for the decreasing reliability of the information provided by the options 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 \( \tau \) increases. For example, for the case of \( t = 61 \) presented above, we go from 1.75 to 0.75 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 our 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 (5), show the same analysis but for OTM options. While confirming the results relative to the behaviour of put options, an opposite trend characterizes the behaviour 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 behaviour. Being the values of \( \alpha_{t,T}^{*} \) linked to them by a fixed proportion, the time series of the precision parameter shows a much higher stability. Table (4) shows summary statistics of the different \( \alpha_{t,T}^{*} \). While starting and ending values are in line, means, medians and standard deviations reflect the different behaviour given by the different underlying used. As a technical detail, to increase the readability of the analysis, and with no loss of generality, we use rounded numbers from the time series of computed \( \alpha_{t,T}^{*} \). Among the two methods, the one that provides better results is when \( \alpha_{t}^{*} \) is linked to the percentage of DOTM options. This is an indirect confirmation of the importance of the information that lie into the deepest tails of the option surface.

9.3 The impact of the implied moments of the conditional physical measure on the EPKs

The good results showed so far are even amplified if we analyze the single operations more closely. The theoretical backbone behind Herrison and Kreps (1979) 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

---

34For the best shape we mean the one that produces a monotonically decreasing function, as required by the neo-classical theory.
35To improve the overall readability of the analysis we omit visually the second time dependence of the precision parameter and we represent the daily precision parameter with just \( \alpha_{t}^{*} \).
36The overall percentage of OTM put options is, as expectable, always above 50% and even higher than the amount of DOTM put options.
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.

In this section we study, graphically and numerically, the statistical properties of the conditional physical distribution and we 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 one. 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 behaviour of any finance phenomenon. As expected, we 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 (5) reports summary statistics of the difference between the fully and partially-conditional physical measures with respect to the "benchmark" unbiased risk neutral one:

\[ \Delta_{t,T}^{\text{Cond.}} = q_{t,T} - p_{t,T}^{\dag} \quad \text{And} \quad \Delta_{t,-\Delta t,T}^{\text{Part. Cond.}} = q_{t,T} - p_{t} - \Delta_{t,T} \]  

Results are on a yearly basis 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 end 2002 beginning 2003 and has big values for the

\[ ^{37} \text{Results confirm and extend broadly the findings of Linn et al. (2014)[36].} \]

\[ ^{38} \text{To prevent possible inconsistencies and following the same time frequency of the empirical models presented above, results are obtained taking the arithmetic mean of the measures computed on a weekly basis. As a result we have respectively 52/53 and 52 values for 2002/2003 and 2004. Daily results confirm the same findings found on a yearly basis and are available upon request.} \]

28
mid and long 2003. Due to the constant up-trend of the S&P500 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 we only report the FHS results all analysis have also been performed using Gaussian innovations. Results confirm the FHS findings and amplify the $\Delta_t$: 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 behaviour of the measures, figure [4] shows what estimated with [49] but setting the time window $t,T$ on a daily basis. The continuous blue(dotted red) line represents the difference between the fully(partially) conditional 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. From the analysis we can confirm our theoretical insights and also 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, we now put a focus on the tails of the same distributions.

Figures [6] and [7] are a focus on the whole spectrum of the left tails of the distributions\footnote{For robustness, we proposed the same experiment with different kernels and different number of points, from 500 to 5000. Best results are obtained with a normal kernel. Aside from the computation time and from a slightly refinement on the shape, an increasing number of points has not a big impact on the final findings. The proposed figures have been computed with a normal kernel and 5000 points. Other experiments are available upon request to the authors.}. The impact on the left tail of the conditional physical distribution is stronger as we 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 it can be expected, a change on the parameter $\alpha_{t,T}$ has a much stronger impact than a change on the risk premium in the tails. 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 $\alpha_t^*$, 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 (8) and (9) 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 options. Also for the right tails, results improve as we increase the time-to-maturity thanks to the higher amount of long term DOTM call options and the values of $\alpha_t^*$.40

To conclude the analysis on the distributions we graph the yearly estimates of the three measures. The fully-conditional physical measure, $p_t^{\text{year}}$, is in red, the partially-conditional physical measure, $p_t^{\text{year}} - \Delta_t$, in green and the risk-neutral measure $q_t^{\text{year}}$ is blue. Plots represent the average of the daily measures divided for years (2002 - 2003 - 2004) and times-to-maturity (short - medium - long). As a first result it is visible the strong departure from the normal distribution, as 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 behaviour in the options literature and the conditional physical measure behaves better than the partially-conditional one. Being the average of the 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 $\alpha_t^*$ and different risk premium. Results are robust and available upon request.

9.4 Focus on a single time $(t)$ and time-to-maturity $(\tau)$

As done for the densities, we analyze the behaviour of the EPKs more closely. Once more, the fully-conditional values achieve better results. For coherence with the previous analysis we use the same days, differently than before we analyze their different times-to-maturity from a closer look. Figure (11) 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 hor-

---

40By construction, results improve for days with a dense amount of DOTM call options, i.e $t = 63$, namely March, 12, 2003.
izontal support, the partially-conditional EPKs explode quite soon. Of course, the further we 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 almost-bounded results on the right side of the functional one needs to cut the support, model the extreme values of the tails with more extreme (but usually parametric) distributions or put stronger a priori assumptions (such as the requirement of non-increasing values).

Among the others, 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 (Christoffersen P. and K. Jacobs(2004)[19]). Here we focus the attention on the latter. A correct modellization of the distributions governing the EPK may answer to important mispricing problems: i.e. the 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 behaviour of the option surface Andersen at al. (2014)[3] 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 overpricing of the short term put options. Using a different approach which makes use of the $\gamma$ parameter, the FHS as innovations and the informative content provided by the options’ implied moments, our model features a more flexible and fully non-parametric ”tail factor” which answers to mispricing of both put and call options. Starting from the former: rational investors are unlikely to buy protection in presence of extremely high costs. As a consequence, too high values on the left portion of the EPK are hardly justified economically since they imply an over-expensive insurance. Even for cases of very low values of the underlying, for which an investor would be prone to spend money to buy protection, a too high price for short (and long) term put options would imply a too high cost for buying insurance.

From the estimated EPKs, a direct way to unveil the presence of possible mispricings is analyzing the values of the expected return obtainable from investing in the primitive asset. The today
expected rate of return over $\tau$ days is defined as:

$$v_{t,t+\tau} = \left( \frac{1}{M_{t,t+\tau}} \right) - 1$$  \hspace{1cm} (50)

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 of the related single estimates of the EPKs. Depending on the degree of liquidity of D/OTM, mispricing appears strongly for both put and call options.

Once more we keep working on $t = 90$. Figure $[12]$ shows, using the above estimated EPKs, the time $t$ expected rate of return over $\tau$ 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 EPKs have a monotonically increasing shape, as required by the neoclassical literature, the partially-conditional functions exhibit a concavity in the central area and extreme values into the tails. The former violation is 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 expects to have a positive return from an insurance and would end up paying an exaggerate cost for protection. Results are strong and show a decreasing magnitudes as $\tau$ increases. As done for the EPKs, To validate the above points, we analyze the mispricing problem from a closer look. The figure for $t = 90$ and $\tau = 31$ shows how both divergences are visible. Starting from values around $S_t = $850 which is close to 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 we go further into the tail.

On the other hand, as a direct consequence of the better estimate of the underlying, both conditional EPKs $v_{t,t+\tau}^{\dagger}$ do not share these problems, thus explaining the overpricing of short-term put options. The above violations are present and persistent for the inter sample in object.

Probably even more interesting since less debated in literature, is the ability of the model to explain possible call options mispricing. While some mispricing is visible also for $t = 90$, due the

\[\text{See Appendix (B) for the details.}\]
low liquidity of DOTM options for that day, the effects are less evident. To better analyze the
capacity of the model to capture mispricing also on the right side, we pick a day where the amount
of DOTM call options is significant, i.e.: U-shaped $t = 63^{42}$. As visible from figure (12), the strong
departure of the partially-conditional expected returns from the monotonically increasing shape
required by the neoclassical literature is strong and persistent for all times-to-maturity. Where the
biased EPKs are U shaped, the relative expected returns show a strong underpricing of call options.
Throughout the entire sample, for all days in which there is enough liquidity of DOTM and OTM,
the partially-conditional EPK leads to a U-shaped functional, thus call option underpricing, while
the fully-conditional EKP, capturing the information provided by the call options, have the right
fraction which correctly goes down monotonically. This is possible, also mathematically, only if the
denominator is not small or null.

Aside from the single examples proposed, we find that, using an information-miscalibrated
model, the mispricing of put and call options is present and persistent. Our results only partially
corroborate Babaogolu et al. (2014)[7]. While we confirm the need of a correct modelling of
the innovations for of a proper estimation of the high moments of the distributions, we 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 our results shows and confirm, from a different angle,
the data speak, our model is very sensitive and flexible depending on the degree and the direction
of the liquidity of the market. As a consequence of its explanatory power, our "tail factor" and
not only confirm the results of Andersen at al. (2014)[3] but it also extends the analysis for the
call options thus confirming the results of Bakashi et al. (2009)[8]. Mispricing are highest for the
partially-conditional EPK with Gaussian innovations, then for the FHS innovations while almost
non existing for the conditional modified EPKs$^{43}$.

**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. We analyze a sce-
nario in which the recent history has been bearish for some time and then, for any reason, the trend
changed suddenly. Empirically, using a non homogenous estimation methods, the obtained EPKs
are strongly U-shaped (days with majority of OTM call options) or, at best, non monotonically

---

$^{42}$For $t = 63$, namely March 12, 2003, the total amount of trades options is 157 divided in: 47 put options of which
53% are DOTM and 110 call options of which the total amount, 100%, are DOTM.

$^{43}$Among the two, the FHS, capturing better the non-normality of the distributions, performs best.
decreasing (days with majority of 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 findings of Jackwerth (2000) and the following papers. It is in fact not by coincidence that the PK puzzle emerges from the S&P 500 data right after the 1987 crash: when the liquidity of put options used as insurance against possible market down-movement grew exponentially. What is missing in their 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-homogenous ratios, hence puzzling EPKs. Our proposed model overcome this issue.

Given our data sample, for the period 2002-2004 the lowest value of $776.76 is touched by the S&P500 on October 10, 2002 ($ = 41). Starting from a value of $949.36 achieved on August 21, 2002, the S&P500 value of $776.76 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. As a consequence of this negative/positive price pattern, $ = 41, is not by chance one of the day of the sample with the highest amount of traded DOTM put and call options (85%). Figure (14) shows the $M_{t-\Delta t,T}$ and $M_{t,T}^\dagger$ estimated with 50,000 simulations, a starting $\alpha_{t,T}^*$ of 2.5 which decreases in time and a R.P.=8%. For $M_{t,T}^\dagger$ all features required by a correct EPK holds while for $M_{t-\Delta t,T}$ two characteristics 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 we give a closer look at each time-to-maturity. Figure (15) shows how, for all $\tau$s, the partially-conditional EPKs present an economically invalid strong U shape. Differently than the first stream of literature which focused on the flex in the central area of the functional, this is what the recent papers call the new PK puzzles. To emphasize the result we plot a vertical line which, starting from $S_t$ highlights the left and right area of the EPKs and a vertical line which identifies the risk-neutral area. Things change using a fully-conditional denominator. Aside from the usual physiological problem at the extremes of the functional, the conditional EPKs produce monotonically decreasing, thus economically valid estimates. As a main result we can state that, if under normal market situations the partially-conditional EPK has, as a main drawback, a strong overpricing of put options, under strong market up/downturns things change dramatically. If in the former case the right side of the EPK correctly converges monotonically to low values, in the latter

\[^{44}\]Opposite results are obtained from a bullish period followed by a downturn (i.e. for $t = 55$).
the classical methodology has problem of convergence that produces strongly U-shaped 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.

Due to the high randomness of the EPK estimation, the unconditional econometric models proposed in literature are simply not enough to capture all the information necessary to produce a valid EPK. This fact emerges strongly during volatile market days. Although the last historic values would be informative of the actual market scenario, the magnitude of the last info is washed out by the long stream of data needed in estimation. This big mass averages out meaningful values. A naive example may confirm what just stated. The artificial Black & Scholes economy assumes constant first and second moments. Under this framework, the PK 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, the physical measure can on average produce valid EPKs.

Both the naive and the historical examples 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. These results confirm and amplify the importance of the high moments of the distributions for the estimation of the EPK and are in line with Agarwal et al. (2008)[1], Bo-Young at al. (2013)[12] and Chabi-Yo (2012)[47] who found that market skewness and kurtosis are priced in the cross section of returns.

9.4.1 The length of the rolling window: when more data doesn’t imply more quality

Given the estimation techniques proposed in this paper, it is worth to mention 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. Under the classical estimation method, for the physical measure, the more we increase the time length of the rolling window the more we gain from the point of view of the degree
of conditionality of the estimation but, at the same time, the more we lose from the point of view of the informative content of the recent most important information.

Due to the structure of the estimation method, it is only by increasing the rolling window that one can increase the degree of conditionality of the estimates. Moreover, using a Gaussian kernel estimation methodology, it is well-known that the mechanics of the estimation improve as the number of inputs grow.

At the same time though, the more one adds data, the less is the effective weight put on each single value and, consequently the less is the informative content that can be extracted by "significative" data points. On the contrary, with our model, as long as the estimated physical parameters are valid, it’s easy to achieve a valid degree of conditionality still using the same or even a shorter rolling window. Thanks to the information present in other assets it is thus possible to save time and increase the final 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, we perform the portmanteau test of Ljung and Box (1978)\[37\] and the Lagrange multiplier (LM) ARCH test (Engle 1982)\[26\]. For both tests, the use of 3500 observations is enough to produce valid and solid results.\[45\]

\[4.2\] Yearly Estimates

As done for the densities we enlarge the time window passing from a single day estimation to a yearly estimation. Following Jackwerth (2000), we perform the estimation of $M_{t,T}$ 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. In all cases we present the wider support possible. To make our results stronger, we perform the same experiment with different calibrations of $\alpha_{t,T}$, of the risk premium and changing the number of simulations. Figure (16) summarizes the main results showing both methodologies duly divided in years (horizontally) and short, medium and long time-to-maturity (vertically). For both cases we 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 related

\[45\] The same test 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.
to the horizontal support, one to the vertical one:

- Confirming the daily results, the fully-conditional EPKs are spread over a wider - if not the entire - horizontal support (domain of gross returns). Differently than the vertical ones, 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 (used from daily to systemic risk management). Confirming the strong informative power of the proposed model on the short and medium time period, major improvements are achieved for the sets of short and medium time-to-maturity. While with the fully-conditional methodology we can quite always go up to 0 (exceptions are the 2003 and 2004 short times-to-maturity which starts form 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 the partially-conditional EPKs show higher noise in the final states, the right sides of the functionals are similar for both methods and extend to almost the same values. This is due to the particular pattern taken by the liquidity of DOTM and OTM call options. Given the amount of traded call options, as seen above, the potentiality of the analysis for the call side is strong for some single days in which the volume of traded DOTM and OTM call option is high enough so that our analysis is able to explain the U-shaped puzzle that emerges using a partially-conditional physical measure.

- Confirming the mispricing problems found 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 $M^{Gauss}_{t,T}$ and 2.5 for $M^{FHS}_{t,T}$, the partially-conditional $M^{Gauss}_{t-\Delta t,T}$ almost arrives to touch the value of 10 for the 2002 long time-to-maturity.

For both domains yearly results confirm all findings obtained on a daily basis.

As a final remark: for all daily and yearly estimates, results may look partially different if the two main random variables, the precision parameter and the risk premium, are would be with

\[46\] The domain of the EPK is even bigger than $[0,10]$. To propose readable plots, the value of 10 is the maximum we report. This is achieved considering only values $< 10$ to be valid EPKs. In reality, the real domain should be $[0, \infty)$. For the figure, the maximum value visible is 4.5.
demonstrate different values. While the literature about the risk premium is enormous (but still unanswered), to our knowledge nobody has ever analyzed the problem of a proper calibration of the precision parameter to estimate a density. As anticipated, it is possible to still follow the Bayesian paradigm and produce statistically more precise or flexible priors and distributions. We keep the analysis of both issues for future researches. In any case, although a different calibration of these variables has an impact in estimation (primarily on the length of the two supports and secondly on the shape of the functional) the difference is usually not so wide for reasonable datasets. In conclusion, the main concept behind the proposed model is robust and untouched: we always obtain better results proposing a fully time-varying EPK with respect to a partially-conditional EPK.

Conclusion

Most of the PK puzzles present in literature are affected by a non-homogeneity bias due to a partially informative physical measure. What is missing are the forward looking beliefs of the investors which cannot be captured using a stream of backward looking stock returns.

To overcome the above problem, we let the data speak as much as possible and we propose a new non-parametric estimation methodology which extracts the missing information from the implied moments of the option surface. The measure, once corrected for a risk premium, is then blended by means of a Dirichlet Process thus making the measure informative and conditional. Taking the present value of the risk-neutral measure over the conditional physical measure produces a now homogenous EPK.

We proposed an empirically intensive analysis which compares, on a daily basis and for different times-to-maturity, the effectiveness of a fully-conditional EPK with respect to a partially-conditional EPK to explain options and PK anomalies. Using the fully-conditional methodology the obtained estimations are monotonically decreasing, thus removing the PK puzzle. The proposed model is used to test in deep the behaviour of the single and aggregate densities and the relative EPKs. Results hold both for daily and yearly time-to-maturity and are in line with the neoclassical theory. From the point of view of the single densities, the behaviour of the fully-conditional objective measures are in line with the risk-neutral ones, thus producing more stable results. Given the distribution of the liquidity of D/OTM options, this is especially true into the left tail of the
distribution, where sentiment matters the most. Using the proposed model is possible to extend the horizontal support of the distribution, thus having a broader set of possible scenarios and reducing the apparent mispricing of OTM and DOTM put and call options. The power of extracting information from option prices is particularly evident and confirmed in periods of market turmoil. While a partially-conditional EPK is puzzling, the conditional one correctly prices the underlying. Results are strong on a yearly and, above all, on a daily basis: a key feature for a proper understanding of day-by-day market operations.
10 Bibliography


A Approximating to the mean and the variance of a ratio using Taylor

To compute the approximated mean and the variance of a ratio, we follow Arnold (1998) and John Wiley & Sons (1980).

Our goal is to compute $E(C)$ and $V(C)$ 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 we perform a Taylor expansion of $c(\cdot)$.

For a generic univariate function $f(x)$ that has $n$ continuous derivatives on a neighbourhood of $\theta$, the Taylor expansion of $f(x)$ is:

$$f(x) \approx \sum_{n=0}^{\infty} \frac{f^n(\theta)}{n!} (x - \theta)^n$$

By the same token, the Taylor expansion of a generic bivariate function $f(x, y)$ is defined as:

$$f(x, y) \approx \sum_{n_1=0}^{\infty} \sum_{n_2=0}^{\infty} \frac{(x - \theta_x)^{n_1} (y - \theta_y)^{n_2}}{n_1! n_2!} \frac{\partial^{n_1+n_2} f(x, y)}{\partial x^{n_1} \partial y^{n_2}}$$

The double infinite series (52) that approximates $f(x, y)$ can also be written as a first order expansion around $\theta$:

$$f(x, y) = f(\theta) + f'_x(\theta)(x - \theta_x) + f'_y(\theta)(y - \theta_y) + \text{remainder}$$

If $\theta = (E(x), E(y))$, the approximated expectation of the function is $E[f(x, y)]$, so that its expansion is:

$$E[f(X, Y)] = f(\theta) + f'_x(\theta)(0) + f'_y(\theta)(0) + O(n^{-1})$$

The bivariate second order Taylor expansion is:

$$f(x, y) = f(\theta) + f'_x(\theta)(x - \theta_x) + f'_y(\theta)(y - \theta_y)$$

$$+ \frac{1}{2} \left\{ f''_{xx}(\theta)(x - \theta_x)^2 + 2(f''_{xy}(\theta)(x - \theta_x)(y - \theta_y) + f''_{yy}(\theta)(y - \theta_y)^2 \right\} + \text{remainder}$$

where $f''_{xx}, f''_{yy}, f''_{xy}$ are the second derivative with respect to $x$ and $y$ and the reminder is defined as.
The expansion can be better approximated as:

\[
E[f(X, Y)] = f(\theta) + \frac{1}{2} \left\{ f''_{xx}(\theta) \text{Var}(X) + 2 f''_{xy}(\theta) \text{Cov}(X, Y) + f''_{yy}(\theta) \text{Var}(Y) \right\} + O(n^{-1}) \tag{60}
\]

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:

\[
E \left( \frac{A}{B} \right) = E(c(A, B)) \tag{61}
\]

\[
\approx \frac{E(A)}{E(B)} - \frac{\text{Cov}(A, B)}{E(B^2)} + \frac{\text{Var}(B) E(A)}{E(B^3)} \tag{62}
\]

By the same token, the variance can be approximated by a first order Taylor expansion:

\[
\text{Var}(f(X, Y)) = E \left\{ [f(x, y) - E(f(X, Y))]^2 \right\} \tag{63}
\]

\[
\approx E \left\{ [f(x, y) - f(E(X), E(Y))]^2 \right\} \tag{64}
\]

\[
= E \left\{ [f(\theta) + f_x'(\theta)(x - \theta_x) + f_y'(\theta)(y - \theta_y) + O(n^{-1})] - \right. \\
\left. [f(\theta) + f_x'(\theta)(0) + f_y'(\theta)(0) + O(n^{-1})] \right\} \\
= E \left\{ [f_x'(\theta)(X - \theta_x) + f_y'(\theta)(Y - \theta_y)]^2 \right\} \\
\approx f'_x(\theta)^2 \text{Var}(X) + 2 f'_x(\theta) f'_y(\theta) \text{Cov}(X, Y) + f'_y(\theta)^2 \text{Var}(Y) \tag{65}
\]

Following the same procedure:

\[
c'_A = B^{-1}, \quad c'_B = -\frac{A}{B^2}
\]

Such that:

\[
\text{Var} \left( \frac{A}{B} \right) \approx \frac{1}{E(B^2)} \text{Var}(A) - \frac{E(A)}{E(B^2)} \text{Cov}(A, B) + \frac{E(A^2)}{E(B^4)} \text{Var}(B) \tag{66}
\]

\[
= \frac{E(A^2)}{E(B^2)} \left[ \frac{\text{Var}(A)}{E(B^2)} - \frac{2 \text{Cov}(A, B)}{E(A)E(B)} + \frac{\text{Var}(B)}{E(B^2)} \right] \tag{67}
\]
B Derivation of the expected rate of return of investing in contingent claims

The time-state preference model of Arrow (1964)\(^4\) and Debreu (1959)\(^5\), introduce the Arrow-Debreu security: a very basic financial instruments also know 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: \(qe^{-rt}\) 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 \(\tau\) days is:

\[
E(u_{t,\tau}) = \frac{p - qe^{-rt}}{qe^{-rt}}
\]

\[= \left(\frac{1}{M_{t,\tau}}\right) - 1 \tag{68}\]

\[= \left(\frac{1}{PK}\right) - 1 \tag{69}\]
<table>
<thead>
<tr>
<th>Year</th>
<th>Mean</th>
<th>St.D.</th>
<th>Mean</th>
<th>St.D.</th>
<th>Mean</th>
<th>St.D.</th>
<th>Mean</th>
<th>St.D.</th>
<th>Mean</th>
<th>St.D.</th>
</tr>
</thead>
<tbody>
<tr>
<td>2002</td>
<td>1.48E-06</td>
<td>1.33E-07</td>
<td>7.53E-03</td>
<td>1.02E-03</td>
<td>0.927</td>
<td>0.005</td>
<td>0.100</td>
<td>0.009</td>
<td>0.984</td>
<td>0.01</td>
</tr>
<tr>
<td>2003</td>
<td>1.33E-06</td>
<td>2.67E-07</td>
<td>7.76E-03</td>
<td>2.21E-03</td>
<td>0.926</td>
<td>0.007</td>
<td>0.107</td>
<td>0.007</td>
<td>0.987</td>
<td>0.01</td>
</tr>
<tr>
<td>2004</td>
<td>1.05E-06</td>
<td>4.60E-08</td>
<td>7.29E-03</td>
<td>2.21E-03</td>
<td>0.930</td>
<td>0.002</td>
<td>0.106</td>
<td>0.003</td>
<td>0.990</td>
<td>0.00</td>
</tr>
</tbody>
</table>

Table 1: Yearly Mean and Standard Deviation 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:

$$\log \frac{S_t}{S_{t-1}} = \mu + \epsilon_t$$

$$\sigma_t = \omega + \beta \sigma^2_{t-1} + \alpha \epsilon^2_{t-1} + 1_t \gamma \epsilon^2_{t-1}$$

Persistence is defined as: $= \alpha + \beta + \gamma/2$
<table>
<thead>
<tr>
<th>Year</th>
<th>( \tilde{\omega} ) Mean</th>
<th>St.D.</th>
<th>( \tilde{\alpha} ) Mean</th>
<th>St.D.</th>
<th>( \tilde{\beta} ) Mean</th>
<th>St.D.</th>
<th>( \tilde{\gamma} ) Mean</th>
<th>St.D.</th>
<th>Persistency</th>
</tr>
</thead>
<tbody>
<tr>
<td>2002</td>
<td>3.14E-06</td>
<td>2.74E-06</td>
<td>9.64E-04</td>
<td>9.05E-02</td>
<td>0.866</td>
<td>0.091</td>
<td>0.216</td>
<td>0.143</td>
<td>0.975</td>
</tr>
<tr>
<td>2003</td>
<td>3.29E-06</td>
<td>4.04E-06</td>
<td>1.31E-03</td>
<td>1.30E-01</td>
<td>0.866</td>
<td>0.130</td>
<td>0.213</td>
<td>0.207</td>
<td>0.973</td>
</tr>
<tr>
<td>2004</td>
<td>2.08E-06</td>
<td>2.96E-06</td>
<td>3.55E-03</td>
<td>1.74E-01</td>
<td>0.832</td>
<td>0.174</td>
<td>0.293</td>
<td>0.301</td>
<td>0.981</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Year</th>
<th>( \tilde{\omega} ) Mean</th>
<th>St.D.</th>
<th>( \tilde{\alpha} ) Mean</th>
<th>St.D.</th>
<th>( \tilde{\beta} ) Mean</th>
<th>St.D.</th>
<th>( \tilde{\gamma} ) Mean</th>
<th>St.D.</th>
<th>Persistency</th>
</tr>
</thead>
<tbody>
<tr>
<td>2002</td>
<td>4.02E-06</td>
<td>5.49E-06</td>
<td>5.84E-03</td>
<td>1.19E-02</td>
<td>0.845</td>
<td>0.093</td>
<td>0.271</td>
<td>0.165</td>
<td>0.985</td>
</tr>
<tr>
<td>2003</td>
<td>3.57E-06</td>
<td>6.23E-06</td>
<td>4.48E-03</td>
<td>8.95E-03</td>
<td>0.829</td>
<td>0.206</td>
<td>0.273</td>
<td>0.327</td>
<td>0.969</td>
</tr>
<tr>
<td>2004</td>
<td>1.69E-06</td>
<td>1.43E-06</td>
<td>2.05E-03</td>
<td>4.95E-03</td>
<td>0.827</td>
<td>0.110</td>
<td>0.315</td>
<td>0.198</td>
<td>0.986</td>
</tr>
</tbody>
</table>

Table 2: Upper panel: Yearly Mean and Standard Deviation 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 (OTM) SPX options with Filtered Historical Innovations (FHS). Optimization method: Simplex.

The GJR GARCH (1,1) model under the risk-neutral measure is:

\[
\log \frac{S_t}{S_{t-1}} = \mu + \epsilon_t
\]

\[
\sigma_t = \tilde{\omega} + \tilde{\beta} \sigma_{t-1}^2 + \tilde{\alpha}\epsilon_{t-1}^2 + \tilde{\gamma}\epsilon_{t-1}^2
\]

Persistence is defined as: \( \tilde{\alpha} + \tilde{\beta} + \tilde{\gamma}/2 \)

Bottom panel: same as the top table but with Gaussian innovations.
<table>
<thead>
<tr>
<th>Moneyness</th>
<th>Maturity</th>
<th>Mean</th>
<th>St.D.</th>
<th>Mean</th>
<th>St.D.</th>
<th>Mean</th>
<th>St.D.</th>
</tr>
</thead>
<tbody>
<tr>
<td>$K/S$</td>
<td>$\tau &lt; 60$</td>
<td>0.77</td>
<td>1.20</td>
<td>2.54</td>
<td>3.20</td>
<td>8.69</td>
<td>7.63</td>
</tr>
<tr>
<td></td>
<td>$60 \leq \tau \leq 160$</td>
<td>38.97</td>
<td>9.59</td>
<td>33.36</td>
<td>7.75</td>
<td>28.37</td>
<td>5.21</td>
</tr>
<tr>
<td></td>
<td>$\tau &gt; 160$</td>
<td>0.99</td>
<td>0.69</td>
<td>0.69</td>
<td>0.68</td>
<td>0.27</td>
<td>0.36</td>
</tr>
<tr>
<td>&lt; 0.85</td>
<td>Put Price ($)</td>
<td>8.43</td>
<td>7.60</td>
<td>19.12</td>
<td>11.63</td>
<td>34.80</td>
<td>15.67</td>
</tr>
<tr>
<td></td>
<td>$\sigma_{BS}$</td>
<td>22.38</td>
<td>7.05</td>
<td>21.88</td>
<td>5.35</td>
<td>21.28</td>
<td>3.91</td>
</tr>
<tr>
<td></td>
<td>Bid-Ask (%)</td>
<td>0.19</td>
<td>0.21</td>
<td>0.10</td>
<td>0.05</td>
<td>0.06</td>
<td>0.03</td>
</tr>
<tr>
<td></td>
<td>Observations (n.)</td>
<td>1798</td>
<td>2357</td>
<td>2845</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>0.85 – 1.00</td>
<td>Put Price ($)</td>
<td>7.45</td>
<td>7.81</td>
<td>15.79</td>
<td>12.60</td>
<td>34.82</td>
<td>18.91</td>
</tr>
<tr>
<td></td>
<td>$\sigma_{BS}$</td>
<td>17.55</td>
<td>5.80</td>
<td>17.31</td>
<td>5.02</td>
<td>17.61</td>
<td>4.00</td>
</tr>
<tr>
<td></td>
<td>Bid-Ask (%)</td>
<td>0.34</td>
<td>0.46</td>
<td>0.17</td>
<td>0.23</td>
<td>0.07</td>
<td>0.04</td>
</tr>
<tr>
<td></td>
<td>Observations (n.)</td>
<td>3356</td>
<td>2136</td>
<td>2314</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1.00 – 1.15</td>
<td>Call Price ($)</td>
<td>0.34</td>
<td>0.43</td>
<td>0.85</td>
<td>1.61</td>
<td>3.96</td>
<td>5.60</td>
</tr>
<tr>
<td></td>
<td>$\sigma_{BS}$</td>
<td>34.87</td>
<td>12.35</td>
<td>23.34</td>
<td>8.27</td>
<td>18.87</td>
<td>4.48</td>
</tr>
<tr>
<td></td>
<td>Bid-Ask (%)</td>
<td>1.81</td>
<td>0.47</td>
<td>1.49</td>
<td>0.71</td>
<td>0.83</td>
<td>0.78</td>
</tr>
<tr>
<td></td>
<td>Observations (n.)</td>
<td>1633</td>
<td>2288</td>
<td>3154</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>&gt; 1.15</td>
<td>Call Price ($)</td>
<td>0.34</td>
<td>0.43</td>
<td>0.85</td>
<td>1.61</td>
<td>3.96</td>
<td>5.60</td>
</tr>
<tr>
<td></td>
<td>$\sigma_{BS}$</td>
<td>34.87</td>
<td>12.35</td>
<td>23.34</td>
<td>8.27</td>
<td>18.87</td>
<td>4.48</td>
</tr>
<tr>
<td></td>
<td>Bid-Ask (%)</td>
<td>1.81</td>
<td>0.47</td>
<td>1.49</td>
<td>0.71</td>
<td>0.83</td>
<td>0.78</td>
</tr>
<tr>
<td></td>
<td>Observations (n.)</td>
<td>1633</td>
<td>2288</td>
<td>3154</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Table 3: Mean, Standard Deviation (St.D.) and Number of observations (n.) for each moneyness/maturity category of the out-of-the-money (OTM) and deep-out-of-the-money (DOTM) SPX options observed on each Wednesdays from January 2, 2002 until December 29, 2004.

Moneyness: ($K/S$). Maturity: $\tau$ is measured in calendar days and divided in short, medium and long time-to-maturity.

$\sigma_{BS}$: Black & Scholes implied volatility using the market prices and inverting numerically the Black & Scholes function.

Bid-ask spread: $100 \times \frac{(\text{Ask price} - \text{Bid price})}{\text{Market price}}$ where the market price is equal to the average of the bid and ask prices.

Filtering criteria: $10 \leq \tau \leq 360$ days, $\sigma_{BS} \leq 70\%$, price $> 0.05$ for a total sample of 29,201 observations.

<table>
<thead>
<tr>
<th>$\alpha_{t,T}^{*}$</th>
<th>DOTM</th>
<th>Mean</th>
<th>Median</th>
<th>St.D</th>
<th>Max. value</th>
<th>Min. Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\alpha_{t,T}^{*}$</td>
<td>DOTM</td>
<td>1.5563</td>
<td>1.4808</td>
<td>0.30961</td>
<td>2.5022</td>
<td>1.0376</td>
</tr>
<tr>
<td>$\alpha_{t,T}^{*}$</td>
<td>OTM</td>
<td>1.6302</td>
<td>1.6138</td>
<td>0.34298</td>
<td>2.5032</td>
<td>1.0264</td>
</tr>
</tbody>
</table>

Table 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.
Risk-neutral Measure - fully-conditional Risk Physical Measure: $\Delta_{t,T}^{\text{Cond.}} = q_{t,T} - p_{t,T}^\dagger$

<table>
<thead>
<tr>
<th>Year</th>
<th>$\tau$</th>
<th>$\Delta_t$ Mean</th>
<th>$\Delta_t$ Var</th>
<th>$\Delta_t$ Skewness</th>
<th>$\Delta_t$ Kurtosis</th>
</tr>
</thead>
<tbody>
<tr>
<td>2002</td>
<td>Short</td>
<td>0</td>
<td>0</td>
<td>0.0530</td>
<td>0.3765</td>
</tr>
<tr>
<td></td>
<td>Medium</td>
<td>0</td>
<td>0</td>
<td>0.0427</td>
<td>0.2592</td>
</tr>
<tr>
<td></td>
<td>Long</td>
<td>0</td>
<td>0</td>
<td>0.0295</td>
<td>0.1636</td>
</tr>
<tr>
<td>2003</td>
<td>Short</td>
<td>0</td>
<td>0</td>
<td>0.0081</td>
<td>0.0777</td>
</tr>
<tr>
<td></td>
<td>Medium</td>
<td>0</td>
<td>0</td>
<td>-0.0050</td>
<td>0.0018</td>
</tr>
<tr>
<td></td>
<td>Long</td>
<td>0</td>
<td>0</td>
<td>0.0033</td>
<td>0.0345</td>
</tr>
<tr>
<td>2004</td>
<td>Short</td>
<td>0</td>
<td>0</td>
<td>0.0943</td>
<td>0.6352</td>
</tr>
<tr>
<td></td>
<td>Medium</td>
<td>0</td>
<td>0</td>
<td>-0.0005</td>
<td>0.0018</td>
</tr>
<tr>
<td></td>
<td>Long</td>
<td>0</td>
<td>0</td>
<td>0.1408</td>
<td>0.8751</td>
</tr>
</tbody>
</table>

Risk-neutral Measure - partially-conditional Risk Physical Measure: $\Delta_{t-T}^{\text{Part. Cond.}} = q_{t,T} - p_t - \Delta_{t-T}$

<table>
<thead>
<tr>
<th>Year</th>
<th>$\tau$</th>
<th>$\Delta_t$ Mean</th>
<th>$\Delta_t$ Var</th>
<th>$\Delta_t$ Skewness</th>
<th>$\Delta_t$ Kurtosis</th>
</tr>
</thead>
<tbody>
<tr>
<td>2002</td>
<td>Short</td>
<td>0</td>
<td>0</td>
<td>0.1583</td>
<td>1.1086</td>
</tr>
<tr>
<td></td>
<td>Medium</td>
<td>0</td>
<td>0</td>
<td>0.0822</td>
<td>0.4956</td>
</tr>
<tr>
<td></td>
<td>Long</td>
<td>0</td>
<td>0</td>
<td>0.0227</td>
<td>0.1503</td>
</tr>
<tr>
<td>2003</td>
<td>Short</td>
<td>0</td>
<td>0</td>
<td>0.0078</td>
<td>0.1406</td>
</tr>
<tr>
<td></td>
<td>Medium</td>
<td>0</td>
<td>0</td>
<td>-0.0377</td>
<td>-0.1361</td>
</tr>
<tr>
<td></td>
<td>Long</td>
<td>0</td>
<td>0</td>
<td>-0.0319</td>
<td>-0.1095</td>
</tr>
<tr>
<td>2004</td>
<td>Short</td>
<td>0</td>
<td>0</td>
<td>0.2885</td>
<td>1.9063</td>
</tr>
<tr>
<td></td>
<td>Medium</td>
<td>0</td>
<td>0</td>
<td>0.3303</td>
<td>2.0117</td>
</tr>
<tr>
<td></td>
<td>Long</td>
<td>0</td>
<td>0</td>
<td>0.3571</td>
<td>1.9243</td>
</tr>
</tbody>
</table>

Table 5: 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, RP=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 $\text{ne} -0.03$ are set equal to 0.
Figure 1: 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 2: Single day \((t = 61)\) all range of times-to-maturity probability density functions, PDF, (left column) and Conditional empirical pricing kernels, EPK, \((M^\dag_{t,t+\tau})\) (right column) with starting \(\alpha_{t,t+24}^* = 1.75\) and decreasing and R.P. = 8%.
Figure 3: Single day \((t = 4)\) all range of times-to-maturity probability density functions, PDF, (left column) and Conditional, \((M_{t,t+\tau}^\dagger)\), partially-conditional empirical pricing kernels, EPK \((M_{t-\Delta t,t+\tau})\) (right column), with 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 4: Daily $\Delta_t$ 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: $\Delta_t$ Mean and $\Delta_t$ Variance. Bottom: $\Delta_t$ Kurtosis and $\Delta_t$ Skewness.
Figure 5: 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 amount of DOTM (left) and OTM (right) options and relative $\alpha_{t,T}$. The former are represented with a continuous blue line and use the left axis scale, the latter with a green dotted line and use the right scale.
Figure 6: 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), \( p_{t,t+\tau} \), fully conditional physical measure (black) \( p_{t,t+\tau} \), and risk-neutral measure (red) \( q^{\pi}_{t,t+\tau} \).

The first column is calculated with \( \alpha_{t,t+24} = 2 \) and decreasing and R.P. = 8%, the second column with decreasing \( \alpha_{t,t+24} = 10 \) 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 7: 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), \( p_{l,t}^{\Delta_t,t+\tau} \), fully conditional physical measure (black) \( p_{l,t}^{\Delta_t,t+\tau} \) and risk-neutral measure (red) \( q_{l,t}^{\Delta_t,t+\tau} \). The first column is calculated with \( \alpha_{l,t+150} = 1 \) and decreasing and R.P. = 8%, the second column with decreasing \( \alpha_{l,t+150} = 8.75 \) and decreasing and R.P. = 8% and the third column with \( \alpha_{l,t+150} = 1 \) 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 8: 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 decreasing \(\alpha_{t,t+24}^* = 10\) 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 9: 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) $p^{t^*}_{\Delta t,t+\tau}$ 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 decreasing $\alpha_{t,t+150} = 8.75$ 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 10: 2002-2003-2004 yearly estimates of the FHS partially-conditional physical measure (green), $p_{t-\Delta t, T}^{\text{Year}}$, conditional physical measure (red) ($p_{t, T}^{\text{Year}}$) and risk-neutral measure (blue) ($q_{t, T}^{*\text{Year}}$) with $\alpha_{t, T}^* = 2$ and R.P. = 4%. Single graphs are divided horizontally in short-medium and long time-to-maturity and vertically for years.
Figure 11: 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, EPK, with $\alpha_{t,t+31}^* = 1.75$, 50,000 simulations 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, EPK, with $\alpha_{t,t+94}^* = 0.75$, 50,000 simulations and R.P. = 4%.
Figure 12: Single day ($t = 90$) all range of times-to-maturity fully and partially-conditional expected return ($\upsilon_{t,t+\tau}$), ($\upsilon_{t-\Delta,t+\tau}$) from investing in a generic contingent claim estimated with $\alpha_{t,t+31}^* = 1.75$ and decreasing, 50,000 simulations and R.P. = 4%.61.
Figure 13: Single day ($t = 63$) all range of times-to-maturity fully and partially-conditional expected return ($\upsilon_{t,t+\tau}^\dagger$, $\upsilon_{t-\Delta_t,t+\tau}$) from investing in a generic contingent claim estimated with $\alpha_{t,t+38}^* = 1.75$ and decreasing, 50,000 simulations and R.P. = 8%.
Figure 14: Single day ($t = 41$) all range of times-to-maturity probability density functions, PDF, (left column), conditional empirical pricing kernels, EPK, $M_{t,t+\tau}$ and partially-conditional EPKs $M_{t-\Delta t,t+\tau}$ (right column) with $\alpha_{t,t+38} = 2.5$ and decreasing, 50,000 simulations and R.P. = 4%.
Figure 15: Single day \((t = 41)\) with focus on the single times-to-maturity \((\tau s = 38/73/164/255/346)\) conditional empirical pricing kernels, EPKs, \(\hat{M}_{t,t+\tau}^{164}\) and partially-conditional empirical pricing kernels, EPKs, \(\hat{M}_{t−\Delta,t+\tau}\) with \(\alpha_{t,t+38}^* = 2.5\) and decreasing, 50,000 simulations and R.P. = 4%.
Figure 16: 2002-2003-2004 yearly estimates of the FHS (green) and Gauss. (red) conditional $M_{t,T}^{\text{spec}}$ and FHS (black dotted) and Gauss. (blue dotted) conditional $M_{t,\Delta t,T}^{\text{spec}}$ 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.