The Pricing Kernel is U-shaped

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

December 20, 2018

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 new pricing kernel estimates imply a downward sloping term structure of risk premia and Sharpe ratios in both good and bad times for the pricing kernel *itself*. A pricing kernel mimicking trading strategy confirms the predictions and yields sizable Sharpe ratios. Furthermore, the U-shaped pricing kernel helps to explain cross-sectional stock return anomalies. The results are robust to numerous variations in the methodology and hold for several major international stock market indices. 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, term structure of risk premia, options, GARCH, change-point model

*I would like to thank David Alexander, Peter Christoffersen, Marc Crummenerl, Bob Dittmar, Mátyás Farkas, Bruce Grundy, Markus Huggenberger, Kris Jacobs, Holger Kraft, Jan Pieter Krahnen, Tom McCurdy, Christoph Meinerding, Christian Schlag, Maik Schmeling, Paul Schneider, Lorenzo Schoenleber, Marti Subrahmanyam, Roméo Tédongap, Julian Thimme, Rüdiger Weber, Amir Yaron, conference participants of the AFA, NFA, DGF, SGF, AFFI and seminar participants at the Universities of Frankfurt (Goethe), Zurich and Toronto (Rotman) 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 (called pricing kernel in the following), where the latter serves as a proxy for aggregate wealth. A large body of literature estimates the pricing kernel using option prices by combining index option data with some model-dependent return forecast. 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. 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 model that nests the previous approaches, but corrects for the forecasting bias, leads to always U-shaped pricing kernel estimates.

To estimate the pricing kernel implied by option prices, two key quantities are required: the risk-neutral index return density implied by option prices and a physical return density forecast. While standard methods exist for the estimation of the first, the second one requires some parametric assumptions. The literature has recognized early on that an important ingredient to predicting return distributions is the volatility forecast. It is well known, e.g., from the vast literature on GARCH models in finance,

[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.

that volatility is time-varying and clustered. While standard volatility models can only capture the first property, I will show that the clustering is very important too.

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 known 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 accompanying, long sample of S&P 500 options over the period 1996-2015, to include the different market phases.

Next, I show that standard methods relying on GARCH models with fixed parameters estimate S-shaped pricing kernels in times of low variance, and U-shaped ones otherwise. Furthermore, when replacing the standard GARCH with the CP-GARCH in the otherwise identical methodology, the S-shaped pricing kernels disappear altogether. The calculation of confidence intervals for the estimated pricing kernels shows that the results are statistically significant.

The paper then shows that the new pricing kernel estimates immediately have several interesting economic properties. While one can study the properties of any pricing kernel (PK) estimate, the results are only meaningful if the estimate is close to the true PK. All applications below are such that they allow to discriminate between good and bad estimates.

First, the new PK estimates set the Euler equation error to zero, while many other estimates fail this necessary condition. This is non-parametric evidence in support of the new approach. A closer look at the timing of the Euler equation errors further highlights the importance of a time-varying pricing kernel.

Second, I use the famous Hansen & Jagannathan (1991) bound to relate the variance of the pricing kernel estimates to the maximum attainable Sharpe ratios. I find that the new estimates imply countercyclical risk premia and Sharpe ratio with a downward sloping term structure in both good and bad times. The term structure of risk premia and Sharpe ratios is discussed vividly since Van Binsbergen et al. (2012), as their results are inconsistent with most leading asset pricing models. Although these results suggest similar patterns for the PK, this must not be the case, as they could also be induced

by state- and horizon-dependent correlations between the assets and the PK. The new results bridge this gap by documenting the analogous patterns for the PK itself.

Third, I introduce the trading of the PK as a synthetic derivative in the options market. The realized returns and Sharpe ratios confirm the predicted properties of the pricing kernel in a non-parametric way. Furthermore, as theory predicts that the pricing kernel itself is the asset with the highest attainable Sharpe-ratio in the economy, it is very comforting that the trading strategy delivers sizable Sharpe ratios of up to 3.3 for a monthly horizon, and above 5 for weekly maturities. By turning around the argument of the Hansen-Jagannathan bound, one can see that the PK, that delivers the higher Sharpe ratios relative to another one, is better and closer to the true PK.

Finally, the U-shaped pricing kernel is relevant beyond the realm of options, as it helps to explain the cross-section of stock returns. Using the example of the low-beta anomaly, I show that the new finding, the upward sloping part of the pricing kernel in the area of high positive returns, contributes significantly to the reduction of the pricing errors.

In the final part of the paper, I demonstrate that the findings are 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 high negative and high positive returns, the variance risk premium causes the projection of the stochastic discount factor on the index returns to be U-shaped. My analysis below shows that the structural breaks are necessary to fit the model to the data. While the approach without breaks fails to match the time-series properties of the empirical pricing kernels, the new model fits the data very well. Also, the same analysis reveals that the introduction of structural breaks increases the fit of the option pricing model considerably. It appears that the cyclical bias in the volatility forecasts carries over to the risk-neutral dynamics as well and makes option prices systematically biased when using standard GARCH. Finally, the model can also generate the empirically observed patterns in risk premia and Sharpe ratios of the pricing kernel.

Overall, the paper provides novel semi-parametric evidence on the time series behavior of the pricing kernel puzzle, which has not been subject to much research. 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.[2] While a few studies document

[2]E.g., Jackwerth (2000), Aït-Sahalia & Lo (2000), Rosenberg & Engle (2002), Liu et al. (2009), Polkovnichenko & Zhao (2013), Figlewski & Malik (2014), Beare & Schmidt (2016),

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.[3] 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.[4] 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. 2016), ambiguity aversion (Cuesdeanu 2016) and other behavioral factors (Figlewski 2018), among others. However, none of the existing models can explain the hump in the S-shape with a risk-factor. 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 brought forward in this paper can reconcile the different empirical and conflicting theoretical results. In particular, this paper shows that S-shaped pricing kernels are the consequence of structural breaks in the volatility process.[5] The new model nests the previous approaches, and this allows me to to pin down the source of the estimation of S-shaped pricing kernels. The analysis shows that a standard GARCH model provides biased multi-period volatility forecasts, and that this is the crucial driver behind the fact that numerous studies find S-shaped pricing kernels. It becomes clear that

Belomestny et al. (2015) Cuesdeanu & Jackwerth (2018), Barone-Adesi et al. (2016) Grith et al. (2016).

[3]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.

[4]This is because the investor would be better off by investing in the region adjacent of the hump. However, this cannot be an equilibrium, since the representative investor by definition has to hold all securities. Hence, prices have to adjust such that the investor is willing to hold all assets. See e.g. Hens & Reichlin (2013) for a more elaborate discussion of this argument.

[5]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. (2016) 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 a stylized S-shaped pricing kernel.

the forecasts by the standard GARCH model revert to the long-run mean too quickly and are not able to capture market phases where volatility is very low for several years. The reverse is, to a lesser extent, true for periods of high volatility. This finding is also a contribution to the econometric literature. Despite the large body of research conducted on modeling and forecasting volatility, this countercyclical pattern has never been documented systematically for standard GARCH. In the case of the pricing kernel estimation, this pattern leads to systematically biased volatility forecasts and therefore to biased forecasts of the physical return distribution. The bias is much more prominent in times of low volatility, which is why S-shaped pricing kernels are observed only during these times. Hence, the estimation of S-shaped pricing kernels appears to be the consequence of a measurement error due to a misspecification of the volatility process.

The intuitive explanation for the results is that the overestimated volatility leads to a return distribution forecast that is too wide, as it has too much probability mass in the tails and too little in the center. Since the pricing kernel is the point-wise ratio of the risk-neutral density over the physical density, these shifts are important. The excess weight in the left tail is not strong enough to change the downward sloping pattern, but the excess weight of 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.

All the aforementioned findings 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.

In sum, the results show that the canonical use of a standard GARCH model with fixed parameters in the PK literature significantly biases the results and causes the often documented S-shaped pricing kernel estimates. These results challenge the existence of a S-shaped pricing kernel, which has almost become consensus in the literature and is considered a stylized fact by some. This provides valuable guidance for theorists when validating the predictions of their models. While for many questions in finance the data we want to explain is by large clear, this is not the case for the pricing kernel puzzle.

Here it is important to first establish the "data", at least 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). On the other hand, the evidence on the behavior and relevance of volatility is of interest to market participants, since volatility is an important quantity, for example in the context of option pricing.

The remainder of the paper proceeds as follows. Section 2 first introduces the change-point GARCH model and then presents the data, estimation methodology, estimation results and the model fit. 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 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 GARCH Model with Structural Breaks to Overcome the Cyclical Forecast Bias

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.[6] Such a CP-GARCH model is laid out in the following. How the model is

[6] 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 to construct a conditional return distribution is presented in Section 3.2.

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 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].[7] 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

used in the CP model below.

[7] 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 I show that the empirical results are fully robust to using e.g. a Student's t-distribution.

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.

2.3 Model Estimation

2.3.1 Data

The data used to estimate both the fixed parameter and switching GARCH are daily S&P 500 log returns (excluding dividends) 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 on 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.

2.3.2 Estimation Methodology

The GARCH model is estiamted 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 (2018) proposes an estimation algorithm for the change-point GARCH model that uses a particle filter for the laten 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 (2018). 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.

2.3.3 Estimation Results

To the best of the author's knowledge, there are no examples of an estimation of change-point Heston-Nandi GARCH model. Hence, 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. Last, 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. For the interested reader, the online Appendix provides a mode detailed analysis of the optimal number of regimes.

The comparison of the estimation results in Table 1 delivers several insights. First, the estimates 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. Within the CP model, periods with low volatility are accompanied by higher values for μ, and also have a higher drift on average, i.e. higher $\mu \cdot h_t$. Furthermore, the estimation over the full period exhibits the typical result that the variance process is almost integrated. The separate CP regimes, on the contrary, all have a degree of integration $(\beta + \alpha\gamma^2)$ lower than in the one regime model.

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

Table 1: Estimation Results of the HN-GARCH model

	FP HN-GARCH	CP-HN-GARCH				
Parameters	'92-15	'92-'96	'96-'03	'03-'07	'07-'12	'12-'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				
Total	19495.9	19691.6				
AIC	-38981.8	-39325.1				
BIC	-38948.4	-39131.0				

Parameter estimates are obtained by optimizing the likelihood on returns. Parameters are daily, long-run volatility is calculated as $\sqrt{long-run\ variance} \cdot 252$. 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.

could not identify them in such a long sample.[8]

Figure 1 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.

[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.

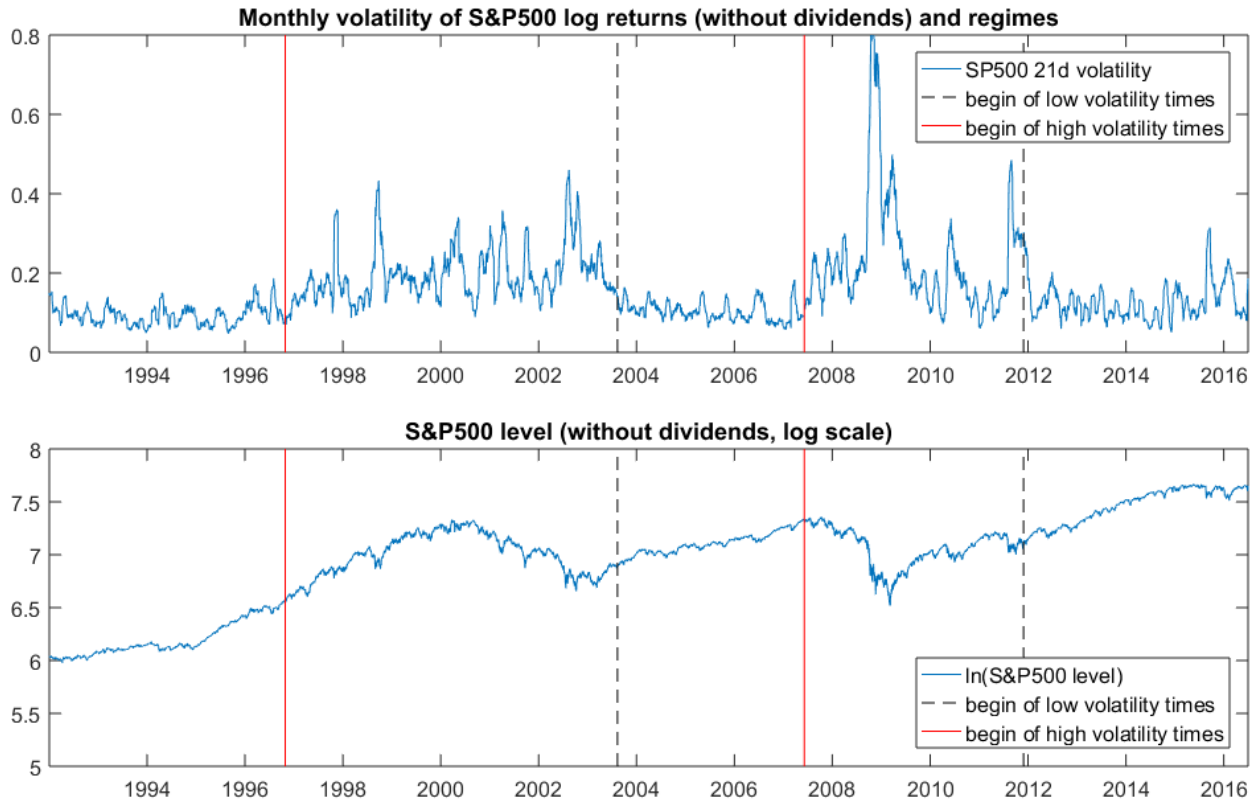


Figure 1: 21-day volatility of daily S&P 500 log returns and log(S&P 500 level)

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

2.4 The CP-GARCH model provides volatility forecasts without a cyclical bias

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 then

the model implied variance for $t+1$ to $t+21$ is calculated.[9] The predicted volatility is calculated as the square root of the predicted variance. The realized volatility is calculated from daily returns R_t as:

$$\sqrt{\sum_{t+1}^{t+21} R_t^2}. \tag{4}$$

Figure 2 displays the result graphically. It compares the realized 21-day volatility to the models' ex ante predicted volatility. 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 because it always 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 statistics corresponding to Figure 2. 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 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 hence 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. This analysis shows that the standard GARCH model does a worse 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.

[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 6.9). To address the second simplification, one would have to filter the probability of being in a certain state at each point in time. This would require running the estimation separately for each day, which is infeasible. However, this quantity is also likely to be low, since in other studies the filtered probability is often very close to one for one state and zero for the others (see e.g. Augustyniak 2014).

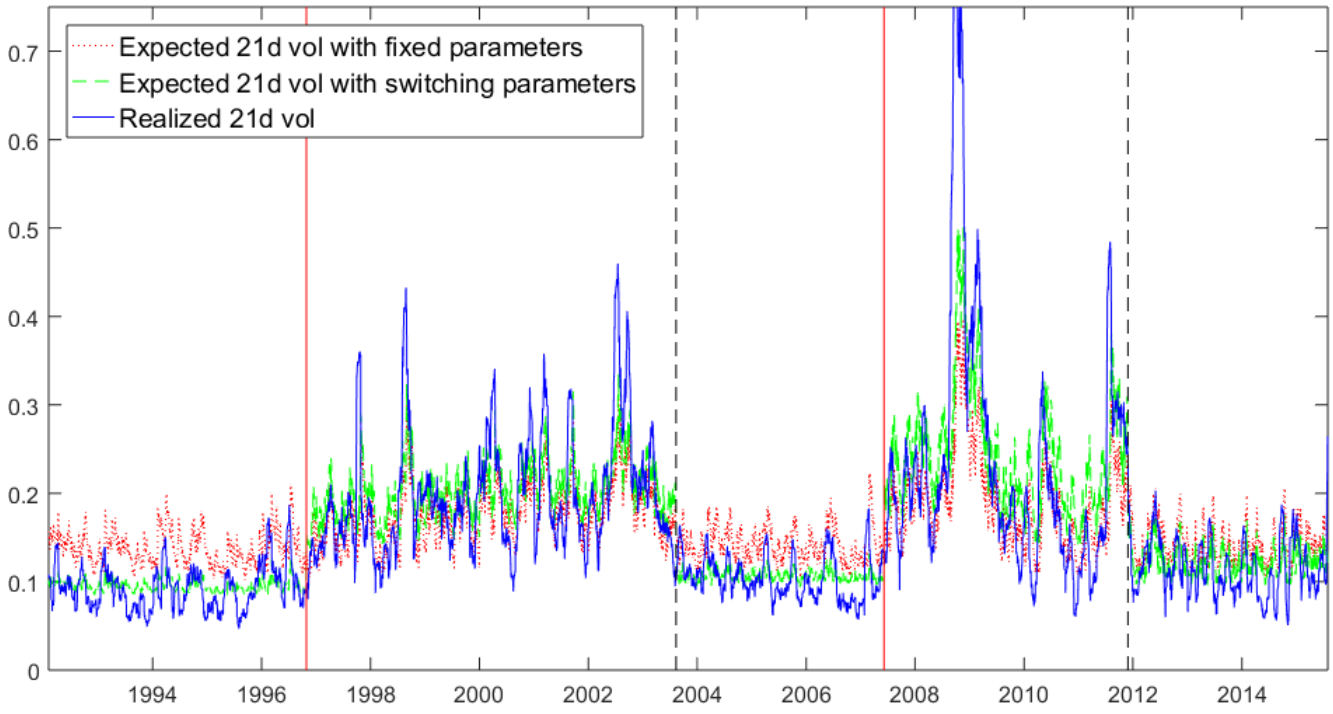


Figure 2: Predicted vs. realized 21-day volatility

The figure shows the annualized 21-day rolling window realized volatility, measured as the square root of the sum of squared 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 expectation and realization are depicted at the same point in time.

With these findings in mind, it becomes clearer why fixed parameter GARCH models have such a high degree of integration and persistence. The higher $\alpha\gamma^2 + \beta$ is, the more slowly the process reverts to its long-run mean. Hence, this parameterization is necessary to make even the one day ahead forecast, which is used in the estimation, to stay at the respective high or low level. Overall, the CP-GARCH model allows to model volatility clustering, while the GARCH with fixed parameters is biased and fails this task.

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 as shown above 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

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 $\text{VIX}/100/\sqrt{12}$.

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.

The documented pattern shed 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 volatility switching 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. For example, from June 2003 until February 2007, the monthly volatility stayed below its long-run mean of 4.79% for about 3.5 years. A standard GARCH model with fixed parameters cannot model this behavior, and consequently its volatility forecasts are systematically biased towards the long-run mean. Despite the large body of literature on modeling and forecasting volatility, to the best of the author's knowledge, this results has not yet been systematically documented.

3 Empirical Pricing Kernels

3.1 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 several ways. 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 the introduction of end-of-quarter, end-of-month and weekly options, and all 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.

3.2 Methodology

The two major quantities that are required to empirically estimate pricing kernels are the risk-neutral and the physical return density. 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 Figlewski (2010). This standard 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

[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).

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 tail that is not covered by the (cleaned) option data. The cumulative risk-neutral return density 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 is semi-parametric and has become popular recently.[12] Many alternative approaches are tested in the robustness section and they deliver very similar results. 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 and hence less parametric than many alternatives. 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 $\sqrt{h(t,t+\tau)}$ forecasted by the GARCH model at the beginning of the month. This yields a series of return shocks:

$$Z(t,t+\tau) = (R(t,t+\tau) - \bar{R})/\sqrt{h(t,t+\tau)}. \tag{5}$$

The conditional distribution is then constructed by multiplying the standardized return shock series Z with the conditional volatility expectation on a given day:

$$\hat{f}(R(t,t+\tau)) = \hat{f}\left(\bar{R} + \sqrt{h(t,t+\tau)}Z\right). \tag{6}$$

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 dis-

[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 directly from applying (16), 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).

tribution 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.

Finally, the log of the empirical pricing kernel (EPK) $\hat{M}$ is calculated as:

$$\ln \hat{M}_{t,t+\tau} = \ln \Big(f^*(R(t,t+\tau)/\hat{f}(R(t,t+\tau)) \Big). \tag{7}$$

3.3 Results for the one month pricing kernel

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 3 shows the time series of the natural logarithm of the estimated pricing kernels using the HN-GARCH model with fixed parameters. Figure 4 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 2.3.3. The lines are not smoothed, to create no false impression of precision. The dotted blue lines at the right end of the pricing kernels depict the ratio of the risk-neutral to the physical cumulative return densities (CDF) in the tail. Each CDF ratio is just one data point, but the line illustrates to which PK the point corresponds. This gives an indication of the behavior of the PK in the right tail. 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 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. The varying wideness of the PKs is to be expected, since the PK has to at least price the risk-free asset and the index correctly. 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 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 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 6.2 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. As the standard GARCH forecasts are biased towards the long-run mean, the 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 2 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 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. Moreover, the PKs seem to be U-shaped at least most of the time, if not for all time periods. Furthermore, it becomes clear that the S-shaped kernels are merely a result of the GARCH model with fixed parameters and are then only found in periods of low variance. The calculation of confidence intervals in Section 6.1 shows that the presented findings are also statistically significant. Furthermore, Section 6.3 shows that the same results can be found for other major international stock market indices. Finally, the robustness section shows that of the results are robust to a variety of changes in data and methods.

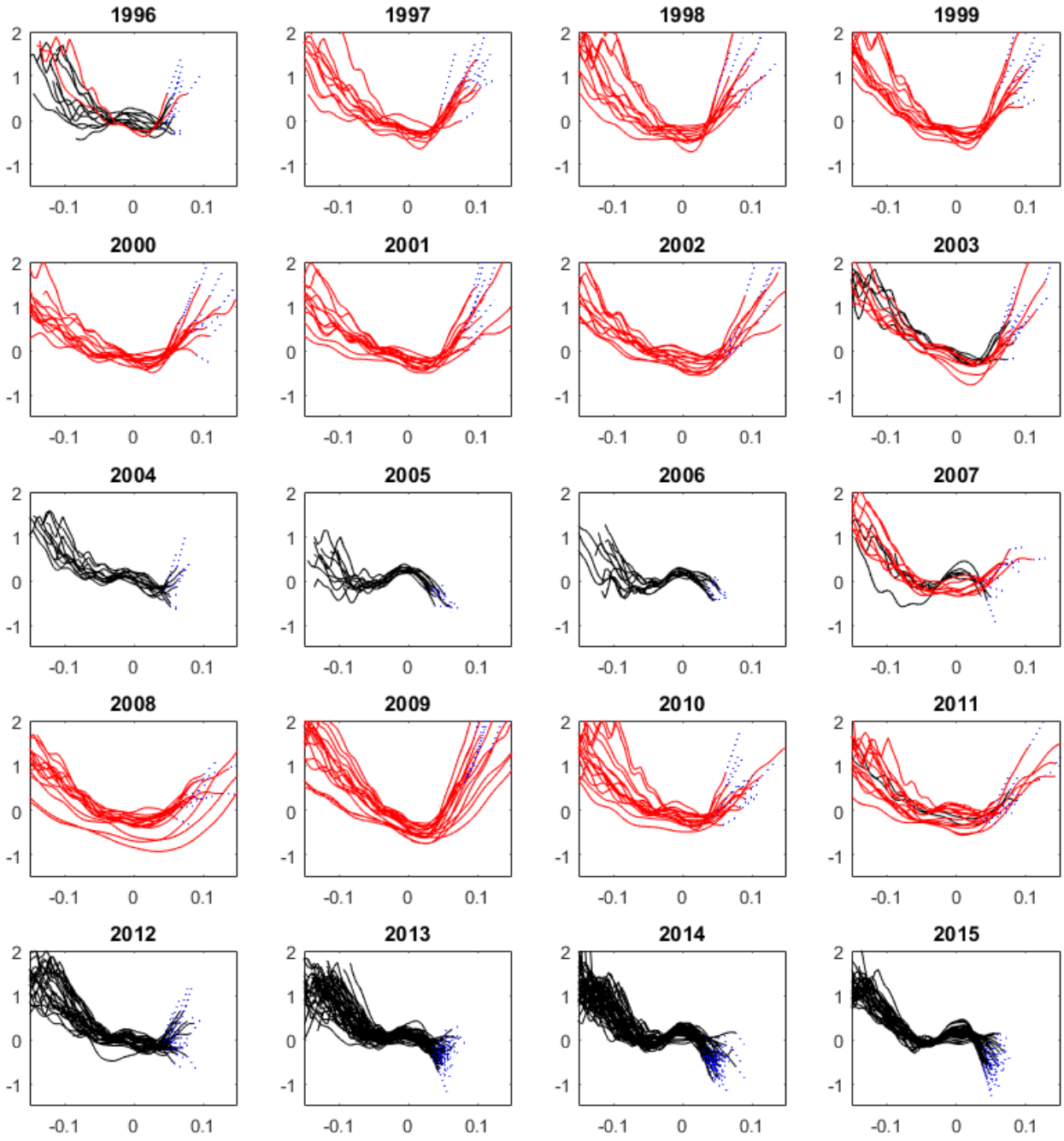


Figure 3: Empirical pricing kernels with fixed parameters in the HN model

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 2.3.3. 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.

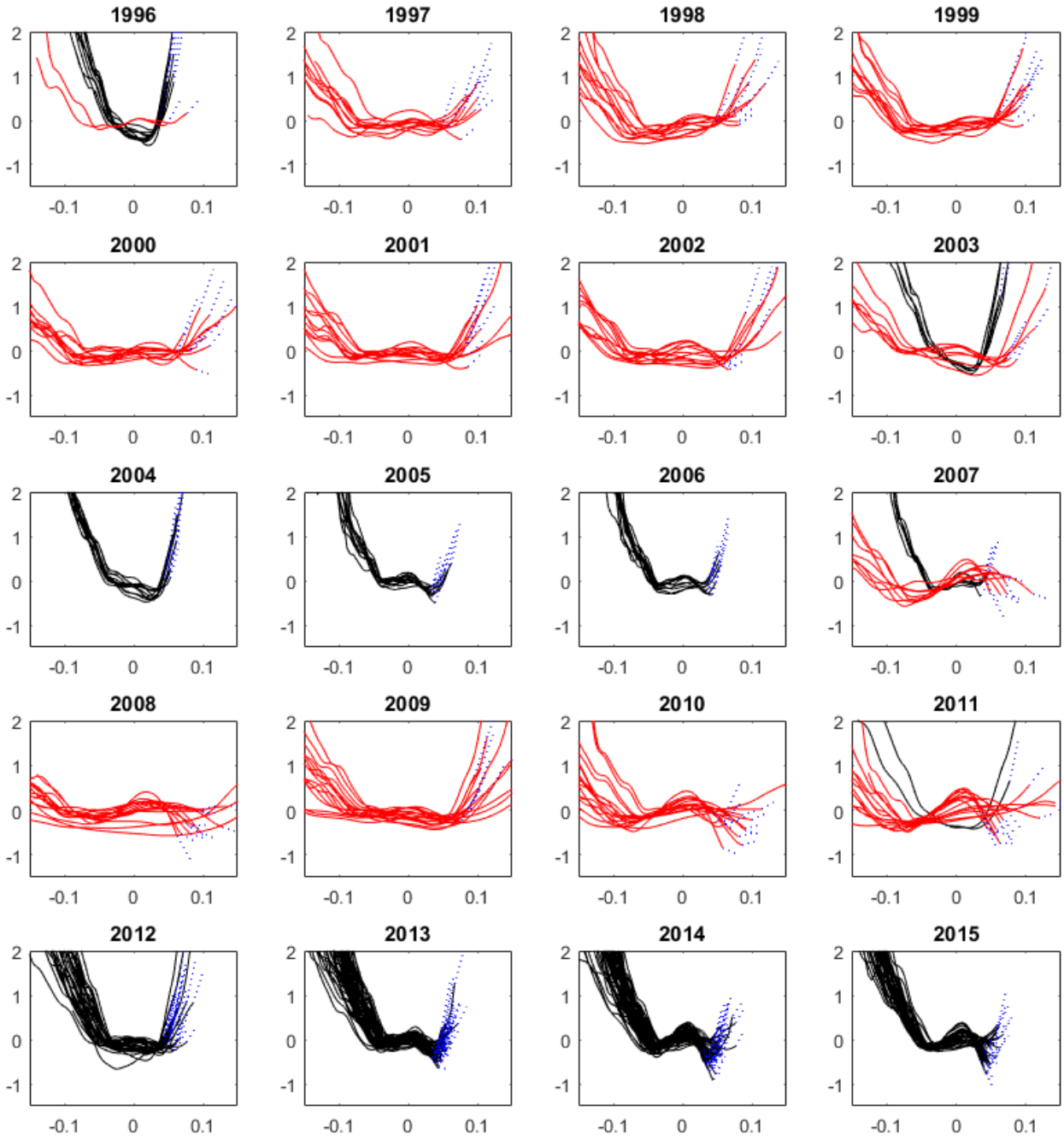


Figure 4: Empirical pricing kernels with CP parameters in the HN model

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 2.3.3. 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.

3.4 How the volatility forecasts drive the result

The introduction of structural breaks into the standard GARCH model changes the results on empirical pricing kernels significantly. The new 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 5 shows that these two distributions are actually very similar. The figure shows the densities of the monthly normalized return shocks (i.e. the Z from Equ. (5)) for the two models. The first subplot 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 5) 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 6 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 5 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 5 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 3 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. The results support what the visual evidence suggests. 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.

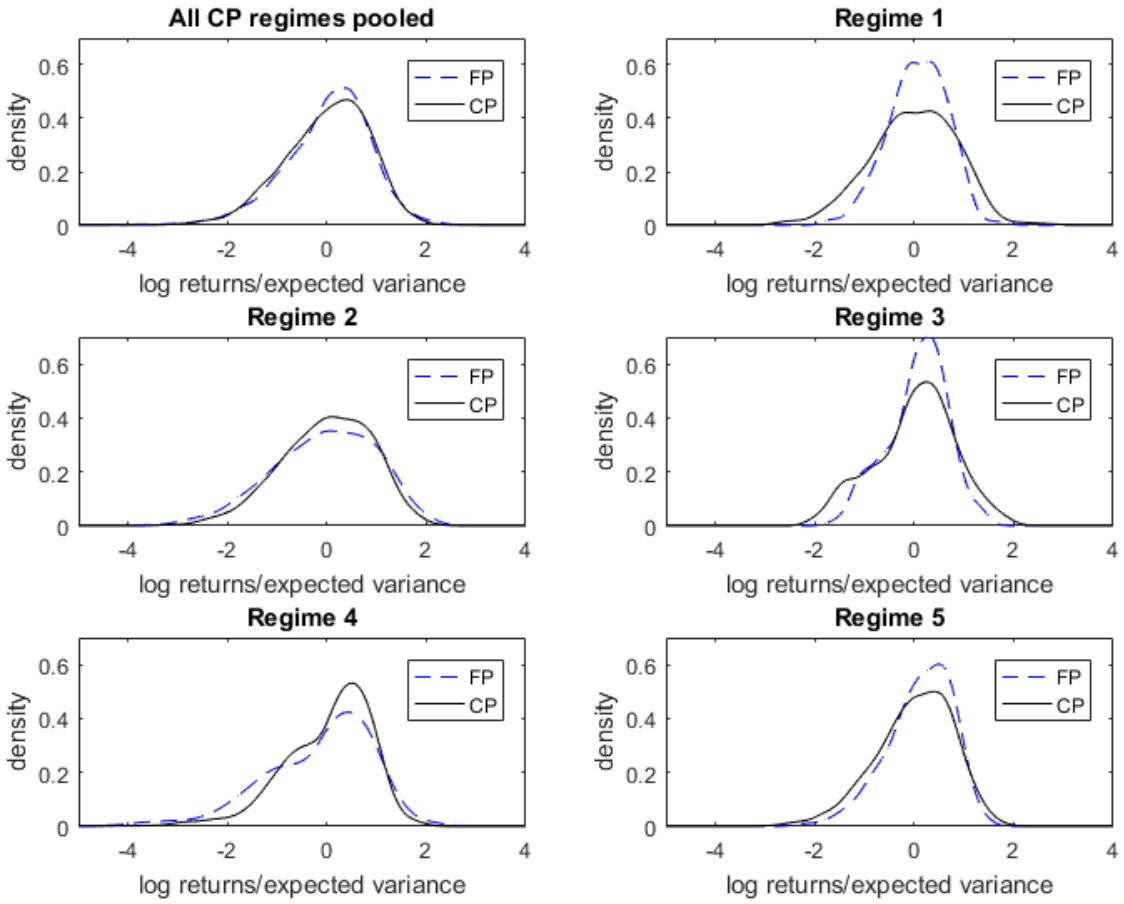


Figure 5: Monthly return shock densities

The figure shows the estimated monthly return shock density (21 days) calculated as in Equ. (5) using the FP GARCH and CP-GARCH. The first sub-plot 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.

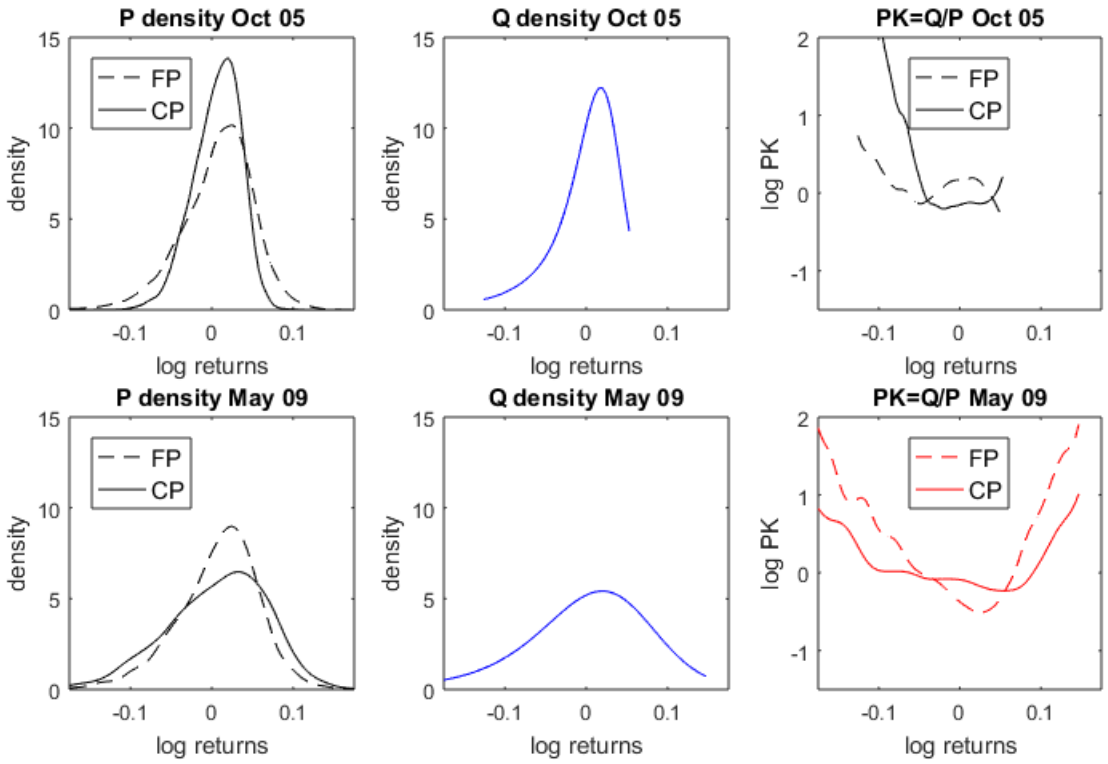


Figure 6: Main mechanism how volatility estimates drive the results

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 3: 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.

4 Economic Properties of the Empirical Pricing Kernels

In the following section, the empirical pricing kernel estimates presented in Section 3.3 are used, with two minor adjustments (see Appendix C.1 for the details). Furthermore, I

will use the level of the implied volatility index to identify times of high and low volatility. As this has a very high overlap with the estimated regimes, the non-parametric way is more intuitive and easier to replicate. For all maturities, the 30 day implied volatility index with the same cutoff value is used. The cutoff values are chosen close to the in-sample medians, with 18 for the VIX, 19 for the FTSE IVI, 23 for the VSTOXX and 21 in case of the VDAX New. If not indicated otherwise, the presented results are for the benchmark specification as in Ch. 3.3, and returns and Sharpe ratios are annualized by $\cdot(365/\text{days to maturity})$ and $\cdot\sqrt{365/\text{days to maturity}}$, respectively. To not to overwhelm the reader, I present only the main results. All additional findings I refer to, but do not show as tables or figures, can be found in the online Appendix.

4.1 Euler equation errors support the model with breaks

Analyzing Euler equation errors is a non-parametric and standard way to test a candidate pricing kernel. The Euler equation follows directly from the fundamental equation of asset pricing, and the true pricing kernel sets the Euler equation errors to zero. The conditional Euler equation is:

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

where R_t is the return of the index and M_t is SDF. In the following, the average (unconditional) Euler equation error is defined as the time-series average of T conditional Euler equation errors:

$$\frac{1}{T}\sum_{t=1}^{T}\Big(\hat{M}_{t,t+\tau} R_{t,t+\tau} - 1\Big), \tag{9}$$

where $\hat{M}$ is the empirical pricing kernel (EPK). An average Euler equation error of zero is a minimum condition that any candidate pricing kernel has to fulfill.

Table 4 shows the results. The first two columns show the Euler equation errors of the empirical PKs from Section 3.3. The other columns show the errors for the three most important methods from the robustness section. In particular, column three shows the errors for the method where the level of the VIX is used as volatility forecast. Columns four and five show the results for the method where the conditional physical return density is obtained from simulating the GARCH model. This is an approach that is often used in the literature and I employ it below both for the fixed and switching parameter

model.[13] The last column shows the results where the realized volatility model of Corsi (2009) is used for forecasting volatility. The first line displays the Euler equation errors and the second line contains the corresponding bootstrapped 90% 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 and fourth line show the average Euler equation errors split into times of high and low volatility.

Table 4: Euler equation errors of empirical pricing kernels

	FP	CP	VIX	FP simul.	CP simul.	Real. Vol
Error	0.052	-0.006	0.007	0.046	0.001	-0.001
[Conf. Bounds]	[.008,	[-.035,	[-.032,	[.013,	[-.028,	[-.032,
bounds]	0.107]	0.026]	0.06]	0.085]	0.03]	0.06]
Error low vol	0.068	0.001	0.019	0.089	0.021	0.019
Error high vol	0.032	-0.016	-0.008	-0.012	-0.026	-0.008

The table shows the average Euler equation errors, together with their 99% confidence intervals as well as the Euler equation errors for the two subsamples of times with high and low volatility (VIX$\lessgtr$18).

The main result is that the standard approaches which use a GARCH model with fixed parameters (column 1 and 4) produce Euler equation errors that are both economically and statistically significant. They therefore fail a necessary condition. The approaches that use a GARCH models with breaks on the contrary have 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 Term structure of risk premia and Sharpe ratios

Next, I turn to the term structure of risk premia and Sharpe ratios, which is discussed vividly since Van Binsbergen et al. (2012), in particular since their results are inconsistent with most leading asset pricing models. Van Binsbergen & Koijen (2017) show that the findings of a downward sloping term structure of risk premia and Sharpe ratios can be documented for a variety of assets and markets (and they also provide a comprehensive

[13] The approach is used e.g. by Rosenberg & Engle (2002), Barone-Adesi et al. (2008) Liu et al. (2009), Barone-Adesi & Dall'o (2012), Cuesdeanu & Jackwerth (2018), Cuesdeanu (2016).

theory review). While these results suggest that the SDF has similar properties, this must not be the case. The time series and term structure properties of the SDF can be different if the assets have state- and horizon-dependent correlations with the SDF.

Therefore, in the following I study the term structure and time series of the pricing kernel *itself*. I document that this asset with the maximum attainable conditional Sharpe ratio in the economy also has a downward sloping term structure of risk premia and Sharpe ratios. Although the pricing kernel is only the projection of the SDF on the index, the results are still very valid. While the full SDF most likely has a higher volatility than its projection, there are no assets to trade it. Given that magnitude of the Sharpe ratios presented below exceeds any value obtainable by standard assets, this is probably the closest one can get to the SDF empirically.

In particular, equipped with the empirical pricing kernels (EPKs) from Section 3.3 one can calculate the bounds of Hansen & Jagannathan (1991) for the expected return and Sharpe ratio. Doing this for the index is a standard exercise, but here I can go one step further. The Hansen-Jagannathan bound for asset i reads:

$$\frac{E(R_{i,t,t+\tau} - R_f)}{\sigma(R_{i,t,t+\tau} - R_f)} = -\rho(R_{i,t,t+\tau}, M_{t,t+\tau}) \cdot \sigma(M_{t,t+\tau}) \cdot R_f. \tag{10}$$

Hence, the asset that is perfectly negative correlated ($\rho = -1$) with the SDF is the asset with the maximum attainable conditional Sharpe ratio.[14] This can be achieved by considering an asset, that pays out the level of the realized PK. Following Bakshi et al. (2010), I call this asset the "pricing kernel strangle" (PKS).[15] In particular, the return of this asset is:

$$R_{PKS,t,t+\tau} = \frac{M_{t,t+\tau}}{E_t^*(M_{t,t+\tau})}. \tag{11}$$

In the following, I discuss the right-hand side of Equ. (10), which is basically the volatility of the pricing kernel. In Section 4.3 I do the complementary and introduce a trading strategy that mimics the left-hand side of Equ. (10).

While the benchmark results above are concerned with the standard one month PK, the analysis can be extended to other horizons.[16] Therefore, I can study the term

[14] The Hansen-Jagannathan bound holds for the SDF and not its projection on the index (the PK). However, as long as one just compares assets which payoff functions are only a function of the market return (as here and in the following), the PK still maximizes the Sharpe ratio.

[15] Bakshi et al. (2010) call the right, upward sloping part of the asset "kernel call", as it resembles the payoff profile of a call. Analogously, the U-shaped full EPKs resemble the payoff profile of a strangle or straddle.

[16] Section 6.8 shows that the estimates for other horizons are also clearly U-shaped.

structure of the pricing kernel, and in particular the term structure of implied risk premia and Sharpe ratios.

Table 5: Average risk premia and Sharpe ratios implied by the EPKs

	1w	2w	1m	2m	3m	1w	2w	1m	2m	3m
	Risk premia					Sharpe ratios				
Panel A: SPX empircal pricing kernels with fixed parameters GARCH										
All dates	10.99	5.05	2.49	1.21	0.93	3.96	2.58	1.78	1.47	1.25
Low vol	6.74	2.40	1.14	0.54	0.47	2.80	1.57	1.08	0.78	0.75
High vol	13.22	6.80	3.54	1.74	1.28	4.57	3.25	2.33	1.95	1.63
Panel B: SPX empircal pricing kernels with change-point GARCH										
All dates	9.63	5.36	3.20	1.54	1.15	3.72	2.80	2.36	1.61	1.45
Low vol	8.70	5.16	3.35	1.59	1.13	3.56	2.76	2.48	1.66	1.48
High vol	10.11	5.49	3.09	1.50	1.16	3.80	2.83	2.27	1.57	1.44

The table shows the model implied excess returns and Sharpe ratios for different maturities for the S&P 500 empirical pricing kernels.

Table 5 reports the results for several horizons, which provides several findings.[17] First, the term structure of both risk premia and Sharpe ratios are clearly downward sloping. This hold not only on average, but also both for calm times and times of financial turmoil. Second, the risk premia and Sharpe ratios are countercyclical, with maybe the exceptions of CP model implied Sharpe ratios. Third, the values are substantial. As discussed below for the trading strategy, the returns are not obtainable, because having a short position in the PKS involves substantial leverage. But the Sharpe ratios are not affected by collateral, if the latter is held in the risk-free asset. Fourth, comparing the Panels A and B it emerges that the EPKs using standard GARCH imply much more countercyclical values, a steeper term structure, and comparably larger values in the short run and smaller ones in the long run. However, the results in Section 4.3 shows that the realized values for the FP model do not match the expected values well, while they do for the CP model. Fifth, the volatility split values for the CP model converge to similar values for longer maturities, while they stay strongly countercyclical for the FP

[17] While it is possible to obtain a full time series of implied risk premia and Sharpe ratios, this plot if very noisy in my case. A substantial part of that noise stems from the time-variation in the probability mass contained in the available risk-neutral densities.

model. In addition it is worth mentioning that all the methods tested in the robustness section have the same predictions.

4.3 Trading the pricing kernel

To complement the previous analysis, I introduce a trading strategy that trades the pricing kernel in the options market. This is a non-parametric way to test the EPK estimates and their properties. It also confirms the accuracy of the model predictions presented before, at least on average.

The pricing kernel strangle (PKS) is a synthetic derivative that pays out the realized level of the PK. 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 the right of Figure 6.

To match the empirical returns to the EPK-implied values, one has to take a short position in the PKS, as one needs a perfect *negative* relation between the PK and the asset with the maximum attainable 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. In any case, it is an interesting theoretically exercise that provides valuable economic insights.

Table 6 shows the obtained, annualized Sharpe ratios from selling the PKS and holding it until maturity. Since theses are key results, the table contains also several international stock market indices, that are otherwise only discussed in the robustness section.[20] There are several main findings. First, the obtained Sharpe ratios are sub-

[18] To reduce the noise from pricing, I calculate the price of this synthetic derivative using the risk-neutral density directly. Nevertheless, the PKS could be approximated well in practice, at least in the absence of transaction costs. This is because the number of traded options clearly exceeds the used number of points on the numerical grid and because the interpolation is done linearly here. Please note also that the approach is model-free, as it is only a derivative with a complex payoff profile that is priced using contemporary option prices, while its payoff is determined by the realized market return.

[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%).

[20] The online Appendix also shows the results for 1.5, 2.5 and 6 months. All three confirm the documented patters. However, for horizons longer than 3 months several empirical problems

stantial. However, they are not completely out of the ballpark of values known for the derivatives market. For example, Broadie et al. (2009) report an annualized Sharpe ratio of 1.01 for selling monthly S&P 500 future OTM puts, while Van Binsbergen & Koijen (2017) report an annualized Sharpe ratio of 1.6 for selling monthly S&P 500 ATM straddles. Nevertheless, the magnitude of the reported Sharpe ratios makes it likely that the trading strategy is actually close to the attainable maximum. In comparison, the annualized monthly Sharpe ratio of the S&P 500 index is 0.34 in the matching sample.

Second, for the CP model, the realized Sharpe ratios match the predicted values presented above well on average, and also for the subsamples. For the FP model however, the realization are substantially lower than the predictions. The values are particularly worse in low volatility times. Since this is also the same for the other indices, these results are included in the online Appendix for brevity. One exception is the very short maturity of 1-2 weeks, where the EPKs from the FP model also performs very well. For these PKs, the difference between the two methods is the lowest, both here and in term of their shape. This is because the volatility forecasting bias has less effect for short maturities and hence the EPKs from the FP model are less S-shaped or even U-shaped in calm times.

Third, holding all else fixed, by large the risk premia and Sharpe ratios are highest for the SPX, followed by FTSE, ESTX and DAX. This order matches the order of number of stocks in the index, which could be a proxy for the amount of unsystematic and unpriced risk in the index.

Fourth, as it comes to the time series, the results confirm the predicted patterns. In times of high volatility the risk premia are clearly higher than in times of low volatility, i.e. are countercyclical, and by large the same hold true for Sharpe ratios. For the SPX and, to lesser extent, also for the FTSE, this seems not to be the case for longer maturities. However, this is driven by some unreasonable EPKs, that are caused by too high volatility forecasts in high volatility times. These EPKs, correspond to the flat or S-shaped pricing kernels visible in Figure 4 as e.g. in 2008. Therefore, I exclude EPKs from the sample where the variance risk premium is positive (forecasted volatility>VIX). This increases the realized values substantially and makes them countercyclical almost always. The results are included in the online Appendix.

Fifth, the term structures of realized excess returns and Sharpe ratios of the PKS

emerge. First, these options are traded significantly less and there are at most four maturities available per year. Second, both the forecasting of volatility and return densities become less reliable.

Table 6: Realized Sharpe ratios and risk premia for the pricing kernel strangle

	1w	2w	1m	2m	3m	1w	2w	1m	2m	3m
	Sharpe ratios					Excess returns				
Panel A: SPX empircal pricing kernels with fixed parameters GARCH										
SPX	11.18	3.95	1.94	1.04	0.63	5.42	1.70	1.12	1.15	0.73
Low vol	7.51	2.75	0.63	0.31	0.35	4.62	3.01	0.50	0.48	0.68
High vol	13.11	4.75	2.97	1.60	0.86	5.94	1.65	1.51	1.61	0.81
Panel B: SPX empircal pricing kernels with change-point GARCH										
SPX	11.58	5.40	3.73	1.96	1.43	6.37	2.33	1.58	2.38	2.27
Low vol	9.78	6.56	4.22	2.12	1.71	5.35	5.43	3.31	2.98	3.31
High vol	12.53	4.62	3.34	1.83	1.83	6.98	1.65	1.14	2.05	1.33
Panel C: International empircal pricing kernels with change-point GARCH										
FTSE	9.38	3.88	2.71	1.40	1.29	5.38	2.52	2.34	1.60	2.19
Low vol	7.15	3.33	1.93	1.00	1.31	3.96	2.34	1.52	1.40	2.53
High vol	12.32	4.57	3.80	1.87	1.25	7.86	2.74	4.16	1.87	1.81
ESTX	9.33	4.36	2.88	1.35	0.74	4.61	3.40	3.10	2.10	1.46
Low vol	3.42	3.55	2.30	1.08	0.58	1.54	3.07	2.85	2.11	1.30
High vol	11.25	4.69	3.20	1.49	0.84	5.98	3.54	3.27	2.14	1.56
DAX	5.48	3.68	2.18	1.19	0.90	2.68	2.48	2.26	1.54	1.73
Low vol	2.57	4.52	1.58	0.87	0.82	1.23	4.18	1.81	1.51	1.67
High vol	7.11	3.25	2.56	1.36	0.97	3.58	1.98	2.55	1.60	1.79

The table reports the annualized Sharpe ratios for a short position in the pricing kernel strangle for maturities of 1 week to 3 months. The values are annualized by $\cdot\sqrt{365/\text{days to maturity}}$. For the second part, only dates are considered, where the physical volatility forecast is below 0.9·volatility index.

are generally downward sloping. This pattern seems to be disrupted for the SPX and the FTSE, but again prevails if the restriction on the variance risk premium is imposed. The slope of the term structure is particularity steep for short maturities and converges to similar values for longer maturities. The two week horizons seems to be an outlier for all indices but the ESTX.

The online Appendix contains some additional results. Among others, it compares the excess returns and Sharpe ratios predicted by the EPKs to the realized ones for all indices, and finds that they are well in line also quantitatively. In addition, similarly strong results are obtained for the methods tested in the robustness section. Furthermore, the corresponding results for the model using FP GARCH show that it performs worse both in terms of results and matching the model's predictions for all indices. The performance of the FP model is particularly bad in times of low volatility. However, in a few cases, trading the PK estimates from the FP GARCH outperforms the CP version. All these cases are in times of high volatility, where all the FP GARCH EPKs are clearly U-shaped. The reason could be that some of the PK estimates from the CP-GARCH are unreasonable flat or S-shaped, due to an overprediction of volatility. Again, this is confirmed by excluding EPKs where the variance risk premium restriction is violated.

Taken together, the realized returns and Sharpe ratios shed new light on key properties of the pricing kernel. Most obvious is that the high Sharpe ratios imply a very high volatility of the SDF, especially in the short run. Also, on average, the volatility of the SDF increases in time, but increases by significantly less than the square root of time. The variance-dependent pricing kernel in Section 5 presents a model that can explain both the U-shape and the empirical patterns in risk premia and Sharpe ratios. Finally, one can reverse the argument of the Hansen-Jagannathan bound and see that the PK that delivers the higher Sharpe ratios relative to another one is better and closer to the true PK.

4.4 The U-shaped pricing kernel helps to explain the low beta anomaly

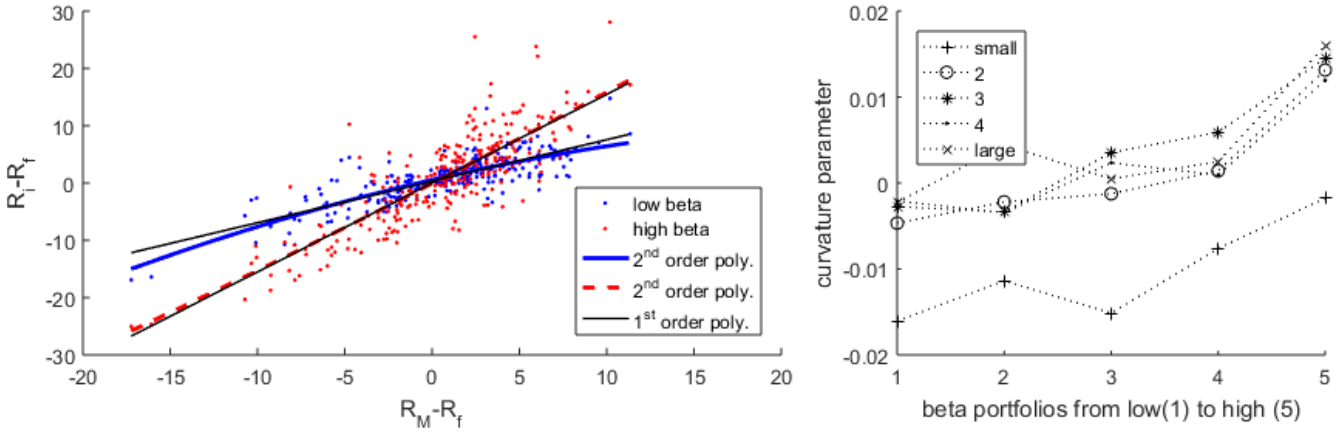


Figure 7: Non-linear relationship between returns of the market and returns of beta-sorted portfolios

The left plot shows the excess returns of the low and high beta portfolio of Kenneth French plotted against the market's excess returns, together with a fitted 1^{st} (each black) and 2^{nd} order polynomial. The right plot shows the estimated curvature parameter for the 2^{nd} order polynomial, using the 25 beta and size portfolios of Kenneth French. The portfolios are equally weighted and the time period is 1996-2015.

The presented empirical pricing kernels also have asset pricing implications beyond the options market, and they can for example help to explain the cross-section of stock returns. The following will show this for the case of the "low beta anomaly", also called "betting against beta", which is widely discussed recently. In short, it is the effect that portfolios comprised of stocks with a high (low) beta have negative (positive) abnormal returns. This anomaly is particularly suited for the pricing kernel presented above, as it is not (yet) an established cross-sectional risk factor, that otherwise might be hard to capture with an SDF that is only a function of the market.

Most existing risk-based explanations for the low beta anomaly focus on the left tail. Therefore, the following analysis will focus on the right tail of the return distribution and show that a U-shaped pricing kernel, and in particular the upward sloping part in the area of high positive returns, can also help to explain the empirical patterns. The intuition is

as follows. A thorough analysis of the conditional correlation of the returns of 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 7 (left) illustrates this pattern. The figure shows the returns of the high and low beta portfolios provided by Kenneth French plotted against the market return, together with a fitted first (only beta with constant) and second order polynomial. In particular, the relationship for the low beta portfolio has a negative curvature relative to a linear relationship. The right graph in Figure 7 generalizes this pattern by showing the curvature parameter of a fitted 2^{nd} order polynomial for all the 25 size and beta sorted portfolios of Kenneth French.

The underperformance of the low beta portfolio for low market returns is well in line with existing explanations, that require a premium for downside risk. However, the underperformance in the case of high market returns implies a *lower* return for models with a decreasing pricing kernel for these states. But if the PK is upward sloping in that area, assets that have lower payoffs in these states are more negatively correlated with the PK, and vice versa. The more negative the correlation then leads to a risk premium and higher average returns.[21] Overall, on an abstract level, one can say that the payoff profile of the low beta portfolio resembles and inverse U, which is then more negatively correlated with the U-shaped PK. The reverse is true for the high beta portfolio.

This explanation is tested quantitatively using the empirical pricing kernel estimates from above. For this, the beta-sorted portfolios are recalculated to match the dates and timing of the option prices. Appendix D contains the technical details. The top of Table 7 shows the in-sample statistics for the portfolios. While the returns are not exactly monotonically decreasing across portfolios, their Sharpe ratios show the typical pattern. The pricing performance of the EPK estimates $\hat{M}$ is tested by calculating the Euler equation errors of the beta-sorted portfolios analogously to Equ. 8-9, where the return of the index is replaced with the return of the beta-sorted portfolios. The errors in Table 7 suggest that the pricing kernel estimated from option prices can also not fully explain the return differences between the beta-sorted portfolios. The pricing errors can be interpreted similar to a factor model as an average alpha.[22]

[21] Similarly, the dip of the U-shaped PK around zero renders these states, where the low beta portfolio outperforms on average, cheap and hence also helps to explain the anomaly.

[22] The errors are not estimated such that they center around zero. Therefore, the apparent overpricing of the portfolios is not necessarily a problem, as the PK was not estimated using the portfolio returns, but rather from option prices. Also, partly due to the low N, neither the pricing error nor the first order differences are statistically significantly different from zero

Table 7: Low beta portfolios and Euler equation errors (monthly, in [%])

Average net return	1.06	1.14	1.15	1.25	1.24
Sharpe ratio	0.19	0.17	0.15	0.14	0.10
Error $\hat{M}$	0.41	0.35	0.29	0.33	0.13
ΔError $\hat{M}$		-0.05	-0.06	0.04	-0.20
Error $\hat{M}^{Mono}$	-5.13	-5.29	-5.39	-5.42	-5.73
ΔError $\hat{M}^{Mono}$		-0.17	-0.09	-0.03	-0.31
Error $\hat{M}^{CAPM}$	0.87	0.75	0.64	0.61	0.29
ΔError $\hat{M}^{CAPM}$		-0.13	-0.11	-0.02	-0.32
ΔError $\hat{M}$-ΔError $\hat{M}^{Mono}$		0.11	0.04	0.07	0.11
[99% Confidence		[0.02,	[-0.01,	[0.02,	[0.02,
bounds]		[0.27,	[0.1,	[0.14,	[0.23,
ΔError $\hat{M}$-ΔError $\hat{M}^{Mono}$		0.07	0.05	0.06	0.12
[99% Confidence		[-0.13,	[-0.04,	[-0.04,	[-0.01,
bounds]		0.32]	0.15]	0.17]	0.37]

The table shows characteristics and the average Euler equation errors for the five beta-sorted portfolios on a monthly basis. $\hat{M}$ denotes the empirical pricing kernel estimate obtained using the CP-GARCH in the benchmark method, and $\hat{M}^{Mono}$ its counterpart that is monotonic in the domain of positive returns. $\hat{M}^{CAPM}$ is the CAPM implied PK. The confidence intervals are obtained from 25,000 pairwise bootstrap draws from the sample of errors differences. The numbers are monthly in percent. The total number of monthly observations are $N = 234$.

Nevertheless, the U-shaped pricing kernel, and in particular the upward sloping part for positive returns can contribute to the pricing of the low beta anomaly. To demonstrate this, I calculated a weakly monotone version of the the estimated pricing kernel, denoted $\hat{M}^{Mono}$, 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. Table 7 show the results of its pricing performance. The discussion focuses on the differences between the portfolios, and less on the level, as the level is distorted by the adjustments (Appendix C.2 provides a detailed discussion of the issue). The results first show that

(reported in online Appendix).

the unexplained return spread between the portfolios widens for the monotone version of the PK. Furthermore, this difference is statistically significantly higher compared to the U-shaped PK. This is the quantitative equivalent of the effect that was described qualitatively before: if the PK was upward sloping in the area of high returns, assets that have low payoffs here require a premium. If the PK is decreasing in the area, the effect would be reversed. The difference is statistically significant here already for a flat PK, and would be even larger for a decreasing one. As a second comparison, I calculate the Euler equation errors of the CAPM.[23] Here, the pricing errors (alphas) are substantially larger. Furthermore, the unexplained return differences are smaller for the EPKs. While the portfolio-wise comparison is at best borderline statistically significant, the high-low comparison is clearly significant.

The presented benchmark analysis uses the equal weighted returns of five beta-sorted portfolios. The results are robust to value weighting, independent double sorts on size and beta and other technical variations, and sometimes get even stronger. Moreover, the EPKs obtained from the FP GARCH model produce sizable and significant pricing errors, similar to those in Section 4.1. Finally, I obtain very similar results for the momentum anomaly with regard to all results discussed above. For brevity, the analysis is included in the online Appendix.

The effect of the upward sloping part of the PK is illustrated here for a selected case. This could be investigated further for other cross-sectional pricing anomalies and also by using standard factor models. As this is beyond the scope of this paper, I leave this for another paper.

5 A Variance-Dependent Pricing Kernel Can Explain the U-shape

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.[24] 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 differences

[23] Please see the online Appendix for the technical details.

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

between the physical and risk-neutral distributions. To be able to evaluate its ability in that regard, it is necessary to fit both distributions using the same, internally consistent, set of parameters.

5.1 Stochastic discount factor

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 kernel 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:

$$M(t) = M(0)\Big(\frac{S_t}{S_0}\Big)^{\phi_{y_t}} \exp\Big(\delta_{y_t} t + \eta_{y_t} \sum_{s=1}^{t} h_s + \xi_{y_t}(h_{t+1} - h_1)\Big), \tag{12}$$

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, (12) collapses to the power utility 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.[25] 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 variance risk premium dominates the equity premium, and the projection is increasing.

Under the assumptions (1), (2), (3) and (12), the risk-neutral dynamics for the HN GARCH model with structural breaks are:

$$\begin{aligned} ln\Big(\frac{S_t}{S_{t-1}}\Big) &= r_t - \frac{1}{2}h_t^* + \sqrt{h_t^*}z_t^*, \\ h_t^* &= \omega_{y_t}^* + \beta_{y_t} h_{t-1}^* + \alpha_{y_t}^*\Big(z_{t-1}^* - \gamma_{y_t}^*\sqrt{h_{t-1}^*}\Big)^2, \end{aligned} \tag{13}$$

[25] For an discussion of the implications and differences of the SDF with and without a variance risk premium see Christoffersen et al. (2013).

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

$$\begin{aligned} h_t^* &= \frac{h_t}{1-2\alpha_{y_t}\xi_{y_t}}, \\ \omega_{y_t}^* &= \frac{\omega_{y_t}}{1-2\alpha_{y_t}\xi_{y_t}}, \\ \alpha_{y_t}^* &= \frac{\alpha_{y_t}}{(1-2\alpha_{y_t}\xi_{y_t})^2}, \\ \gamma_{y_t}^* &= \gamma_{y_t} - \phi_{y_t}. \end{aligned} \tag{14}$$

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

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 in the sample period, the option series with a maturity closest to 30 days is selected. From that maturity, the 15 most actively traded options are used. This results in 15,171 option prices with a maturity between 17 and 53 days. The full details on the data cleaning are reported in Appendix A.

5.3 Identification of breaks points

Theoretically it would be possible to perform a maximum-likelihood estimation of the change-point model over the full sample or returns and option prices. 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 estimation in Section 2.3 in the return process only, and estimate the model separately in each regime for the change-point version. This is equivalent of treating each period as a separate sample.

5.4 Model fit and properties

Table 8 presents the estimation results. The physical parameters are now different from the ones presented in Table 1. The requirement to simultaneously fit the option prices requires a balancing of the risk-neutral and physical dynamics. 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 captures the sizeable negative variance risk premium.

The likelihood of the CP model is significantly higher than the likelihood of the model with fixed parameters. Also the two information criterion, that put a penalty on the number of parameters, prefer the model with breaks. The physical likelihood from returns of the CP model is only slightly higher than of the FP model. However, the option likelihood increases significantly. This is further studied in Section 5.6.

Table 8: 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-'12	'12-'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				
Total	47,687	53,461				
From returns	19,359	19,560				
From options	28,328	33,901				
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 calculated as $\sqrt{long-run\ variance \cdot 252}$. For each model, the total likelihood value at the optimum is reported as well as the value of the returns component at the optimum and the option component at the optimum. 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.

5.5 Model implied pricing kernels

Figure 8 and present the model implied pricing kernels for the GARCH with fixed parameters (left) for the CP-GARCH (right). Each plotted line is obtained by first simulating 100,000 paths under the physical measure using the parameter estimates from Table 8, 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 exhibit S-shapes in many periods. On the contrary, the CP version of the model matches the empirical results 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 corresponding EKPs. In contrast, the CP version exhibits the same strong time variation in their wideness as their empirical counterparts. In addition, for the CP model, the model implied pricing kernels do appear similar within one regime. This also matches the empirical findings well, where the empirical shape within one regime is rather stable.

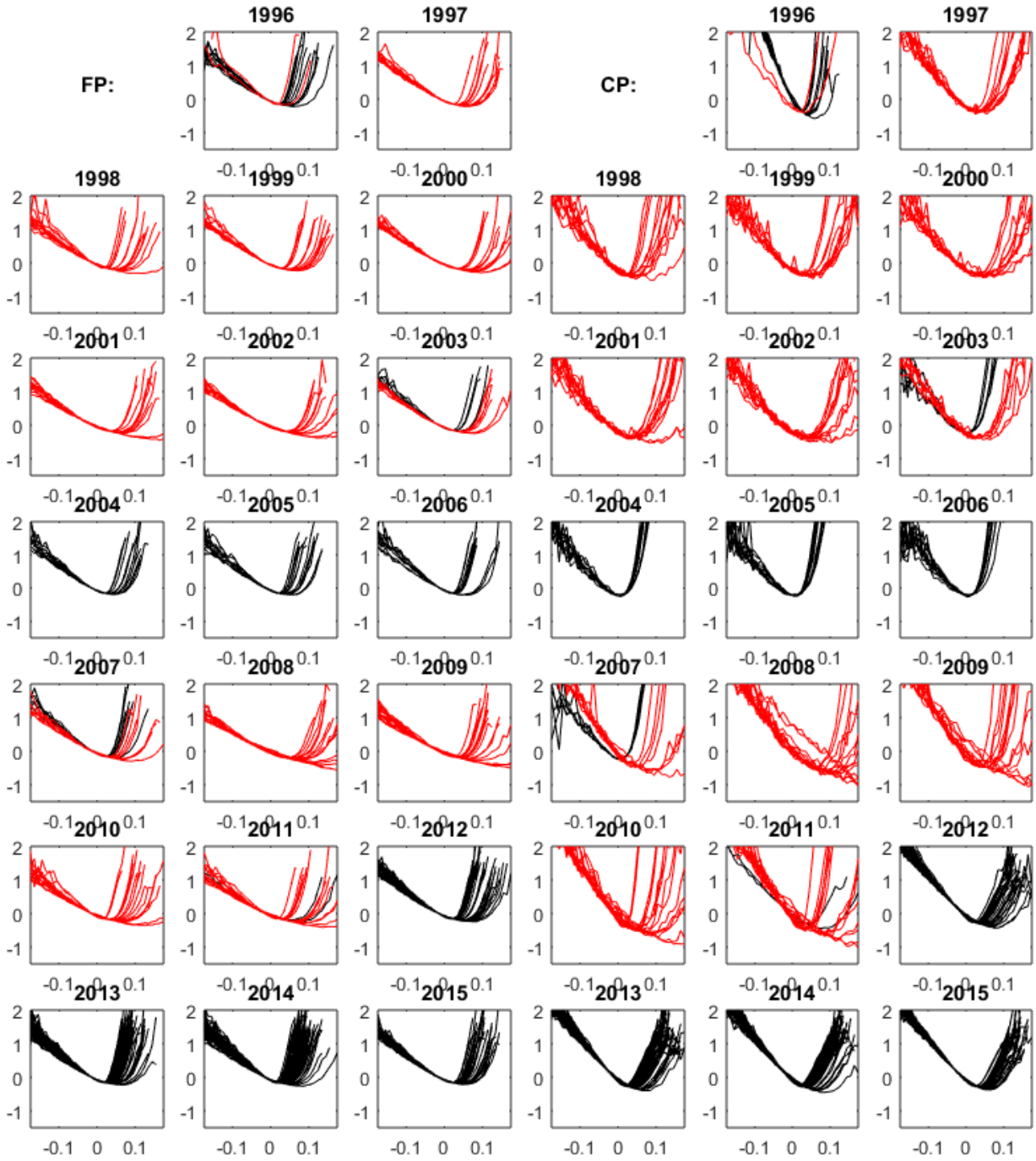


Figure 8: Model implied pricing kernels with FP and CP parameters

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. 2.3.3. Log-returns are on the horizontal axis. The horizon is one month.

5.6 Option pricing fit

Table 8 shows that the model with the breaks has a substantially higher likelihood than the standard GARCH. The major part stems from the better option pricing fit, i.e. a reduction in the option pricing errors. Figure 9 displays the time series of the option pricing errors of the model, which provides further insights. At each date, the vega weighted pricing errors are not squared, but averaged. Pricing errors are market prices minus model prices, hence a negative pricing error occurs when the model overprices the option, and vice versa.[26] The average pricing errors exhibit an interesting time series pattern: in times of low volatility the FP model frequently overprices the actual data. For example, between 2004 and 2006, there is a period of several years where options are constantly overpriced by the model. To a lesser extent, the reverse is true for high volatility periods. For example, at the end of the 1990ies and around 2010 there are periods of several years where options are constantly under-priced by the model. The pricing errors of the CP model on the contrary have a much lower time-clustering of the pricing errors. These observations suggest that the FP GARCH option pricing model has a systematic bias in its prices. This is also visible in the autocorrelation of the errors in the last plot in Figure 9: while both model have significant autocorrelation in the pricing errors for few lags, the autocorrelation of the CP model is always lower and dies out more quickly.

The reason for this bias is most likely the same bias in forecasted volatility that was documented above. Over the average option maturity of one month, the forecasted risk-neutral volatility also reverts back to its long-run mean. This leads to an overestimation of volatility in periods of low volatility, and vice versa, relative to the market believes. This then results in the over-pricing and under-pricing, respectively. This complements the findings of Pan (2002) for stochastic volatility models with jumps. The study documents systematic over- (under-)pricing on days of low (high) volatility, although only for options with maturity of more than 60 days.

In sum, one can conclude that the bias in multi-period volatility forecasts of the GARCH model with fixed parameters also carries over to the risk-neutral dynamics as well. When the GARCH model is enriched by structural breaks, the bias is removed. The combined results show the structural breaks are important for both the physical and risk-neutral dynamics.

[26] The used parameters are the risk-neutral parameters from Table 8. The same results can be obtained when the risk-neutral GARCH parameters are estimated using only option price data and no return data.

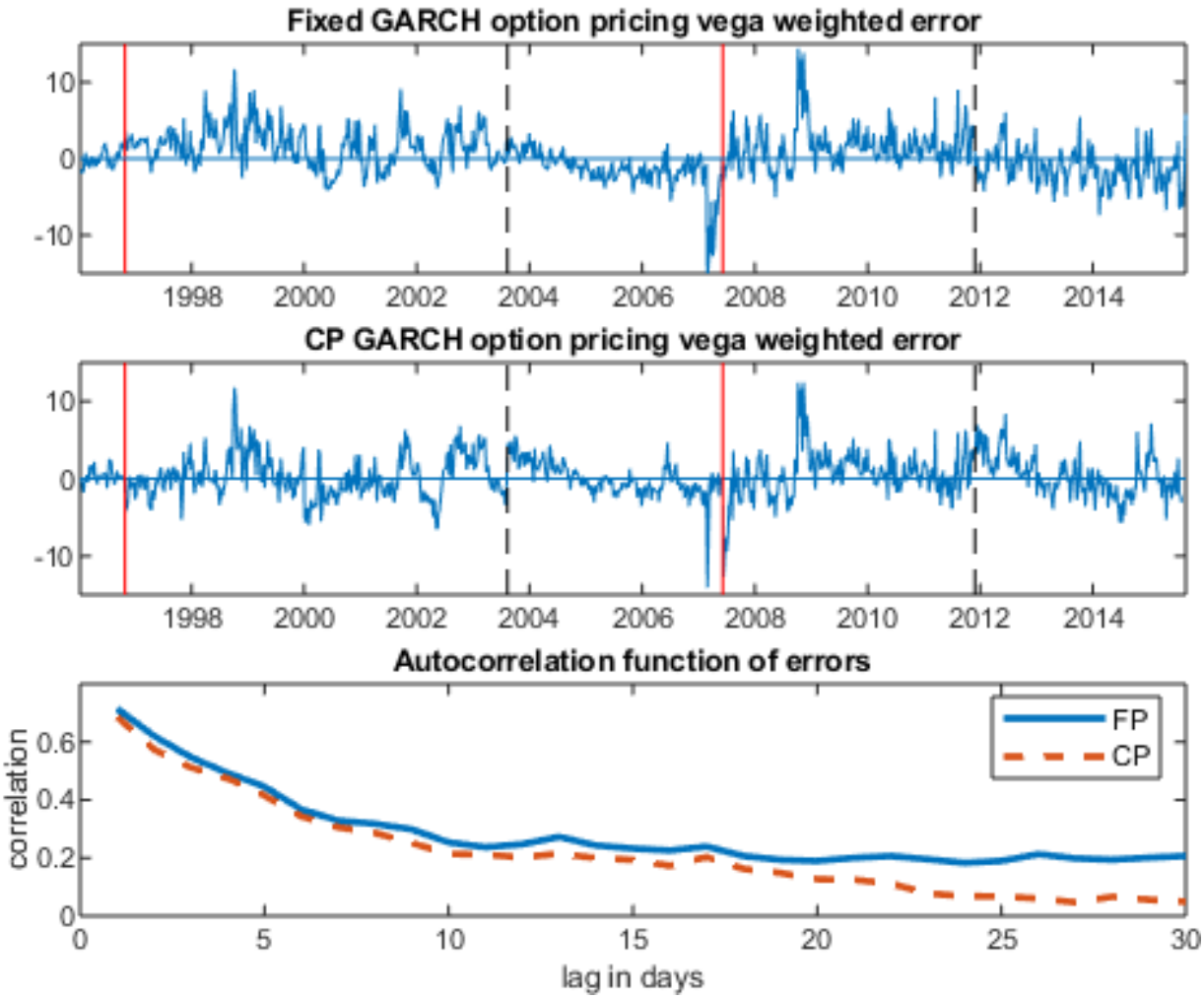


Figure 9: Time series of average option pricing erros

The figure shows the time series of average vega weighted option pricing errors per date of the Heston-Nandi GARCH model with fixed and CP parameters in the first two plots. Black (red) vertical lines indicate the beginning of a low (high) variance regime. The last plot shows the autocorrelation of the errors.

5.7 Term structure and time series of the pricing kernel

Table 9 shows the risk premia and Sharpe ratios for trading the pricing kernel implied by the estimated GARCH model with the variance-dependent pricing kernel. The used parameters are those from Table 8, and high and low volatility regimes are identified by the model. All specifications are chosen such that they match the empirical results from

Table 6.

The results in table 9 show that the FP model implied values do not match the observed patterns. While the values are generally countercyclical, they are much too low, and the term structure is rather flat. On the contrary, the CP model matches the empirical results well. Risk premia and Sharpe ratios are countercyclical and approach similar values for longer maturities. For the risk premia, the term structure is downward sloping, with a very steep slope for short maturities. In the model, this is caused by the mean-reverting property of the variance: as there is a high short-run risk of changes in the variance, this risk fades out in the long run. The Sharpe ratios of the pricing kernel straddle by large also have a downward sloping term structure. However, the slope is not as steep as observed in the market. A higher slope and also higher values for short maturities could probably be obtained by adding e.g. jumps in returns or volatility to the model.

Table 9: Model-implied risk premia and Sharpe ratios

	1w	2w	1m	2m	3m	1w	2w	1m	2m	3m
	Panel A: FP GARCH					Panels B: CP-GARCH				
Risk premia pricing kernel strangle										
Low vol	1.07	0.86	0.77	0.76	0.77	3.60	2.86	2.44	2.05	1.84
High vol	1.50	1.25	1.13	1.06	1.02	11.54	8.21	5.88	3.90	3.11
Sharpe ratios pricing kernel strangle										
Low vol	1.04	0.94	0.90	0.93	0.96	1.98	1.85	1.77	1.81	1.82
High vol	1.22	1.12	1.09	1.12	1.15	4.05	3.73	3.69	3.67	3.71

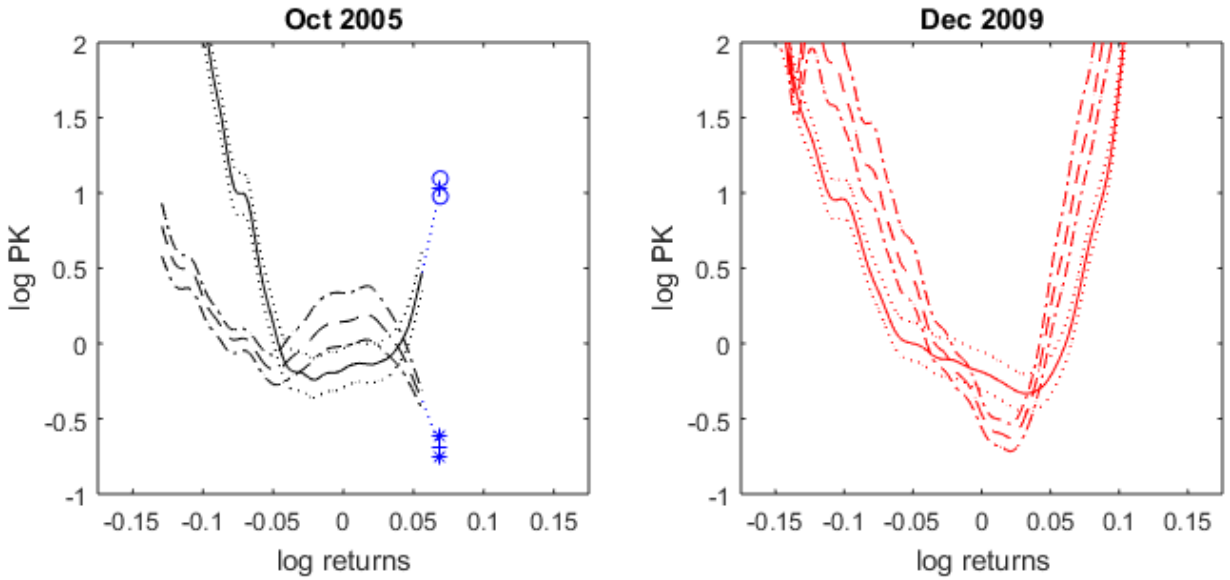
The table shows the risk premia and Sharpe ratios for trading the PK implied by the estimated GARCH models with the variance-dependent pricing kernel. The used parameters are those form Table 8. The values are obtained by simulating 500,000 paths of the model at each date.

6 Robustness

6.1 Confidence intervals for the estimated 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

Figure 10: Confidence intervals for EPK



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

a CP-GARCH model. This section introduces confidence intervals for the estimated pricing kernels in order to test the differences statistically. The 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 method of Aït-Sahalia & Duarte (2003), and the variation of $f()$ is calculated relying on asymptotic results from Härdle et al. (2014). Appendix F contains the full details of the approach.

Figure 10 illustrates the results by plotting the 90% confidence intervals for two selected years. The plot shows that the confidence intervals are fairly tight. However, testing for global shape differences is not straightforward, because the different estimates always intersect. Therefore I rather test for differences at certain, prominent points that are decisive for the global shape. In particular, I test whether the disappearance of the hump of the S-shape is statistically significant, and for properties of the right end and tail of the EPKs. The calculated statistics over all dates are summarized in Table 10.

First, I test whether the hump of the S-shape, when it appears in the method with the FP GARCH, is significantly different from the CP estimate. A hump is defined as a the local maximum of the estimated PK in the area of returns between -5% and 5%,

that is surrounded by lower values on both sides. The test statistic is whether the 90% lower bound of the FP method is higher than the 90% upper bound of the method with the breaks. 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 10: Test results on differences between estimated pricing kernels

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

The table reports the results for the statistical test on differences between the pricing kernel estimates. The test statistic for the hypothesis $P\hat{K}_A > P\hat{K}_B$ is $P\hat{K}_A^{lower90\%bound} > P\hat{K}_B^{upper90\%bound}$. The table reports the percentage of dates in the 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. 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.

6.2 Verifying U-shape using 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 6.1 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 returns on options 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 the non-parametric, average (payoff-weighted) risk-neutral density,i.e. the call prices, relative to the non-parametric, average physical density, i.e. the realized returns, and hence is a function of the PK.

To show that this effect stems not only from returns from times of high volatility (where the U-shape is less ambiguous), I replicate their approach using only call option data from the low volatility regimes. For this, from all call options from the respective time period (and exactly the same that are used above for the PK estimation), those are selected that are closest to the target moneyness of the bin. Then their returns are calculated using the corresponding settlement prices. Finally, average returns for each moneyness group are calculated and the results are displayed in Table 11. The confidence intervals are bootstrapped as in Bakshi et al. (2010). The target moneyness ranges from 0%, 1%,...,9%, and 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 11 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). Only the 4% or 5% and the 7% or 8% group appear to violate the monotonicity, with either the first having too low returns, or the later having too high returns, or both. Among the negative average returns, only those in the 7% to 9% bins are significantly different from zero. However, most pair-wise differences between groups are positive and many of them are statistically different from zero.[27] In sum, the results are consistent with a U-shaped pricing kernel. They show

[27]Note that the statistical significance of the differences could be increased by using less bins with larger difference in moneyness. While the presented results use only the monthly cycle of

non-parametrically that also in times of low volatility the PK is very likely to be 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 11: Returns of OTM call options

% OTM	1%	2%	3%	4%	5%	6%	7%	8%	9%
Return [%]	9.24	3.8	-6.8	-15.8	-10.6	-32.5	-90.5	-87.5	-88.2
90% Conf.	[-9, 28]	[-18.3, 27.2]	[-34.2, 23.4]	[-52.8, 28]	[-66, 56.9]	[-84.2, 35.3]	[-97.8, -80.3]	[-98.1, -73.6]	[-100, -75.4]
	0%-1%	1%-2%	2%-3%	3%-4%	4%-5%	5%-6%	6%-7%	7%-8%	8%-9%
Differences:	3.1	5.3	10.7	8.9	-5.2	21.9	58.0	-3.1	0.7
90% Conf.	[-1.7, 7.7]	[-1.7, 12.2]	[0.1, 20.6]	[-5.2, 29.6]	[-31.3, 17.7]	[-6.9, 29]	[1.3, 13.9]	[0.1, 5.6]	[0.2, 8.5]

The reported average returns (in percent) are of call options on the S&P 500 index over 01/1996-08/2015 that are closest to the target moneyness and that are from periods of low volatility, as defined in Section 2.3.3. Call returns are calculated using actual settlement prices. Moneyness is calculated as: [strike/(index level at price date)]. The average time to maturity is 30 days. The 90% confidence intervals for average returns are obtained from 25,000 i.i.d. bootstrap draws from the sample of option returns (in square brackets). To test for difference in returns, 25,000 pairwise bootstrap samples of returns are drawn.

6.3 International evidence

Most studies on the pricing kernel puzzle use US option data, and in particular often 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.[28] 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.

For brevity, this section presents only the main results for the FTSE 100, as it is the second most studies underlying in the PK literature. The missing details, as e.g. the

AM settled options, the results are similar if all options are used or the level of the VIX is used to identify times of high and low volatility.

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

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 in Section 2.3, only with option and return data for the FTSE 100.[29] The one month GBP LIBOR rate was used 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 would strongly select 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 11 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 high 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.

[29] The number of regimes is always chosen to be five for better comparison. However, for most indices this is also the optimal number of regimes. Furthermore, the starting date of the time-series, 02.01.1992, is also not "optimized", meaning, the end of the previous regime was not estimated with maximum likelihood, and the same start date is used to match the S&P 500 case.

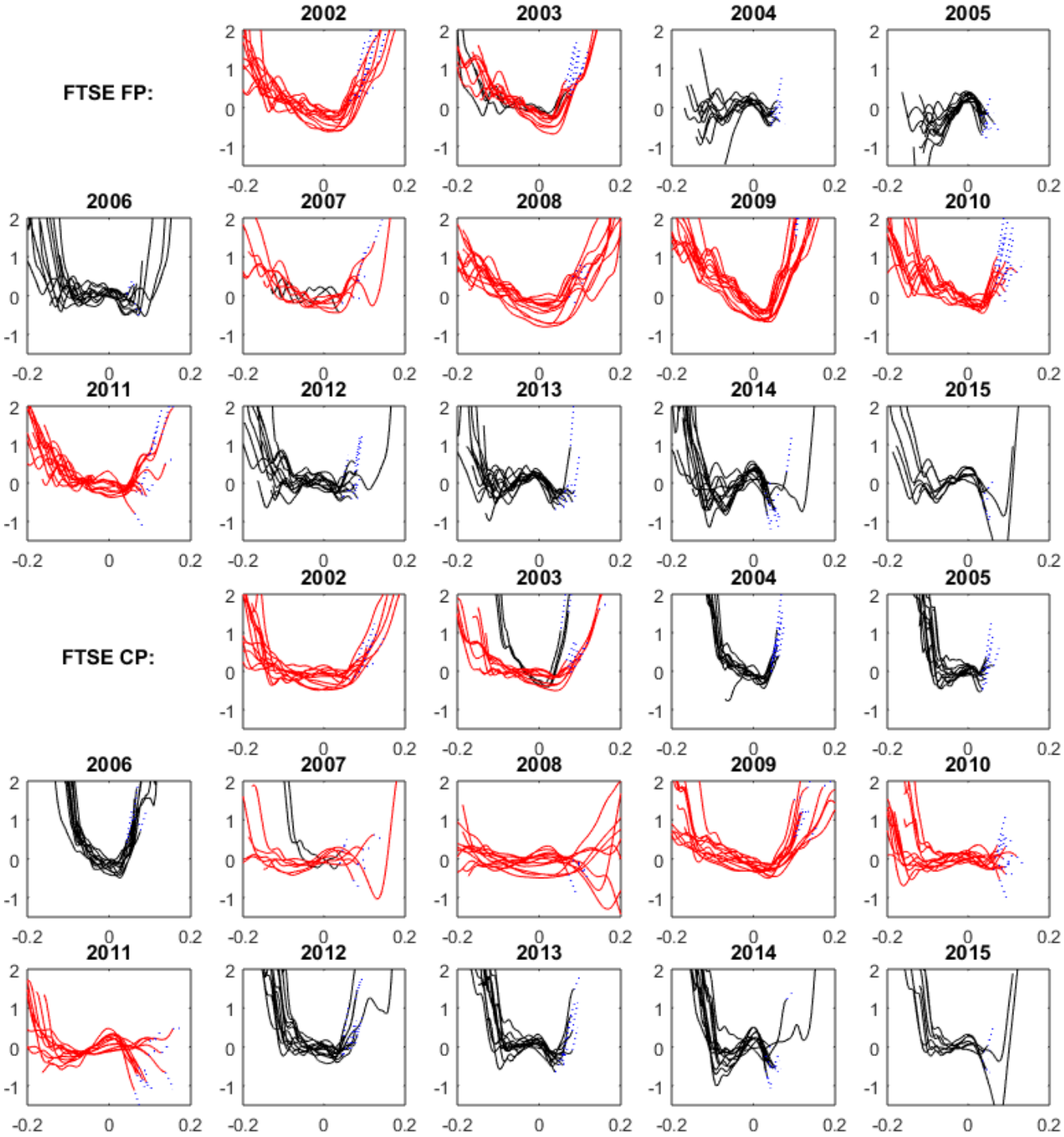


Figure 11: Empirical pricing kernels with FP and CP for the FTSE 100

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 2.3.3. 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.

6.4 Using VIX volatility forecasts

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 problem 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 (5) and (6). 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 may 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 12 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.

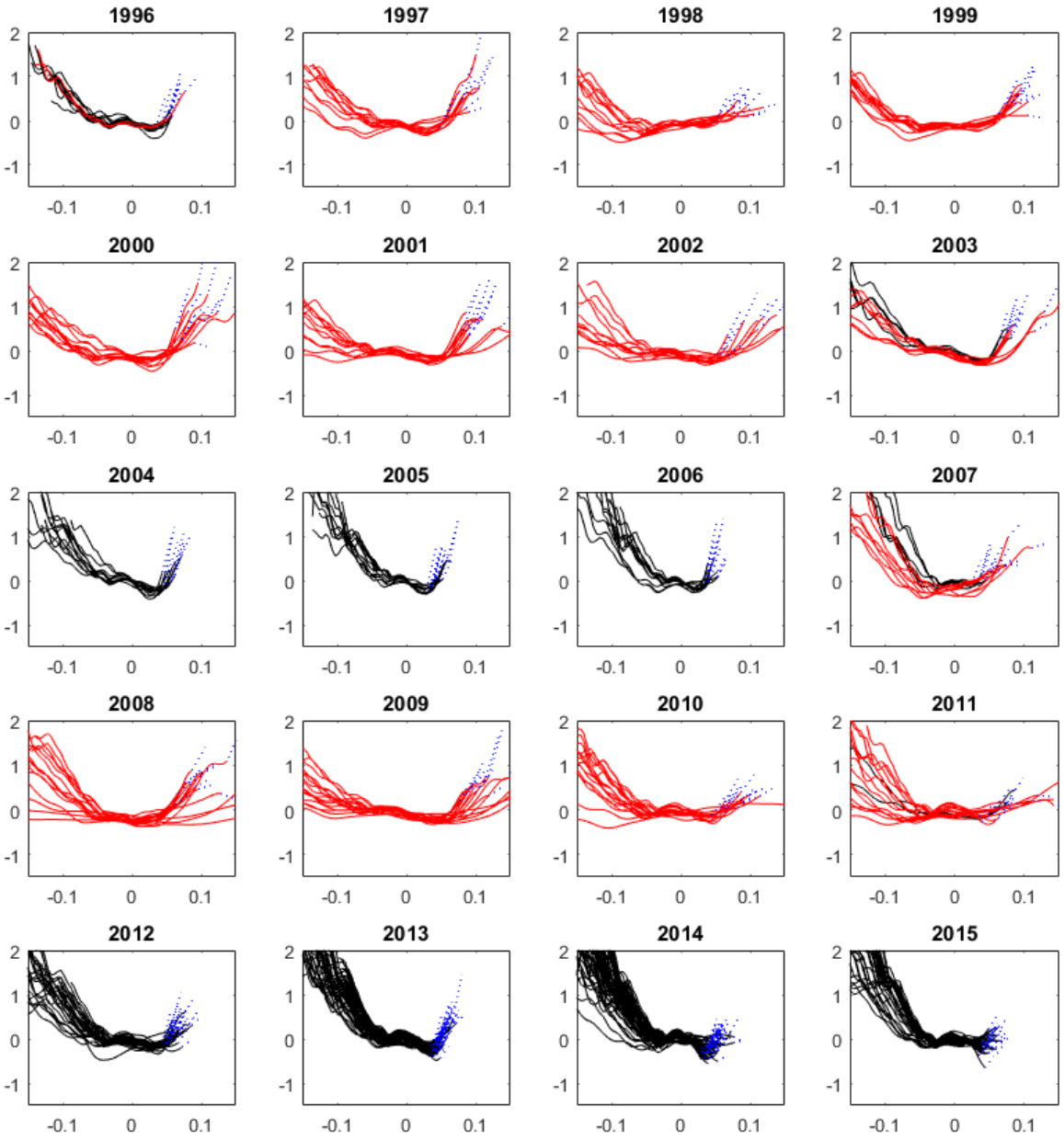


Figure 12: Empirical pricing kernels using the VIX as volatility forecast

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 2.3.3. 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.

6.5 Expected physical return distribution from GARCH simulations

A popular alternative to obtain a conditional forecast of the physical return distribution is to simulate it directly from the GARCH model.[13] 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 switching 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.

6.6 GARCH innovation follow Student's t-distribution

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.

6.7 Using realized volatility forecasts from the Corsi model

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.[30]

The following summarizes the main results, while the online Appendix provides technical details, estiamtion 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 (5) and (6), 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 fully support this conjecture. Lastly, the wiggles around zero are the result of an unsmooth empirical return shock density. Figure 3 in the online Appendix illustrates this. Furthermore, smoothing these wiggles would give a PK estimate that typically is well within the confidence intervals. Overall, using the HAR model at least partly confirms the U-shape of the the pricing kernels. The results also point in the direction of different volatility regimes, that are hard to capture with standard models.

6.8 Different maturities

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 6.5 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.

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

6.9 Including the switching probability

The benchmark analysis ignores the probability to switch into another regime, since this probability is very small. 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 a red (black) regime to an average low (high) variance regime. The 2% are still a rather high estimate, 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 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.

6.10 Different GARCH 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 following analysis shows that the results 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. The detailed results can be found in the online Appendix. In sum, this suggests that the key driver of the results, namely the biased volatility forecast, prevails in any GARCH specifications.

6.11 Rolling window GARCH

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 run 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, 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 so smooth results.

6.12 Relation of the results to other methods in the existing literature

The methods used so far to derive an estimate of the conditional physical return distribution are the methods used in the majority of the existing literature on option implied pricing kernels. There are a few more methods, that e.g. assume a specific parametric function of the return distribution. However, these methods most likely are also affected by the volatility bias documented above. The online Appendix provides a more detailed discussion.

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 different questions in asset pricing. One of the key implications is the implied time series and term structure of risk premia. An important step in this regard is to trade the pricing kernel in the options market. This is a novel way to test the quality and the properties of pricing kernel estimates non-parametrically. The obtained Sharpe ratios show that the volatility of the pricing kernel is very high, especially in the short run. Furthermore, the

term structure of risk premia and Sharpe-ratios is generally downward sloping, and the values are countercyclical. 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.

Finally, the empirical results are matched very well by a variance-dependent pricing kernel, but only when the model contains structural breaks. It furthermore emerges that the breaks are equally important for modeling the risk-neutral GARCH dynamics, which are otherwise also cyclically biased.

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. 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. 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 option literature. The following gives a detailed list of the option data cleaning steps applied.

A.1 Pricing Kernel Option Data

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
    - 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 Wednesday, 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 (15) and (16) to recover the risk-neutral return distribution.

$$1 - F^*(S_{t+\tau}) = -\exp(r\tau)\Big[\frac{\partial C_{BS}(S_t, X, \tau, r, \hat{\sigma}(S_t, X))}{\partial X}\Big]_{|X=S_{t+\tau}}. \tag{15}$$

$$f^*(S_{t+\tau}) = \exp(r\tau)\Big[\frac{\partial^2 C_{BS}(S_t, X, \tau, r, \hat{\sigma}(S_t, X))}{\partial X^2}\Big]_{|X=S_{t+\tau}}. \tag{16}$$

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.

B.1 Note on the matching physical return distribution

Since the option data contains several slightly different times to maturity (from 29 up to 32 days) and thereby also different numbers of expected trading days, several different 'monthly' returns are calculated, one for each observed number of expected trading days. Expected trading days are the number of working days minus the holidays between the date of the option price and the maturity date.[31] Each option price date is then matched with the correct length of the 'monthly' returns.

C Adjustments to empirical pricing kernel for applications

C.1 General

In Section 4 the empirical pricing kernel estimates presented in Section 3.3 are used, with a few minor adjustments. First, where necessary, the PK is interpolated linearly between the two grid points adjacent to the realized index return. Furthermore, an upper bound of 30 is put on the empirical PK estimates. The main reason is to reduce some noise in the estimates, as the maximum PK point is in the order of magnitude of 10^6. This cap is rarely applied, and this happens only in the tails. Furthermore, it makes the ratio of price to maximal payoff of the PK strangle, as well as its price to margin ratio, more similar to the one of a deep out-of-the-money put. Finally, increasing the bound slightly increases the botht he implied and realized risk premia and the Sharpe ratios, but does not change the main findings. Lastly, the ratio of the CDFs is used for the tails. In Section 4.1, 4.3 and 4.4 only the PK 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. Option settlement prices are obtained from the website of the CBOE.

[31] The majority of options in the used sample (closing prices) are AM settled (SPX), 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 results below get even stronger.

C.2 Making the pricing kernel monotone

In Section 4.4 the pricing kernel is made weakly monotonically decreasing in the domain of positive returns in the following way. First, the minimum value of the PK estimate in the domain of positive returns is found. Then all points of the PK to the right of the return corresponding to the minimum are replaced with that minimum, including the CDF ratio in the right tail. This implies that the PK is flat to the right of its minimum.

Consequently, the PK is often "too low" on average. This can be solved by shifting it upwards by adding a constant a. The effect on the pricing errors is be the following: $E\big((M^{mono} + a) \cdot R\big) = E\big(M^{mono} \cdot R\big) + E(a \cdot R)$. Taking the time series average gives $a \cdot \bar{R}$ for the second term, i.e. a times the sample average return. Since the differences in returns are small on a monthly basis, the effect on the pricing errors is also small. Furthermore, the effect would go in favor of the results, as the average returns decrease from the low to the high beta portfolio, and hence the spread between the portfolio would increase.

Alternatively, a_t can be calculated time-varying such that $E_t\big((M^{mono}_{t+\tau} + a_t) = E_t\big(\hat{M}_t\big)\big)$. As long as there is no time series correlation between a_t and the return, the error is the same as for the constant a case.

Both are tested, and the results are qualitatively the same, and quantitatively stronger, as predicted.

D Construction of low beta portfolios and returns

Following Bali et al. (2014) the ex ante beta is estimated using a rolling window of one year of daily returns. The stock returns are obtained from CRSP, the risk-free rate and the market factor are form the website of Kenneth French. Stocks are excluded if the have less then 180 valid return observations per year. Lastly, stocks with an ex ante price below 1$ (5$ as robustness) are excluded. For the beta quintile break points only the betas 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 beta, 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 beta estimation periods, using either only beta-sorting or independent double sorts on size and beta 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 the preceding day 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.

E Data for the model estimation

The steps for the data cleaning is as above. The selection of the option data is different and follows by large Christoffersen et al. (2013). For each Wednesdays in the sample period, the option series with a maturity closest to 30 days is selected. From that maturity, the 15 most actively traded options are used. This results in 15,171 option prices with a maturity between 17 and 53 days. In contrast to Christoffersen et al. (2013) 15 instead of six options are chosen per date, to broaden the moneyness range for each date and to have more deep out of the money options in the sample.

F 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 $\hat{f}^{*}$[32] I follow Aït-Sahalia & Duarte (2003), that 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

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

the same specification and also $N = 5000$ simulations. Since the liquidity in the sample used here is higher, this should be a conservative approach. Next, I follow the standard methodology from Section 3.2 above. The 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:

$$\sqrt{n_p h_{n_p}}\{\hat{f}(x) - f(x)\} \xrightarrow{\mathcal{L}} N(0, f(x) \int K^2(u)du), \tag{17}$$

where $\mathcal{L}$ denotes convergence in lax, 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:

$$P\Big(\frac{\hat{f}^*(x)}{\hat{f}(x)}\Big) \geq \hat{PK}^{upper}(x) = (\alpha/2), \tag{18}$$

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

$$P\Big(\frac{\hat{f}^*(x)}{\hat{f}(x)}\Big) \geq \hat{PK}^{lower}(x) = (1-\alpha/2), \tag{19}$$

where

$$\Big[P\Big(\frac{\hat{f}^*(x)}{\hat{f}(x)}\Big) \geq \hat{PK}(x)\Big] = \frac{1}{N} \sum_{\hat{f}^*(x)_{min}}^{\hat{f}^*(x)_{max}} \int_{-\infty}^{\hat{f}^*(x)/\hat{PK}(x)} \hat{f}(x)dx. \tag{20}$$

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

The distribution of the risk-neutral density estimate is always the same, regardless of which method is used to estimate the physical density. Therefore, differences in the confidence intervals of different methods at the same point in time only arise from differences in the physical density estimate.

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.

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. & Dall'o, H. (2012), 'Is the price kernel monotone?', *ACRN Journal of Finance and Risk Perspectives* **1**(2), 43–68.

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* .

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.

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.

Figlewski, S. (1994), 'Forecasting volatility using historical data'.

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* .

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. (2016), '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.

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.

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.

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.

Pan, J. (2002), 'The jump-risk premia implicit in options: Evidence from an integrated time-series study', *Journal of financial economics* **63**(1), 3–50.

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* pp. 407–425.

Sichert, T. (2018), 'Maximum likelihood estimation of change-point garch models', *Working Paper, Goethe University Frankfurt* .

Van Binsbergen, J., Brandt, M. & Koijen, R. (2012), 'On the timing and pricing of dividends', *American Economic Review* **102**(4), 1596–1618.

Van Binsbergen, J. H. & Koijen, R. S. (2017), 'The term structure of returns: Facts and theory', *Journal of Financial Economics* **124**(1), 1–21.

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