The Pricing Kernel is U-shaped

Tobias Sichert*
sichert@finance.uni-frankfurt.de

January 9, 2020

Job Market Paper

Abstract

Numerous studies find S-shaped pricing kernels, which is conflicting with standard theory. In contrast to that, based on a novel GARCH model with structural breaks, I show that the pricing kernel is consistently U-shaped. The results are robust to variations in the methodology and hold for several major international stock market indices. The new U-shaped pricing kernel estimates help to explain cross-sectional stock return anomalies. Furthermore, a pricing kernel mimicking trading strategy yields sizable Sharpe ratios. Finally, the empirical results can be explained well by a model with a variance-dependent pricing kernel, but only if structural breaks are included in the model.

JEL classification: G10, G12, G13

Keywords: Pricing kernel, stochastic discount factor, pricing kernel puzzle, volatility forecasting, options, GARCH, cross section of returns, anomalies

*I would like to thank David Alexander, Daniel Andrei, Peter Christoffersen, Marc Crummenerl, Bob Dittmar, Mátyás Farkas, Mathieu Fournier (AFA discussant), Michael Hasler, Burton Hollifield, Markus Huggenberger, Kris Jacobs, Holger Kraft, Jan Pieter Krahnen, Tom McCurdy, Christoph Meinerding, Chay Ornthanalai, Christian Schlag, Maik Schmeling, Paul Schneider, Lorenzo Schoenleber, Mike Simutin, Marti Subrahmanyam, Roméo Tédongap, Julian Thimme, Rüdiger Weber, Amir Yaron, conference participants of the AFA, NFA, TADC, DGF, SGF, AFFI and seminar participants at the Universities of Carnegie Mellon (Tepper), Frankfurt (Goethe), Toronto (Rotman) and Zurich (UZH) for valuable comments and suggestions. An earlier version of this paper was circulated under the title "Structural Breaks in the Variance Process and the Pricing Kernel Puzzle".

1 Introduction

The stochastic discount factor is the central object of interest in modern asset pricing. It conveys valuable information about the assessment of risks by investors and tells us how real-world probabilities are transformed into risk-neutral probabilities. In models with a representative investor, it additionally relates to the agent's marginal utility and therefore speaks about preferences.

A natural way to get closer to the object of interest and to learn about these fundamental economic questions is to look at the projection of the stochastic discount factor on returns of a broad market index as a proxy for aggregate wealth. This projection is called the pricing kernel in the following. A large body of literature estimates the pricing kernel as the ratio of risk-neutral return destiny, usually inferred from option prices, over physical return densities obtained from a statistical model chosen by the researcher. While many classical theories, like the CAPM, predict that the pricing kernel is monotonically decreasing in returns, empirical estimates show that this is not necessarily the case. This stylized fact is called the pricing kernel puzzle and was first documented by Jackwerth (2000), Aït-Sahalia & Lo (2000) and Rosenberg & Engle (2002), and has since then been confirmed by many others.[1] Most studies document that the pricing kernel plotted against returns has the shape of a rotated S, meaning, that it is generally downward-sloping but has a hump around zero, as illustrated in Figure 1. However, these findings are incompatible with standard, rational asset pricing theory.

This paper argues that the findings of S-shaped pricing kernels are spurious and caused by a previously unrecognized bias in the volatility forecasts. By overestimating future volatility in calm market periods, the standard estimation methodology leads to S-shaped pricing kernels in calm times, and U-shaped ones otherwise. Using a new volatility model that nests the previous approaches, but corrects for the forecasting bias, leads to consistently U-shaped pricing kernel estimates.

To develop the argument, first note that for estimating the pricing kernel, two key quantities are required: the risk-neutral index return density and a physical return density forecast. The first quantity is inferred from option prices and accepted standard estimation methods exist. However, the second quantity requires some parametric assumptions. The literature has recognized early on that an important ingredient to

[1] The literature on the pricing kernel puzzle has become too large to fully describe here. For more details see e.g. Cuesdeanu & Jackwerth (2018), who provide an excellent and comprehensive overview on the existing empirical, theoretical and econometric literature on the pricing kernel puzzle.

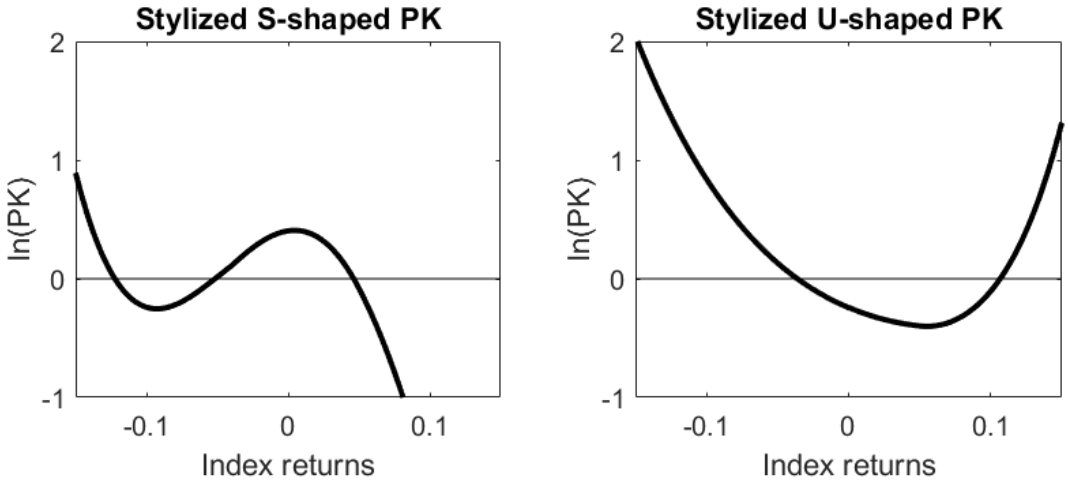


Figure 1: Stylized Pricing Kernels

predicting return distributions is the volatility forecast. It is well known, e.g., from the vast literature on GARCH models in finance, that volatility is time-varying and clustered. In contrast to conventional knowledge, I will show that the workhorse GARCH model cannot capture the degree of volatility clustering observed in the data, which then impacts the PK estimation.

First, I estimate a change-point (CP) GARCH model via maximum likelihood to identify points where the parameters of the GARCH process change. The potential existence of parameter changes in GARCH processes has been documented and discussed in the econometric literature for many years (e.g., Diebold 1986). I suggest a new GARCH model with structural breaks, and in 25 years of S&P 500 return data I estimate five different regimes that exhibit significantly different volatility dynamics. In particular, this regime-switching structure is able to capture the clustering of volatility by identifying phases where market volatility remains below its long-term average for many years, which is not possible using a standard GARCH model. Therefore I also use the matching long sample of S&P 500 options over the period 1996-2015, to include the different market phases.

Next, I show that standard estimation methods, which virtually all rely on GARCH models with fixed parameters, yield S-shaped pricing kernels in times of low variance, and U-shaped ones otherwise. Furthermore, when replacing the standard GARCH with the CP-GARCH while leaving everything else unchanged, the S-shaped pricing kernels disappear altogether. This change in shape is statistically significant.

To give an intuitive explanation for the shift from S- to U-shape, recall that the pricing kernel is the point-wise ratio of the risk-neutral density over the physical density. The risk-neutral density is the same for any approach and the change comes from the differences in physical densities. The key driver is that the overestimated volatility leads to a return distribution forecast that is too wide, as it exhibits too much probability mass in the tails and too little in the center. The excess weight in the left tail is not strong enough to change the downward slope, but the excess weight in the right tail makes the estimated pricing kernel slope downward instead of upward. The corresponding lack of probability mass in the center in turn causes the locally increasing part, and hence the typical S-shape.

The paper then shows that the new pricing kernel estimates have several interesting economic properties and yield new insights with respect to a number of key questions in asset pricing. First, I study Euler equation errors, which is a non-parametric and standard way to test a candidate pricing kernel. Euler equation errors can be interpreted economically as pricing errors, and are commonly referred to as "alphas". I show that the new PK estimates price the S&P 500 index correctly, while other standard estimates yield large alphas of 4% per month and higher. A closer look at the timing of the Euler equation errors further highlights the importance of a time-varying pricing kernel.

Furthermore, the U-shaped PKs estimated from the options market help to explain the cross-section of stock returns. The key implication of the U-shaped PK is that payoffs in states of the world where the index return is large and positive, are expensive. To test this prediction quantitatively and in order to isolate the impact of the upward sloping part, I compare the pricing performance of my U-shaped PK estimates to a monotonically decreasing version of my estimates. For the latter, graphically speaking, I find the lower point of the "U" and make it monotonic by replacing the increasing part with a flat line. In addition, I use the monotonically decreasing CAPM and the Fama & French (1993) 3 factor model as a benchmark. As test assets I use several standard portfolio sorts such as value, size, momentum, beta, industry and volatility. The results show that the U-shaped PK reduces the portfolio alphas by at least 25% and up to 95%, compared to either the monotonic version of the PK or the two factor models, respectively. While most of the asset pricing literature is concerned with "downside-risk", the new results suggest that there is also a sizable "upside-risk" premium. To incorporate the U-shaped PK into mainstream asset pricing, I show how one can add the "U" to the FF3 model. Augmenting the standard FF3 model with the normalized, squared market return as a fourth factor reduces the alphas of the industry, beta and

volatility portfolios to almost zero.

Third, I show that the PK can be traded as a synthetic derivative in the options market. This has the natural interpretation as the mean-variance optimal trading strategy in index options, since the famous Hansen & Jagannathan (1991) bound predicts that the pricing kernel itself is the asset with the highest attainable Sharpe ratio in the economy. The trading strategy based on the new PK estimates indeed delivers sizable annualized Sharpe ratios of up to 3.1. On the contrary, S-shaped PK estimates yield relatively low returns and Sharpe ratios. By reversing the Hansen-Jagannathan argument, one can conclude that the PK delivering the higher Sharpe ratios is closer to the true PK.

Fourth, the new PK estimates are a good predictor for future excess market returns (equity risk premium). Especially for short horizons it outperforms other well-known predictors such as the variance risk premium (Bollerslev et al. 2009) or the SVIX (Martin 2016). The new predictor has the appealing feature that the predictive relationship is directly motivated by asset pricing theory and can be updated continuously.

In the final part of the paper, I demonstrate that the U-shaped pricing kernel is consistent with the explanation brought forward by Christoffersen et al. (2013). The authors suggest a variance-dependent stochastic discount factor, which is increasing in variance and decreasing in returns. Since volatility is high both for large negative and large positive returns, the variance risk premium causes the projection of the stochastic discount factor on the index returns to be U-shaped. I extend the model by combining their variance-dependent stochastic discount factor with the change-point GARCH model and estimate the model on option prices and returns jointly. The results show that only the model with the structural breaks can capture the time-variation of the PK in the data.

In term of robustness, all the aforementioned findings on the pricing kernel estimates are also documented for several other major international equity indices, namely the FTSE 100, EuroStoxx 50 and DAX 30. Furthermore, the empirical results are robust to numerous variations in the methodology. While the benchmark analysis is kept as non-parametric as possible, the robustness section includes the popular approach, where the physical density is obtained directly from a GARCH model simulation. Moreover, a VIX-based volatility forecast is studied as well as the Corsi (2009) realized volatility model based on high frequency data. Lastly, I test different popular GARCH model specifications, consider various time horizons and also vary several other methodological details.

Overall, the paper provides novel semi-parametric evidence on the time series behavior of pricing kernel estimates, which has not been subject to much research.[2] Many studies use a sample of option data that are much shorter than the 20 years used here, and almost all studies document the typical S-shape.[3] While a few studies document U-shaped pricing kernels (Christoffersen et al. 2013, Grith et al. 2013), they at the same time find S-shapes in some periods of their sample, and do not address this conflict. However, the two shapes are theoretically incompatible. Although theoretical models can explain either one of the shapes, neither can explain both.[4] In addition, the S-shape is incompatible with an economy with one representative investor and rational expectations, which is the backbone of most modern theoretical asset pricing models (Hens & Reichlin 2013). It is therefore not surprising that theoretical explanations for the S-shape have to turn to heterogeneous investors (e.g. Ziegler 2007), market incompleteness (Hens & Reichlin 2013), probability misestimation (e.g. Polkovnichenko & Zhao 2013), reference-dependent preferences (Grith et al. 2017), ambiguity aversion (Cuesdeanu 2016) and other behavioral factors (Figlewski 2018), among others. However, none of the models can offer a risk-based explanation of the hump coming with the S-shape of the PK. In contrast, the U-shape can be explained by the variance risk premium (Christoffersen et al. 2013), which itself is empirically well established (e.g. Carr & Wu 2009). Hence, any model that generates a variance risk premium can at least qualitatively also generate a U-shaped pricing kernel. Altogether, in the existing literature the two different shapes seem to pose two different puzzles. Their potential co-existence would be a challenge for theory and would pose a third puzzle.

The results in this paper challenge the existence of an S-shaped pricing kernel, which has almost become consensus in the literature and is sometimes even considered a stylized fact. Correctly estimating the shape of the PK provides valuable guidance for theorists

[2]The relevance of regime switches for the pricing kernel puzzle has been advocated in the literature before, but in a theoretical and not empirical context. Chabi-Yo et al. (2008) and Grith et al. (2017) develop models with a representative agent and state dependent preferences, that feature pricing kernels which are monotonically decreasing in returns in each state. Their non-monotonic weighting of the different linear pricing kernels results in an *unconditional* S-shaped pricing kernel.

[3]E.g., Jackwerth (2000), Aït-Sahalia & Lo (2000), Rosenberg & Engle (2002), Liu et al. (2009), Polkovnichenko & Zhao (2013), Figlewski & Malik (2014), Beare & Schmidt (2016), Belomestny et al. (2015), Cuesdeanu & Jackwerth (2018), Barone-Adesi et al. (2016), and Grith et al. (2017).

[4]This is true for both potential ways of coexistence: the shapes could alternate over time, or a combination of the two could be present at the same point in time, resembling a W-shape.

when validating the predictions of their models. It established the "data", in the sense of a stylized property that theory can build on. In sum, the results help to identify the kind of asset pricing model required to explain the joint pricing of options and the index, which is still considered a major challenge in finance (Bates 1996).

The remainder of the paper proceeds as follows. Section 2 first introduces the change-point GARCH model and then presents the data, estimation methodology and estimation results. Next, it analyzes the volatility forecasts of the model. Section 3 shows the empirical pricing kernels obtained with the new model and contrasts them with the standard findings. It furthermore provides a detailed analysis of how the different GARCH models and volatility forecasts drive the results. Section 4 contains several economic applications of the new pricing kernel estimates. Section 5 presents the partial equilibrium model that explains the empirical findings. Section 6 conducts several robustness checks, including the international evidence, and Section 7 concludes. The Appendix collects methodological details. In addition, I provide a detailed online Appendix, which contains additional results and can be found on my website.

2 A change-point GARCH Model

2.1 Motivation

Three quantities are required to estimate conditional pricing kernels (PKs) empirically: the risk-free rate, conditional risk-neutral probabilities and conditional physical (objective) probabilities. The estimation of the first is an easy task and, since the discounting effects over typical horizons of around one month are low, it is not a crucial parameter in any case. The estimation of the second quantity is not straightforward, but established and well understood methods exist. The remaining third quantity, however, the conditional physical probability, is not easily quantifiable and requires a minimum of parametric assumptions. The method of conditioning the return distribution estimate is later shown to be crucial for the results.

Some of the first studies on the pricing kernel puzzle use a kernel density estimation of past raw index returns (e.g. Aït-Sahalia & Lo 2000). Many other studies agree that it is important to condition the estimate on current market volatility (see e.g. Jackwerth 2000, Rosenberg & Engle 2002, Beare & Schmidt 2016). Almost all studies use a GARCH model for this, since it is the workhorse model for stochastic volatility in finance. However, some econometric papers (e.g. Diebold 1986, Mikosch & Stărică 2004)

suggest that a standard GARCH model with fixed parameters does not fit a long time series very well. The high degree of variance persistence, in particular for long time series, has been questioned. It is argued that estimated dynamics close to a unit root process are caused by changes in the parameters of the GARCH process, which are ignored if the model is specified with fixed parameters. Hence, one potential solution is to allow for structural breaks where the parameters of the GARCH model may change. Among others, the studies of Bauwens et al. (2014), Augustyniak (2014) and Klaassen (2002) show that GARCH models with switching parameters outperform the standard model both in- and out-of-sample.

In the empirical analysis below, I show that the GARCH model with structural breaks has another important advantage over the standard model. Despite its high degree of persistence, the multi-period variance forecasts of the standard GARCH model always revert to the long run mean rather quickly. This leads to a systematic overestimation or underestimation of future volatility when there are extended periods of time, where the volatility is constantly below or above its time-series average (volatility clustering). For the applications at hand, this will have decisive implications. I use the CP-GARCH model in the benchmark analysis because it has the particular advantage that it nests most existing methods, as opposed to several alternative models in the robustness section that can also (partly) overcome the bias.

2.2 Dynamics of the CP-GARCH model

One way to make the standard GARCH model more flexible is to use a change-point (CP) model.[5] Such a CP-GARCH model is laid out in the following. Section 3.2 shows how the model is used to construct a conditional return distribution.

There are two prominent GARCH models often used for modeling the dynamics of stocks as well as for option pricing. The first is the NGARCH model of Duan (1995), the second is the Heston-Nandi (HN) GARCH model of Heston & Nandi (2000). The

[5] A close relative of a change-point model is the probably more popular Markov-switching (MS) model, also called regime-switching GARCH. For several reasons the CP model is preferred here over the MS model. The most important one is that paths simulated from the MS GARCH estimates of Augustyniak (2014) exhibit unreasonable dynamics. In particular, a significant number of paths have volatility levels that vastly exceed any ever observed level in the data. This is because in a two state MS GARCH model, the high variance state (over) fits the extreme positive and negative daily returns that occur from time to time. Therefore, one would probably need at least three or four states to produce reasonable dynamics. A reliable estimation of this model would be very difficult, and the number of regimes would be similar to the five regimes used in the CP model below.

main analysis uses the HN-GARCH model because it conveniently allows for closed form option pricing, and the robustness section shows that the results also hold for the NGARCH model.

The dynamics of the asymmetric HN-GARCH model with structural breaks are:

$$ln\Big(\frac{S_t}{S_{t-1}}\Big) = r_t + \Big(\mu_{y_t} - \frac{1}{2}\Big)h_t + \sqrt{h_t}z_t, \tag{1}$$

$$h_t = \omega_{y_t} + \beta_{y_t}h_{t-1} + \alpha_{y_t}\Big(z_{t-1} - \gamma_{y_t}\sqrt{h_{t-1}}\Big)^2, \tag{2}$$

$$z_t \sim N(0,1), \tag{3}$$

where S_t is the stock's spot price at time t, r_t is the daily continuously compounded interest rate, z_t are return innovations and h_t is the conditional variance, and y_t is an integer random variable taking values in [1, K + 1].[6] The latent state process y_t is first order Markovian with the absorbing and nonrecurrent transition matrix

$$\mathbf{P} = \begin{bmatrix} p_{11} & 1-p_{11} & 0 & \dots & 0 & 0 \\ 0 & p_{22} & 1-p_{22} & \dots & 0 & 0 \\ \dots & \dots & \dots & \dots & \dots & \dots \\ 0 & 0 & 0 & \dots & p_{KK} & 1-p_{KK} \\ 0 & 0 & 0 & \dots & 0 & 1 \end{bmatrix}.$$

This transition matrix characterizes a change-point model (CP-GARCH) with K breaks. A standard GARCH model with fixed parameters (FP) can be obtained by setting $K = 0$.

The economic interpretation of the change-point model is that there are different regimes in the market, and when they end, fundamentals change. These changes are so dramatic, that the standard and already dynamic model cannot capture them, but a full new parameterization of the model is required in each regime. The estimation below shows that the identified regimes are several years long and are related to business cycles.

[6] I use the conventional normal innovations here. Many papers in the econometric literature argue that the shock in GARCH models are not normally distributed. However, the assumption is required to be able to solve the model in Section 5. Nevertheless, in Section 6.6.1 I show that the empirical results are fully robust to using e.g. a t-distribution.

2.3 Estimation

2.3.1 Data and Methodology

The stock data used to estimate both the fixed parameter and switching GARCH are daily S&P 500 log returns from 02.01.1992 to 31.08.2015. The sample is chosen to match the available option data from 02.01.1996 to 31.08.2015. The earlier start date is used because the analysis of an even longer sample shows that the regime, that prevails in 1996, starts around January 1992. As a robustness check, the fixed parameter model is also estimated over the longer sample from 02.01.1986 to 30.06.2016 and the obtained parameters are very similar.

The GARCH model is estimated via maximum likelihood. For the Heston-Nandi GARCH model with fixed parameters a classical likelihood function based on daily returns is used. GARCH models with switching parameters on the other hand are notoriously difficult to estimate as a result of the path dependence problem. Sichert (2019) proposes an estimation algorithm for the change-point GARCH model that uses a particle filter for the latent state variable, and both the Monte Carlo expectation-maximization algorithm and the Monte Carlo maximum likelihood method in the optimization step to obtain the maximum likelihood estimator (MLE). This hybrid algorithm, called Particle-MCEM-MCML, is based on the algorithms proposed by Augustyniak (2014) and Bauwens et al. (2014). The main steps of the algorithm are repeated in the Online Appendix. For a more detailed discussion of the approach as well as empirical studies the reader is referred to Sichert (2019). To identify the optimal number of breaks the algorithm is run with the number of breaks $K = 2, ..., 9$. Then the optimal number of breaks is chosen by the algorithm using the Bayesian information criterion.

2.3.2 Results

Since there are no examples of an estimation of change-point Heston-Nandi GARCH model, a more detailed analysis seems appropriate. Table 1 presents the estimation results. In the upper panel, the first column gives the parameters for the standard GARCH, while the remaining columns contain the CP parameters for each regime. The panel in the middle shows the degree of integration of each regime's variance process and annualized long-run volatility. The lower panel of the table shows the log-likelihoods of the estimates. The log-likelihood of the CP-GARCH model was calculated using the particle filter methodology with 100,000 particles as in Bauwens et al. (2014), which is accurate to the first decimal place. Finally, two standard information criteria, namely

the Akaike information criterion (AIC) and the Bayesian information criterion (BIC) are provided. The optimal number of regimes is five. The identified break dates are 01.01.1992, 28.10.1996, 12.08.2003, 07.06.2007 and 29.11.2011, each date being the first date of the new regime. These five regimes capture the economic relevant states, while more regimes are prone to overfitting the data.[7]

The comparison of the estimation results in Table 1 show that there are distinct variance regimes in the CP model. The long-run variances differ significantly across the regimes, while the variance of the fixed GARCH fits the average variance. In particular, the long-run variances in the first, third and fifth regime are significantly lower than the sample average of 16.6%, and the variance of regime two and four are much higher.

Finally, when comparing the likelihood of the FP with the CP model, it becomes apparent that the second one fits the data much better. This is of course expected if one adds more parameters to the model. However, the two information criteria both correct the log-likelihood for the number of parameters which are used. The comparison of these measure across models also strongly suggests that the CP-GARCH is better. An additional likelihood-ratio test also clearly prefers the model with five regimes.

A few words are called for to address potential data mining concerns. First, note that the average duration of one state in the CP model is about 5.5 years, which is fairly long. Second, the parameters in the above estimations are structurally different in several regards. Furthermore, the dynamics across the regimes are very different as well as the long-run variances. If there would not be any structural changes, the estimation could not identify them in such a long sample.[8]

Figure 2 illustrates the identified regimes by plotting the break dates together with the level and 21-day realized volatility of the S&P 500 index. By visual inspection alone one can recognize clear patterns of low and high volatility, which are accompanied by good and low to moderate aggregate stock market returns, respectively. The estimated regimes capture these periods very well. The first high volatility regime contains extreme market events at the LTCM collapse and the market downturn of the early 2000s. The second high volatility regime compasses the recent financial crisis and its aftermath.

[7] The Online Appendix provides a mode detailed analysis of the optimal number of regimes.

[8] It is standard in the pricing kernel literature to estimate the GARCH model over the full sample. Furthermore, the analysis focuses on the comparison of the standard GARCH with fixed parameters versus the CP-GARCH, and both are estimated over the same sample.

Table 1: Estimation Results of the HN-GARCH models

	FP HN-GARCH	CP-HN-GARCH				
Parameters	'92-15	'92-'96	'96-'03	'03-'07	'07-'11	'11-'15
ω	3.01E-19	2.24E-06	1.91E-06	5.36E-06	4.41E-10	3.51E-06
α	4.34E-06	1.37E-06	6.13E-06	8.05E-07	7.66E-06	2.15E-06
β	0.821	0.801	0.786	0.301	0.772	0.273
γ	188.9	269.1	164.7	836.8	161.1	542.8
μ	2.256	9.149	1.090	8.243	-0.380	8.650
p_{jj}		0.99918	0.99941	0.99898	0.99911	1
Properties						
$\beta + \alpha\gamma^2$	0.9762	0.900	0.9522	0.8641	0.9703	0.9074
Long-run volatility	0.166	0.096	0.206	0.107	0.255	0.124
Log-likelihood	FP	CP	Likelihood-ratio test			
Total	19495.9	19691.6	Test statistic			401.4
AIC	-38981.8	-39325.1	1% critical value (χ^2_{24})			43.0
BIC	-38948.4	-39131.0	0.1% critical value (χ^2_{24})			51.2

Parameter estimates are obtained by optimizing the likelihood on returns. Parameters are daily, long-run volatility is annualized. For each model, the total likelihood value at the optimum is reported. The volatility parameters are constrained such that the variance is positive ($0 \leq \alpha < 1,\ 0 \leq \beta < 1,\ \alpha\gamma^2 + \beta < 1,\ 0 < \omega$). The Akaike information criterion (AIC) is calculated as $2k - 2\ln(L^R)$ and Bayesian information criterion (BIC) is calculated as $\ln(n)k - 2\ln(L^R)$, where n is the length of the sample and k is the number of estimated parameters. The test statistic of the likelihood ratio test is $-2(LL_{FP} - LL_{CP})$.

2.4 Volatility forecasts from the CP-GARCH model

The key property of the change-point GARCH model for this study is that it overcomes the cyclical bias in multi-period GARCH models that is present for the standard GARCH model with fixed parameters. Therefore, I next study the ex ante predicted 21-day volatility of each model specification for each day in the sample, and compare it to the ex post realized volatility. These multi-periods volatility forecasts are of interest, because one month (21 trading days) is a typical horizon in the pricing kernel literature and therefore the benchmark maturity in the empirical section below. For the comparison, the estimated parameters are used to filter the volatility up to a point in time t, and

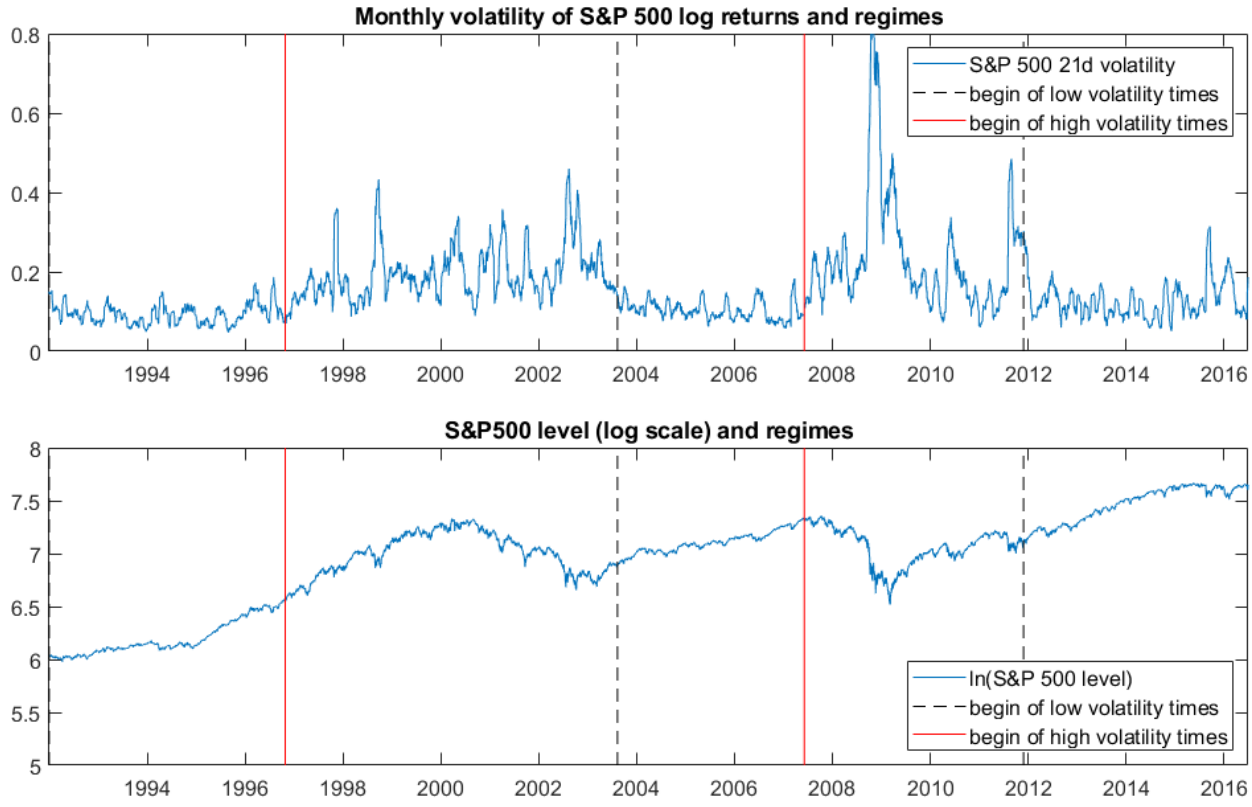


Figure 2: Monthly volatility of the S&P 500 returns and S&P 500 level

The figure shows the monthly volatility (top, calculated as annualized 21-day rolling window square root of the sum of squared log returns) and the natural logarithm of the level of the S&P 500 Index (bottom).

then the model implied variance for $t+1$ to $t+21$ is calculated.[9] The realized volatility is conventionally calculated as sum of squared daily returns.

Figure 3 displays the result graphically. The time series for the prediction is lagged by 21 days in the plot, such that for each point in time, the ex ante expectation is compared to the ex post realization. It is clearly visible that the fixed parameter GARCH constantly over-predicts volatility in times of low variance (regime 1, 3 and 5). This is

[9] In this analysis the probability to switch into another regime is ignored, as well as the uncertainty about which regime currently prevails. The first one has minor effects, since the MLE for the switching probability is in the magnitude of 1/1000 to 1/1500 (see also Section 6.3.3). The impact of the second simplification is typically small as well, since the filtered state probabilities are usually close to one. The exact quantification, however, would require running the estimation separately for each day, which is computationally infeasible. Nevertheless, the main argument below, that an unbiased volatility forecast is crucial, still holds. Furthermore, the robustness section includes several alternative approaches that are not state-dependent.

because it reverts back to the long-term mean too quickly and cannot capture extended periods with a below average volatility. To a lesser extent, the reverse is true for the high variance regimes, where the one state GARCH mostly under-predicts volatility. On the contrary, the CP-GARCH is much closer to the realized volatility in each case.

Table 2 shows the corresponding statistics. The first line contains the realized volatility, while the following lines contain the average predicted volatility for the fixed parameters (FP) and CP-GARCH as well as the conventional root-mean-square error (RMSE). The numbers match the visual findings. The FP model is always biased towards the long-run mean and hence severely over-predicts volatility in times of low volatility (by 30% up to 50%), and vice versa. The CP model does match average numbers very well and also has a lower RMSE, often much lower. In the Online Appendix I show that very similar results can be obtained when splitting the sample on other criteria, as e.g. the beginning of period VIX or GARCH forecast. In addition, the over- and underprediction pattern of the FP model is confirmed by standard predictive (Mincer-Zarnowitz) regressions (tabulated in the Online Appendix). In sum this analysis shows that the standard GARCH model does a poor job in predicting future volatilities. The model performs particularly badly in times of low variance, which will be important in the estimation of empirical pricing kernels.

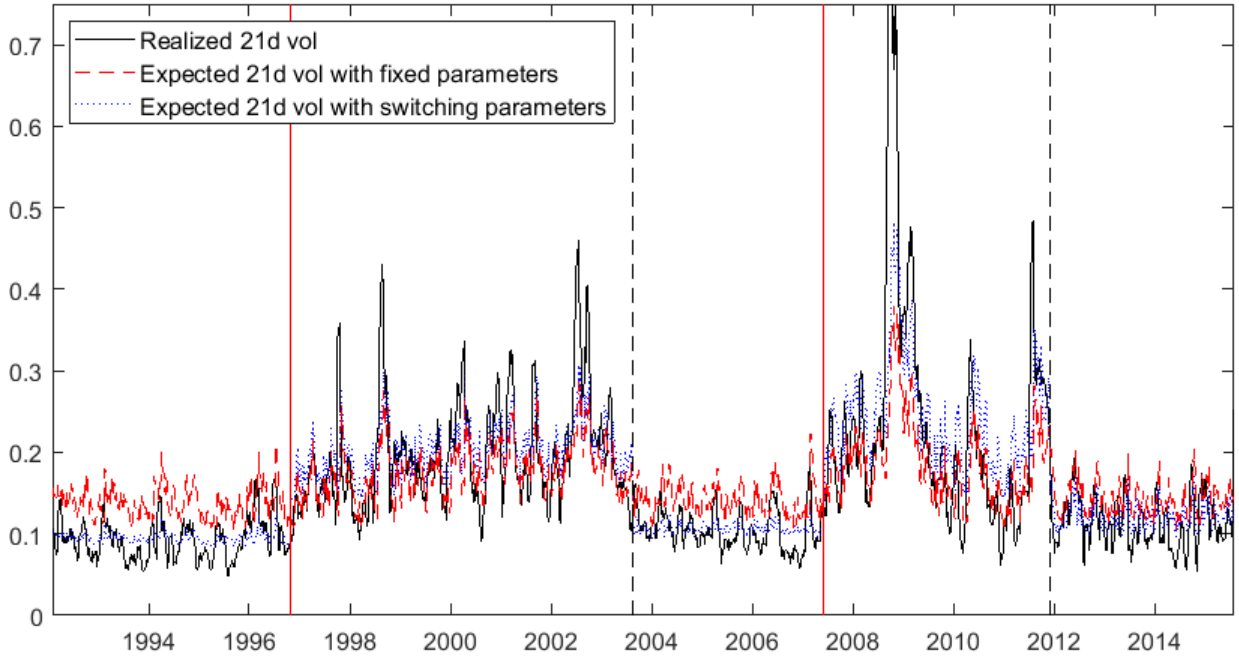


Figure 3: Predicted vs. realized 21-day volatility

The figure shows the annualized 21-day rolling window realized volatility, measured as the square root of the sum of squared daily returns, as well as the ex ante expected volatility implied by the FP and CP-GARCH model. The predicted variance is lagged by 21 days, such that ex ante expectation and ex post realization are depicted at the same point in time.

Table 2: Predicted vs. realized 21-day volatility

	'92-'96	'96-'03	'03-'07	'07-'11	'11-'15
Average realized 21d volatility	.0268	.0573	.0300	.0683	.0334
FP Avg. predicted 21d volatility	.0391	.0511	.0409	.0543	.0413
CP Avg. predicted 21d volatility	.0270	.0586	.0308	.0701	.0354
FP RMSE predicted 21d volatility	.0155	.0167	.0129	.0350	.0120
CP RMSE predicted 21d volatility	.0079	.0157	.0066	.0302	.0085
Average VIX	.0406	.0709	.0406	.0777	.0450

The table shows the average realized 21-day volatility across the different regimes, as well as the average ex ante predicted volatility by both the FP and CP-GARCH model and the root-mean-square error (RMSE) of the predictions. The monthly VIX is calculated as $\text{VIX}/100/\sqrt{12}$.

One could of course object that rational expectations during times of low market

volatility could have been higher than the ex post realized ones. It seems a very unlikely event, however, that the rational expectations exceeded the realization as extremely for such a long time period. The following comparison illustrates this in another way: the FP GARCH volatility forecast in calm periods is virtually as high as the VIX in the same time period. The VIX is the non-parametric risk-neutral expectation of the future volatility derived from option prices. It is typically significantly higher than the physical expectation, since it includes the variance-risk-premium. Hence, on average, the physical volatility forecast should be significantly below its risk-neutral counterpart given the large magnitude of the variance risk premium documented in the literature (e.g. Carr & Wu 2009). Several alternative volatility models as well as the VIX as a purely forward-looking volatility forecast are used in the robustness section. Finally, these results also hold for all horizons considered below (1 week to 3 months), as well as for the other equity indices.

On a more general note, the documented pattern sheds new light on the volatility clustering, which is considered a stylized fact of volatility in the financial econometrics literature. So far, this behavior is usually interpreted as another manifestation of volatility persistence. The CP-GARCH model on the contrary makes the process switch between alternating high and low volatility regimes. These regimes capture the empirical fact that volatility is above or below its long-run mean for many years.

3 Empirical Pricing Kernels

3.1 Option data

The empirical analysis uses out-of-the-money S&P 500 call and put options that are traded in the period from January 01, 1996 to August 31, 2015. This is the full sample period available from OptionMetrics at the time of writing. The option data is cleaned further in the standard way. For each expiration date in the sample, the data of the trading date is selected which is closest to the desired time to maturity (e.g. 30 days for one month).[10] Prior to 2008 there are only 12 expiration days per year (third Friday of each month), but afterwards the number of expiration dates increased significantly with

[10]For each time horizon that is analyzed in this paper, the desired time to maturity was set such that it would be Wednesday data. It is common to use Wednesday data, because it is the day of the week that is least likely to be a holiday and also less likely than other days to be affected by day-of-the-week effects (such as Monday and Friday).

the introduction of end-of-quarter, end-of-month and weekly options, and all of them are included. Next, only options with positive trading volume are considered and the standard filters proposed by Bakshi et al. (1997) are applied. The full details of the data cleaning process are listed in Appendix A. Table 3 presents presents descriptive statistics for the option data by moneyness. The implied volatility (IV) shows the typical volatility smirk pattern. The volume and open interest data shows that the option contracts are highly liquid and traded for all moneyness levels .

Table 3: Option Data Descriptive Statistics

	Total number of contracts	Average IV (%)	Average volume	Average open interest	Average spread
$K/S_0 <= .9$	4,311	33.8	1,375	13,331	0.63
$.9 < K/S_0 <= .925$	1,823	23.4	1,572	13,151	0.68
$.925 < K/S_0 <= .95$	2,160	20.5	1,574	12,568	0.76
$.95 < K/S_0 <= .975$	2,242	18.5	1,912	13,447	0.95
$.975 < K/S_0 <= 1.00$	2,297	16.7	2,563	11,938	1.20
$1.00 < K/S_0 <= 1.025$	1,080	16.8	2,174	9,721	1.73
$1.025 < K/S_0 <= 1.05$	2,201	12.4	1,465	9,862	0.69
$1.05 < K/S_0 <= 1.075$	981	15.2	1,558	11,240	0.66
$1.075 < K/S_0 <= 1.10$	364	20.0	1,734	15,274	0.76
$K/S_0 > 1.1$	361	32.2	1,411	20,045	1.15
All	19,044	21.2	1,756	12,462	0.83

The table present descriptive statistics for the S&P 500 option data. The data is Wednesday closing OTM options contracts from January 10, 1996 to August 18, 2015.

3.2 Approach

The conditional empirical pricing kernel (EPK) $\hat{M}_{t,t+\tau}\big(R_{t,t+\tau}\big)$ as a function of the returns of the index $R_{t,t+\tau}$ is calculated as:

$$\hat{M}_{t,t+\tau}\big(R_{t,t+\tau}\big) = f^*\big(R_{t,t+\tau}\big)/\hat{f}\big(R_{t,t+\tau}\big). \tag{4}$$

Hence, the two major quantities that are required to empirically estimate pricing kernels are the risk-neutral return density f^* and the physical return density f. The chosen approaches allow to stay as non-parametric as possible, but provide evidence on the conditional density.

The method to estimate the risk-neutral density $f^*(R_{t,t+\tau})$ follows the standard appraoch of Figlewski (2010). This approach is known to work well and is used by numerous studies, and Appendix B contains the technical details. The obtained densities are truly conditional because they reflect only option information from a given point in time. Also note that the risk-neutral estimates are not influenced by any assumption of structural breaks.

In the following, the risk-neutral probabilities are only estimated where option data exists, and the implied volatility curve is not extrapolated. I choose to deviate from Figlewski (2010) in this point, as, on the one hand, any extrapolation or tail fitting is potentially unreliable, and the shape of the PK in the tail would crucially depend on guessing the right parametric distribution for the tails. On the other hand, the data on average covers a cumulative probability of 95.5% at the one month horizon and therefore the main results can be shown without any tail probabilities.[11] Instead, I present the ratio of the cumulative return density in the right tail that is not covered by the (cleaned) option data. Formally:

$$\hat{M}_{t,t+\tau}^{\text{right tail}} = \Big(1 - F^*\big(R_{t,t+\tau}^{max}\big)\Big) / \Big(1 - \hat{F}\big(R_{t,t+\tau}^{max}\big)\Big), \tag{5}$$

where R^{max} if the return corresponding to the highest available strike that passes the data filters and F denotes the cumulative density function (CDF). The risk-neutral CDF can also be obtained from option prices non-parametrically. Dividing this quantity by its physical counterpart gives one data point for the right tail. This point provides an indication of the behavior of the pricing kernel in the tail. It can be interpreted as the average PK in that region.

The adopted approach for obtaining the conditional physical density of returns $\hat{f}$ is semi-parametric and has become popular recently.[12] The starting point is a long daily time series of the natural logarithm of one month returns from January 02 1992 to August 31 2015. First, the monthly return series is standardized by subtracting the sample mean return $\bar{R}$ and afterwards dividing by the conditional one month volatility

[11] Refraining from "completing the tails" does not influence the estimation of the risk-neutral probabilities over the range of available option strikes. The risk-neutral probability is obtained non-parametrically, and no additional treatment as e.g. kernel fitting or scaling is necessary. Therefore, the standard approach to exclude option prices with low prices (best bid below \$3/8) is at least innocuous and probably leads to an increase of the precision of the derivation of the risk-neutral probabilities.

[12] Christoffersen et al. (2013) use the same method, and similar methods in related settings are used e.g. in Barone-Adesi et al. (2008) and Faias & Santa-Clara (2017).

$\sqrt{h_{t,t+\tau}}$ forecasted by the GARCH model at the beginning of the month. This yields a series of monthly return shocks Z:

$$Z_{t,t+\tau} = (R_{t,t+\tau} - \bar{R})/\sqrt{h_{t,t+\tau}}. \tag{6}$$

The conditional distribution $\hat{f}_t$ is then constructed by multiplying the standardized return shock series Z with the conditional monthly volatility expectation $h_{t,t+\tau}$:

$$\hat{f}_t(R_{t,t+\tau}) = \hat{f}\left(\bar{R} + \sqrt{h_{t,t+\tau}}Z\right). \tag{7}$$

Hence, for each date in the sample a different conditional density is estimated. The difference arises from the conditional volatility expectation, while the shape of the distribution is always the same. For both models the full return time series is used, and for the CP model the GARCH parameters of the respective regime are used.

The benchmark method for the physical return density is chosen because it has several distinct advantages. First, it is flexible enough to incorporate the volatility forecasts of several other models, which is done in the robustness section. Second, it allows me to explicitly detect the main driver of the results by comparative statics. Finally, it is only semi-parametric as it requires a minimum of parametric assumptions, preserves the empirical patterns for moments higher than two and last but not least has a good fit to the empirical distribution. Nevertheless, many alternative approaches are tested in the robustness section and they deliver very similar results.

3.3 Results

This section documents the shape of the conditional pricing kernel using the non-parametric method for estimating the risk-neutral and the semi-parametric method to estimate the physical conditional densities described above. The one month horizon is chosen to be the benchmark analysis, since it is the most studied horizon in the literature on empirical pricing kernels and a maturity with very liquid option contracts. The robustness section shows that the results also hold for other typical horizons. Figure 4 shows the time series of the natural logarithm of the estimated pricing kernels using the HN-GARCH model with fixed parameters. Figure 5 displays the same for the CP-GARCH. The scale of the horizontal axis is log returns. The coloring indicates times with high volatility (red) and times with low volatility (black), as defined in Section 2.3.2. The lines are not smoothed, to avoid creating a false impression of precision. The

dotted blue lines at the right end of the pricing kernels depict $\hat{M}_{t,t+\tau}^{\text{right tail}}$. Each CDF ratio is just one data point, and the blue line connects the point to the corresponding PK. This gives an indication of the behavior of the PK in the right tail, as it can be interpreted as the value of the average PK in the tail. The value on the horizontal axis for the CDF ratio is chosen as the return of the highest traded strike (end of domain of $\hat{f}_t$) plus 0.013 (0.02) log-return points in times of low variance (high variance).

When comparing the two plots, several observations emerge. The first plot mostly exhibits U-shaped pricing kernels in times of high volatility, while the PKs in times of low volatility have the typical S-shape. The finding that the latter pattern prevails in times of low volatility was never documented systematically nor for such a long time series. It is quite striking to see how well the different market regimes, depicted with red and black graphs in the plot, separate S-shaped pricing kernels from U-shaped ones. Although the estimation of the break points uses only return data and no option data, it classifies the changes in the PK estimates obtained from standard methodology very well.

However, when the GARCH parameters are not fixed but structural breaks are introduced, the kernels in times with low volatility are predominantly U-shaped. In times of high volatility, the estimated PKs are now more noisy, but still mostly U-shaped. Since the PK has to at least price the risk-free asset and the index correctly, the varying wideness of the PKs is to be expected. If the physical distribution expectation becomes more disperse, the PK must change in order to price the asset correctly, and vice versa. Furthermore, the PK estimates from the CP model are closer together in the plots than their FP counterparts. This suggests that they are closer to documenting a stable relationship over time. Lastly, the observation that the estimated PKs in times of low volatility are very steep at their left end is reasonable in economic terms. If the market return in these times would be very low, this would very likely be accompanied by a large increase of variance and a severe worsening of economic conditions, and possibly even with a regime shift. Such an adverse event should be expensive to insure against.

Two further comments on the shape of the PKs in the CP version are warranted. A first objection might arise from the unclear direction of the plots at the right end, especially in periods of low variance. Note that this ambiguity clearly increases from the beginning to the end of a calm period (regime 1, 3 and 5). Therefore, a likely explanation is that after several years of strong bull markets, the probability of further large positive returns is lower. The adopted approach, however, cannot incorporate such a specific conditional expectation, since the shape of the distribution (i.e. mean, skew

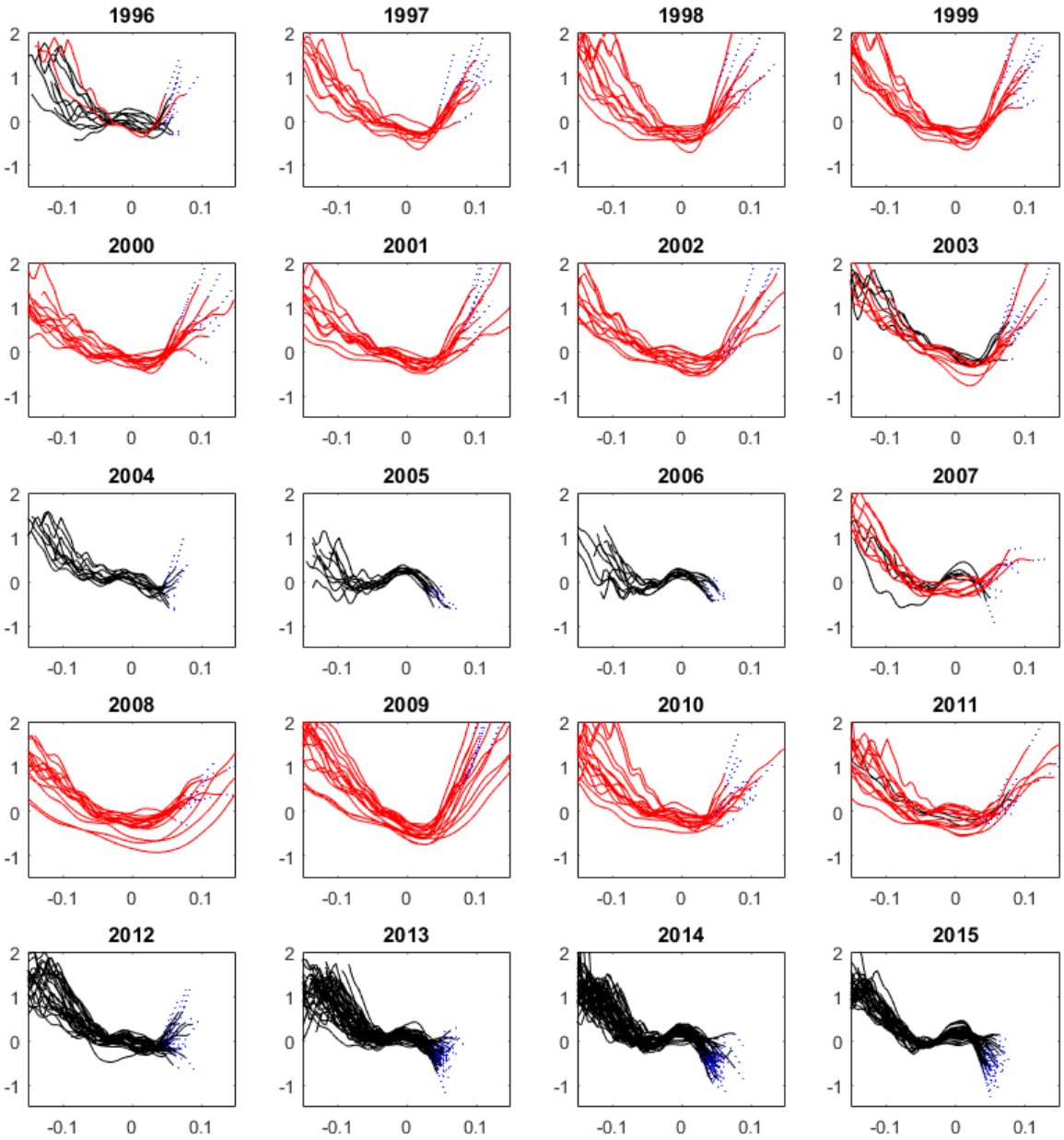


Figure 4: Empirical pricing kernels with fixed parameters in the HN model

The figure shows the natural logarithm of estimated pricing kernels obtained when using the Heston-Nandi model with fixed parameters. Red (black) depicts times with high (low) variance, as defined in Section 2.3.2. Log-returns are on the horizontal axis. The horizon is one month. The dotted blue line connects the points, which depict the ratio of the CDFs of the tail, with the corresponding pricing kernels.

and kurtosis) is always the same and only the wideness (volatility) is conditional. This is supported by the findings of Giordani & Halling (2016), who document that returns

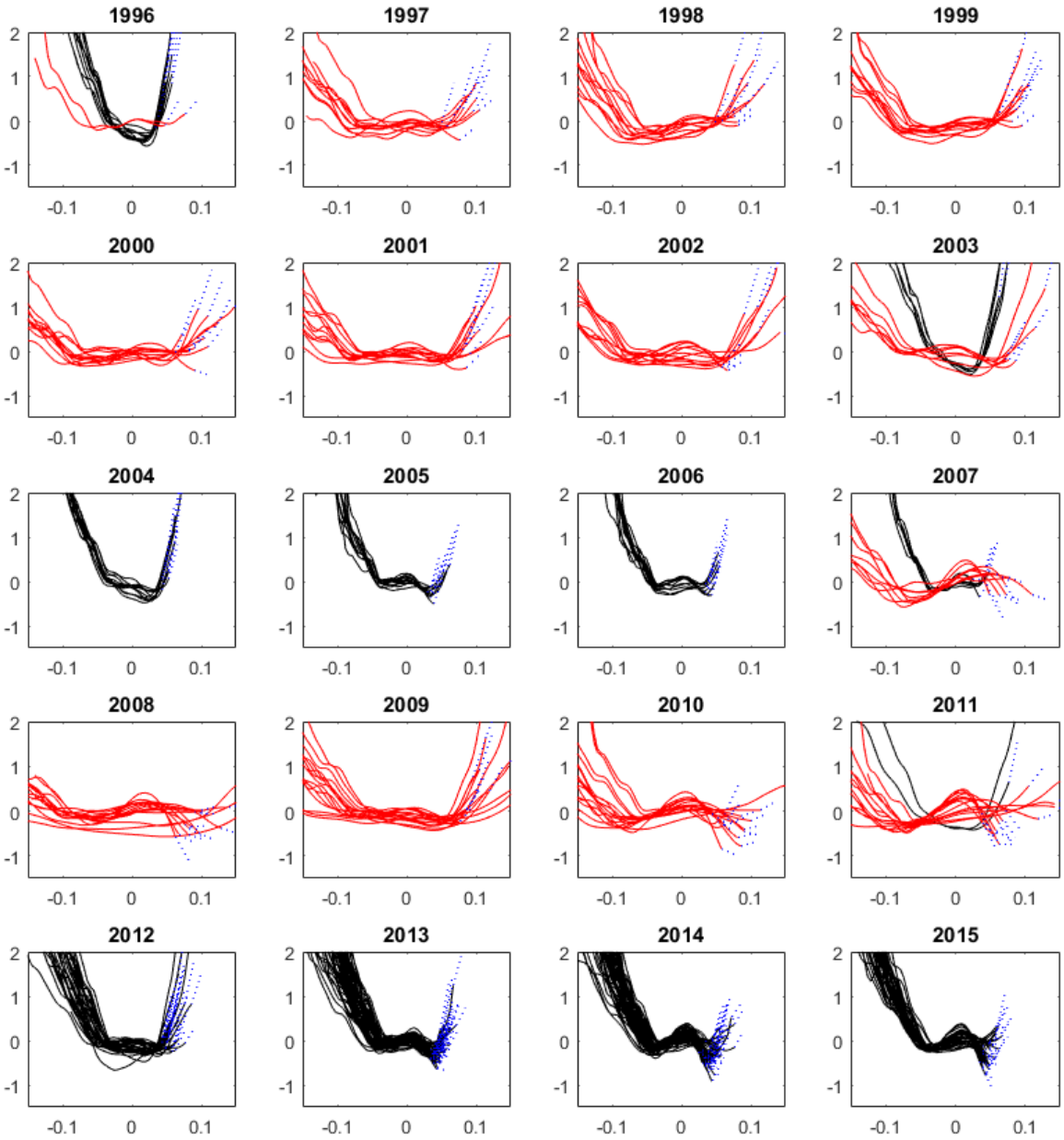


Figure 5: Empirical pricing kernels with CP parameters in the HN model

The figure shows the natural logarithm of estimated pricing kernels obtained when using the Heston-Nandi model with CP parameters. Red (black) depicts times with high (low) variance, as defined in Section 2.3.2. Log-returns are on the horizontal axis. The horizon is one month. The dotted blue line connects the points, which depict the ratio of the CDFs of the tail, with the corresponding pricing kernels.

are more negatively skewed when valuation levels are high. Furthermore, this pattern of decreasing steepness of right-hand end of the estimated PK over the course of a

regime is also observed for all robustness checks below. All alternatively tested methods have in common that the skewness and all higher moments of the return density only depend on the volatility. In contrast, the skewness of the risk-neutral return distribution clearly decreases over the course of the last low volatility regime. Figure 2 in the Online Appendix illustrates this. However, there exists no established method to model this potential time series pattern in the physical expectation.

In addition, the point where the PK starts to increase again is rather deep OTM. It is possible that these strikes are not traded, or best bid is below \$3/8, which is the cut-off point in the data cleaning. Both arguments are supported by the finding that the lower the highest available strike is, i.e. the right-hand end of the line, the lower the right-hand end of the PK line is. Furthermore, Table 2 above shows the model still slightly over-predicts the volatility in calm periods, especially in the last regime with on average 6%. Over-predicted volatility is the key driver that generates the typical S-shape, as discussed in detail in the next section. Therefore this can help to explain why the PKs in the last regime are the most ambiguous ones. Finally, Section 6.1 shows fully non-parametrically that the pricing kernel is at least on average upward sloping in times of low volatility by analyzing OTM call option returns.

The second comment refers to the PK estimates in high volatility regimes with the CP model, which are more noisy and sometimes exhibit a pronounced hump around zero. Similar to above, this is again mostly observed at the end of a high volatility regime. The reason is again that the GARCH model has the tendency to overfit high volatility states at the expense of low volatility states. In the same ways as the standard GARCH forecasts are biased towards the long-run mean, the CP-GARCH forecasts in high volatility times are also biased towards their high long-run mean. This long-run mean is significantly influenced by the extreme returns, which are mostly observed at the beginning of the high variance regime. Figure 3 shows that there are also periods with relatively low volatility within these periods, as for example most of the year 2011. However, the GARCH forecasts are not able to capture these periods. In fact, it even overestimates the average volatility of these periods, as can be seen from Table 2. Hence, the mechanism that causes these slightly S-shaped estimates here is the same that causes the S-shape when one uses the standard GARCH methodology, as discussed in the next section.

Overall, one can conclude that the rather simple modification of the methodology leads to a large change in findings. The application of a more accurate volatility forecast makes the prominent finding of a hump around zero returns in the empirical pricing

kernel vanish.

3.4 Importance of the volatility forecasts

The introduction of structural breaks into the standard GARCH model changes the results on empirical pricing kernels significantly. Naturally, the question emerges what drives this result. The following analysis will show that the forecasted volatility is the key driver. Furthermore, the key mechanism how the overestimation of volatility leads to the estimation of S-shaped pricing kernels is dissected.

The CP model has three potential channels that can influence the results: the filtered volatility, the different conditional return distributions and the forecasted volatility. Untabulated results show that the first factor has little impact since the filtered volatilities are very similar for both FP GARCH and CP-GARCH. The second factor has also surprisingly little influence. One could expect that the different volatility forecasts of the models lead the different distributions of standardized return shocks. But the top left plot in Figure 6 shows that these two distributions are actually very similar. The figure shows the densities of the monthly normalized return shocks (i.e. all Z from Eq. (6)) for the two models. The first sub-plot shows all return shocks of the full time series and the following ones only the return shocks of the respective sub-periods. While the regime-specific shock densities are very different, the aggregate density, which is used for the empirical study, is very similar.

The last remaining channel, the different volatility forecasts, turns out to be the major driver of the results. At each point in time, the physical return density forecast is constructed by multiplying the respective return shock density (top left Figure 6) with the conditional volatility forecast. Since the densities are very similar, the key difference is the more realistic volatility forecast. In times of low volatility, the upward-biased monthly volatility forecasts of the FP GARCH create a physical density that is too wide and has too much probability mass in the tails, and too little in the center.

Figure 7 illustrates the main mechanism how the volatility forecasts effect the empirical pricing kernel estimates. The top row shows the actual data for October 2005, and the bottom row shows the data for May 2009. The first column contains the physical return forecasts. It is clearly visible, how in times of low volatility, the return density for the FP model has more probability mass in the tails and less in the center, relative to the CP counterpart, while the reverse is true for the high volatility times. The second column shows the risk-neutral return density, and the last column shows the EPKs. One

can nicely see how the overestimated volatility for the FP model influences the shape of the estimated PK in 2005. The overweight of probability mass in the left tail flattens the EPK, but it is still downward sloping. The lack of probability mass in the center causes the hump in the middle. Finally, the fatter right tail makes the EPK downward-sloping on the right end. In 2009, the underestimation of the volatility makes the pricing kernel steeper, but does not change the estimation qualitatively.

The small differences between the FP and CP return shock density depicted top left in Figure 6 actually counteract this mechanism. The FP density has slightly more mass at the mode, which should reduce the hump.

The sub-plots two to six in Figure 6 point to another interesting finding. First, for the FP model, the densities of the respective regimes are substantially different from the one of the full sample. An apparent pattern is that the densities in times of low variance are much tighter, while the reverse is true for the other times. This pattern is barely visible in the respective densities for the CP model. The application of a model with better volatility forecasts leads to estimated shock densities that are very similar across time. The difference is caused solely by the different volatility forecasts. In the low variance regimes, in the FP version the monthly returns are divided by an upward-biased volatility forecast and hence produce a very narrow shock distribution. Similarly, in times of high variance, where monthly returns are also much more volatile, these returns are standardized by a downward-biased volatility forecast.

To further evaluate the similarity of the shock distributions I conduct a formal test. Table 4 presents the p-values of a Kolmogorov-Smirnov-Test. The null hypothesis is that the shock distribution of one regime is not different from the shock distribution of the full sample. For the model with fixed parameters the shock distribution of the regimes are significantly different from the shocks of the full sample. For the CP model the null of equality cannot be rejected for the three low variance regimes 1, 3 and 5, nor for regime 2, and only be rejected for regime 4, which contains the financial crisis. This is interpreted as support of the approach, and especially for the inclusion of breaks.

Overall, the estimation of very homogeneous monthly shock distributions is a very interesting side result, and not least because this is solely attributable to the different volatility forecasts. This gives rise to the possibility of finding a time-invariant distribution for stock returns, which is left for further research.

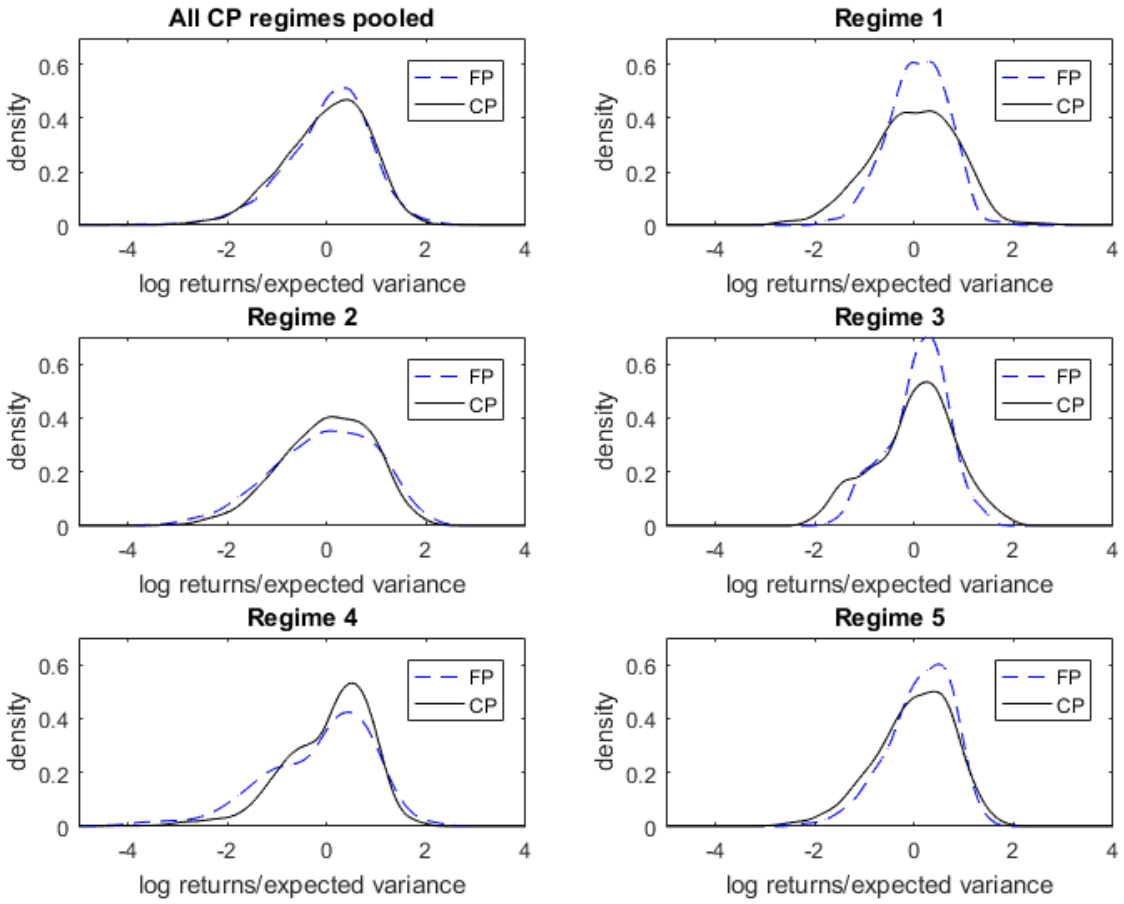


Figure 6: Monthly return shock densities

The figure shows the estimated monthly return shock density (21 days) calculated as in Eq. (6), for both the FP GARCH and CP-GARCH model. The first subplot depicts the case where the shocks of all periods are pooled together, while the remaining ones only contain the shocks of the respective regimes in timely order.

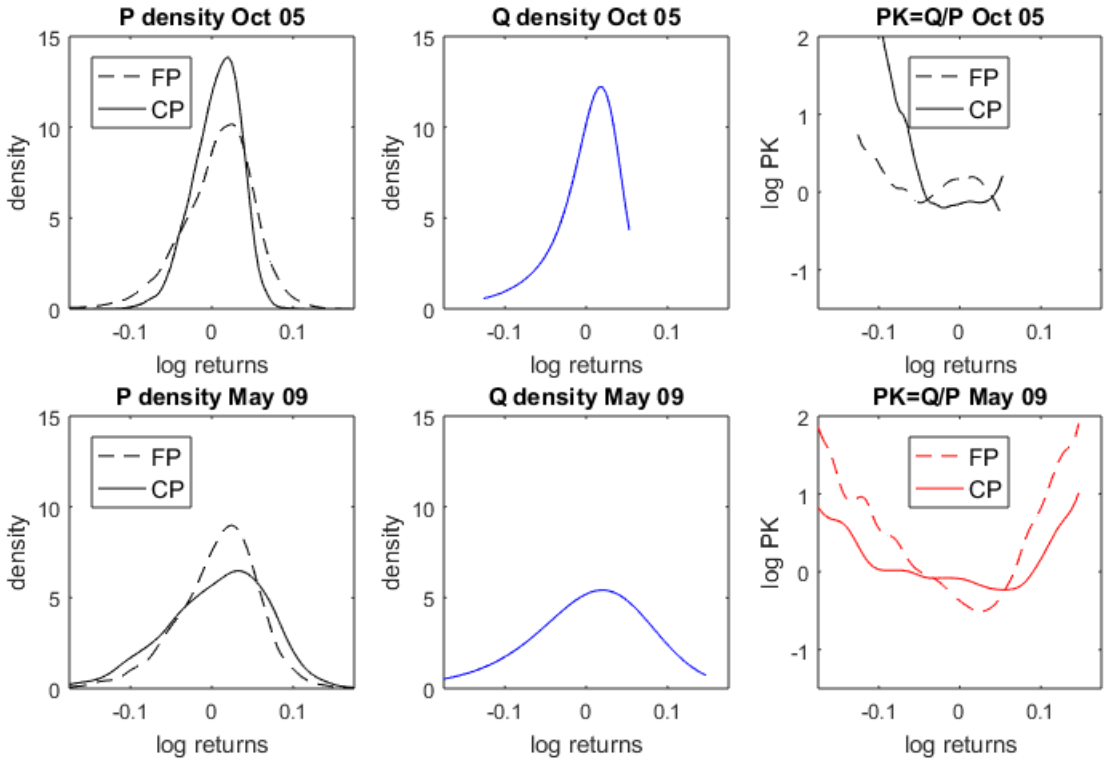


Figure 7: Main mechanism how volatility estimates drive the results

The top row shows data from October 2005 and the bottom row shows data for May 2009. The first column displays the estimated conditional monthly return density at the given date, while the second column displays the risk-neutral monthly return density at the same date. The last column illustrates the estimated pricing kernels.

Table 4: Kolmogorov-Smirnov-Test of equal distribution of monthly return shocks

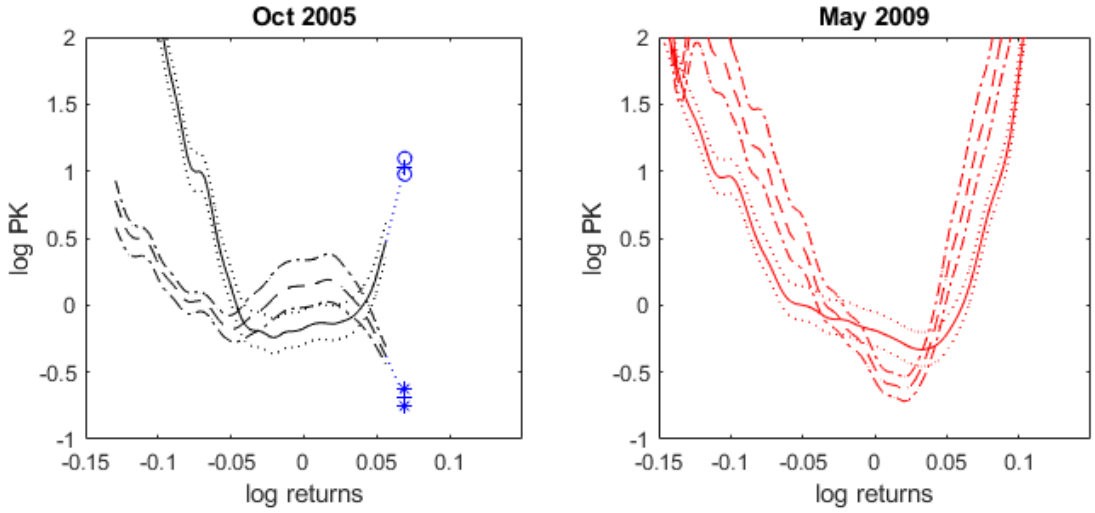
	'92-'96	'96-'03	'03-'07	'07-'11	'11-'15
FP GARCH	1.5E-07	7.2E-08	4.2E-05	5.2E-09	5.7E-05
CP-GARCH	0.051	0.072	0.126	0.001	0.128

The table shows the p-values of a Kolmogorov-Smirnov-Test. The null hypothesis is that the shock distribution of one regime is not different from the shock distribution of the full sample.

3.5 Confidence intervals for the empirical pricing kernels

The empirical results above suggest that there is a significant change in findings between the pricing kernel estimation that relies on FP GARCH compared to the one that employs a CP-GARCH model. This section introduces confidence intervals for the estimated

Figure 8: Confidence intervals for the empirical pricing kernels



The figure shows pricing kernel estimates from October 2005 and May 2009, together with their pointwise 90% confidence intervals.

pricing kernels in order to test the differences statistically. The pointwise confidence intervals characterize the precision of the PK estimates for a given value of the index return. The confidence intervals are calculated by sampling from the distributions of $f()$ and $f^*()$ independently. The variation of $f^*()$ is obtained using the popular method of Aït-Sahalia & Duarte (2003), and the variation of $f()$ is calculated relying on the asymptotic results from Härdle et al. (2014). Appendix D contains the full details of the approach.

Figure 8 illustrates the results by plotting the 90% confidence intervals for the two dates from Figure 7. The plot shows that the confidence intervals are fairly tight. However, testing for global shape differences is not straightforward, because the different estimates always must intersect. Therefore I rather test for differences at certain, prominent points that are decisive for the global shape. The calculated statistics are summarized in Table 5.

First, I test whether the disappearance of the hump of the S-shape from the FP GARCH is statistically significant. To this end, a hump is defined as a the local maximum of the estimated PK in the area of returns between -5% and 5%. The test statistic for the hypothesis $\hat{M}^A > \hat{M}^B$ is $\hat{M}^{A,lower\ 90\%\ bound} > \hat{M}^{B,upper\ 90\%\ bound}$. At 96.64% of the dates where a hump exists, the hump of the FP method is statistically significantly different from the CP counterpart. Since the majority of humps appears in times of low

volatility, the second line reports the same test only for regime 1, 3 and 5.

Table 5: Test results on differences between estimated pricing kernels

	Percentage of dates with statistical significant results	N
$\hat{M}^{FP}$ 'Hump' $> \hat{M}^{CP}$	96.6%	149
$\hat{M}^{FP}$ 'Hump' $> \hat{M}^{CP}$, only Regime 1, 3, 5	97.0%	135
Right tail of $\hat{M}^{FP}$ < right tail of $\hat{M}^{CP}$, only Regime 1, 3, 5	98.7%	239
Right tail of $\hat{M}^{FP}$ > right tail of $\hat{M}^{CP}$, only Regime 2, 4	82.8%	157
Right tail of $\hat{M}^{FP} < 1$, only Regime 1, 3, 5	77.0%	239
Right tail of $\hat{M}^{CP} > 1$, only Regime 1, 3, 5	63.2%	239

The table reports the results for the statistical test on several differences between the pricing kernel estimates. The middle column reports the percentage of dates in the (sub-)sample where the test statistic is true. The last column reports the number of dates which are tested.

Second, I test whether the ratios of the CDF in the right tail are significantly different by comparing the respective 90% confidence bounds. At 98.7% of the dates in times of low volatility the CDF tail probability ratio of the method that uses the CP-GARCH is statistically significantly higher than its counterpart that uses the FP GARCH. On the contrary, in times of high volatility 82.8% of the CP estimates are lower than the FP estimates.

Finally, I test whether the CDF tail ratio is higher or lower than 1 in times of low volatility. If the probability ratio in the tail is larger than 1, this suggests that the PK is upward sloping in that region, and vice versa. The test returns that 63.2% of all CP estimates are statistically significantly larger than one, while 77.0% of their FP counterparts are smaller than one.

4 Economic Properties of the Empirical Pricing Kernels

In the following section, the empirical pricing kernel estimates presented in Section 3.3 are used, but only the PKs estimated from the monthly AM expiration cycle (SPX) are used. This gives a series of (almost) non-overlapping 30 day periods, and, more importantly, it avoids a sample selection problem, as the vast increase in expiration dates happened towards the end of the sample. Realized pricing kernels, denoted $\check{M}_{t,t+\tau}$, are

calculated by using the the functional form of the conditional EPKs $\hat{M}_{t,t+\tau}(R_{t,t+\tau})$ to map the corresponding realized index return $R_{t,t+\tau}$ into the level of the realized PK. High and low variance are as identified in Section 2.3.2.[13]

4.1 Euler equation errors

Analyzing Euler equation errors is a non-parametric and standard way to test a candidate pricing kernel. Euler equation errors can be interpreted economically as pricing errors, and are commonly referred to as "alphas". To start with, the conditional Euler equation for any asset j is:

$$1 = E_t[M_{t,t+\tau} R^j_{t,t+\tau}], \tag{8}$$

where R_t is the return of the asset j (the index with superscript I in the following) and M_t is the (true) SDF.

Empirically, we can only measure unconditional pricing errors. The pricing error of the asset is defined as the difference between its historical mean return and the risk-premium implied by the candidate pricing kernel $\hat{M}_t$. In the following, the unconditional Euler equation error for the index α^I is calculated as:

$$\hat{\alpha}^I = \frac{1}{T} \sum_{t=1}^{T} \Big(\check{M}_{t,t+\tau} R^I_{t,t+\tau} - 1 \Big), \tag{9}$$

where $\check{M}$ is the (realized) empirical pricing kernel (EPK). The realized PKs are calculated using the functional form of the conditional EPKs to map the realized index return $R_{t,t+\tau}$ into the level of the realized PK. An average Euler equation error of zero is a necessary condition that a valid candidate pricing kernel should fulfill.

Table 6 shows the results. The first two columns show the alphas of the empirical PKs from Section 3.3. The other columns show the alphas for two important methods from the robustness section. In particular, columns three and four show the results when in the benchmark method the level of the VIX or a rolling window GARCH ares used as volatility forecast, all else equal. The first line displays the Euler equation errors and the second line contains the corresponding bootstrapped 99% confidence intervals in square brackets. They are obtained from N=25,000 i.i.d. bootstraps. The draws are i.i.d., since the Ljung-Box test at lags up to 30 rejects autocorrelation in the time series. The third

[13] The conditional results below are very similar if the ex ante VIX level is used to identify the regimes. Using the criterion: 'ex ante VIX$\lessgtr$ in-sample median' produces an overlap in estimated high an low volatility regimes of 91.1%.

and fourth line show the average Euler equation errors split into times of high and low volatility.

Table 6: Euler equation errors (α's) of empirical pricing kernels for the index (in %, per month)

	$\hat{M}^{FP}$	$\hat{M}^{CP}$	$\hat{M}^{VIX}$	$\hat{M}^{Rolling\ Window\ GARCH}$
α^I	5.2***	-0.6	0.7	4.0*
[99% Conf. bounds]	[0.8, 10.7]	[-3.5, 2.6]	[-3.2, 6.0]	[-1.2, 10.9]
α^I low vol	6.8***	0.1	1.9	4.5
α^I high vol	3.2	-1.6	-0.8	3.4

The table shows the average Euler equation errors α^I for different EPKs $\hat{M}$, together with their 99% confidence intervals as well as the Euler equation errors for the two subsamples of times with high and low volatility (as defined in in Section 2.3.2). The symbols ***, ** and * indicate that values are significantly different from zero at the 1%, 5%, and 10% significance levels, respectively. The confidence intervals are obtained from 25,000 bootstrap draws from the sample of errors. The superscript denotes the model used to forecast volatility in the otherwise unaltered benchmark method.

The main result is that the standard estimation approaches, which use a GARCH model with fixed parameters, fails to price the index correctly. Its average pricing errors of 5.2% per month are economically significant, especially in comparison to the average monthly index gross (excess) return in the sample of 0.7% (0.5%). In addition, it is the only EPK that has an α that is statistically different from zero. The approach that use a GARCH model with breaks on the contrary produces errors that are virtually zero. The third and fourth line show that this is also the case if only times of high or low volatility are studied. For the standard model, the last two lines reveal that it performs particularly bad in times of low volatility.

4.2 Explaining cross-sectional stock return anomalies

4.2.1 Intuition

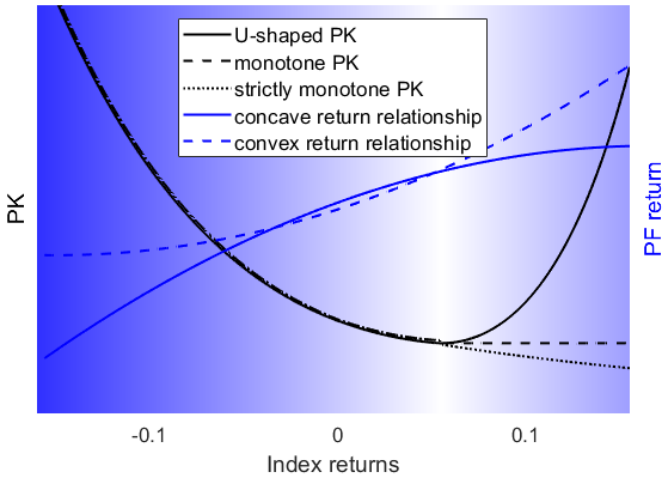


Figure 9: Stylized PKs and return relationships

The plot illustrates a stylized U-shaped PK, monotone PK and strictly monotone PK, together with a stylized concave and convex relationship between asset returns and index returns.

The characteristic part of the U-shaped pricing kernel is the increasing region for high positive index returns. This implies that assets, which pay off well in these bad states of the world, are an insurance and hence earn lower returns on average. While a large body of literature analyzes the downside risk premium, the U-shaped PK implies that there is also an *upside risk premium*. This upside risk premium will be the focus in the following.[14]

Figure 9 illustrates the U-shaped PK together with two hypothetical payoff profiles. The first one is concave and the second one is convex, each as a function of the market

[14]The idea is potentially related to the literature on co-skewness, as e.g. in Harvey & Siddique (2000), and in recent parallel research Schneider et al. (2019). While a U-shaped PK and a co-skewness risk premium are consistent with each other, neither one must imply the other. Furthermore, the empirical tests differ in two key aspects. First, the analysis here does not rely on a return-based factor, but rather tests a functional form of a pricing kernel directly. In addition, the tested pricing kernel is estimated without using any cross-sectional return information. Second, the analysis extends the literature by explicitly testing whether there is an upside risk premium. This is important, as the standard factor based approach only measures the combined effects of the downside and upside risk premium.

return. The convex payoff profile can be seen as a hedge, as it pays off relatively well in times when the PK is high , i.e. the dark shaded areas. The reverse is true for the concave return profile.
More formally, consider that the expected excess return on any asset j is:

$$E(R^{j,e}) = -Cov(R^{j,e}, M) \cdot R_f. \tag{10}$$

Hence, the more negative the covariance of the asset's excess return with M, the higher the expected return. In the area of positive market returns, the increasing PK is more negatively (positively) correlated with the concave (convex) return relationship, and hence this relationship commands an additional premium (discount). If the PK would be decreasing in this region, this implication would be reversed.

So far, the convex and concave return pattern were hypothetical. The next section shows that this is a real world phenomenon that can be observed for several well-known anomalies.

4.2.2 Non-linear relationship between market and portfolio returns

The standard CAPM-β analysis suggests a linear relationship between stock or portfolio returns and the index. On the contrary, I will show that this relationship can be non-linear at the example of the low beta anomaly.

The "low beta anomaly" describes the empirical fact that portfolios comprised of stocks with a high (low) CAPM beta have low (high) abnormal returns. A thorough analysis of the correlation of the returns on the beta-sorted portfolios and the market shows that the low-beta portfolio tends to underperform the (beta adjusted) market both in times of very negative and very positive market returns, while the reverse is true for the high beta portfolio. Figure 10 illustrates this pattern. The two left plots show the returns of the high and low beta portfolios provided by Kenneth French, plotted against the market return, where each point in one return observation. In addition, the figure depicts a first order (only beta with constant) and second order polynomial, fitted to the portfolio returns as a function of the market return. One can see that the relationship of the low beta portfolio has a negative curvature relative to a linear relationship. The right graph in Figure 10 generalizes this pattern by showing the curvature parameter of a fitted second order polynomial for all the five beta sorted portfolios together with the CAPM alpha of the portfolios. The relationship is almost perfectly negative, with a correlation coefficient of -98.2%. Put differently, the higher the alpha, the more concave

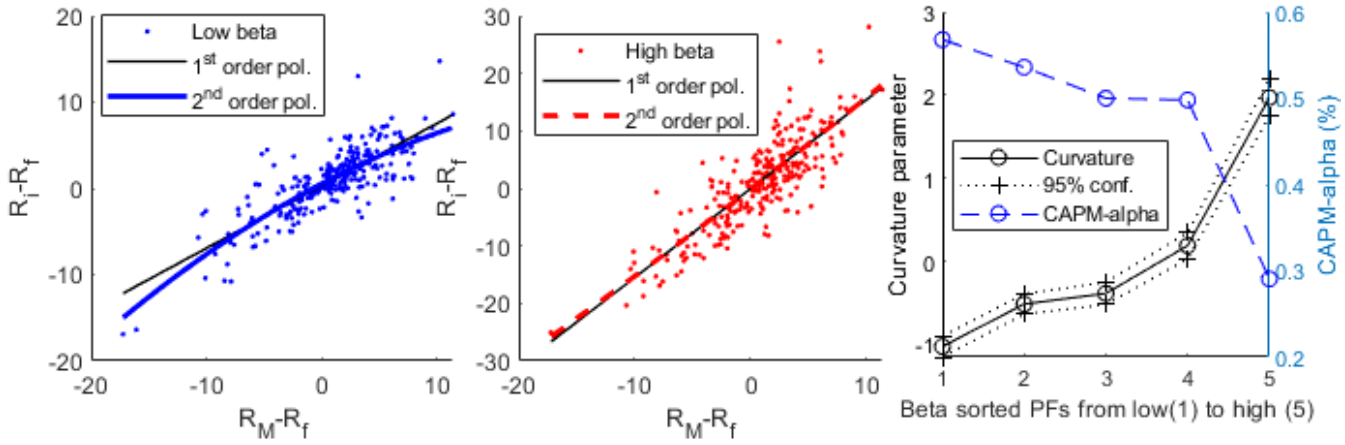


Figure 10: Non-linear relationship between returns of the market and returns of beta-sorted portfolios

The left two plots shows the monthly excess returns of the low and high beta portfolio of Kenneth French plotted against the market's excess returns, where each point is one return observation. In addition, they depict a fitted 1st and 2nd order polynomial for each portfolio. The right plot shows for all five beta portfolios the estimated curvature parameters of the 2nd order polynomial together with its 95% confidence bounds, as well as the CAPM-alphas of the portfolios (second vertical axis). The portfolios are equally-weighted and the time period is 1996-2015.

is the relationship between the portfolio and the market return, and the lower the alpha, the more convex. The confidence bounds for the curvature parameters show that this relationship also statistically robust.

This finding squares with the intuition outlined above. In the presence of a U-shaped PK the concave relationship earns a premium for both the underperfomance in high and low return states relative to a linear (CAPM-β) relationship. Empirically, we observe that the low beta portfolio has such a concave relationship, and also earns a premium relative to the high beta portfolio.

This concave-convex return relation pattern documented in detail here for the low beta case can actually be found for many more anomalies. In the next section, I show that this is also present for momentum, volatility, idiosyncratic volatility and industry portfolios. Hence one can consider this a new stylized fact that is important in combination with a U-shaped PK.

4.2.3 Pricing performance

Most existing risk-based explanations for the low beta anomaly, and for many anomalies in general, focus on the left tail. On the contrary, the following analysis focuses on the right tail of the return distribution and shows that a U-shaped pricing kernel, and in particular the upward sloping part in the area of high positive returns, also helps to explain the anomalies.

For this, I next test the idea outlined in the previous paragraphs quantitatively, using the empirical pricing kernel estimates from Section 3.3. To the best of my knowledge, this is the first attempt in the literature that an EPK estimated from option prices is applied to the cross-section of stock returns. Therefore, no standard methodology exists, but nevertheless the main steps should be straightforward.

As test assets, I use several standard portfolio sorts, namely the established size, value (book-to-market), momentum, and industry portfolios, together with the recently discussed beta, volatility and idiosyncratic volatility sorts. To compensate for the short sample, for each sort, five portfolios are formed to obtain a less volatile return spread between the portfolios. All portfolios are value weighted.

The standard portfolios sorts are recalculated to match the dates and timing of the option prices. Appendix F contains the technical details. The pricing performance of the EPKs $\hat{M}$ (estimated using the CP-GARCH model in the benchmark method) is tested by calculating the alphas of the portfolios analogously to Eq. 8-9, where the return of the index is replaced with the portfolio returns.[15] To keep the main discussion short, I focus on the conventional high-low portfolio. The full pricing performance on all portfolios together with summary statistics on the portfolios can be found in the Online Appendix.

To isolate the pricing contribution of the upward sloping part, I calculated a weakly monotone version of the estimated pricing kernel, denoted $\hat{M}^{mon}$, and compare the pricing performance to the unadjusted $\hat{M}$. The monotone PK estimate is obtained by first finding the minimum level of $\hat{M}$ in the domain of positive returns, and then replacing all PK values to the right of the minimum with the minimum value, i.e., it is made flat (monotone) to the right. This results in a weakly monotone PK as the dashed line in Figure 9. This approach is the simplest and most conservative way to make the PK monotonically decreasing, as one doesn't have to specify the alternative

[15] The upside of analyzing pricing errors is that one does not need to specify any conditional correlations and risk premia. The highly time-varying EPKs suggest that using a conditional correlation is important, and using unconditional estimates as e.g. rolling betas is not a good approximation (which is confirmed by preliminary tests).

parametrically.[16]

Finally, to compare the results to the popular benchmark approaches, I also calculate the α's of the CAPM pricing kernel and the PK of the the Fama & French (1993) 3 factor model (FF3). The first is a linearly decreasing function of the market return, and also the FF3 PK is virtually monotonically decreasing in the index, as the other two factors are almost orthogonal to the market factor.

Table 7: Pricing results for several standard portfolio sorts

	correlation (curvature, α^{CAPM})	α H-L (%, p.a.)				Differences in α H-L (%, p.a.)		
		$\hat{M}$	$\hat{M}^{mon}$	$\hat{M}^{CAPM}$	$\hat{M}^{FF3}$	$\hat{M}$-$\hat{M}^{mon}$	$\hat{M}$-$\hat{M}^{CAPM}$	$\hat{M}$-$\hat{M}^{FF3}$
Size	0.07	3.8	3.7	3.0	1.0	0.1	0.7	2.8*
Value	0.34	3.2	2.3	1.4	-3.0	1.0	1.8	6.2***
Industry	-0.87	3.5	5.0*	5.6**	6.2***	-1.5***	-2.1**	-2.7***
Mom	-0.62	-1.3	1.1	4.3	5.9	-2.4**	-5.6***	-7.3***
Beta	-0.71	0.2	3.8	6.3	5.5	-3.6***	-6.2***	-5.3**
Vol	-0.68	-0.5	3.1	6.6	5.0	-3.6***	-7.1***	-5.5***
IVOL	-0.60	0.9	3.9	6.9	5.9	-3.0***	-6.0***	-5.0***

The table shows the pricing results of the EPKs $\hat{M}$ (estimated using the CP-GARCH in the benchmark method), its monotonic counterpart $\hat{M}^{mon}$, the CAPM ($\hat{M}^{CAPM}$) and the Fama & French (1993) 3 factor model ($\hat{M}^{FF3}$) for value-weighted portfolio sorts. The first column contains the correlation of the portfolios' curvature parameters with the CAPM-alphas for each sort. The next four columns show the alphas of the high-low portfolios. The long and short PFs are chosen such that it produces a positive premium (i.e. for Size, Beta, Vol and IVOL the ordering is PF1-PF5; for Value and Mom the ordering is PF5-PF1; industry PFs are sorted ascending according to their average return). The last three columns depict the differences in the α H-L between $\hat{M}$ and $\hat{M}^{mon}$, $\hat{M}^{CAPM}$ and $\hat{M}^{FF3}$, respectively. The symbols ***, ** and * indicate that values are significantly different from zero at the 1%, 5%, and 10% significance levels, respectively. The significance levels are obtained from 25,000 (pairwise) bootstrap draws from the sample of alphas (differences in alphas). The numbers are annualized in percent. The total number of monthly observations is $N = 234$.

The pricing results are shown in Table 7. The first column summarizes the cross-sectional correlation between the curvature parameters and the CAPM-alphas for each

[16] In addition, to make the α's of different PKs comparable, they are anchored such that they price the CRSP index correctly. As this does not affect the results, I delegate the technical details to Appendix E.

sort. For all sorts but size and value there is the strong pattern that the portfolio with the high abnormal returns have a more concave (inverse-U-shaped) relationship with the index, and vice versa.[17]

The remaining columns summarize the alphas of the H-L portfolio for the different PKs, as well as the differences in alphas between $\hat{M}$ and $\hat{M}^{mon}$, $\hat{M}^{CAPM}$ and $\hat{M}^{FF3}$, respectively. The numbers show that the U-shaped PK does not price the size and value portfolio any better than the monotone version or the factor models. This suggests that these anomalies cannot be explained by a non-monotonic PK, but confirm the conventional understanding that they are risk-factors that are orthogonal to the market risk.

On the contrary, the results for the industry, momentum, beta, vol and IVOL portfolios are highly encouraging. The U-shaped PK reduces the H-L α significantly. Relative to the monotone PK, they are lower by at least 22%, and up to 95%. Against the CAPM, the reduction is at least 25%, and reaches up up to 92%, and relative to the FF3 the reduction is between 5% and 95%. The last three columns of Table 7 show that this reduction in pricing errors is also statistically significant.

In terms of robustness, the results are similarly strong for double sorts on size and even stronger for equally-weighted portfolios. For equally-weighted portfolios, the correlation between the curvature parameter and the CAPM-alphas for Mom, beta, Vol and IVOL is between -95% and -99%. The industry portfolios are an exception, as the negative correlation between alphas and curvature only is present for value-weighting, and reverses for equally-weighting. In addition, the results are similar if the CP-GARCH model in the PK estimation is replaced with alternative volatility forecasts from the robustness section 6.3, as e.g. the VIX. The Online Appendix reports the full pricing statistics for each portfolio sort, together with all confidence intervals. It is worth pointing out that the better pricing performance of the U-shaped PK vs. the monotone one can be observed already for portfolio-wise differences, without even resorting to the high-low portfolio.

When comparing the EPK of CP model to the FP model, the EPKs obtained from the FP GARCH model produce sizable and significant pricing errors. They are in magnitude very similar to those for the index in Section 4.1, and are all statistically significantly different from zero. Hence, on this ground alone, the EPKs estimated using the standard GARCH model must be rejected. In addition, the unexplained return differences are

[17] Boguth et al. (2011) also document the concave-convex return relationship for the momentum portfolios, but their factor model implies a monotonically decreasing pricing kernel.

larger relative to the EPKs from the CP model.

In sum, given that the PK is estimated from option data and without any cross-sectional information, the results are astonishingly strong. The results suggest that some anomalies are only anomalies for linear models, and do not exist if the PK is allowed to be more general. The systematic relationship of higher abnormal returns - higher curvature can be investigated further. For example, one could try to identify "upside" risk factors in standard factor models or use GMM to estimate a non-linear PK. If the PK is estimated such that it explicitly prices above portfolios, one can reduce the PF alphas even closer to zero. The next section presents such an approach to show the link between the U-shaped PK and factor models. Nevertheless, these is still scope for future research.

4.2.4 Adding the U-shape to factor models

How can we integrate the U-shaped PK into standard factor pricing models? To answer this, I suggest a simple modification and illustrate this at the example of the conventional FF3 model. In particular, I propose to "add the U" to the FF3 model:

$$M^{FF3+U} = M^{FF3} + a\frac{(R^I_{t,t+\tau} - b)^2}{h_{t,t+\tau}} - E\Big[a\frac{(R^I_{t,t+\tau} - b)^2}{h_{t,t+\tau}}\Big], \tag{11}$$

where a and b are constants to be estimated, $R^I_{t,t+\tau}$ is the realized return on the S&P 500 index and $h_{t,t+\tau}$ is a conditional ex ante variance forecast for the S&P 500. The "U" is the is the second term and the expectation (third) term normalizes the PK. The idea of this approach is based on two insights from above. First, the U-shaped PK is important to capture several stock-return anomalies, and the simplest way to model this is to use the squared market return. Second, Figure 5 highlights the time-varying wideness of the U-shaped PK, and the conditioning volatility in the denominator captures this. As a volatility forecast, I use either the forecast from the CP-GARCH model, or the VIX, as the latter is model free and easily accessible for other researchers.

I estimate the parameters a and b such that the new PK jointly minimizes the $(\alpha^{H-L})^2$ of all test portfolios and Table 8 presents the results. The numbers show that augmenting the model with the normalized, squared market return improves the pricing performance of the FF3 model for most portfolios. For the industry, beta, Vol and IVOL portfolio, the otherwise sizable alphas are reduced by 40% to 80% in absolute terms (ignoring the sign changes). The last two columns of Table 8 show that these improvements are also statistically significant. With the exception of momentum, these

Table 8: Pricing results of the augmented FF3 model

	α H-L (%, p.a.)			Differences in α H-L (%, p.a.)	
	$\hat{M}^{FF3}$	$\hat{M}^{FF3+U^{CP}}$	$\hat{M}^{FF3+U^{VIX}}$	$\hat{M}^{FF3+U^{CP}} - \hat{M}^{FF3}$	$\hat{M}^{FF3+U^{VIX}} - \hat{M}^{FF3}$
Size	1.0	0.5	1.1	-0.5	0.1
Value	-3.0	-2.2	-2.1	0.8	0.9
Industry	6.2***	3.0	3.6	-3.2***	-2.6***
Momentum	5.9	5.7	6.3	1.7	2.3
Beta	5.5	-1.4	-1.1	-7.3**	-7.0***
Volatility	5.0	-1.4	-1.1	-6.9***	-6.6***
IVOL	5.9	1.0	1.1	-4.0***	-3.9***

The table shows the pricing results of the Fama & French (1993) 3 factor model ($FF3$), and the augmented FF3 pricing kernel ($FF3 + U$) as in Eq. (11) for value-weighted portfolio sorts. The superscript CP and VIX indicate that the conditional volatility forecast $h_{t,t+\tau}$ is obtained either from the CP-GARCH model or the VIX. The pricing errors are for the high-low portfolios. The long and short PFs are chosen such that it produces a positive premium (i.e. for Size, Beta, Vol and IVOL the ordering is PF1-PF5; for Value and Mom the ordering is PF5-PF1; industry PFs are sorted ascending according to their average return). The symbols ***, ** and * indicate that values are significantly different from zero at the 1%, 5%, and 10% significance levels, respectively. The significance levels are obtained from 25,000 (pairwise) bootstrap draws from the sample of alphas (differences in alphas). The numbers are annualized in percent. The total number of monthly observations is $N = 234$.

are the portfolios with the negative correlation between curvature and CAPM-alpha in Table 7. While most portfolios are priced well by the same a and b parameter, the momentum portfolio does require a different set of parameters. Nevertheless, the new PK also does not worsen the pricing of the momentum portfolio. In line with the results above, the improvement in alphas from the standard FF3 model to the augmented one is very small. Finally, one can reduce the alphas to virtually zero if the new PK is estimated to price only the industry, beta, Vol and IVOL portfolio. But the discipline in the estimation by pricing all portfolios shows again that here is an upside risk factor that is relevant for some portfolios, but not all.

While previous studies also have worked with the squared market return as a factor, I find that the normalization in the denominator by the conditional volatility is crucial. Untabulated results show that without the denominator, there is no reasonable set of parameters that achieves even a decent pricing performance. In addition, I again test

whether the upward sloping part of the new PK^{FF3+U} has a relevant contribution to the pricing performance. The results show that when using only the monotone version of the PK^{FF3+U}, the pricing performance worsens even more than above for $\hat{M}^{mon}$ above.

4.3 Trading the pricing kernel

Finally, I introduce a trading strategy that trades the pricing kernel in the options market. The famous Hansen-Jagannathan bound states that the asset, which pays the negative of the SDF, is the asset with the maximum conditional Sharpe ratio in the economy. Here, this is a synthetic derivative which uses the functional form of the EPKs and pays out the realized level of the EPK. The natural interpretation is that it represents the optimal trading strategy in index options in in terms of Sharpe ratio maximization. Since the U-shaped EPKs resemble a strangle, I call this asset the pricing kernel strangle (PKS). Using the cross-section of options one can mimic this optimal derivatives position very well.[18] Practically this means that one combines several puts and calls at a given point in time, to obtain a payoff profile that mimics the lines of the EPKs, as they are e.g. depicted in Figure 7. Formally, the return on a long position in the PKS is:

$$R_{t,t+\tau}^{PKS} = \frac{\check{M}_{t,t+\tau}}{E_t^*(\hat{M}_{t,t+\tau}(R_{t,t+\tau}))}. \tag{12}$$

The optimal strategy is to take a short position in the PKS, as one needs a perfect *negative* relation between the PK and the asset with the maximum Sharpe ratio. In the following, I calculate the excess return on a short position in the PKS as the negative of the long position. In practice, selling the PKS requires to post significant amount of collateral, as it involves large short option positions. While this would reduce the returns significantly[19], it does not affect the Sharpe ratios. Therefore, I focus on the analysis of the latter. One can always construct this synthetic derivative, but only the true PK estimate will maximize the Sharpe ratio of the asset. In practice trading the PKS this would involve significant transaction costs. Hence this is a merely theoretically exercise,

[18] Here, an upper bound of 30 is put on the empirical PK estimates. The main reason is that it reduces noise in the estimates, as the maximum EPK point is in the order of magnitude of 10^6. This cap is rarely applied and only in the tails. A second benefit is that this makes the ratio of price to required collateral similar to a put option. Furthermore, to reduce the noise from pricing, I calculate the price of this synthetic derivative using the risk-neutral density directly.

[19] The ratio of required collateral for the PKS price is in the magnitude of 15-20:1. As a robustness, I calculate the returns to a fully collateralized position in the PKS. The results are very similar in all regards, only the returns are significantly reduced (by about 93%).

but it nevertheless provides valuable economic insights.

Table 9: Sharpe ratios and returns (monthly, in%) of the PKS

	$\hat{M}^{FP}$	$\hat{M}^{CP}$	S&P 500	Short ATM Straddle
Realized return - R_f	16.0	30.6	0.48	10.4
Sharpe ratio	0.32	0.45	0.097	0.13
Thereof in low vol times	0.14	0.95	0.081	0.05
Thereof in high vol times	0.43	0.33	0.108	0.21

The table reports the results for a short position in the pricing kernel strangle (PKS) on a monthly basis, as well as for the S&P 500 index and a short ATM straddle on the S&P 500.

The results in Table 9 offer three findings. First, the obtained Sharpe ratios are indeed very high. However, they are not completely out of the ballpark of values known from the derivatives market. For example, Broadie et al. (2009) find a monthly Sharpe ratio of 0.29 for selling monthly S&P 500 future OTM puts. Furthermore, the Sharpe ratios reported here clearly exceed the values for the index and a short ATM straddle. The latter is included as a comparison as it is well known to deliver high Sharpe ratios and also has some resemblance with the U-shaped PKs.

Second, the Sharpe ratios of the trading strategy that uses the EPKs obtained from the CP model are on average much larger than those from the SP model. In particular, in times of low volatility, where the FP EPKs are mostly S-shaped, the differences are striking. The PKS employing CP estimates delivers very high Sharpe ratios of 0.95 per month (or $0.95 \cdot \sqrt{12} \approx 3.1$ p.a), while the FP version performs poorly. In times of high volatility on the contrary the FP estimates actually perform better. Note that the FP estimates here are typically U-shaped. This again points in the direction that some of the CP estimates in high vol times are too flat or S-shaped due to a too high volatility forecasts. Untabulated results show that excluding these dates does indeed improve these returns and Sharpe ratios.

Third, from studying the payoff profiles of the EPKs in Figure 5 it becomes clear that the PKS involves significant short positions in the tail. Together, this provides a rationale for the well documented high returns and Sharpe ratios of selling puts.[20] The Sharpe ratio is high because selling a put, and in particular deep OTM puts, already

[20] See e.g. Jones (2006), Driessen & Maenhout (2007), Broadie et al. (2009) or Santa-Clara & Saretto (2009).

captures an important part of the payoff of the PKS, which is the asset with the highest attainable Sharpe ratio.

4.4 Predicting market returns

From the Euler equation, one can obtain the following expression for conditional expected excess return of the market $E_t(R^{I,e}_{t,t+\tau})$:

$$E_t(R^{I,e}_{t,t+\tau}) = -\rho(R^{I,e}_{t,t+\tau}, M_{t,t+\tau})\sigma^I_{t,t+\tau}\sigma^M_{t,t+\tau}R_{f,t,t+\tau} \leq \sigma^I_{t,t+\tau}\sigma^M_{t,t+\tau}R_{f,t,t+\tau}, \tag{13}$$

where σ^I and σ^M are the volatility of the index and the PK, respectively. Hence theory suggests that one can use the available empirical estimates of σ^I and σ^M to predict the the market risk premium.[21] As forecasting the future correlation $\rho()$ of $R^{I,e}$ and M comes with additional uncertainty, I use the upper bound on the right-hand side as predictor. I then follow the standard approach as e.g. in Bollerslev et al. (2009) and run predictive regressions of the form:

$$R^I_{t,t+\tau} - R_{f,t,t+\tau} = \alpha + \beta(\sigma^I_{t,t+\tau}\sigma^M_{t,t+\tau}R_{f,t,t+\tau}) + \epsilon_t. \tag{14}$$

Table 10: Predictive regressions

EPK:	$\hat{M}^{CP}$			$\hat{M}^{VIX}$			$\hat{M}^{Rolling\ window\ GARCH}$		
Horizon	$\hat{\alpha}$	$\hat{\beta}$	$R^2(\%)$	$\hat{\alpha}$	$\hat{\beta}$	$R^2(\%)$	$\hat{\alpha}$	$\hat{\beta}$	$R^2(\%)$
1 mo	-0.015	0.75	3.00	-0.011	0.85	1.47	-0.006	0.42	1.05
	(0.0061)	(0.22)		(0.0080)	(0.46)		(0.0063)	(0.27)	
2 mo	-0.014	0.60	2.13	-0.015	0.81	1.06	-0.016	0.70	2.32
	(0.0199)	(0.64)		(0.0232)	(1.03)		(0.0164)	(0.59)	
3 mo	-0.043	1.15	0.77	-0.004	0.39	1.23	-0.021	0.75	2.14
	(0.0495)	(1.06)		(0.0625)	(1.94)		(0.0416)	(1.10)	

The table reports the results for predictive regressions of future S&P 500 market returns on the volatility of the PK and the index (Eq. (13)). Robust standard errors accounting for overlapping observations following Hodrick (1992) are reported in parentheses.

The results in Table 10 suggest that the predictor is particularly strong for short

[21] I thank Mathieu Fournier for pointing out this direction.

horizons of one and two months. To assess the predictive power, I compare the results to two other successful and theoretically motivated predictor variables, namely the variance risk premium as in Bollerslev et al. (2009) and the SVIX as in Martin (2016). The R^2 is high for the one and two month horizon relative to 0.34% and 0.86% obtained for the SVIX, and the 1.07% obtained for the variance risk premium at the monthly horizon (two months not reported). Furthermore, the significance level of the β coefficient also peaks at the shortest horizon, in contrast to the other studies, where the significance peaks are at longer maturities.

In terms of coefficients, the null hypothesis is $\alpha = 0$ and $0 < \beta \leq 1$, which is not rejected at any horizon. The level of β can be interpreted as the negative average correlation of the index returns with the empirical PK. The negative α's suggest that the conditional correlation $\rho()$ and the predictors co-move negatively, i.e. that the correlation is more negative in times of high volatility, and vice versa. To show that the high predictive power stems from the volatility of the EPKs, I also perform the regression with only σ^I as predictor. This delivers insignificant beta estimates and R^2s below 0.2% (results available in the online Appendix).

Table 10 also reports the results for two important methods from the robustness section, which perform similarly well as the benchmark method. This corroborates that the forecasting ability does not rely on a particular volatility forecasts, but rather stems from the high predictive power of the volatility of the volatility σ^M of the EPKs.[22]

5 A Variance-Dependent Pricing Kernel with CP-GARCH

An important question is how the empirically observed U-shaped pricing kernel can be explained economically. Christoffersen et al. (2013) show that the variance risk premium can rationalize the U-shape.[23] This section shows that structural breaks are necessary to make a variance-dependent stochastic discount factor fit the empirical results.

In the following, first the model is introduced and then its empirical investigation presented. An important purpose of the model is to successfully capture the pricing kernel, i.e. the difference between the physical and risk-neutral distributions. To be able

[22] The $\hat{M}^{FP}$ again constantly performs worse than the presented estimates and is hence omitted for brevity.

[23] A similar idea can be found in Chabi-Yo (2012).

to evaluate its ability in that regard, it is necessary to fit both distributions using the same, internally consistent, set of parameters.

5.1 Model

To bridge the gap from the physical to the risk-neutral probabilities a stochastic discount factor (SDF) is required. In their original model Heston & Nandi (2000) use the SDF of Rubinstein (1976). In a log-normal context, this is equivalent to using the Black-Scholes formula for one-period options. Instead, following Christoffersen et al. (2013), the following SDF is assumed here:

$$M(t) = M(0)\Big(\frac{S_t}{S_0}\Big)^{\phi_{y_t}} \exp\Big(\delta_{y_t} t + \eta_{y_t} \sum_{s=1}^{t} h_s + \xi_{y_t}(h_{t+1} - h_1)\Big), \tag{15}$$

where the parameters δ and η govern the time preference, while ϕ and ξ govern the respective aversion to equity and variance risk. When variance is constant, (15) collapses to the power utility case from Rubinstein (1976) and the Black-Scholes model. With $\xi = 0$ the variance risk premium is zero, and with $\xi > 0$ the variance risk premium is negative.With $\phi > 0$ and $\xi > 0$, the SDF is monotonically decreasing in returns and monotonically increasing in variance. The projection of the SDF on the index returns is U-shaped, as can be seen below. The reason is that volatility is not only high for large negative returns, but also for large positive returns. For large positive returns, the negative variance risk premium dominates the positive equity premium, and the projection is increasing.

Under the assumptions (1), (2), (3) and (15), the risk-neutral dynamics for the HN-GARCH model with structural breaks are:

$$\begin{aligned} ln\Big(\frac{S_t}{S_{t-1}}\Big) &= r_t - \frac{1}{2}h_t^* + \sqrt{h_t^*}z_t^*, \\ h_t^* &= \omega_{y_t}^* + \beta_{y_t} h_{t-1}^* + \alpha_{y_t}^*\Big(z_{t-1}^* - \gamma_{y_t}^*\sqrt{h_{t-1}^*}\Big)^2, \end{aligned} \tag{16}$$

where z_t^* has a standard normal distribution under the risk-neutral measure, and

$$\begin{aligned} h_t^* &= \frac{h_t}{1 - 2\alpha_{y_t}\xi_{y_t}}, \\ \omega_{y_t}^* &= \frac{\omega_{y_t}}{1 - 2\alpha_{y_t}\xi_{y_t}}, \\ \alpha_{y_t}^* &= \frac{\alpha_{y_t}}{(1 - 2\alpha_{y_t}\xi_{y_t})^2}, \\ \gamma_{y_t}^* &= \gamma_{y_t} - \phi_{y_t}. \end{aligned} \tag{17}$$

The model is estimated via maximum likelihood, by jointly maximizing the likelihood over the observed returns (physical dynamics) and option prices (risk-neutral dynamics). The approach is as in Christoffersen et al. (2013) and for the details I refer to their paper.

5.2 Data and estimation

As before, the data is out-of-the-money S&P 500 call and put options that are traded in the period from January 01, 1996 to August 31, 2015. For each Wednesdays the option series with a maturity closest to 30 days is selected. From that maturity, the 15 most actively traded options are used, which covers a broad range of moneyness. This results in 15,171 option prices with a maturity between 17 and 53 days.

Table 11 presents the estimation results.[24] The requirement to simultaneously fit the option prices and returns requires a balancing of the risk-neutral and physical dynamics. Therefore, the physical parameters are now different from the ones presented in Table 1. Nevertheless, as above, the standard GARCH with fixed parameters has an average long-run volatility, while the CP model has regimes that capture times of high and low volatility. The risk-neutral volatility is higher than the physical volatility, which implies a sizable variance risk premium.

The likelihood of the CP model is significantly higher than the likelihood of the model with fixed parameters. Furthermore, both information criteria, that put a penalty on the number of parameters, prefer the model with breaks, as well as the likelihood-ratio test. Note that the option likelihood increases significantly. This shows that the breaks

[24] Theoretically it would be possible to perform a maximum-likelihood estimation of the change-point model over the full sample of returns and option prices jointly. However, this is practically infeasible due to the computational burden that mostly stems from the option pricing part. Therefore, I use the breaks identified in the return process from the estimation in Section 2.3, and estimate the model separately in each regime for the change-point version.

are also important for the risk-neutral dynamics.[25]

[25]The Online Appendix provides a more detailed analysis of the option pricing performance.

Table 11: Estimation results of the joint estimation of the HN-GARCH model

	FP HN-GARCH	CP-HN-GARCH				
Physical parameters	'92-15	'92-'96	'96-'03	'03-'07	'07-'11	'11-'15
ω	1.90E-14	8.97E-08	2.20E-10	6.87E-07	1.82E-11	2.08E-12
α	1.56E-06	2.89E-07	2.47E-06	1.78E-06	1.42E-06	4.57E-06
β	0.713	0.699	0.826	0.733	0.696	0.820
γ	417.1	999.0	238.6	341.7	441.0	174.4
μ	4.327	17.016	5.895	0.109	10.471	7.825
Risk-neutral parameters						
ω^*	2.22E-14	1.26E-07	3.49E-10	9.64E-07	2.87E-11	2.50E-12
α^*	2.13E-06	5.71E-07	6.25E-06	3.50E-06	3.53E-06	6.63E-06
$\beta^* = \beta$	0.713	0.699	0.826	0.733	0.696	0.820
γ^*	360.8	723.0	153.9	243.8	286.7	151.4
ξ	46,248	499,023	75,168	80,716	128,518	18,541
Properties						
$\beta + \alpha\gamma^2$	0.9839	0.988	0.9667	0.9408	0.9728	0.9592
$\beta^* + \alpha^*\gamma^{*2}$	0.9896	0.9980	0.9741	0.9412	0.9864	0.9722
Long-run volatility	0.156	0.089	0.137	0.102	0.115	0.168
$\sqrt{h^*/h} = \sqrt{(1-2\alpha\xi)}$	1.081	1.186	1.261	1.184	1.255	1.097
Log-likelihood	FP	CP	Likelihood-ratio test			
Total	47,687	53,461	Test statistic			11549
From returns	19,359	19,560	1% critical value (χ^2_{29})			49.6
From options	28,328	33,901	0.1% critical value (χ^2_{29})			58.3
AIC	-95,362	-106,862				
BIC	-95,314	-106,624				

Parameter estimates are obtained by optimizing the likelihood on returns and options jointly. Parameters are daily, long-run volatility is annualized. For each model, the total likelihood value at the optimum is reported as well as the value of the return component and the option component. The volatility parameters are constrained such that the variance is positive ($0 \leq \alpha < 1,\ 0 \leq \beta < 1,\ \alpha\gamma^2 + \beta < 1,\ 0 < \omega$). The Akaike information criterion (AIC) is calculated as $2k - 2\ln(L^R + L^O)$ and Bayesian information criterion (BIC) is calculated as $\ln(n)k - 2\ln(L^R + L^O)$, where n is the length of the sample and k is the number of estimated parameters. The test statistic of the likelihood ratio test is $-2(LL_{FP} - LL_{CP})$.

5.3 Results

5.3.1 Model implied pricing kernels

Figure 11 presents the model implied pricing kernels for the GARCH with fixed parameters (left) and for the CP-GARCH (right). Each plotted line corresponds to the date of one of the EPKs from the Figures 4 and 5. The lines are obtained by first simulating 100,000 paths under the physical measure using the parameter estimates from Table 11, and then calculating the stochastic discount factor and its projection on the index return. The comparison of the plots shows how the structural breaks are necessary to match the empirical results. First, the standard GARCH always implies U-shaped PKs, while its empirical counterparts in Figure 4 exhibit S-shapes in many periods. On the contrary, the CP version of the model matches the empirical results from Fiigure 5 astonishingly well. Second, the model implied PKs of the standard GARCH have little variation in their wideness across time, which contradicts the strongly time-varying EKPs. In contrast, the CP version exhibits a similar time variation in their wideness as their empirical counterparts. In addition, for the CP model, the model implied PKs do appear similar within one regime. This again matches the empirical findings well, where the estimated shape within one regime is relatively stable.

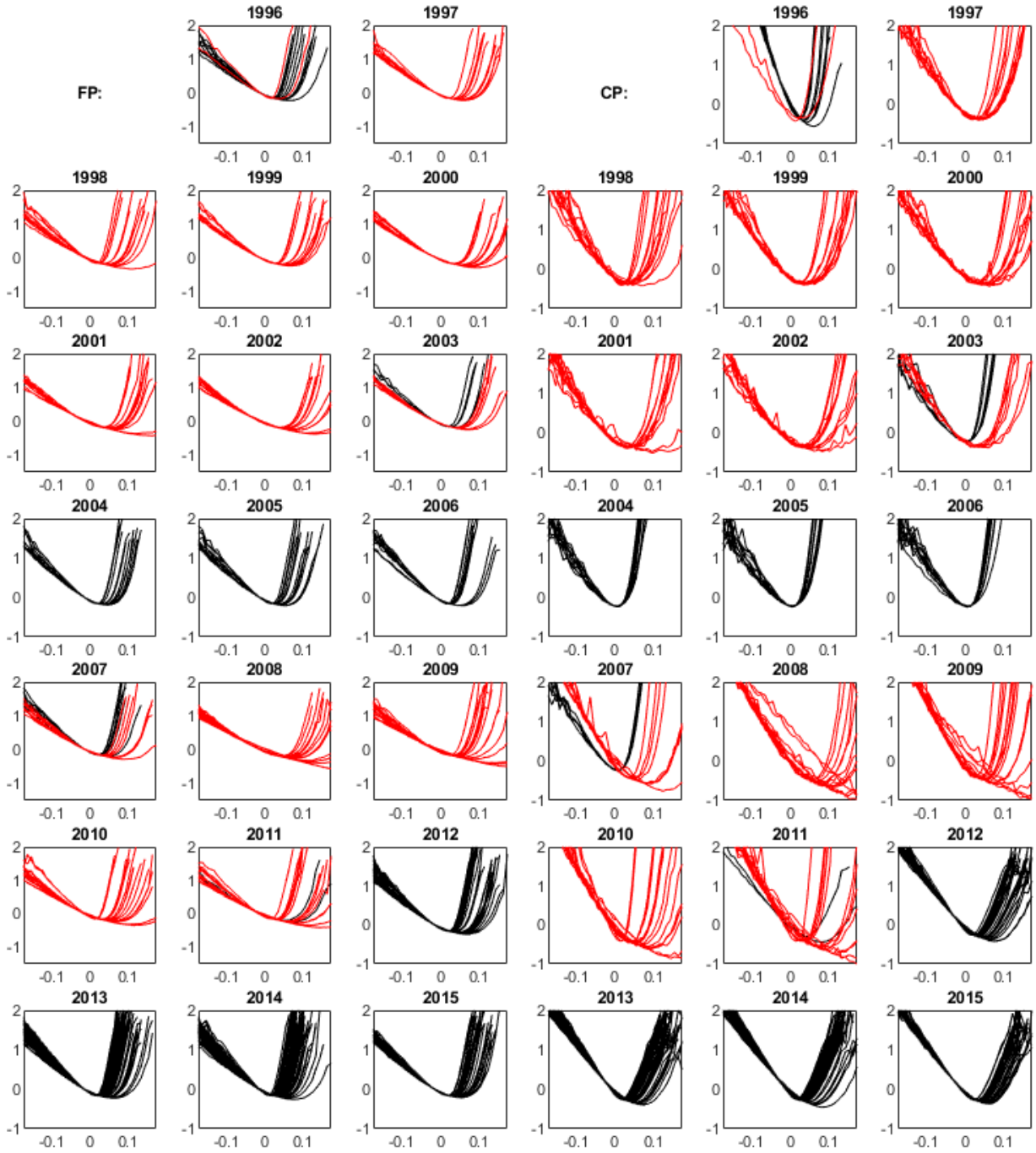


Figure 11: Model implied pricing kernels with FP and CP parameters

The figure shows the theoretical pricing kernel in the Heston-Nandi GARCH model with fixed (left) and CP (right) parameters. Red (black) depicts times with high (low) variance, as defined in Ch. 2.3.2. Log-returns are on the horizontal axis. The horizon is one month.

6 Robustness

6.1 Option returns

The results on the estimated empirical pricing kernels in Section 3.3 are partially ambiguous regarding the shape. In particular, towards the end of the low volatility regimes, it is unclear whether the PK is upward or downward sloping. The statistical tests in Section 3.5 do confirm that the majority of pricing kernel estimates with the CP model are upward sloping, but do not allow to test for global shape characteristics. I therefore take an alternative approach. Bakshi et al. (2010) show how to use option returns to examine the shape of the pricing kernel. In particular, a U-shaped PK implies that the returns of OTM call options decrease in their strike beyond a certain point. This can be studied by sorting the options according to their moneyness and then calculating average returns per moneyness bin. Their results show that the PK is U-shaped on average for their full sample from 1988-2008. The method is fully non-parametric, but only speaks about the average, i.e. unconditional pricing kernel. Intuitively, the method analyzes an object similar to the inverse of the PK: physical probabilities over risk-neutral probabilities. In particular, the first is the realized payoff and the second is the ex ante call price:

$$E_t(R_{t,t+\tau}^{Option\ i}) = \frac{E_t((S_{t+\tau} - K_i)^+)}{E_t^*((S_{t+\tau} - K_i)^+)} = \frac{E_t((S_{t+\tau} - K_i)^+)}{E_t((S_{t+\tau} - K_i)^+ M_{t,t+\tau})}. \tag{18}$$

To show that the findings of Bakshi et al. (2010) are not only driven by option returns from times of high volatility where the U-shape is less ambiguous, I replicate their approach using only call option returns from only the low volatility times. For this, from all call options from the respective time period, those are selected that are closest to the target moneyness of the bin. Then option returns are calculated using the corresponding settlement prices. Finally, average returns for each moneyness group are calculated and the results are displayed in Table 12. The confidence intervals are bootstrapped as in Bakshi et al. (2010). As in the original paper, I use 2% moneyness steps starting at 1% OTM. The sorting ends at 9% OTM, because the number of traded options decreases significantly here.

If average call returns decrease when their moneyness increases, this shows that the (average) pricing kernel is increasing with returns. The returns documented in Table 12 have exactly this pattern. The noisy nature of option returns is well known (e.g. Broadie et al. 2009), but the results here have both a higher statistical and economic significance

than in Bakshi et al. (2010). All the negative average returns are significantly different from zero at conventional levels. In addition, all pair-wise differences between groups are negative and two of them are statistically different from zero. In sum, the results are highly consistent with a U-shaped pricing kernel. They show non-parametrically that also in times of low volatility the unconditional PK is upward sloping in returns in the domain of high positive returns. This does not rule out that some conditional PKs are not increasing, but in a sense, the majority of them should be increasing.

Table 12: Returns (monthly, in %) of OTM call options

% OTM	1%	3%	5%	7%	9%
Average return	10.2	-11.5*	-28.8*	-54.6*	-84.1***
99% Confidence	[-12.6	[-35.1	[-61.0	[-83.8	[-97.7
	19.0]	9.7]	19.6]	30.6]	-0.2]
		3% - 1%	5% - 3%	7% - 5%	9% - 7%
Differences:		-21.7***	-17.3*	-25.8	-29.6
99% Confidence		[5.7	[-17.7	[-130.6	[0.0
		40.2]	70.1]	58.6]	161.5]

The reported average returns (monthly, in percent) are of call options on the S&P 500 index in the sample from 01/1996 to 08/2015 that are closest to the target moneyness and that are from periods of low volatility, as defined in Section 2.3.2. Call returns are calculated using actual settlement prices. Moneyness is calculated as K/S_0. The average time to maturity is 30 days. The symbols ***, ** and * indicate that values are significantly different from zero at the 1%, 5%, and 10% significance levels, respectively. The significance levels are obtained from 25,000 (pairwise) bootstrap draws from the sample of returns (return differences).

6.2 International evidence

Most studies on the pricing kernel use US option data, and in particular usually use options written on the S&P 500 index. However, some studies present international evidence, mostly on the German DAX 30 index and the British FTSE 100, and most of them document S-shaped pricing kernels.[26] Therefore, I repeat all the above analysis for several major international equity indices, namely the FTSE 100, EuroStoxx 50 and DAX 30. In general, the results are the same as above in all regards.

[26] Cuesdeanu & Jackwerth (2018), Section 3.2. to 3.4., provide a great and comprehensive overview of these studies.

For brevity, this section presents only the main results for the FTSE 100, as it is the second most studied underlying in the PK literature. The missing details, as e.g. the parameter estimates and volatility prediction details, as well as all the results for the other indices are presented in full depth in the Online Appendix. The used methodology is exactly as above, only with option and return data for the FTSE 100 and the one month GBP LIBOR rate as risk-free rate.

The estimated model shows the same patterns and properties as the model estimated with S&P 500 data. The change-point GARCH model exhibits alternating high and low variance regimes which are accompanied by low and high drift, respectively. The degree of integration of the variance equation in each regime is lower than in the standard GARCH model. Finally, the increase of total log-likelihood from the standard GARCH with fixed parameters to the change-point GARCH is substantial, and also both the Bayesian and Akaike information criterion strongly prefer the CP model. Furthermore, the volatility forecasts of the standard GARCH model with fixed parameters show a strong cyclical bias, while the forecasts of the CP-GARCH capture the different volatility regimes very well. Again, this result can also be obtained without relying on estimating the structural breaks, but by using the FTSE volatility index or the standard GARCH to identify times of high and low volatility.

Finally, Figure 12 shows the EPKs estimated with the standard methodology (upper part) and with the change-point model (lower part). The analysis delivers results that are very similar to the S&P 500 case. First, it is clearly visible that the standard estimation methodology delivers S-shaped PKs in times of low variance (black), and U-shaped ones otherwise (red). On the contrary, the estimation methodology that employs the CP-GARCH mostly delivers U-shaped PK estimates. The last regime, and in particular the year 2014 seems to be an exception. The details on the volatility time series however show, that the volatility in 2014 was so low, that also the CP model has problems capturing it and severely overpredicts the volatility.

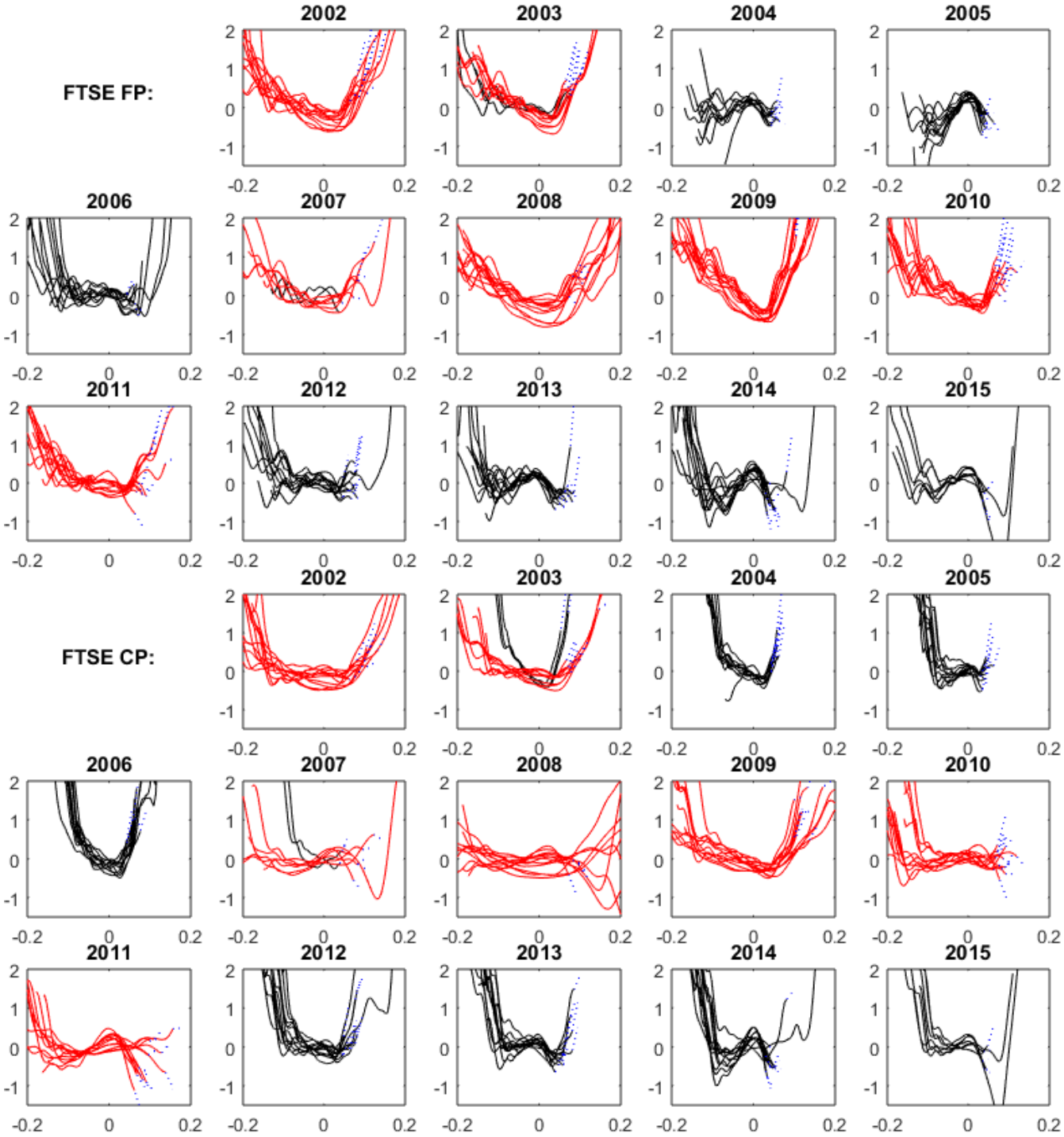


Figure 12: Empirical pricing kernels with FP and CP for the FTSE 100

The figure shows the natural logarithm of estimated pricing kernels obtained when using the Heston-Nandi model with fixed (upper half) and CP parameters (lower half). Red (black) depicts times with high (low) variance, as defined in Section 2.3.2. Log-returns are on the horizontal axis. The horizon is one month. The dotted blue line connects the points, which depict the ratio of the CDFs of the tail, with the corresponding pricing kernels.

The Online Appendix shows that very similar results can be obtained for the EuroStoxx 50 (which has not been studied in the literature so far) and the DAX 30. This shows that the results presented so far are generalizable to other indices. The only attenuation is that the results get less strong the fewer stocks the index contains. In parallel, also the likelihood of the GARCH model decreases. The reason could be that the ratio of systematic to idiosyncratic volatility in the index is likely to decrease, i.e. that the amount of unpriced risk increases.

6.3 Alternative volatility forecasts

6.3.1 VIX

The above analysis points out that it is important to use a precise and unbiased conditional volatility forecast. So far at least the volatility forecast was model-dependent. The VIX is a non-parametric and conditional volatility forecast that is well understood in finance, and the flexibility of the benchmark method allows me to incorporate it into the approach. However, it comes with the caveat that it is not the expected future volatility, but the expected future volatility under the risk-neutral measure. This means that it includes one or several risk premia and is typically much higher than the physical expectation. Nevertheless, I present an analysis where the level of the VIX replaces the model-implied ex ante expected volatility in Eq. (6) and (7). This means that the level of the VIX is used both for normalizing the monthly returns and for rescaling. Chang et al. (2012) show that for certain applications, the bias caused by using risk-neutral moments can be quite small. This means here, there are risk premia both in the numerator and the denominator, and the two effects by large cancel out.

The results are presented for two reasons. First, the VIX is a well understood measure and can at least be used as a robustness check. Second, the VIX is a truly conditional expectation that in particular reflects market expectations. This is the reason why other studies also use it as a measure for the expected physical variance (e.g. Figlewski & Malik 2014). Figure 13 presents the results. The estimated PKs are U-shaped throughout the entire time series and without any major humps. Again it is evident that the steepness of the right-hand end of the PK estimates decreases over the course of the regimes.

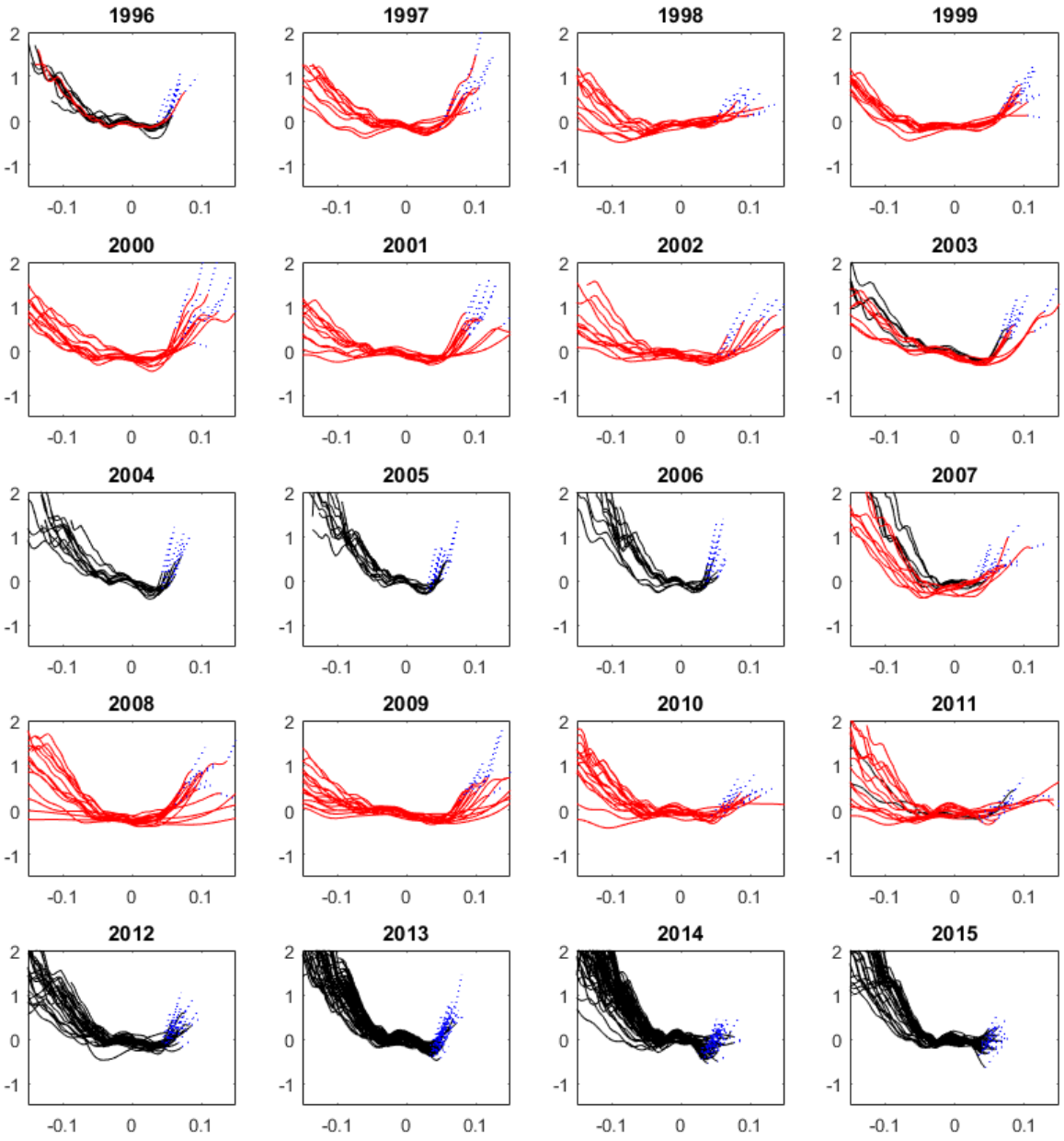


Figure 13: Empirical pricing kernels using the VIX as volatility forecast

The figure shows the natural logarithm of estimated pricing kernels obtained by using the VIX as volatility forecast. Red (black) depicts times with high (low) variance, as defined in Section 2.3.2. Log-returns are on the horizontal axis. The horizon is one month. The dotted blue line connects the points, which depict the ratio of the CDFs of the tail, with the corresponding pricing kernels.

6.3.2 Realized volatility

Corsi (2009) proposes an appealing model for realized volatility, that is structurally very different to the GARCH models. It has become popular due to its parsimony, straightforward estimation and good empirical performance. The model uses high frequency data (typically five minute intervals) to estimate realized volatility. Since both using realized volatility from high-frequency data and the Corsi model have become popular, I include the latter as a robustness check.[27]

The following summarizes the main results, while the Online Appendix provides the model's technical details, estimation results and a more detailed analysis. First, the volatility forecasts of the heterogeneous autoregressive (HAR) model of Corsi (2009) also have a cyclical bias. Compared to the standard GARCH model, the bias is smaller in magnitude and more pronounced in high volatility times than in calm periods. Next, I incorporate this volatility model into my pricing kernel estimation methodology, by employing its volatility forecast in Eq. (6) and (7), while everything else remains unchanged. The estimated pricing kernels appear U-shaped most of the time. Again, the pattern of decreasing steepness of the right-hand side of the PK estimates is present. As above, the reason is that this approach cannot handle time-variation in the higher moments. Furthermore, the wideness of the PK estimates across time is very similar. This is probably caused by the general underestimation of volatility by the HAR model, especially in times of high volatility. The Euler equation errors in Section 4.1 support this conjecture. It appears that the HAR model, which is specified in terms of volatility, has the tendency to overfit low volatility states at the expense of high volatility states. This is reversed relative to GARCH models, which a specified in terms of variance. The results again point in the direction of different volatility regimes, that are hard to capture with standard models.

6.3.3 Including the switching probability

The benchmark analysis ignores the probability to switch into another regime, since this allows the analysis to be more tractable and the probability is very small anyway. As a first robustness, a version with a crude control for a regime switching probability is estimated. The highest switching probability is close to 0.001. With this switching probability, the probability of the state to switch at all over a time period of 21 days is $1 - 0.999^{21} = 2.08\%$. The analysis is repeated using a 2% switching probability from

[27] I thank the Oxford-Man Institute (Heber et al. 2009) for providing the data on their website.

a red (black) regime to an average low (high) variance regime. The 2% are still upper bound, because even if the state switches, it can switch at any of the 21 days and hence only parts of the final density come from the later state. The results show that the impact is only marginal and the omission of the switching probability cannot explain the results.

The second robustness fully includes the switching probabilities in the forecasting of the variance. The forecast is then used both in the construction of the shock densities and the conditional return density, which comes at the cost that the approach is significantly more involved. Again, the results show that the impact is only marginal and the omission of the switching probability is a reasonable simplification. Both results and the technical details are included in the Online Appendix.

6.4 Simulated return distributions

A popular alternative to obtain a conditional forecast of the physical return distribution is to simulate it directly from the GARCH model. I implement this approach for both the fixed parameter and change-point HN-GARCH, using the parameters from Table 1. For each density, I simulate 100,000 paths, starting from the filtered variance. The results confirm the previous findings: for the FP model, one observes U-shaped PKs in times of high variance and strongly S-shaped ones in times of low variance. Notably, this pattern is even more pronounced here than in the analysis above. When turning to the CP-GARCH model, the finding of S-shaped PKs disappears. To save space, the corresponding plots are included in the Online Appendix.

In addition, the results in Section 4.1 show that the PK estimates, which employ the fixed parameter GARCH model, violate the unconditional Euler equation. The pricing errors are particularly large in times of low volatility, which casts additional doubt on the validity of the S-shape. It also shows that there is a problem by the physical density forecasts obtained from the FP GARCH model especially in these regimes.

6.5 Different horizons

Studies on empirical pricing kernels typically use maturities between two weeks and two months. To show that the results are not specific to the one month horizon, I repeat the analysis with maturities of two weeks, six weeks and two months. For each horizon, both the results for the benchmark method with empirical shocks and for the simulated GARCH kernel method from Section 6.4 are reported. The graphs show that the results

are robust against changes in the analyzed horizon. All the observations above can be found for the other maturities and are equally strong. The results can be found in the Online Appendix.

6.6 Alternative GARCH specifications

6.6.1 t-distributed innovations

Numerous studies argue that the shocks in GARCH models do not follow a normal distribution. A popular alternative is to use t-distributed shocks. I implement this as a robustness check and find that the results a fully robust to the alternative model specification. In sum, the fit of the GARCH model improves for the CP model, but worsens for the FP model estimated over the full sample. Furthermore, the shape of the EPKs is almost unaltered, and the economic properties are quantitatively very similar. More detailed results are included in the Online Appendix.

6.6.2 NGARCH model

An alternative to the so far employed Heston-Nandi GARCH model that is also used for option pricing is the NGARCH model of Duan (1995). The results in Online Appendix show that the findings are robust to using an alternative GARCH model specification. The major difference in the dynamics of the two GARCH models are the drift term in the return equation and the "alpha-term" in the variance equation. Most other popular GARCH models use one of these two specifications, and mainly differ in how the "leverage-effect" is modeled.

All the above analysis is repeated using the NGARCH model and the results remain the same as above. This does not only hold for the pricing kernel estimates, but also for the (biased) volatility forecasting, the properties of the shock densities as well as all the robustness checks. In sum, this suggests that the key driver of the results, namely the biased volatility forecast, prevails in any GARCH specifications.

6.6.3 Rolling window estimation

The analysis above points out that the mean reversion property of standard GARCH estimated over a long time series biases the volatility forecasts, especially in times of low volatility. One way to address this could be to use a rolling window GARCH with a relatively short window length. To test this, I conduct an explicit data mining exercise.

First, I estimated the HN-GARCH model on the first trading day of each calendar month using the past N=[63, 84, 105, 126, 189, 252, 315, 378, 504, 630] daily returns, and then use the obtained parameters to filter and forecasts variance in that month. The forecast horizon was 21 days. To keep the analysis comprehensible, I present the key findings, while more detailed results are included in the Online Appendix.

First, it appears that the estimated parameters vary substantially, and are at the boundary of the parameter space a significant number of times. As a consequence, for example, if one simulates the GARCH model, many of the obtained return distributions appear unreasonable. Second, the volatility forecasting with rolling window GARCH works worse than for the change-point GARCH, but better than for the standard GARCH which is estimated over the full sample (measured by RMSE). This is possibly surprising, and contradicts studies which find that the GARCH forecasts are the better the longer the estimation sample is, both for short- and long-term forecasts (e.g. Figlewski 1994). The volatility forecasts are best for a 126d-252d estimation window. Third, the pricing kernel estimates are mostly U-shaped. Compared to the estimates with the CP-GARCH they have less humps in the period 2008-2010, but have significant humps when leaving high vol phase, e.g. 2003, 2011-12 and they are unreasonable tight at beginning of high vol phase: 96/97, 2007. At these times, the forecasts are most unreasonable, since they are based on observations of the recent very calm or volatile period, when conditions change suddenly.

However, the estimated PKs using rolling window GARCH all violate the unconditional Euler equation. The average value is larger than one, and this is statistically significant. It seems that the error emerges mostly in times of high variance, possibly from the PK being "too tight/not wide enough". The lowest error was obtained for the 126d window, followed by 189d and 252d. These specification also produce the most "reasonable" PK estimates, in the sense that they appear most regular.

Together, this points into the direction that a variance estimate that is based (only) on the recent past can partly solve the problem. However, the problem of sudden changes remains. This problem is nicely addresses by the specification with the VIX as volatility forecasts, and is probably the reason why the latter specification produces rather smooth results.

7 Summary and Conclusion

Numerous empirical studies agree that the pricing kernel derived from option prices is not monotonically decreasing in index returns, but disagree whether it is U-shaped or S-shaped. The paper shows that the finding of S-shaped pricing kernels, which is conflicting with most asset pricing models, is spurious. The reason is the canonical use of standard GARCH models that, when estimated over a long time series, lead to volatility forecasts that systematically biased towards the long-run mean, especially in times of low volatility. Therefore, this paper identifies structural breaks in the volatility process of the market index using a change-point GARCH model. The model has the advantage that it can capture market phases where the volatility is very high or low for extended periods of time. The results brought forward show that the breaks are not only a technical econometric issue, but rather change economic results. Moreover, from all the different tested methods one picture emerges: as long as the method to forecast volatility avoids the previously unknown cyclical volatility forecast bias, the S-shaped pricing kernel estimates disappear.

The paper also shows how the new pricing kernel estimates can be applied to several key questions in asset pricing. First, they allow to construct a pricing kernel mimicking trading strategy in the options market. The obtained Sharpe ratios are high for U-shaped PKs, but not for S-shaped ones. In addition, these results provide a rational for the observed high returns and Sharpe ratios from selling puts, as these two strategies are related. But there are also implications outside the world of option, as the U-shaped pricing kernel for example helps to explain the cross-section of stock returns. In particular, the upward sloping part of the U-shaped PK commands a risk premium for high return states. Testing this for several standard return anomalies confirms that this is very relevant for pricing e.g. beta or momentum sorted portfolios. Lastly, the PK estimates provide a good and theoretically motivated predictor of future market returns.

Finally, the empirical results are matched very well by a variance-dependent pricing kernel, but only when the structural breaks are included in the model. The model generates a U-shaped projection of the stochastic discount factor on the returns of the index due to the variance risk premium. Including the breaks into the model is necessary to match both the shape and time-variation of the pricing kernel in the data.

The results in this paper can be extended and generalized in a number of ways. First, there are likely to exist further areas of research in finance where the introduction of structural breaks into the GARCH process might lead to a change in findings. Second,

one can study more general pricing kernels and richer dynamics. When using option data, the particular challenge is to fit both the observed returns and option prices as pointed out by Christoffersen et al. (2013). Finally, the results provide an important benchmark to more general asset pricing models. The findings suggest that a model should generate a U-shaped projection of the stochastic discount factor on the returns of the market index. Not only because this matches the documented shape, but also because the U-shaped property is important for various asset pricing questions. Identifying risk factors which make states with high returns risky and expensive for investors is economically very interesting.

Appendix

A Data Cleaning

The data cleaning follows the standard approach in the literature:

1. Remove all quotes that: have zero trading volume on the day of the price quote, have best bid is below \$3/8, are more than 20 points in-the-money and violate standard no-arbitrage bounds.
2. For each month, find the option series that has a remaining time to maturity closest to the desired time to maturity. For the benchmark one month horizon this is 30 days, and in the robustness section this is 9 days (one week), 16 days (two weeks), 44 days (six weeks), 60 days (two months) and 90 days (three months). The time is chosen such that the data is from Wednesdays, if available.

B Estimation of Risk-Neutral Density

The exact steps for the estimation of the risk-neutral density from options prices:

1. Clean the data as described above.
2. Get risk-free rate from OptionMetrics, and interpolate linearly for the correct maturity.
3. Get the dividend yield data from OptionMetrics, and interpolate linearly for the correct maturity (Using the implied dividend yield from at the money call and put pair leads to similar results, but the dividend yield estimates are more noisy).

4. Transform mid-prices into implied volatilities using Black and Scholes (1973). In the region of +/- 20 points from at-the-money, take a weighted average (by volume) of put and call implied volatilities. The results remain unaltered if the implied volatility provided by OptionMetrics is used.

5. Fit a 4th order polynomial to the implied volatilities over a dense set of strike prices, and convert back into call option prices using Black-Scholes. (Using a second order polynomial delivers virtually the same results).

6. Numerically differentiate the call prices using (19) and (20) to recover the risk-neutral return distribution.

$$1 - F^*(S_{t,t+\tau}) = -\exp(r\tau)\Big[\frac{\partial C_{BS}(S_t, X, \tau, r, \hat{\sigma}(S_t, X))}{\partial X}\Big]_{|X=S_{t,t+\tau}}. \tag{19}$$

$$f^*(S_{t,t+\tau}) = \exp(r\tau)\Big[\frac{\partial^2 C_{BS}(S_t, X, \tau, r, \hat{\sigma}(S_t, X))}{\partial X^2}\Big]_{|X=S_{t,t+\tau}}. \tag{20}$$

7. Finally, if on a selected date, the probability mass of the risk-neutral measure within the range of traded range is below 0.8 (i.e. the CDF between the lowest and highest traded strike on that date), the date is discarded for all the pricing kernel related analysis.

C Note on the matching maturity

The physical return density in Section 3.2 uses the number of expected trading days between the date of the option price and expiry. Expected trading days are the number of working days minus the holidays between the date of the option price and the maturity date (i.e. not including unexpected trading suspensions). The majority of options in the sample are AM settled, meaning, the settlement price is determined by the opening price on the expiration date. Therefore, the chosen approach slightly overestimates the time to maturity. Overestimating the time horizons also leads directly to a higher variance estimate. If the AM settled options would be treated as if settled PM the previous day (as some studies do), the U-shape get even stronger due to lower volatility forecasts.

D Calculation of Confidence Intervals

In order to calculate confidence intervals for the estimated pricing kernels, measures for the precision of both the risk-neutral and physical density estimates are required. To obtain the asymptotic distribution of $\hat{f}^*$.[28] I follow Aït-Sahalia & Duarte (2003), who add noise to the observed option prices. Aït-Sahalia & Duarte (2003) calibrate the noise to the typical intraday variation of S&P 500 option prices. In particular, the noise is uniformly distributed, and the values range from 3% for deep in the money options (moneyness of 1.2) to 18% for deep out of the money options (moneyness of 0.8). I use the same specification and also $N = 5000$ simulations. Since the liquidity in the sample used here is higher, this is a conservative approach. Next, I follow the standard methodology from Section 3.2 above. The noisy option prices are converted into Black-Scholes implied volatility and the implied volatility is smoothed using a fourth order polynomial. Then, call prices are calculated using a dense grid of strikes and finally these prices are differentiated once and twice to obtain the risk-neutral CDF and risk-neutral density, respectively. Repeating this $N = 5000$ times delivers a distribution of $\hat{f}^*$ for each point on the grid.

Next, for the precision of the physical density estimate $\hat{f}$, I rely on the asymptotic result of Härdle et al. (2014). They show that:

$$\sqrt{n_p h_{n_p}}\{\hat{f}(x) - f(x)\} \xrightarrow{\mathcal{L}} N(0, f(x) \int K^2(u)du), \tag{21}$$

where $\mathcal{L}$ denotes convergence in law, x is a point on the grid, $f(x)$ is the true density, n_p is the number of observed returns and $K(\cdot)$ is a kernel function with bandwidth h_{n_p}.

Finally, to obtain the $(1-\alpha)\%$ confidence interval of the estimated pricing kernel for each point on the grid, I numerically solve for $\hat{PK}^{upper}(x)$ in:

$$P\Big(\frac{\hat{f}^*(x)}{\hat{f}(x)}\Big) \geq \hat{PK}^{upper}(x) = (\alpha/2), \tag{22}$$

and for $\hat{PK}^{lower}(x)$ in:

$$P\Big(\frac{\hat{f}^*(x)}{\hat{f}(x)}\Big) \geq \hat{PK}^{lower}(x) = (1-\alpha/2), \tag{23}$$

[28] I use the notation $\hat{f}^*$ here for what was denoted f^* above, to differentiate between the estimated density and the true, unknown quantity.

where

$$\Big[P\Big(\frac{\hat{f}^*(x)}{\hat{f}(x)}\Big) \geq \hat{PK}(x)\Big] = \frac{1}{N} \sum_{\hat{f}^*(x)_{min}}^{\hat{f}^*(x)_{max}} \int_{-\infty}^{\hat{f}^*(x)/\hat{PK}(x)} \hat{f}(x)dx. \tag{24}$$

The calculation for $\hat{F}^*$ is done analogously.

E Pricing kernels for the cross-section

E.1 Anchoring the PK

It is necessary to anchor the EPKs for the cross-sectional return analysis in Section 4.2, since otherwise $E(\hat{M}^{mon}) < 1/R_f$. As the EPK of the tail of the return distribution is not available, one can not use R_f as anchor. Therefore, the EPKs are anchored (adjusted) such that they price the CRSP (R^I here) value-weighted market index. The PK are adjusted by adding a constant a:

$$a = 1 - \frac{1}{T} \sum_{t=1}^{T} \Big(\tilde{M}_{t,t+\tau} R^I_{t,t+\tau}\Big), \tag{25}$$

where $\tilde{M}$ is the unadjusted PK. For the $\hat{M}^{CP}$ a is very small, i.e. around 0.0004, but for $\hat{M}^{CP,mon}$ it is around 0.05. Note that this only anchors the α's around zero by shifting the level of the α's, but has no effect on the differences in α's between portfolios.

E.2 Pricing kernel in the CAMP

In the CAPM, the pricing kernel is linear in the market index return R^I, and given by:

$$M^{CAPM} = a - b \cdot R^I, \tag{26}$$

where:

$$a = \frac{1}{R_f} + b \cdot \mu_I, \tag{27}$$

$$b = \frac{\mu^I_M - R_f + 1}{R_f \sigma^2_I}, \tag{28}$$

and R_f is the gross risk free rate. To be consistent, I use the CRSP value weighted index and μ_I and R_f are calculated as the sample mean. The realized PK is then calculated

as:

$$M_{t,t+\tau}^{CAPM} = a - b \cdot R_{t,t+\tau}^{I}. \tag{29}$$

F Portfolios construction

The beta, momentum, volatility, and IVOL portfolios are calculated using CRSP data and opening prices (see below). The industry, size and value sorts use the daily return data from the website of Kenneth French. They are therefore not opening, but closing returns on the option expiration date. Using the other sorts, I verify that this has very little impact on the results. Lastly, as the industry portfolios do not have a monotonic return pattern, they are resorted according to their in-sample returns.

The stock returns are obtained from CRSP, the risk-free rate and all factors are form the website of Kenneth French. Stocks are excluded if they have less then 180 valid return observations per year. Lastly, stocks with an ex ante price below 1$ (5$ as robustness), exchange code other than 1, 2, 3 and 31 and share code other then 10 and 11 are excluded.

Following Bali et al. (2014), the ex ante beta is estimated using a rolling window of one year of daily returns. Momentum is calculated as the return from month -12 to month -1. Following Ang et al. (2006), volatility is measured as thee sum um 21 daily returns, and similarly, IVOL as the sum of squared 21 daily residuals of the Fama-French 3 factor model. For the quintile break points only the characteristics of NYSE traded stocks are considered, as suggested by Kenneth French. This aims at mitigating the impact of the large number of small stocks that are predominantly traded as NASDAQ. The monthly portfolio returns are calculated to match the dates of the option prices and expiration dates. Only option for the monthly AM expiration cycle (SPX) are used, which gives a series of (almost) non-overlapping 30 day returns.

For the independent double sorts on size and the respective characteristics, the market values from the CRSP database are used. The results are robust against using equal-weighted or value-weighted portfolio returns, a 1$ or 5$ exclusion criterion, different estimation periods, using either only univariate sorting or independent double sorts on size and the characteristics and against using the in-sample size quintiles or the size quintiles provided by Kenneth French.

Lastly, the options expiry "AM" on the third Friday of each month, i.e. the settlement price is the opening price. Hence to be exact, I calculate "open returns" by using the closing return provided by CRSP and divide this by the intraday return of that day,

where I follow the approach of Polk et al. (2018). As a robustness, the results remain unchanged if the closing returns either from the expiry date or the preceding day are used. For size, value and industry sorts, only closing returns for the expiry date are used.

The confidence intervals for the pricing errors are bootstrapped. Since the Ljung-Box test at lags up to 30 rejects autocorrelation in the time series of the pricing errors, the calculation uses 25,000 i.i.d. bootstrap draws.

References

Aït-Sahalia, Y. & Duarte, J. (2003), 'Nonparametric option pricing under shape restrictions', *Journal of Econometrics* **116**(1-2), 9–47.

Aït-Sahalia, Y. & Lo, A. W. (2000), 'Nonparametric risk management and implied risk aversion', *Journal of Econometrics* **94**(1), 9–51.

Ang, A., Hodrick, R. J., Xing, Y. & Zhang, X. (2006), 'The cross-section of volatility and expected returns', *Journal of Finance* **61**(1), 259–299.

Augustyniak, M. (2014), 'Maximum likelihood estimation of the markov-switching garch model', *Computational Statistics & Data Analysis* **76**, 61–75.

Bakshi, G., Cao, C. & Chen, Z. (1997), 'Empirical performance of alternative option pricing models', *The Journal of Finance* **52**(5), 2003–2049.

Bakshi, G., Madan, D. & Panayotov, G. (2010), 'Returns of claims on the upside and the viability of u-shaped pricing kernels', *Journal of Financial Economics* **97**(1), 130–154.

Bali, T., Engle, R. & Murray, S. (2014), 'Empirical asset pricing: The cross section of stock returns: An overview', *Wiley StatsRef: Statistics Reference Online* pp. 1–8.

Barone-Adesi, G., Engle, R. F. & Mancini, L. (2008), 'A garch option pricing model with filtered historical simulation', *Review of Financial Studies* **21**(3), 1223–1258.

Barone-Adesi, G., Mancini, L. & Shefrin, H. (2016), 'Sentiment, risk aversion, and time preference', *Available at SSRN* .

Bates, D. S. (1996), '20 testing option pricing models', *Handbook of Statistics* **14**, 567–611.

Bauwens, L., Dufays, A. & Rombouts, J. V. (2014), 'Marginal likelihood for markov-switching and change-point garch models', *Journal of Econometrics* **178**, 508–522.

Beare, B. K. & Schmidt, L. D. (2016), 'An empirical test of pricing kernel monotonicity', *Journal of Applied Econometrics* **31**(2), 338–356.

Belomestny, D., Ma, S. & Härdle, W. K. (2015), 'Pricing kernel modeling', *Available at SSRN* .

Boguth, O., Carlson, M., Fisher, A. & Simutin, M. (2011), 'Conditional risk and performance evaluation: Volatility timing, overconditioning, and new estimates of momentum alphas', *Journal of Financial Economics* **102**(2), 363–389.

Bollerslev, T., Tauchen, G. & Zhou, H. (2009), 'Expected stock returns and variance risk premia', *The Review of Financial Studies* **22**(11), 4463–4492.

Broadie, M., Chernov, M. & Johannes, M. (2009), 'Understanding index option returns', *The Review of Financial Studies* **22**(11), 4493–4529.

Carr, P. & Wu, L. (2009), 'Variance risk premiums', *Review of Financial Studies* **22**(3), 1311–1341.

Chabi-Yo, F. (2012), 'Pricing kernels with stochastic skewness and volatility risk', *Management Science* **58**(3), 624–640.

Chabi-Yo, F., Garcia, R. & Renault, E. (2008), 'State dependence can explain the risk aversion puzzle', *Review of Financial Studies* **21**(2), 973–1011.

Chang, B.-Y., Christoffersen, P., Jacobs, K. & Vainberg, G. (2012), 'Option-implied measures of equity risk', *Review of Finance* **16**(2), 385–428.

Christoffersen, P., Heston, S. & Jacobs, K. (2013), 'Capturing option anomalies with a variance-dependent pricing kernel', *Review of Financial Studies* **26**(8), 1963–2006.

Corsi, F. (2009), 'A simple approximate long-memory model of realized volatility', *Journal of Financial Econometrics* **7**(2), 174–196.

Cuesdeanu, H. (2016), 'Empirical pricing kernels: A tale of two tails and volatility?', *Available at SSRN* .

Cuesdeanu, H. & Jackwerth, J. C. (2018), 'The pricing kernel puzzle: Survey and outlook', *Annals of Finance* **14**(3), 289–329.

Diebold, F. X. (1986), 'Modeling the persistence of conditional variances: A comment', *Econometric Reviews* **5**(1), 51–56.

Driessen, J. & Maenhout, P. (2007), 'A portfolio perspective on option pricing anomalies', *Review of Finance* **11**(4), 561–603.

Duan, J.-C. (1995), 'The garch option pricing model', *Mathematical Finance* **5**(1), 13–32.

Faias, J. A. & Santa-Clara, P. (2017), 'Optimal option portfolio strategies: Deepening the puzzle of index option mispricing', *Journal of Financial and Quantitative Analysis* **52**(1), 277–303.

Fama, E. F. & French, K. R. (1993), 'Common risk factors in the returns on stocks and bonds', *Journal of financial economics* **33**(1), 3–56.

Figlewski, S. (1994), 'Forecasting volatility using historical data', *Available at SSRN* .

Figlewski, S. (2010), 'Estimating the implied risk neutral density for the us market portfolio', *Volatility and Time Series Econometrics: Essays in Honor of Robert Engle* pp. 323–353.

Figlewski, S. (2018), 'Risk neutral densities: A review', *Available at SSRN* .

Figlewski, S. & Malik, M. F. (2014), 'Options on leveraged etfs: A window on investor heterogeneity', *Available at SSRN* .

Giordani, P. & Halling, M. (2016), 'Up the stairs, down the elevator: valuation ratios and shape predictability in the distribution of stock returns', *Available at SSRN* .

Grith, M., Härdle, W. K. & Krätschmer, V. (2017), 'Reference-dependent preferences and the empirical pricing kernel puzzle', *Review of Finance* **21**(1), 269–298.

Grith, M., Härdle, W. & Park, J. (2013), 'Shape invariant modeling of pricing kernels and risk aversion.', *Journal of Financial Econometrics* **11**(2), 370 – 399.

Hansen, L. P. & Jagannathan, R. (1991), 'Implications of security market data for models of dynamic economies', *Journal of political economy* **99**(2), 225–262.

Härdle, W. K., Okhrin, Y. & Wang, W. (2014), 'Uniform confidence bands for pricing kernels', *Journal of Financial Econometrics* **13**(2), 376–413.

Harvey, C. R. & Siddique, A. (2000), 'Conditional skewness in asset pricing tests', *The Journal of Finance* **55**(3), 1263–1295.

Heber, G., Lunde, A., Shephard, N. & Sheppard, K. (2009), 'Oxford-man institute's realized library, version 0.2'.

Hens, T. & Reichlin, C. (2013), 'Three solutions to the pricing kernel puzzle', *Review of Finance* **17**(3), 1065–1098.

Heston, S. L. & Nandi, S. (2000), 'A closed-form garch option valuation model', *Review of Financial Studies* **13**(3), 585–625.

Jackwerth, J. C. (2000), 'Recovering risk aversion from option prices and realized returns', *Review of Financial Studies* **13**(2), 433–451.

Jones, C. S. (2006), 'A nonlinear factor analysis of s&p 500 index option returns', *The Journal of Finance* **61**(5), 2325–2363.

Klaassen, F. (2002), 'Improving garch volatility forecasts with regime-switching garch', *Empirical Economics* **27**(2), 363–394.

Liu, X., Shackleton, M. B., Taylor, S. J. & Xu, X. (2009), 'Empirical pricing kernels obtained from the uk index options market', *Applied Economics Letters* **16**(10), 989–993.

Martin, I. (2016), 'What is the expected return on the market?', *The Quarterly Journal of Economics* **132**(1), 367–433.

Mikosch, T. & Stărică, C. (2004), 'Nonstationarities in financial time series, the long-range dependence, and the igarch effects', *Review of Economics and Statistics* **86**(1), 378–390.

Polk, C., Lou, D. & Skouras, S. (2018), 'A tug of war: overnight versus intraday expected returns', *Journal of Financial Economics* .

Polkovnichenko, V. & Zhao, F. (2013), 'Probability weighting functions implied in options prices', *Journal of Financial Economics* **107**(3), 580–609.

Rosenberg, J. V. & Engle, R. F. (2002), 'Empirical pricing kernels', *Journal of Financial Economics* **64**(3), 341–372.

Rubinstein, M. (1976), 'The valuation of uncertain income streams and the pricing of options', *The Bell Journal of Economics* **7**(2), 407–425.

Santa-Clara, P. & Saretto, A. (2009), 'Option strategies: Good deals and margin calls', *Journal of Financial Markets* **12**(3), 391–417.

Schneider, P., Wagner, C. & Zechner, J. (2019), 'Low risk anomalies?', *Journal of Finance, Forthcoming* pp. 19–50.

Sichert, T. (2019), 'Structural breaks in garch processes do matter', *Working Paper, Goethe University Frankfurt* .

Ziegler, A. (2007), 'Why does implied risk aversion smile?', *Review of Financial Studies* **20**(3), 859–904.