

# Structural Breaks in the Variance Process and the Pricing Kernel Puzzle

Tobias Sichert\*

sichert@finance.uni-frankfurt.de

January 1, 2018

## Abstract

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. This is not only empirically inconsistent, but the two observations also seem theoretically incompatible. In particular, the S-shape is conflicting with most modern asset pricing models. By providing novel time series evidence, this paper reconciles the so far conflicting empirical results. I show that the finding of S-shaped pricing kernels is spurious and is removed by including structural breaks in the data generating process into the estimation. In the sample period from 1992-2015 I identify five different high or low variance regimes. Conditional on the regime, the obtained pricing kernels appear U-shaped, while the S-shaped pricing kernels consistently disappear. The results are robust to numerous variations in the methodology. The empirical results can be explained by a variance-dependent pricing kernel, with structural breaks as a necessary component. Lastly, the results show that the fit of the option pricing model increases substantially when breaks are introduced.

**Keywords:** Asset pricing, pricing kernel puzzle, stochastic discount factor, options, GARCH, change-point model

---

\*The author would like to thank Marc Crummenerl, Bob Dittmar, Bruce Grundy, Markus Huggenberger, Holger Kraft, Jan Pieter Krahen, Christoph Meinerding, Christian Schlag, Paul Schneider, Marti Subrahmanyam, Roméo Tédongap, Julian Thimme, Amir Yaron, participants of the 24th Annual Meeting of the German Finance Association in Ulm, and seminar participants at Goethe University Frankfurt for valuable comments and suggestions.

# 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. 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), Ait-Sahalia & Lo (2000) and Rosenberg & Engle (2002) and, since then, has been confirmed by many others.<sup>1</sup> 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.<sup>2</sup> The top left plot of Figure 1 illustrates the typical S-shape. Yet, some studies find a U-shaped pricing kernel (Christoffersen et al. 2013), as illustrated in the bottom left plot, and others find even both S- and U-shapes in their sample (Grith et al. 2013).

However, the two shapes are theoretically incompatible. Although many theoretical models can explain either one of the shapes, neither can explain both.<sup>3</sup> 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.<sup>4</sup> It is therefore not surprising that theoretical explanations for the S-shape

---

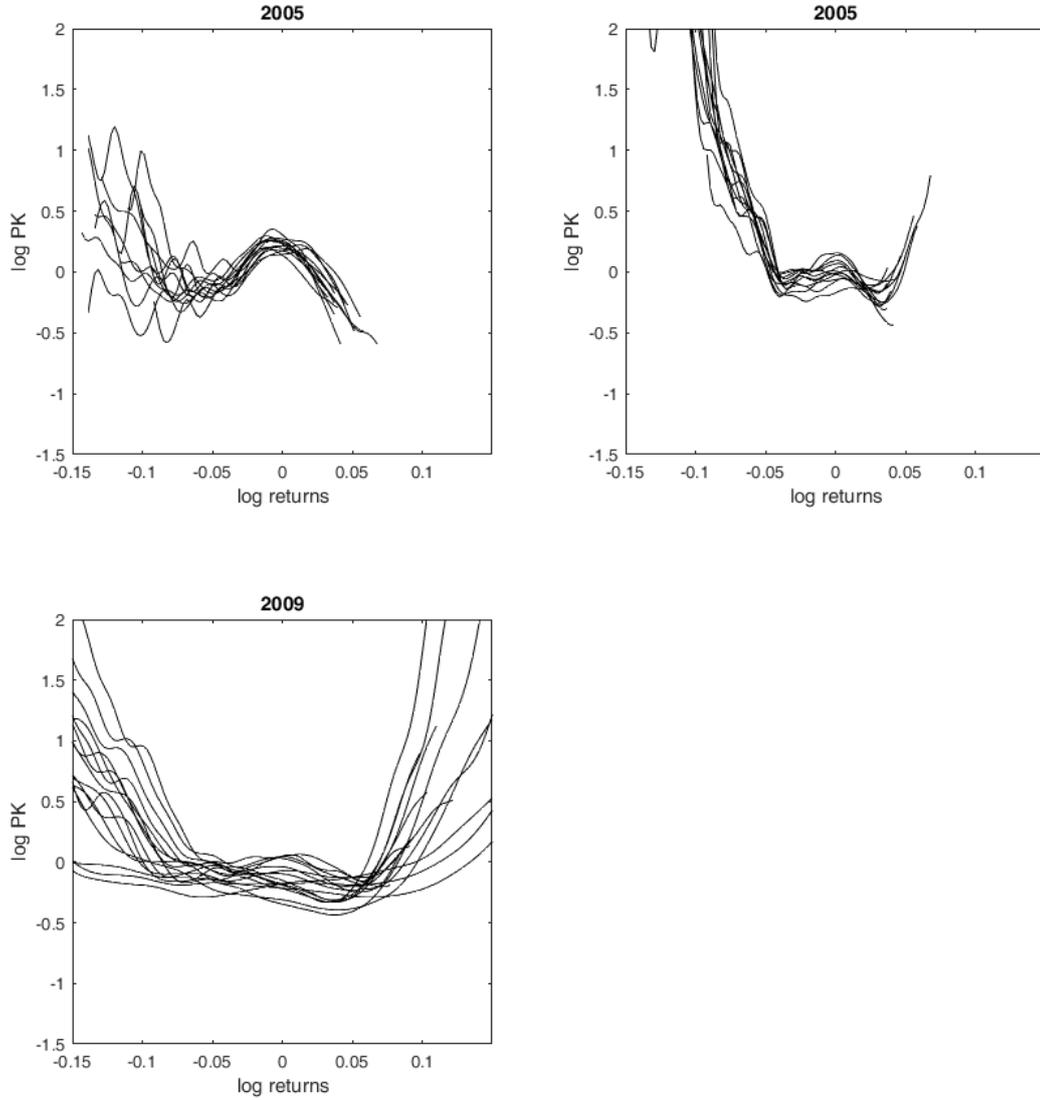
<sup>1</sup>The literature on the pricing kernel puzzle has become too large to fully describe here. For more details see e.g. Cuesdeanu & Jackwerth (2016), which provide a great and comprehensive overview on the existing empirical, theoretical and econometric literature on the pricing kernel puzzle.

<sup>2</sup>E.g., Jackwerth (2000), Ait-Sahalia & Lo (2000), Rosenberg & Engle (2002), Liu et al. (2009), Polkovnichenko & Zhao (2013), Figlewski & Malik (2014), Beare & Schmidt (2016), Belomestny et al. (2015) Cuesdeanu & Jackwerth (2016), Barone-Adesi et al. (2016) Grith et al. (2016).

<sup>3</sup>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, like a W-shape.

<sup>4</sup>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.

Figure 1: Selected representative results



The figure shows the results for two representative years. The plots contain the natural logarithm of estimated pricing kernels. The left column shows the results obtained with a standard GARCH model with fixed parameters. The right column shows the results for the change-point GARCH model. Log-returns are on the horizontal axis. The horizon is one month. 2005 is a typical low-volatility year, while 2009 is a typical high-volatility year.

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), or ambiguity aversion (Cuesdeanu 2016). 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)). Branger et al. (2011) show that this explanation can also be obtained in a general equilibrium framework with Epstein-Zin preferences and stochastic volatility. 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.

This paper attempts to reconcile the different empirical and conflicting theoretical results. First, it provides novel evidence on the time series behavior of the pricing kernel puzzle, which has not been subject to much research. In particular, this paper shows that structural breaks in the data generating process are the reason why some researchers find S-shaped pricing kernels. 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. This is shown by comparing the results of a novel estimation technique with the (nested) standard estimates. Second, the paper demonstrates that the structural breaks are also necessary to obtain a robust economic explanation for the results. In particular, a variance dependent pricing kernel matches the empirical findings only if structural breaks are included in the model.

To estimate the pricing kernel from 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 recognized early on that the key ingredient to predicting return distributions is the volatility forecast. It is well known, e.g., from the vast literature on GARCH models in finance, that volatility is time-varying and clustered. While standard volatility models can only capture the first property, I will show that the clustering is important too.

First, I estimate a change-point (CP) GARCH model to identify points where the parameters of the GARCH process change. The potential existence of structural breaks in GARCH processes has been known in the econometric literature for many years (e.g. Diebold (1986)). Although the vast majority of studies on empirical pricing kernels apply

a GARCH model to condition the estimation on contemporary market expectations, the robustness to breaks has never been analyzed. 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.

Next, the paper studies the estimation of empirical pricing kernels in a long sample of S&P 500 options over the period 1996-2015. The analysis demonstrates that standard methods relying on GARCH models with fixed parameters, tend to estimate S-shaped pricing kernels in times of low variance, and U-shapes otherwise. Furthermore, when replacing the standard GARCH with the CP-GARCH in the otherwise identical methodology S-shaped pricing kernels disappear altogether. Figure 1 illustrates these findings for two representative years. In 2005, a typical low-volatility year, the standard estimates (top left) are S-shaped, and become U-shaped when the CP-GARCH model is used (top right). In 2009, a typical high-volatility year, the standard estimates are U-shaped, and they remain U-shaped with the new methodology.

The analysis further shows that a standard GARCH model provides biased multi-period volatility forecasts and that this is the crucial driver behind the fact that many studies find S-shaped pricing kernels. It turns out that 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 over extended periods of time. The reverse is, to a lesser extent, true for periods of high volatility. This leads to systematically biased volatility estimates 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.

The intuitive explanation for the results is that the overestimated volatility leads to a return distribution forecast that is too wide, and has too much probability mass in the tails and too little in the center. The excess weight in the left tail is not strong enough to change the downward 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.

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 a 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 the final part, I discuss several asset pricing implications of the results. First, in contrast to the results of some studies that at best apply to an unconditional average estimate of the pricing kernel, I examine the conditional pricing kernel. A simple example demonstrates how results concerning the unconditional pricing kernel can be misleading in the light of the new empirical evidence.

Second, I show 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 stochastic discount factor is non-monotonic. The high negative variance risk premium causes the projection of the stochastic discount factor on the index returns to be U-shaped. The 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 bias in the multi-period volatility forecasts carries over to the risk-neutral dynamics as well and makes option prices systematically biased when using fixed parameters. 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).

Overall, the paper provides novel semi-parametric evidence on the time series behavior of the pricing kernel puzzle. It shows that the canonical use of a standard GARCH model with fixed parameters significantly biases the results and the often documented S-shaped pricing kernels are not robust to the application of a GARCH process with structural breaks. These results challenge the existence of S-shaped pricing kernels, which has almost become consensus in the literature and is considered a stylized fact by some. Furthermore, the results show that the presence of structural breaks is relevant for several other economic applications. The findings are relevant for both researchers and practitioners. On the one hand, they significantly reduce the complexity of the pricing kernel puzzle by ruling out the typical S-shape. This provides valuable guidance for theorists when validating the predictions of their models. 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 introduces the change-point GARCH model. Section 3 presents the data, estimation methodology, estimation results and the model fit. Section 4 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 5 presents asset pricing implications and a model that explains the empirical findings. Section 6 conducts several robustness checks and Section 7 concludes. The Appendix collects methodological details, algorithms and additional tables and plots.

## 2 A GARCH Model with Structural Breaks

### 2.1 From standard GARCH to change-point GARCH

Three quantities are required to estimate conditional pricing kernels (PKs) empirically: the risk-free rate, conditional risk-neutral probabilities and conditional physical (objective) probabilities.<sup>5</sup> The estimation of the first is an easy task and, since the discounting effects over typical horizons of one or a few months are low, it is not a crucial parameter in any case. The estimation of the second quantity is not straightforward, but well established and understood methods exist. The estimation of the conditional risk-neutral probabilities from option prices is described in Section 4.2. The remaining third quantity, however, the conditional physical probability, is not easily quantifiable and requires a minimum of parametric assumptions. The chosen method to condition the constructed estimate of the return distribution is later shown to be the force that drives the results.

Some of the first studies on the pricing kernel puzzle use a kernel density estimation of past raw index returns on the S&P 500 (e.g. Jackwerth (2000), Aït-Sahalia & Lo (2000)). Many other studies agree that it is important to condition the estimate on current market volatility (see e.g. Rosenberg & Engle (2002) or Beare & Schmidt (2016)),

---

<sup>5</sup>There are different ways to study empirical pricing kernels implied by option prices. Some approaches, as for example the use of option return data, only allow to study an unconditional pricing kernel. Studying a conditional pricing kernel, however, has at least two advantages. Firstly, one is able to look at time series properties. Secondly, an unconditional pricing kernel, which is an average of all conditional ones, could dampen or average out any local increases or decrease (see also Section 5.1). There is also a group of studies that assume a specific functional form of the pricing kernel. This usually restricts the shape one can find and sometimes also hampers the study of conditional kernels.

and almost all studies use a GARCH model for this. 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.

## 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.<sup>6</sup> Such a CP-GARCH model is laid out in the following. How the model is used to construct a conditional return distribution is presented in Section 4.2.

There are two prominent GARCH models often used for modelling 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.<sup>7</sup>

The dynamics of the standard HN-GARCH model are:

$$\ln\left(\frac{S_t}{S_{t-1}}\right) = r_t + \left(\mu - \frac{1}{2}\right)h_t + \sqrt{h_t}z_t, \quad (1)$$

---

<sup>6</sup>A related and similar specification is a Markov switching model. See Appendix A for a discussion for the relation of the CP model to the MS model in the context of the present study.

<sup>7</sup>The paper includes both versions for several reasons. Hsieh & Ritchken (2005) provide evidence that the NGARCH model fits S&P 500 option prices better, especially out-of-the-money contracts. Furthermore, the NGARCH model also fits the physical return distribution much better, as can be seen by comparing the estimation results provided in Sections 5.2.6 and 6.6. However, the HN-GARCH model allows for a closed form solution of the option price, while the NGARCH model requires a numerical solution, usually based on Monte Carlo simulations. This is of considerable relevance when estimating the model based on option prices. As a compromise, the main analysis of the paper uses the HN-GARCH model. Some of the analysis is repeated in the robustness section using the NGARCH model, and the results are very similar for both models.

$$h_t = \omega + \alpha \left( z_{t-1} - \gamma \sqrt{h_{t-1}} \right)^2 + \beta h_{t-1}, \quad (2)$$

$$z_t \sim N(0, 1), \quad (3)$$

where  $S$  is the stock's spot price,  $r$  is the daily continuously compounded interest rate,  $z$  are return innovations and  $h$  is the conditional variance. The long-run variance of the HN-GARCH model is:

$$E[h_t] = \frac{\omega + \alpha}{1 - \beta - \alpha\gamma^2}, \quad (4)$$

and the expected variance over  $T$  days is:

$$E_0 \left[ \sum_{t=1}^T h_t \right] = T \cdot E[h_t] + (h_1 - E[h_t]) \frac{1 - (\beta + \alpha\gamma^2)^T}{1 - (\beta + \alpha\gamma^2)}. \quad (5)$$

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

$$\ln\left(\frac{S_t}{S_{t-1}}\right) = r_t + \left(\mu_{y_t} - \frac{1}{2}\right)h_t + \sqrt{h_t}z_t, \quad (6)$$

$$h_t = \omega_{y_t} + \beta_{y_t}h_{t-1} + \alpha_{y_t} \left( z_{t-1} - \gamma_{y_t} \sqrt{h_{t-1}} \right)^2, \quad (7)$$

$$z_t \sim N(0, 1),$$

where  $y_t$  is an integer random variable taking values in  $[1, K + 1]$ . 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 can be obtained by setting  $K = 0$ .

The economic interpretation of the change-point model is that there are different regimes in the market, and when they end, fundamentals change. These changes are so dramatic, that the standard and already dynamic model cannot capture them, but a full new parametrisation of the model is required in each regime. The estimation below

shows that the identified regimes are on average 5-6 years long and are closely related to business cycles.

## 3 Data, Model Estimation and Model Fit

### 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 a longer sample shows that the regime that prevails in 1996 starts around January 1992. As a robustness check, the fixed parameter model is also estimated over the longer sample from 02.01.1986 to 30.06.2016 and the obtained parameters are very similar. The risk-free rate is obtained from OptionMetrics and the term structure is interpolated linearly.

### 3.2 Likelihood functions

#### 3.2.1 Standard GARCH

For the Heston-Nandi GARCH model with fixed parameters a classical likelihood function based on daily returns is used. The conditional density function of the daily returns is normal, so:

$$f(R_t|h_t) = \frac{1}{\sqrt{2\pi h_t}} \exp\left(-\frac{(R_t - r_t - (\mu - \frac{1}{2})h_t)^2}{2h_t}\right), \quad (8)$$

where  $R_t$  ( $r_t$ ) denotes the observed daily log stock return (daily continuously compounded risk-free rate) at time  $t$ . The return log-likelihood is:

$$\ln L^R = -\frac{1}{2} \sum_{t=1}^T \left\{ \ln(2\pi h_t) + \left( R_t - r_t - \left( \mu - \frac{1}{2} \right) h_t \right)^2 / h_t \right\}. \quad (9)$$

The physical variance is filtered from returns using:

$$\begin{aligned}
h_t &= \omega + \beta h_{t-1} + \alpha \left( z_{t-1} - \gamma \sqrt{h_{t-1}} \right)^2, \text{ where} \\
z_t &= \left[ R_t - r_t - \left( \mu - \frac{1}{2} \right) h_t \right] / \sqrt{h_t}, \\
h_0 &= \frac{\omega + \alpha}{1 - \beta - \alpha \gamma^2}.
\end{aligned} \tag{10}$$

The optimal parameters are obtained from:

$$\Theta = \{\omega, \alpha, \beta, \gamma, \mu\} = \arg \max_{\Theta} \ln L^R. \tag{11}$$

### 3.2.2 Change-point GARCH

For the change-point Heston-Nandi GARCH the likelihood of the observations, denoted by  $g(R|\Theta)$ , is:

$$\begin{aligned}
g(R|\Theta) &= \sum_y g(R, y|\Theta) = \sum_y g(R|y, \Theta) p(y|\Theta) = \\
&= \sum_y \left[ \prod_{t=1}^T \frac{1}{\sqrt{2\pi h_t}} \exp \left( - \frac{(R_t - r_t - (\mu_{t_t} - \frac{1}{2})h_t)^2}{2h_t} \right) \right] p(y|\Theta),
\end{aligned} \tag{12}$$

where  $R = \{R_1, \dots, R_T\}$  is the vector of returns and  $y = \{y_1, \dots, y_T\}$  is the vector of regimes. The exact calculation of (12) is infeasible, but it is possible to obtain a good approximation for it, which is discussed next.

## 3.3 Estimation Methodology

Sichert (2017) proposes an estimation algorithm for the change-point GARCH model that uses particle filters, and both the Monte Carlo expectation-maximization algorithm and the Monte Carlo maximum likelihood method to obtain the maximum likelihood estimator (MLE).<sup>8</sup> This hybrid algorithm, called Particle-MCEM-MCML, is based on

---

<sup>8</sup>GARCH models with switching parameters are notoriously difficult to estimate as a result of the path dependence problem. This means that, due to the recursive nature of the GARCH process, the conditional variance at any given point in time depends on the entire sequence of regimes visited up to that point. To calculate the full likelihood function one would have to integrate over all possible regime paths when computing the likelihood function. This is infeasible since the number of possible paths grows linearly in the number of observations in the case of

the algorithms proposed by Augustyniak (2014) and Bauwens et al. (2014). The main steps of the algorithm are repeated in Appendix D. For a more detailed discussion of the approach as well as empirical studies the reader is referred to Sichert (2017). To identify the optimal number of breaks the algorithm is run with the number of breaks  $K = 2, \dots, 10$ . Then the optimal number of breaks is chosen by the algorithm using the Bayesian information criterion. Using the Aikake information criterion would deliver the same result. The optimal number of breaks

## 3.4 Model fit and properties

### 3.4.1 Estimation results

To the best of the author’s knowledge, there are no examples of an estimation of change-point (or Markov-switching) Heston-Nandi GARCH model. Hence, a more detailed analysis seems appropriate. Table 1 presents the estimation results. In the upper part, the first column gives the parameters for the standard GARCH, while the remaining columns contain the CP parameters for each regime. The section in the middle of the table shows the degree of integration of each variance process as well as the annualized long-run volatility implied by the parameters. The lower part 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 breaks is five, and the identified break dates are 01.01.1992, 28.10.1996, 12.08.2003, 07.06.2007 and 29.11.2011, each being the first date of the new regime.

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. The two information criteria both correct the log-likelihood for the number of parameters which are used. Comparing these measure across models also strongly suggests that the CP-GARCH is better (lower values of AIC and BIC are

---

one break, and much faster for more breaks (but not exponentially, as in the Markov switching case). Therefore, the first proposed estimation methods relied on simplification procedures that circumvent the path dependence problem (Gray (1996), Klaassen (2002)). Bayesian methods were, among others, proposed by Bauwens et al. (2010), Bauwens et al. (2014) and Billio et al. (2016). The approach of Bauwens et al. (2014) furthermore allows to calculate the likelihood function using a particle Markov chain Monte Carlo method.

better). Furthermore, the estimates show that there are distinct variance regimes. The long run variance differs significantly across the regimes, while the variance of the fixed GARCH more or less fits the average variance. Lastly, 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  $\beta + \alpha\gamma^2$  lower than the one regimes model.

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, which becomes especially clear when considering the lower ‘degree of integration’, i.e. the value of  $\beta + \alpha\gamma^2$ . Furthermore, the dynamics across the regimes are very different as well as the long-run variances. If there would not be any structural changes, the estimation could not identify them in such a long sample.<sup>9</sup>

Figure 2 illustrates the identified regimes by plotting the break dates together with the level and 21 day realized volatility of the S&P 500 index. By visual inspection alone it becomes immediate that there are 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 bust of the dotcom bubble. The second high volatility regime contains the recent financial crisis and its aftermath.

---

<sup>9</sup>It is standard in the related literature to estimate the GARCH model over the full sample, even though this could introduce a potential bias. Second, the analysis focuses on the comparison of the standard GARCH with fixed parameters versus the CP-GARCH, and both are estimated over the same sample.

Table 1: Estimation Results of the HN-GARCH model

Parameters	FP HN-GARCH		CP-HN-GARCH			
	'92-15	'92-'97	'97-'03	'03-'07	'07-'12	'12-'15
$\omega$	3.01E-19	2.24E-06	1.91E-06	5.36E-06	4.41E-10	3.51E-06
$\alpha$	4.34E-06	1.37E-06	6.13E-06	8.05E-07	7.66E-06	2.15E-06
$\beta$	0.821	0.801	0.786	0.301	0.772	0.273
$\gamma$	188.9	269.1	164.7	836.8	161.1	542.8
$\mu$	1.756	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						
Total	19495.9	19691.6				
AIC	-38981.8	-39325.1				
BIC	-38948.4	-39131.0				

Parameter estimates are obtained by optimising the likelihood on returns. Parameters are daily, long run volatility is calculated as  $\sqrt{\text{long-run variance} \cdot 252}$ . For each model, the total likelihood value at the optimum is reported. The volatility parameters are constrained such that the variance is positive ( $0 \leq \alpha < 1$ ,  $0 \leq \beta < 1$ ,  $\alpha\gamma^2 + \beta < 1$ ,  $-\alpha < \omega$ ). 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.

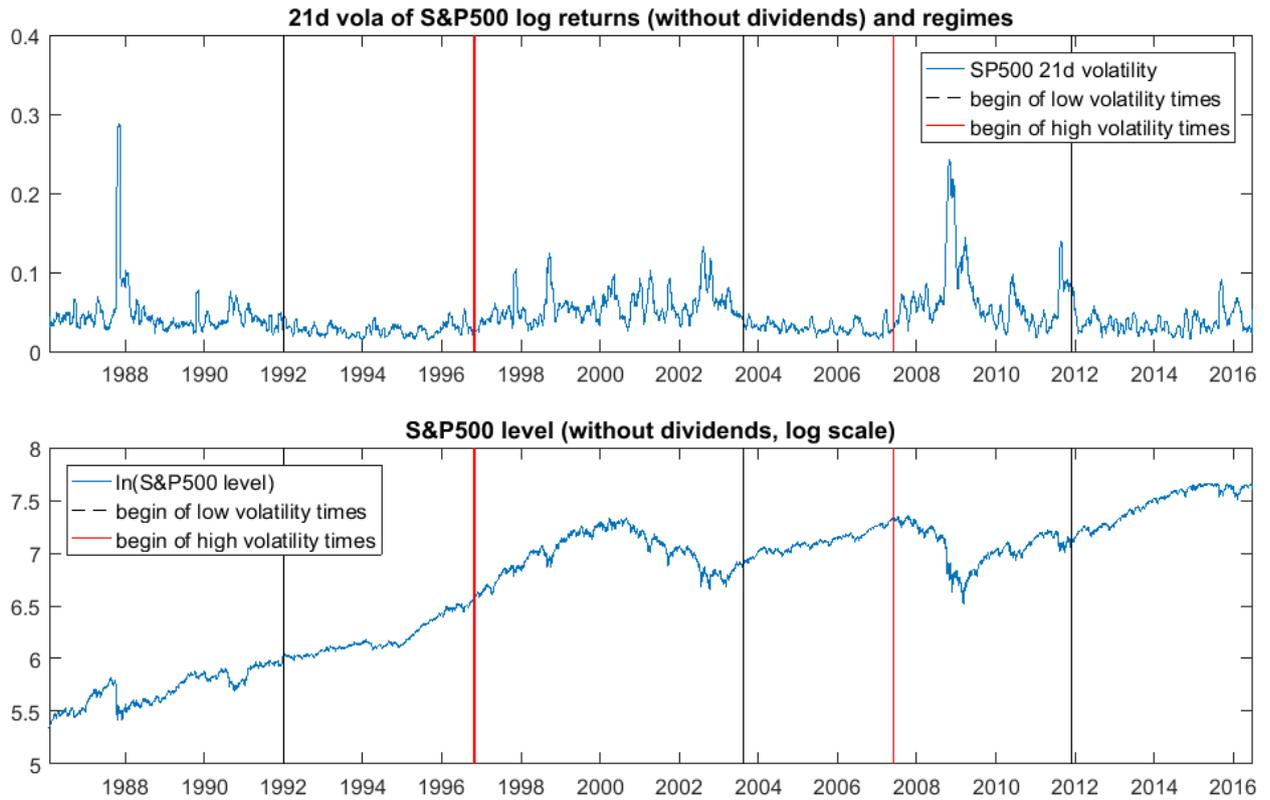


Figure 2: 21 day volatility of S&P 500 log returns and  $\log(\text{S\&P 500 level})$

The figure shows the 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).

Black (red) vertical lines indicate the beginning of a low (high) variance regime.

### 3.4.2 Volatility forecasts of the model

To further assess the model fit, 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. This multi-periods volatility forecast is of interest, because one month (21 trading days) is a typical horizon of interest 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 using (27) up to a point in time

$t$ , and then the model implied variance for  $t + 1$  to  $t + 21$  is calculated using (4).<sup>10</sup> The predicted volatility is calculated as the square root of the predicted variance.<sup>11</sup> realized volatility is calculated as:

$$\sqrt{\sum_{t+1}^{t+21} R_t^2}. \quad (13)$$

Figure 3 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. 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 3. 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, and vice versa. The CP model does match average numbers very well and hence has a lower RMSE, often much lower. This analysis shows that the standard GARCH model does

---

<sup>10</sup>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 6.5). 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 to run 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)).

<sup>11</sup>Strictly speaking, the expected future volatility is not the same as the square root of the expected variance. Unfortunately, no closed form solution for the expected volatility exists. Untabulated numerical simulations show, however, that the two numbers are very similar. Furthermore, the comparison of predicted variances to realized variances gives very similar results. Lastly, one might argue that the true comparison to the used estimate for realized volatility is the expected volatility of returns, i.e. including the risk-free rate and the drift term. Using this measure does have a very small effect on the results, since the daily  $r_t$  and  $\mu \cdot h_t$  are very low compared to daily volatility (order of magnitude of 100).

a poor job in predicting future volatilities over a longer period of time. The model performs particularly badly in times of low variance, but also in the other periods. This indicates that the parameter estimates have the tendency to fit the extreme returns and trade this off with a worse fit for the calm periods.

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 parametrisation 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 extended periods of low variance, 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 expectations exceeded the realization as extremely as shown above for such a long time period. Another way to illustrate this is by the following comparison. The FP GARCH forecast of volatility in calm periods is on average as high as the VIX in the same time period. The VIX is the non-parametric risk-neutral expectation of the future volatility derived from option prices. It is typically significantly higher than the physical expectation, since it includes the variance-risk-premium. Hence, on average the physical volatility forecast should be much below its risk-neutral counterpart given the size of the variance risk premium documented in the literature (e.g. in Carr & Wu (2009)). Several alternative models and also a forward-looking volatility forecast are used in the robustness section.

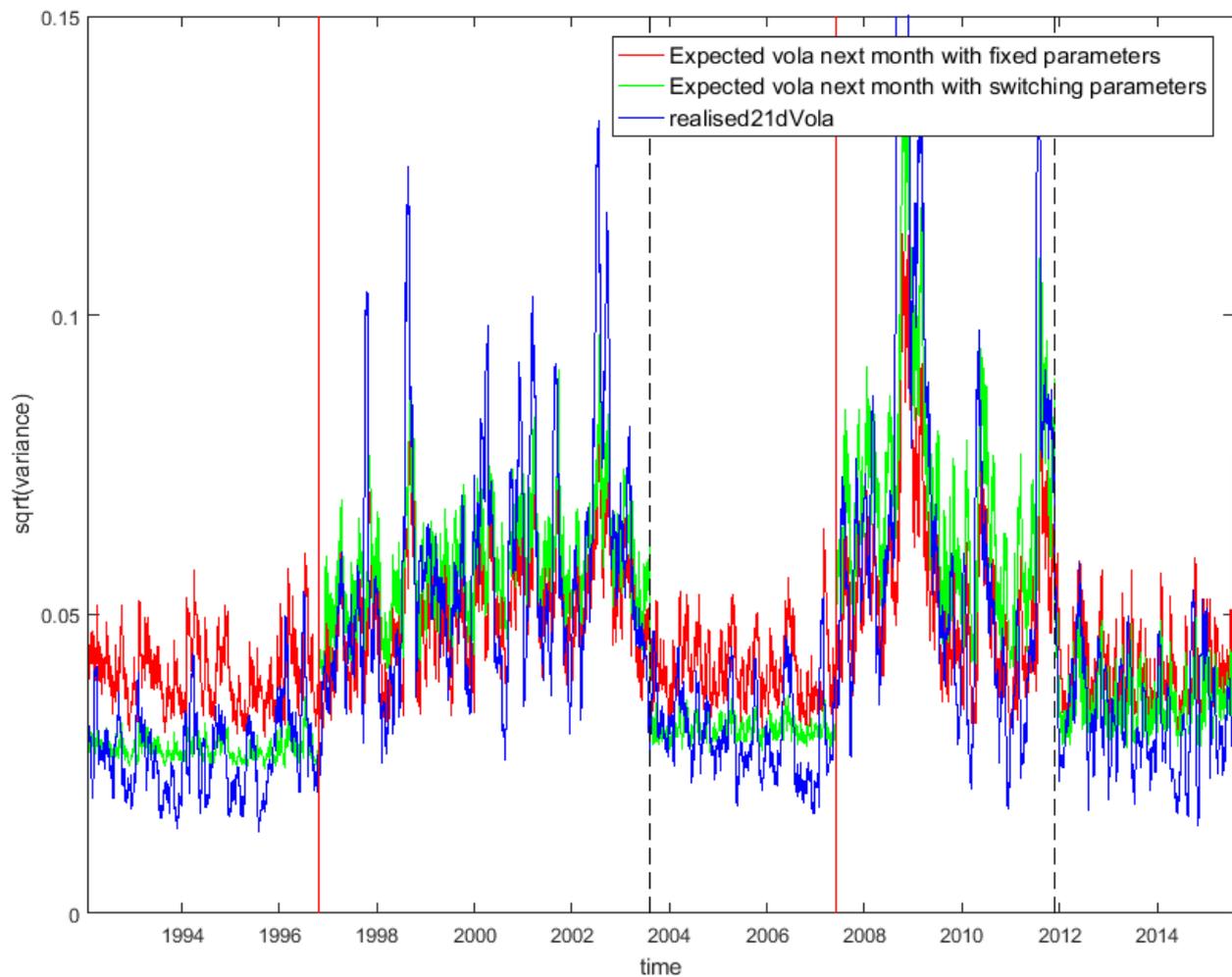


Figure 3: Predicted vs. realized 21 day volatility

The figure shows the 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.

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

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.

## 4 Empirical Pricing Kernels

In this section, I first discuss the option data used in the empirical analysis. The subsequent analysis then focuses on the shape of empirical pricing kernels implied by option data. In particular, the sensitivity of the results to the GARCH model specifications is studied. Subsequently, I analyze in detail the channel how the GARCH models impact the result.

### 4.1 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).<sup>12</sup> 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-ofquarter, 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.<sup>13</sup>

<sup>12</sup>For each time horizon that is analyzed here and in the robustness section, 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).

<sup>13</sup>For the full details on the data cleaning see the Appendix B.

## 4.2 Methodology

The two major quantities that are required to empirically estimate pricing kernels are the risk-neutral and the physical return density. The approach adopted here closely follows the approach used by Christoffersen et al. (2013), which in turn is close to other previous studies. The approach is chosen because it stays as non-parametric as possible, but provides evidence on the conditional density.

The approach to estimate the risk-neutral density is standard and has the following steps.<sup>14</sup> Starting from the entire cross-section of options on a given day, first a fourth-order polynomial for implied Black-Scholes volatility as a function of moneyness is estimated.<sup>15</sup> Using this estimated polynomial, next a grid of implied volatilities corresponding to a dense grid of strikes is calculated. Then, call prices are calculated using the Black-Scholes formula. The risk-neutral interest rate is obtained from OptionMetrics and linearly interpolated. The risk-neutral density can then be calculated using the result of Breeden & Litzenberger (1978):

$$f^*(S_T) = \exp(r\tau) \left[ \frac{\partial^2 C_{BS}(S_t, X, \tau, r, \hat{\sigma}(S_t, X))}{\partial X^2} \right]_{|X=S_T}. \quad (14)$$

Finally, in order to plot the density against log returns rather than future spot prices, the probabilities are transformed. The obtained densities are really conditional because they reflect only option information from a given point in time. Note that here the risk-neutral probabilities are only estimated (and later plotted) where data exists, and the implied volatility curve is not extrapolated. This is chosen since on the one hand, any extrapolation or tail fitting is potentially unreliable, and on the other hand the data on average covers a cumulated probability of 95.5% at the one month horizon and therefore the main results can be shown without any tail probabilities. Refraining from "completing the tails" does not influence the estimation of the risk-neutral probabilities

---

<sup>14</sup>This approach is very close to the one proposed by Figlewski (2010). Numerous studies use a similar approach, that all smooth and fit implied volatility instead of prices. It is generally understood that fitting implied volatility is more reliable than fitting prices directly. Small differences between approaches are in the degree of the polynomial used (typically degree of two or four) and the filters applied to the option price data. Figlewski's suggestion to fit the tails is not considered here, because it is potentially unreliable. The tails are usually hard to fit, and this holds both for the risk-neutral and the physical part. Furthermore, the general shape of the pricing kernel becomes clear even without the tails.

<sup>15</sup>Appendix C contains the details on the calculation of the implied volatility. The results remain unaltered if the implied volatility provided by OptionMetrics is used.

over the range of available option strikes.<sup>16</sup> Also note that the risk-neutral estimates are non-parametric and are not influenced by any assumption of structural breaks.

To provide some information about the behavior of the PK in the right tail, a different approach is adopted. Below, I present the ratio of the cumulative return density in the tail that is not covered by the (filtered) option data. The cumulative risk-neutral return density, denoted  $F^*(S_T)$ , can also be obtained from option prices non-parametrically:

$$1 - F^*(S_T) = -\exp(r\tau) \left[ \frac{\partial C_{BS}(S_t, X, \tau, r, \hat{\sigma}(S_t, X))}{\partial X} \right]_{|X=S_T}. \quad (15)$$

Setting  $S_T$  equal to the highest available strike in the data delivers an estimate for the cumulative risk-neutral probability in the right tail. 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 approach for obtaining the conditional physical density of returns is semi-parametric and has become more popular recently.<sup>17</sup> Many alternative approaches are tested in the robustness section and they deliver very similar results. The benchmark method for the 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. Last, but not least it is only semi-parametric and thereby 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  $\sqrt{h(t, T)}$  expected at the beginning of the month and calculated using (5). This yields a series of

---

<sup>16</sup>The risk-neutral probability is obtained directly from applying (14), and no additional treatment as e.g. kernel fitting or scaling is necessary. Therefore, the standard approach to exclude option prices with very low prices (best bid below 0.5\$) is at least innocuous and probably leads to an increase of the precision of the derivation of the risk-neutral probabilities. This is because the bid-ask spreads of options with very low prices are usually very large, and the mid-price is likely to be not the true price. Including the probably distorted prices would influence the implied volatility interpolation, which would influence prices and this would finally influence the results.

<sup>17</sup>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).

return shocks:

$$Z(t, T) = (R(t, T) - \bar{R}) / \sqrt{h(t, T)}. \quad (16)$$

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

$$\hat{f}(R(t, T)) = \hat{f}\left(\bar{R} + \sqrt{h(t, T)}Z\right). \quad (17)$$

Hence, for each date in the sample a different conditional density is estimated. The difference arises from the conditional volatility expectation, while the shape of the distribution is always the same. For both models the full return time series is used. For the change-point model the volatility forecasting is performed using the parameters of the respective regime.<sup>18</sup>

Since the option data contain several slightly different times to maturity 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.<sup>19</sup> Each option price date is then matched with the correct length of the ‘monthly’ returns.

### 4.3 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. Each pricing kernel is then calculated as the ratio of the current risk-neutral to the physical density, and then this ratio is divided by the risk-free rate.<sup>20</sup> The one month horizon is chosen for the benchmark analysis, since it is the most studied horizon in the literature on empirical pricing kernels and a maturity with very liquid option contracts. The robustness section shows that the results also hold for other typical horizons. Figure 4 shows the natural logarithm of the

---

<sup>18</sup>Here again the probability of switching into a different state is ignored to keep the analysis simpler and more tractable. The robustness section studies the impact of including this.

<sup>19</sup>This is called expected number of trading days here, because there happen to be unexpected closings of the exchanges due to extreme events. Hence, the calculation is not based on the number of actual trading days, but those that could reasonably be expected.

<sup>20</sup>In log-space the division by the risk-free rate is just a small parallel shift of the curve downwards. Therefore, the documented PKs are easily comparable to studies that report just the ratio of risk-neutral to physical probability.

estimated pricing kernels using the HN-GARCH model with fixed parameters. Figure 5 displays the same for the CP-GARCH. The scale of the horizontal axis are log returns. The colouring indicates times with high volatility (red) and times with low volatility (black), as defined in Chapter 5.2.6. 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 behaviour of the PK in the right tail. It can be interpreted as the value of the average PK in the tail. The x value for the CDF ratio is (arbitrarily) chosen as the return of the last traded strike plus 0.013 (0.02) log-return points in times of low variance (high variance).

When comparing the two plots, several points 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 is noted already by Grith et al. (2013), but was never documented systematically nor for such a long time series. However, when the GARCH parameters are not fixed, 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 predominantly 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 two assets, 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 makes a lot of sense 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.

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. 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 are constant. Figure 16 in Appendix G shows the skewness of the risk-neutral return distribution during the last low volatility regime. The skewness clearly decreases over the course of the regime. 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 \$0.50, which is the cut-off point in the data cleaning.<sup>21</sup> 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 1 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, Chapter 4.6 provides an argument that the pricing kernel is upward sloping in times of low volatility at least on average.

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 is biased towards the long-run mean, the GARCH in high volatility times is also biased towards its high long-run mean, which is significantly influenced by the extreme returns. Figure 3 shows that there are also periods with relatively low volatility within these periods. However, the GARCH is 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 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 led 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

---

<sup>21</sup>Cuesdeanu (2016) uses a similar argument in a different setting.

not for all time periods. Furthermore, it becomes clear that the S-shaped kernels are merely a results of the GARCH model with fixed parameters and are then only found in periods of low variance. The robustness of the results to a variety of changes in data and methods is shown below.

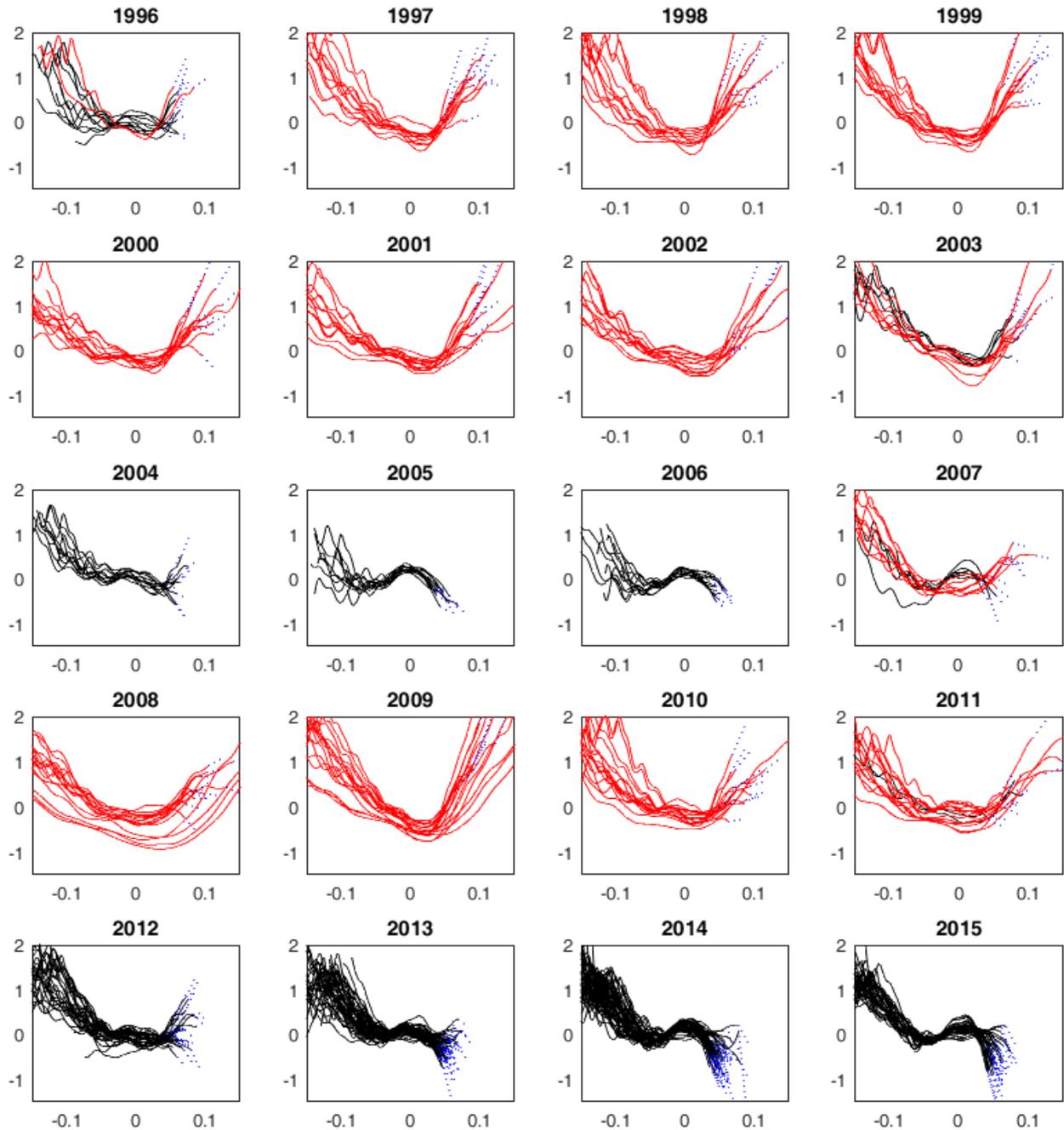


Figure 4: Empirical pricing kernels with fixed parameters in the HN model

The figure shows the natural logarithm of estimated pricing kernels obtained from using the Heston-Nandi model with fixed parameters. Red (black) depicts times with high (low) variance, as defined in Chapter 5.2.6. Log-returns are on the horizontal axis. The horizon is one month. The blue line connects the points, which depict the ratio of the CDFs of the tail, with the corresponding pricing kernels.

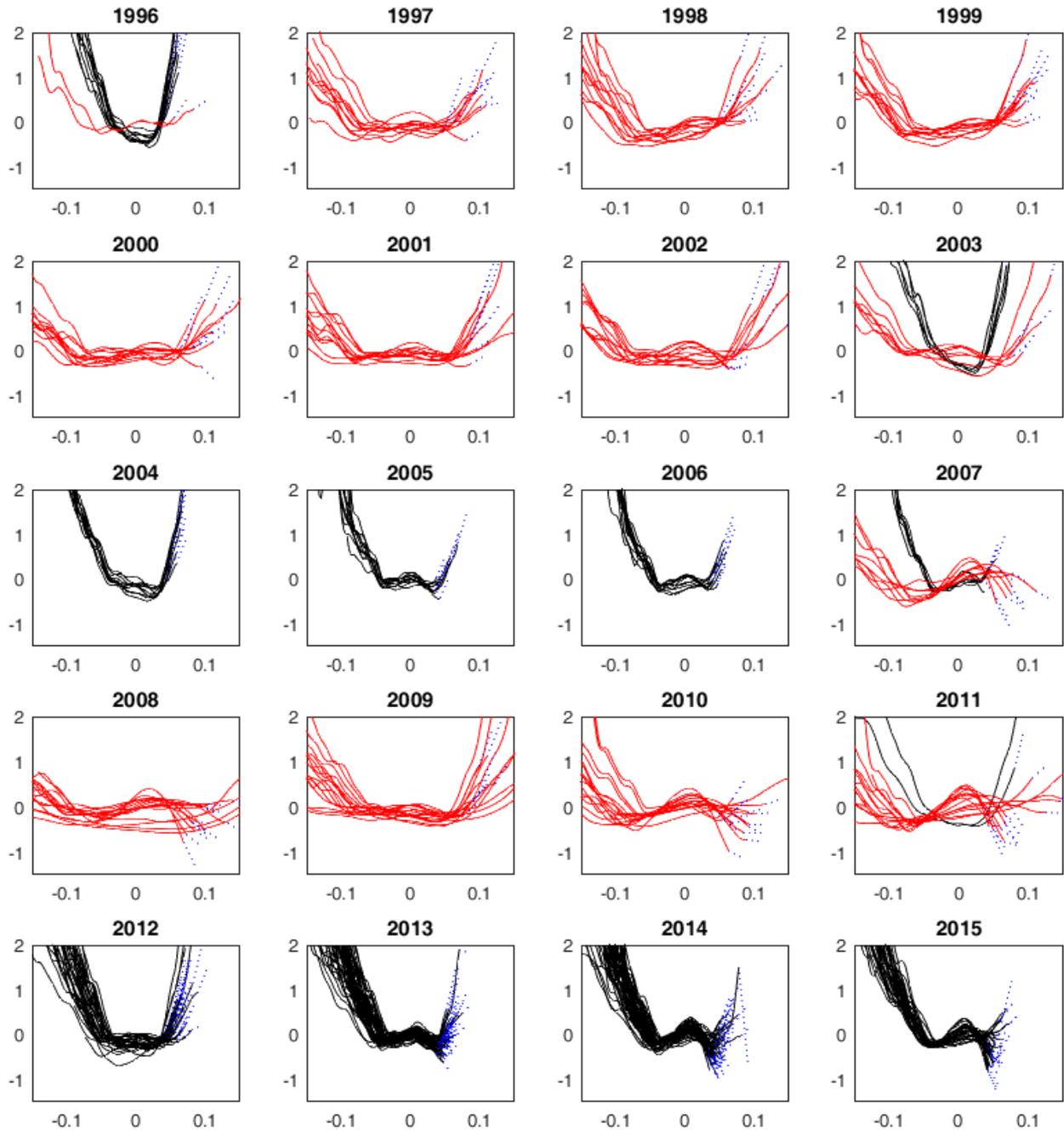


Figure 5: Empirical pricing kernels with CP parameters in the HN model

The figure shows the natural logarithm of estimated pricing kernels obtained from using the Heston-Nandi model with CP parameters. Red (black) depicts times with high (low) variance, as defined in Chapter 5.2.6. Log-returns are on the horizontal axis. The horizon is one month. The blue line connects the points, which depict the ratio of the CDFs of the tail, with the corresponding pricing kernels.

## 4.4 How the volatility estimates 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 and that the filtered volatilities are similar for both FP GARCH and CP-GARCH. The second factor has also surprisingly little influence. One could expect that the different volatility forecasts of the models lead the different distributions of standardized return shocks. But the top left plot in Figure 6 shows that these two distributions are actually very similar. The figure shows the densities of the monthly normalized return shocks (i.e. the  $Z$  from Equ. (16)) for the two models. The first sub-plot shows all return shocks of the full time series, and the following ones only the return shocks of the respective sub-periods. While the regime-specific shock densities are very different, the aggregate, which is used for the empirical study, is very similar.

The last remaining factor, the different volatility forecasts, turn 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 (which are depicted top left in Figure 6) with the conditional volatility forecast. Since the densities are very similar, the key difference is the more realistic volatility forecast. 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. The left downward-sloping part is so steep that it is still downward-sloping despite the fatter tail. However, the lack of probability mass in the center causes the hump in the middle. Finally, the fatter right tail makes the PKs downward-sloping on the right end.

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

Figure 7 summarizes this mechanism graphically. The top plot contains the (similar) shock distributions of the two models and is the same as the top left plot in Figure 6. In the following, the left plot shows the actual data for October 2005, and the right plot shows the data for December 2009. The second row contains the physical return forecasts. They are obtained by multiplying the respective shock density with the corresponding volatility forecast and adding the mean return. 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 third row contains the risk-neutral return density derived from option prices. The last row shows the ratio of the risk-neutral to the physical return density. One can clearly see how the overestimated volatility for the FP model influences the shape of the estimated PK in 2005. In 2009, the underestimation of the volatility makes the pricing kernel steeper, but does not change the estimation qualitatively.

The sub-plots two to six in Figure 6 point to another interesting finding. First, for the FP model, the densities of the respective regimes are substantially different from the one of the full sample. An apparent pattern is that the densities in times of low variance are much tighter, while the reverse is true for the other times. This pattern is barely visible in the respective densities for the CP model. The application of a model with better volatility forecasts, however, 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 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 and the method to obtain the conditional physical density.

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.

The above analysis has shown how the shortcomings of a standard GARCH bias the estimation of the physical return density and how this shapes the estimated pricing kernels. To be able to observe this effect, it is necessary to use a long time series of both

returns and option prices, which cover different market phases with both high and low volatility.

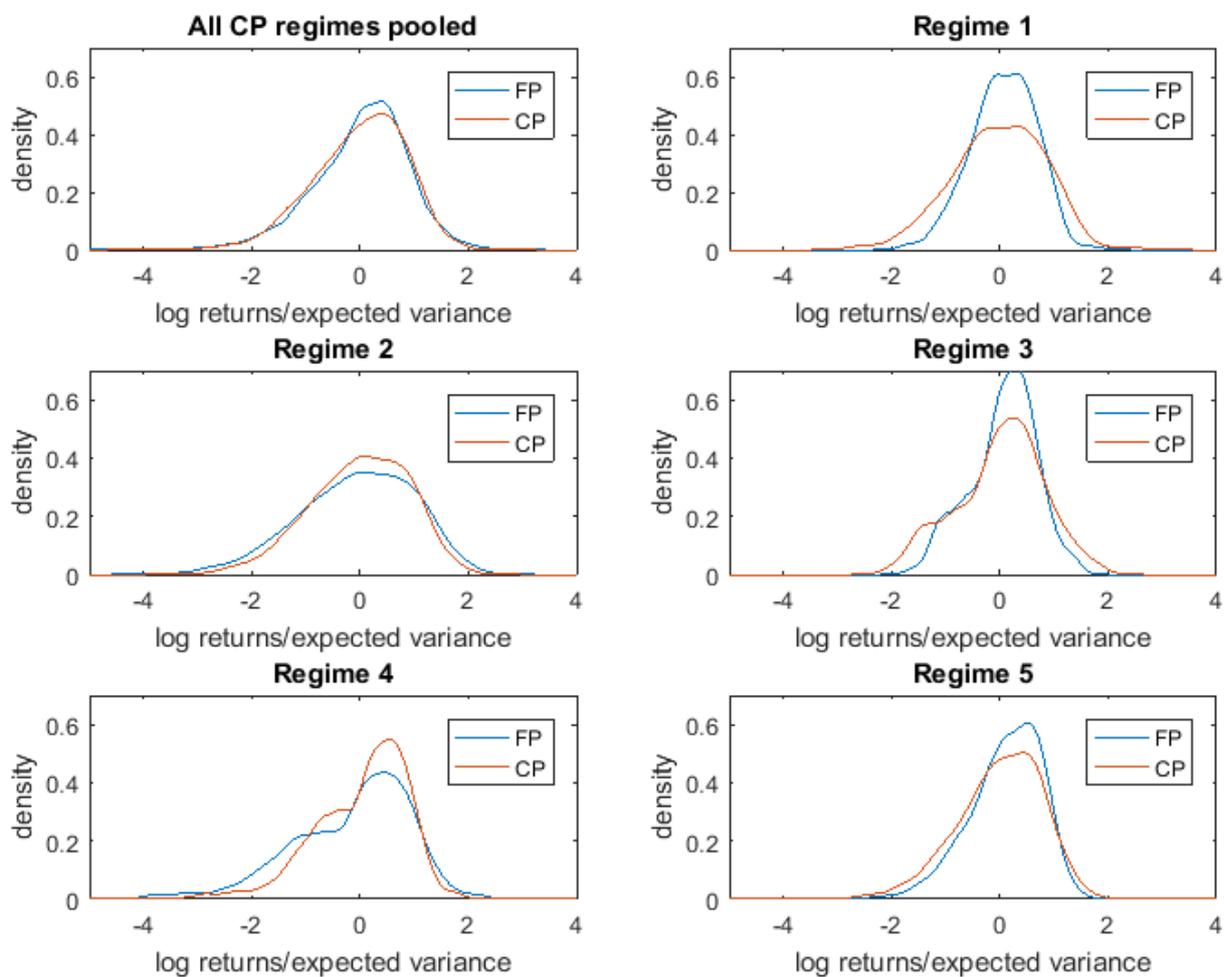


Figure 6: Monthly return shock densities

The figure shows the estimated monthly return shock density (21 days) calculated as in Equ. (16) 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.

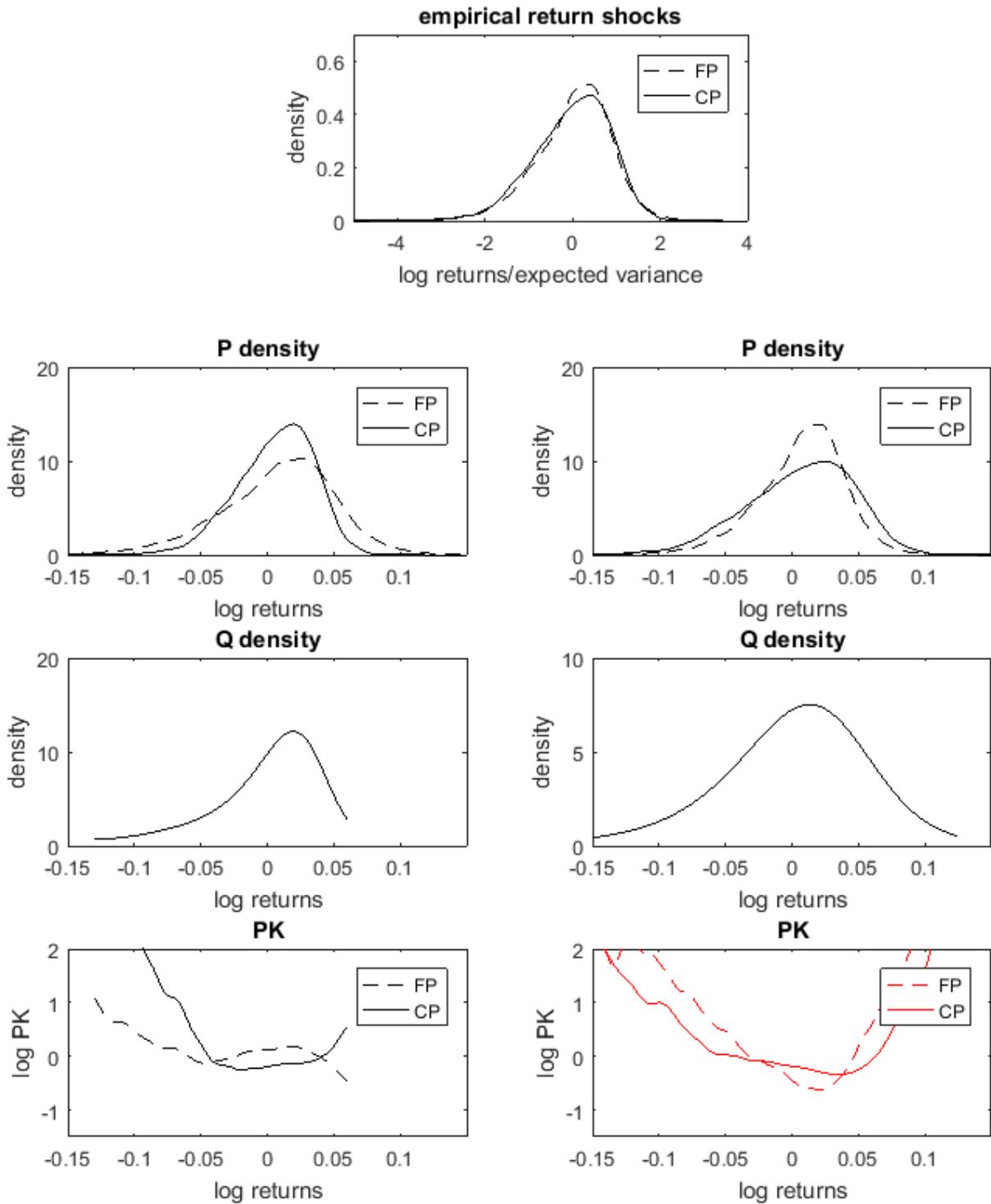


Figure 7: Main mechanism how volatility estimates drive the results

The top row contains the monthly return shock densities of the FP GARCH and CP-GARCH model for the full time series. From the second to the last row, the left plots show data from October 2005<sup>31</sup> and the right plots show data for December 2009. The displays row shows the estimated conditional monthly return density at the given date, while the third row displays the risk-neutral monthly return density at the same date. The last row illustrates the estimated pricing kernels, which are the ratio of the the respective densities.

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.

## 4.5 Euler equation errors

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 unconditional Euler equation errors to zero.

In the following, the unconditional Euler equation error is defined as:

$$e_R = E[M_{t+1}R_{t+1}] - 1, \quad (18)$$

where  $R_t$  is the return of the index and  $M_t$  is the (empirical) pricing kernel, i.e. the projection of the stochastic discount factor on the returns of the index.<sup>22</sup> A unconditional Euler equation error of zero is a minimum condition that any candidate pricing kernel should fulfill. To avoid unreliable extrapolation, I exclude observations where the realisation is outside the domain where the empirical PK is defined.

Table 4 shows the results. The first two columns show the Euler equation errors of the empirical PKs from Section 4.3, both with fixed and switching parameters. The other columns show the errors for two methods from the robustness section. In particular, columns three and four show the errors 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. The last column shows the results where the realised volatility model of Corsi (2009) is used for forecasting volatility. The first line displays the Euler equation

<sup>22</sup>Note that the nature of the projection makes it unlikely that the empirical pricing kernel estimates perform well with other test assets. This is because any cross-sectional variation that is additional to the markets (as e.g. value or size) should by definition be orthogonal to the market risk.

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, 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, as defined in Chapter 5.2.6.

Table 4: Euler equation errors of empirical pricing kernels

	FP empirical	CP empirical	FP GARCH	CP GARCH	Real. Vol
Error	0.067	0.003	0.054	-0.001	0.023
90% Confidence	[0.032,0.106]	[-0.03,0.043]	[0.027,0.084]	[-0.021,0.02]	[-0.014,0.066]
Error low vol	0.044	-0.007	0.087	-0.004	0.019
Error high vol	0.103	0.02	0.006	0.005	0.032

The table shows the unconditional Euler equation errors, together with the 90% confidence interval of the error and the Euler equation errors for the two sub-samples of times with high and low volatility (as defined in Chapter 5.2.6). The 90 % confidence intervals for the errors are obtained from 25,000 bootstrap draws from the sample of errors (in square brackets). The first (second) column shows the errors for the empirical pricing kernels with fixed (switching) parameters from Section 4.3. The third and fourth column are for the model where the physical return density is obtained from simulating the GARCH model (see also robustness Section 6.1). The last column is for the model where the volatility forecasts are obtained from the realised volatility model of Corsi (2009) (see also robustness Section 6.3).

The main result is that the standard approaches which use a GARCH model with fixed parameters (column 1 and 3) 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 reveal 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 differently in these sub-periods, although no clear pattern emerges.

## 4.6 Verifying U-shape using option returns

The results on the estimated empirical pricing kernels in Section 4.3 are partly 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. Unfortunately, it is not

possible to test this statistically, or to calculate confidence bounds.<sup>23</sup> 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 after 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. Note that the method is fully non-parametric, but only speak about the average, i.e. unconditional pricing kernel.

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 5 (in percent). The confidence bounds are bootstrapped as in Bakshi et al. (2010). The target moneyness ranges from 0%, 1%,...,7%, and the sorting ends at 7% 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 documented returns in the sample 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% 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 4% and 7% bin are significantly different from zero. However, all except one pair-wise differences between groups are positive and most of them are statistically significant. In sum, the results are strongly consistent with a U-shaped pricing kernel. They show that also in times of low volatility, at least on average, 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. This helps to clarify the ambiguity of the empirical results above.

---

<sup>23</sup>Beare & Schmidt (2016) and (Härdle et al. 2014) develop a method for a statistical test and for confidence bounds, respectively. However, they both use very restrictive assumptions, which are not compatible with my approach.

Table 5: Average returns of S&P500 index call options in times of low volatility

% OTM	1%	2%	3%	4%	5%	6%	7%
Average [%]	13.0	7.2	-6.4	-26.0	-25.3	-27.0	-51.1
90% Conf.	[-0.2,26.4]	[-8.5,23.6]	[-25.4,13.7]	[-49.1,-0.8]	[-58.5,12.3]	[-68,22]	[-88,-4.9]
	0% - 1%	1% - 2%	2% - 3%	3% - 4%	4% - 5%	5% - 6%	6% - 7%
Differences:	2.4	5.8	13.6	19.7	-0.7	1.7	24.1
90% Conf.	[-1.1,5.8]	[0.9,10.7]	[6.3,20.6]	[8.4,30.6]	[-9,19.1]	[-8.2,19.7]	[6.8,64]

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 Chapter 5.2.6. Call returns are calculated using settlement prices. Moneyness is calculated as: strike/(index level at price date). The average time to maturity is 30 days, and always in the interval [29,32]. The 90 % confidence intervals for average returns are obtained from 25,000 bootstrap draws from the sample of option returns (in square brackets). To test for difference in average returns across strikes, first 25,000 pairwise bootstrap samples of returns to options are drawn.

## 5 Asset Pricing Implications

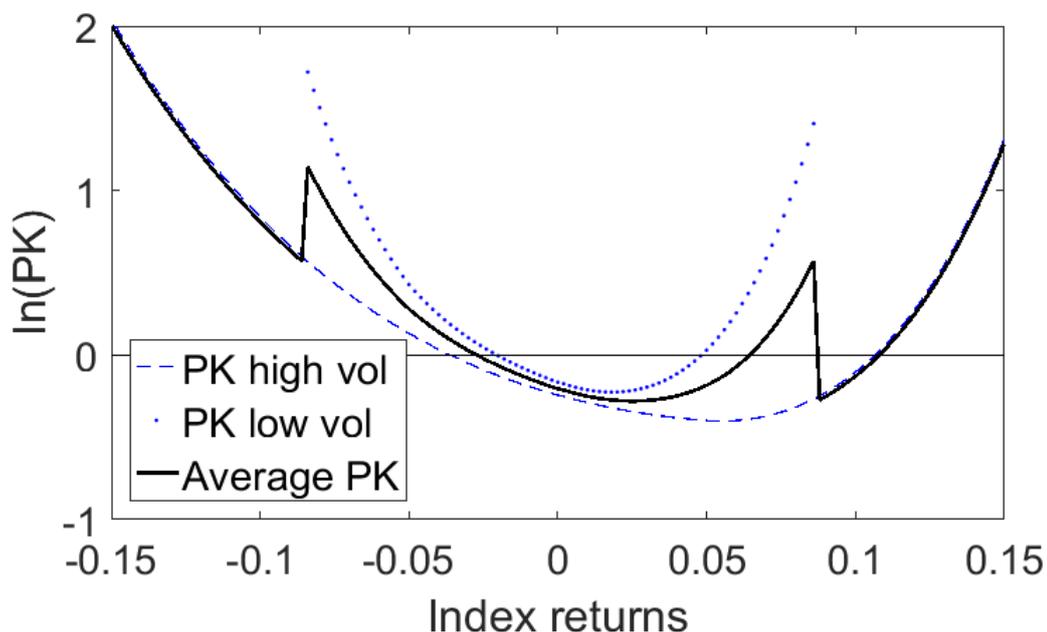
### 5.1 Implications for the unconditional pricing kernel

The results presented so far have important implications for the unconditional pricing kernel. This is on the one hand relevant since there are several papers that explicitly or implicitly estimate the unconditional pricing kernel. On the other hand, this is of interest for economic models that imply only an unconditional pricing kernel.

One of the major empirical results above is that the pricing kernel is always U-shaped and furthermore its wideness is strongly time-varying. The latter behavior is a necessary property, since the PK has to adjust to the time-varying volatility of the physical return distribution such that it prices at least the index and the risk-free rate correctly. If one estimates an unconditional pricing kernel, this is in most methods analogous to averaging all conditional kernels. However, when the wideness of the pricing kernel is time-varying, this leads to misleading results. Figure 8 illustrates this with a simple example, which consists of two stylized U-shaped PKs (dashed lines) and their average (solid line). The wider PK represents high volatility times (dashed line) and the tighter one low volatility times (dotted line), and both are roughly in line with the empirical results. The tighter one, however, ends where the option data typically ends in calm periods. The option

data ends where option prices get below typical thresholds of e.g. 0.50\$ or are not traded. Unavailable data is a restriction that no method can circumvent.<sup>24</sup> If one calculates the average of these two PKs where they are defined, one obtains an unconditional PK, that has puzzling shapes and “humps” in unexpected areas. If one generalizes this idea further and allows for more PKs with varying wideness, the “humps” would become smoother, but the result is still hard to interpret. In sum, in the presence of U-shaped PKs, an unconditional pricing kernel is hard to detect or even does not exist. This in turn emphasizes the relevance to study the conditional pricing kernel.

Figure 8: Average pricing kernel in the presence of U-shaped pricing kernels



The figure shows two stylized U-shaped pricing kernels, which are in line with the data, and their average. The tighter PK represents times of low variance (dotted line), while the wider PK represents times of high variance (dashed line). The solid line is the average of the two PKs where they are defined.

<sup>24</sup>Even if the data would be available or extrapolatable, the pricing kernel is very likely to either increase towards infinity, or drop to zero, if the risk-neutral probabilities would be zero. Either case would blur the average pricing kernel.

## 5.2 A variance-dependent pricing kernel

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 finding. This section shows that the structural breaks are necessary to make a variance-dependent stochastic discount factor fit the empirical results. For this, I first replicate the results of Christoffersen et al. (2013) and then introduce the structural breaks.

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 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. As pointed out by Christoffersen et al. (2013), such an exercise has been attempted only by a very limited number of studies.<sup>25</sup> Moreover, note that it is possible and frequently done to fit option prices and at the same time ignore the fit to returns completely by only fitting the volatility state variable.

### 5.2.1 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:

$$M(t) = M(0) \left( \frac{S_t}{S_0} \right)^\phi \exp \left( \delta t + \eta \sum_{s=1}^t h_s + \xi (h_{t+1} - h_1) \right), \quad (19)$$

where the parameters  $\delta$  and  $\eta$  govern the time preference, while  $\phi$  and  $\xi$  govern the respective aversion to equity and variance risk.<sup>26</sup> With  $\xi = 0$  the variance risk premium is zero, and with  $\xi > 0$  the variance risk premium is negative.<sup>27</sup> With  $\phi > 0$  and  $\xi > 0$ , the SDF is monotonically decreasing in returns and monotonically increasing in variance.

---

<sup>25</sup>For a discussion of the existing literature in this area see Christoffersen et al. (2013), beginning of their Chapter 3.

<sup>26</sup>When variance is constant, (19) collapses to the power utility from Rubinstein (1976) and the Black-Scholes model.

<sup>27</sup>For an discussion of the implications and differences of the SDF with and without a variance risk premium see Christoffersen et al. (2013).

The projection (pricing kernel) of the SDF on the return space 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 (19), the risk-neutral dynamics for the HN GARCH model are:

$$\begin{aligned} \ln\left(\frac{S_t}{S_{t-1}}\right) &= r - \frac{1}{2}h_t^* + \sqrt{h_t^*}z_t^*, \\ h_t^* &= \omega^* + \beta h_{t-1}^* + \alpha^* \left(z_{t-1}^* - \gamma^* \sqrt{h_{t-1}^*}\right)^2, \end{aligned} \tag{20}$$

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

$$\begin{aligned} h_t^* &= \frac{h_t}{1 - 2\alpha\xi}, \\ \omega^* &= \frac{\omega}{1 - 2\alpha\xi}, \\ \alpha^* &= \frac{\alpha}{(1 - 2\alpha\xi)^2}, \\ \gamma^* &= \gamma - \phi. \end{aligned} \tag{21}$$

For the proof, see Appendix B of Christoffersen et al. (2013).

The risk-neutral dynamics are different from the physical dynamics and the transition from one to the other is described by the stochastic discount factor. For a given set of physical parameters, the parameter  $\xi$  governs the transition from the physical to the risk-neutral parameters.

### 5.2.2 Data

The data used here is similar to the one above. Again, 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. For the full details on the data cleaning see the Appendix B. This results in 15,171 option prices with a maturity between 17 and 53 days.

### 5.2.3 Joint likelihood function

First, the joint likelihood function is presented, which consists of an option-based component and a return-based component. The conditional density function of the daily returns is normal. For the HN-GARCH model:

$$f(R_t|h(t)) = \frac{1}{\sqrt{2\pi h_t}} \exp\left(-\frac{(R_t - r_t - (\mu - \frac{1}{2})h_t)^2}{2h_t}\right). \quad (22)$$

The return log likelihood is:

$$\ln L^R = -\frac{1}{2} \sum_{t=1}^T \left\{ \ln(2\pi h_t) + \left( R_t - r_t - \left( \mu - \frac{1}{2} \right) h_t \right)^2 / h_t \right\}. \quad (23)$$

For the likelihood of the option prices, first define the Black-Scholes Vega (BSV) weighted option valuation errors as:

$$\varepsilon_i = \left( C_i^{Mkt} - C_i^{Mod} \right) / BSV_i^{Mkt}, \quad (24)$$

where  $C_i^{Mkt}$  denotes the market price of the  $i^{th}$  option,  $C_i^{Mod}$  denotes the model price, and  $BSV_i^{Mkt}$  denotes the Black-Scholes vega of the option at the market-implied level of volatility. Under the assumption of i.i.d. normal pricing errors, the option log likelihood reads:

$$\ln L^O = -\frac{1}{2} \sum_{i=1}^N \left\{ \ln(2\pi) + \ln(s_\varepsilon^2) + \varepsilon_i^2 / s_\varepsilon^2 \right\}, \quad (25)$$

where  $s_\varepsilon^2$  is estimated using the sample analog  $\hat{s}_\varepsilon^2 = \frac{1}{N} \sum_{i=1}^N \varepsilon_i^2$ .

The joint optimisation problem is now:

$$\max_{\Theta, \Theta^*} \ln w^O L^O + \ln L^R, \quad (26)$$

where  $\Theta = \{\omega, \alpha, \beta, \gamma, \mu, \xi\}$  are the physical parameters and  $\Theta^*$  the risk-neutral parameters that are mapped from  $\Theta$  using (21). The physical variance is filtered from returns using:

$$\begin{aligned}
h_t &= \omega + \beta h_{t-1} + \alpha \left( z_{t-1} - \gamma \sqrt{h_{t-1}} \right)^2, \text{ where} \\
z_t &= \left[ R_t - r - \left( \mu - \frac{1}{2} \right) h_t \right] / \sqrt{h_t}, \\
h_0 &= \frac{\omega + \alpha}{1 - \beta - \alpha \gamma^2},
\end{aligned} \tag{27}$$

and the risk-neutral variance is computed using:

$$h_t^* = h_t / (1 - 2\alpha\xi).$$

The risk-free rate is obtained from OptionMetrics, the term structure is interpolated linearly and used both for option pricing and return filtering.

With this set-up, the estimation fits the physical and risk-neutral dynamics jointly. The transition between the two is governed by parametrically specified pricing kernel. The parameters of the pricing kernel are estimated simultaneously with the parameters of the dynamics.

#### 5.2.4 Identification of breaks points

Theoretically it would be possible to perform a maximum-likelihood estimation of the change-point version of the estimation that includes option data. However, this is practically infeasible due to the computational burden. Therefore, I use the breaks identified in the estimation in Chapter 3, and estimate the model separately in each regime for the change-point version. This is equivalent of treating each period as a separate sample.

#### 5.2.5 Model fit and properties

Table 6 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 regimes of high and low volatility. The risk-neutral volatility is higher than the physical volatility.

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 (lower AIC/BIC are better). 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 5.3

### 5.2.6 Model implied pricing kernels

Figure 9 and 10 present the model implied pricing kernels. The first one contains the PKS for the GARCH with fixed parameters, and the second for the CP-GARCH. Each plotted line is obtained by first simulating 100,000 paths under the physical measure using the parameter estimates from Table 6, and then calculating the stochastic discount factor and its projection on the index return. The comparison of the two 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 very well. Second, the model implied PKS of the standard GARCH have a very similar shape across time, and there is little variation in their wideness. The empirical kernels, however, are strongly time-varying. The CP version on the other hand exhibits the same strong time variation in their wideness, and fits the empirical counterpart well.

Table 6: Estimation Results of the Joint Estimation of the HN-GARCH model

Physical parameters	FP HN-GARCH		CP-HN-GARCH			
	'92-15	'92-'97	'97-'03	'03-'07	'07-'12	'12-'15
$\omega$	-1.81E-06	4.74E-06	-2.56E-06	3.52E-06	-2.34E-06	1.03E-06
$\alpha$	3.24E-06	2.19E-06	5.99E-06	2.45E-06	4.67E-06	1.67E-06
$\beta$	0.877	0.380	0.891	0.550	0.866	0.463
$\gamma$	184.1	453.6	114.1	349.8	151.5	524.4
$\mu$	3.396	8.736	1.323	8.835	0.808	9.736
$p_{jj}$		0.99918	0.99941	0.99898	0.99911	1
Risk-neutral param.						
$\omega^*$	-2.06E-06	7.77E-06	-3.30E-06	5.74E-06	-2.85E-06	1.30E-06
$\alpha^*$	4.23E-06	5.88E-06	9.97E-06	6.50E-06	6.96E-06	2.64E-06
$\beta^*$	0.877	0.380	0.891	0.550	0.866	0.463
$\gamma^*$	164.3	282.1	89.6	220.3	124.8	424.5
$\xi$	19185	89267	18727	78927	19384	61690
Properties						
$\beta + \alpha\gamma^2$	0.9864	0.830	0.9689	0.8496	0.9736	0.9213
$\beta^* + \alpha^*\gamma^{*2}$	0.9906	0.8484	0.9710	0.8655	0.9749	0.9387
Long-run volatility	0.163	0.101	0.167	0.100	0.149	0.093
Long-run volatility*	0.240	0.151	0.240	0.151	0.203	0.127
Log-likelihood						
Total	31597.7	33061.3				
From returns	19492.5	19530.1				
From options	12105.1	13531.2				
AIC	-63183.4	-66062.7				
BIC	-63139.0	-65840.7				

Parameter estimates are obtained by optimizing the likelihood on returns and options jointly. Parameters are daily, long run volatility is calculated as  $\sqrt{\text{long-run variance} \cdot 252}$ . 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 ( $0 \leq \alpha < 1$ ,  $0 \leq \beta < 1$ ,  $\alpha\gamma^2 + \beta < 1$ ,  $-\alpha < \omega$ ). 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.

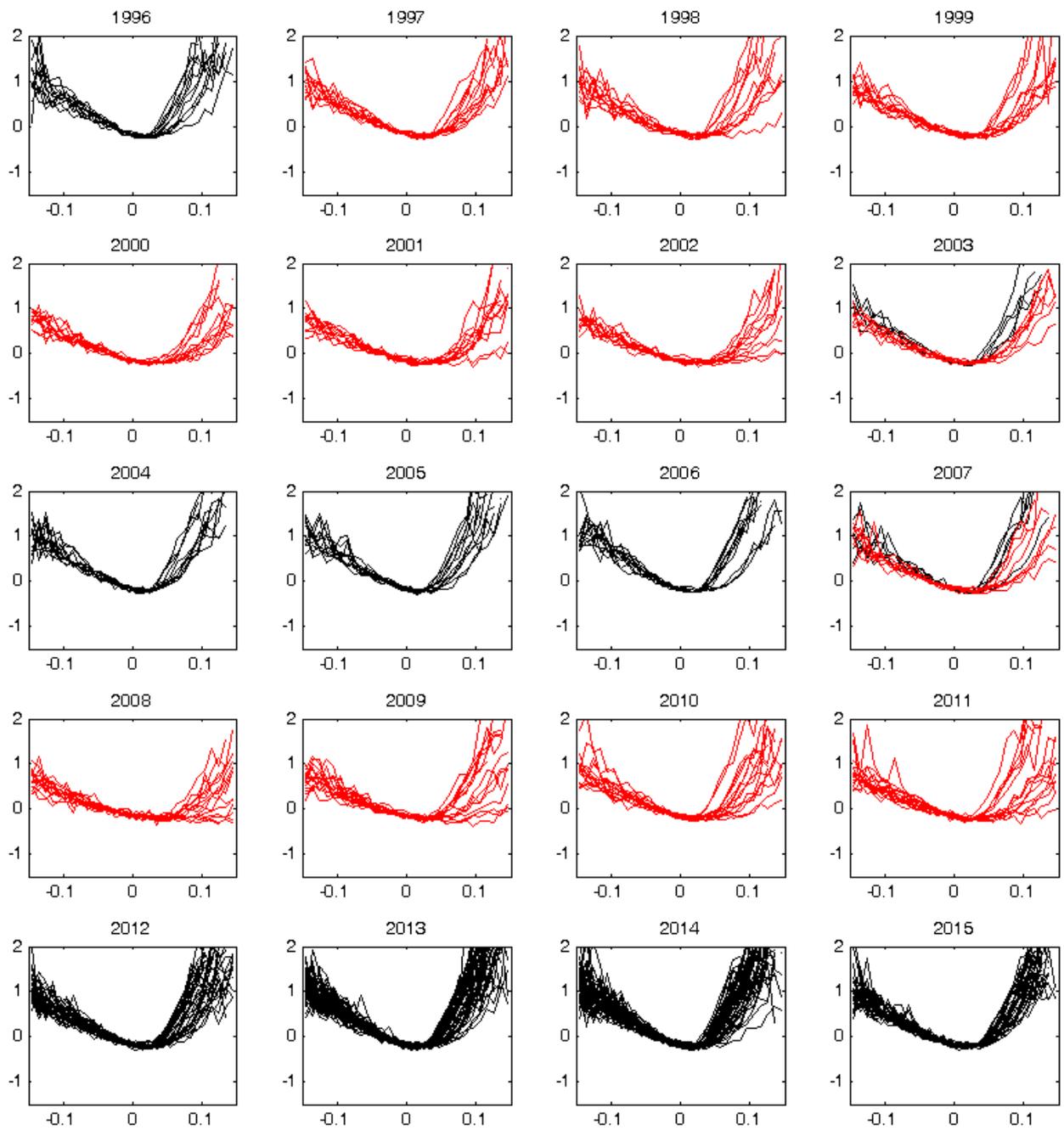


Figure 9: Theoretical pricing kernels with fixed parameters in the HN model

The figure shows the theoretical pricing kernel in the Heston-Nandi GARCH model with fixed parameters. Red (black) depicts times with high (low) variance, as defined in Ch. 5.2.6. Log-returns are on the horizontal axis. The horizon is one month.

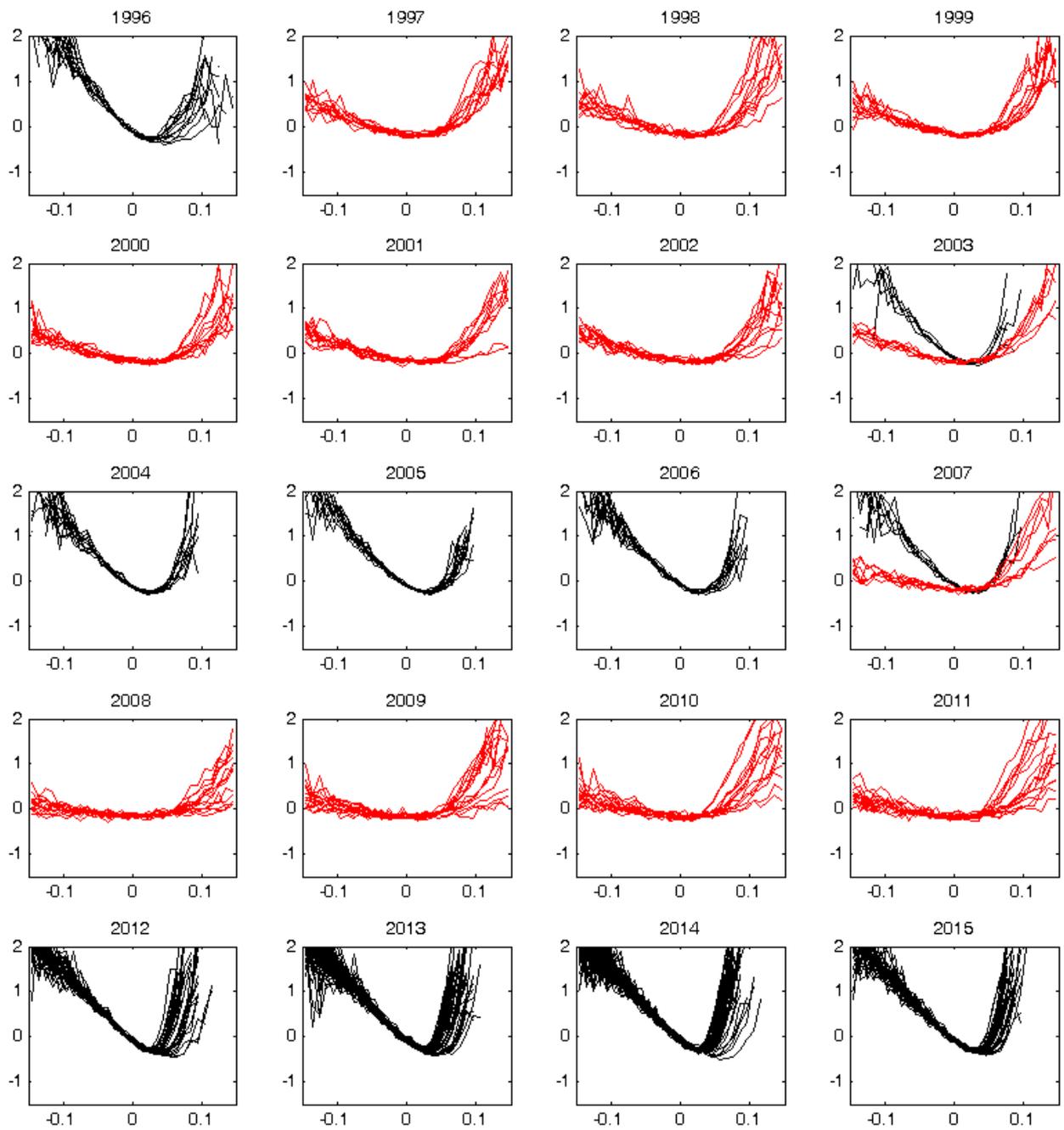


Figure 10: Theoretical pricing kernels with CP parameters in the HN model

The figure shows the theoretical pricing kernel in the Heston-Nandi GARCH model with CP parameters. Red (black) depicts times with high (low) variance, as defined in Ch. 5.2.6. Log-returns are on the horizontal axis. The horizon is one month.

### 5.3 Option pricing fit

The variance-dependent pricing kernel has a good fit to the estimated empirical pricing kernels when the breaks are included. Part of the better fit is that the change-point physical dynamics match the volatility regimes better, as discussed above. In the estimation, the theoretical pricing kernel is fitted both to the physical and risk-neutral dynamics jointly. This section analyzes how the option pricing part contributes to the model performance.

In Table 6, the estimated higher likelihood from option prices of the CP model relative to the FP model already indicates that the pricing error is reduced significantly by the breaks. Further insights can be obtained from the time series pattern of the pricing errors. Figure 11 displays the time series of the option pricing errors of the model. At each date, the vega weighted pricing errors are not squared, but averaged. Pricing errors are market prices minus model prices, hence when the pricing error is negative, the model overprices the option, and vice versa. The used parameters are the risk-neutral parameters from Table 6. 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 underpriced by the model. The pricing errors of the CP model on the contrary do not display any systematic pattern of over- or underpricing. They also have lower time-clustering of the pricing errors. These findings suggest that the FP GARCH option pricing model has a systematic bias in its prices.

The reason for this bias is most likely the same bias in forecasted volatility that was extensively 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 overpricing and underpricing, respectively.

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 results show the structural breaks are important for both the physical and risk-neutral dynamics. In sum, they are equally relevant for the model variance-dependent

pricing kernel to match the data.

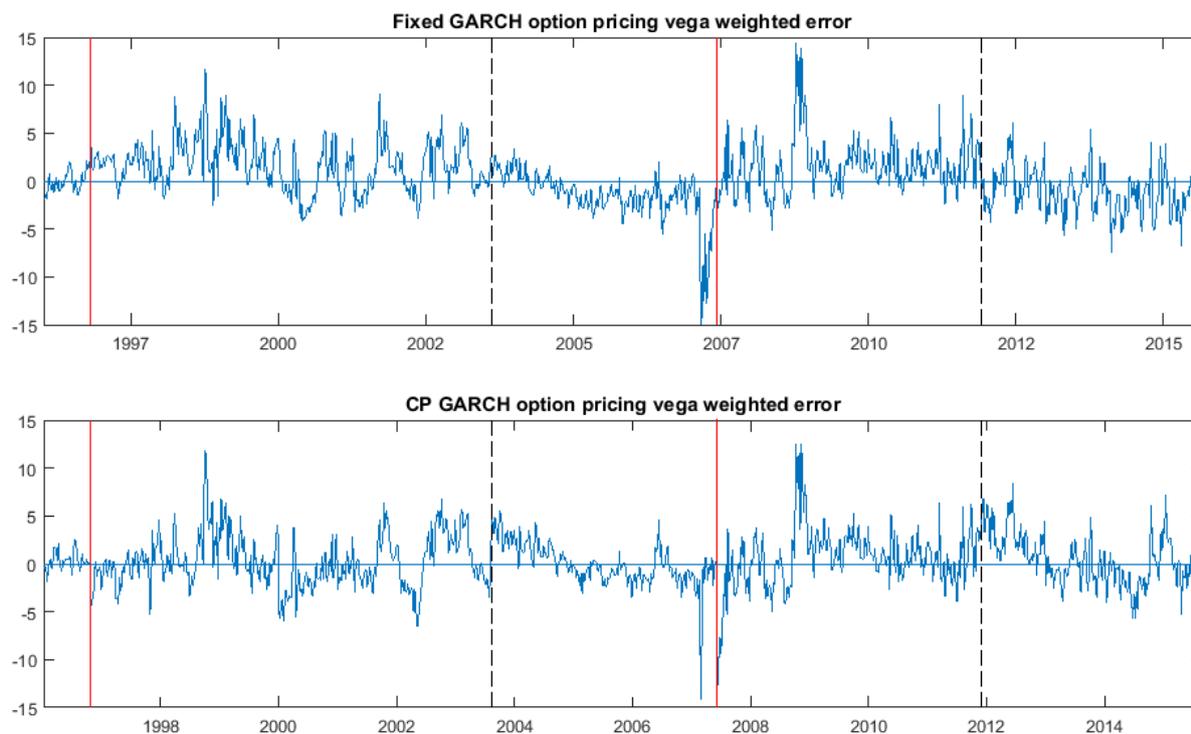


Figure 11: Time series of average option pricing errors

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

## 6 Robustness

### 6.1 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.<sup>28</sup> In the following, this approach is adopted and implemented once using the FP GARCH estimates and once using the

<sup>28</sup>e.g. Rosenberg & Engle (2002), Barone-Adesi et al. (2008) Liu et al. (2009), Barone-Adesi & Dall'o (2012), Cuesdeanu & Jackwerth (2016), Cuesdeanu (2016) .

CP-GARCH estimates as in Table 1. The physical expectation is obtained from a kernel estimation applied to the monthly returns from 10,000 simulated paths of the HN-GARCH model with daily normal innovations starting from the ex ante filtered variance. Figure 12 and 13 show the results. In the first plot, one observes the typical pattern: 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.

In some low volatility time periods the estimated PKs now have a region where they are non-decreasing, but not increasing anymore. The previous observation that the PK estimates become less U-shaped towards the end of a regime carries over also to this analysis. Presumably, the fully conditional risk-neutral density adjusts more quickly to changes in the shape of the distribution, while the physical estimate can not incorporate this. A systematic analysis of the source of the different shapes as in the benchmark case is not possible. Since the shape of the returns density depends on the interplay of all parameters, the identification of the driving force is not possible. However, the results above strongly suggest that the reason is the biased volatility forecast of the GARCH with fixed parameters. Since the forecasted volatility studied in 5.2.6 is the expected sum of the future path, this bias will be present in the simulated density too.

Furthermore, it appears that the GARCH models often do not have a good fit to the empirical returns. Figure 17 in the Appendix G illustrates this by comparing the return densities obtained from the GARCH models with the empirical returns. For the full time series, a long simulation (1 million days) with the FP GARCH model is used to calculate monthly returns. This should ideally converge to the full time series of the data. The top left figure shows that the FP GARCH model on average has too much probability mass in the tails and too little in the center. This generates the typical S-shape. However, this is only true in the average, and does not convey information on when and how this effect is present conditionally, since the true conditional distribution is unobservable.

For the regimes, the CP model is also simulated for 1 million days and then monthly returns are calculated. For the FP version, 100,000 paths are simulated with the fixed parameters, but starting from the long-run variance of the CP model, which is close to the average variance of the regimes. The comparison only within one regime should be taken with a grain of salt, because it is not clear whether the ex post selected sample is what ex ante should be expected. In any case it becomes clear that the GARCH model does not have a great fit to the observed data, and the fixed parameter version less so

than the CP version.

Lastly, the results in Section 4.5 show that the estimated PKs that 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.

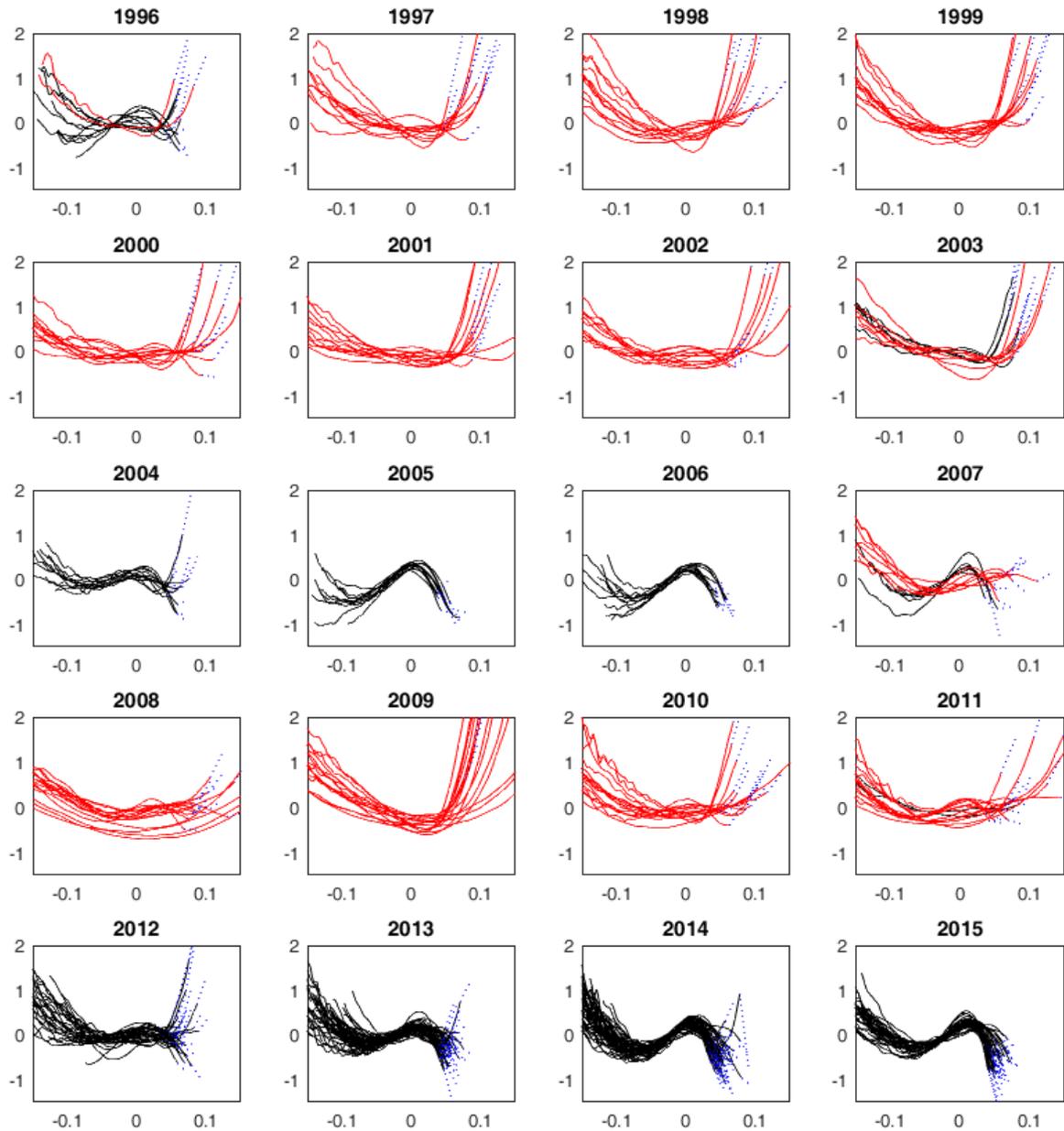


Figure 12: Empirical pricing kernels with fixed parameters in the HN model with simulated densities

The figure shows the natural logarithm of estimated pricing kernels obtained from using the Heston-Nandi GARCH model with FP parameters. The physical density is obtained from the simulation of the GARCH model. Red (black) depicts times with high (low) variance. Log-returns are on the horizontal axis. The horizon is one month. The blue line connects the points, which depict the ratio of the CDFs of the tail, with the corresponding pricing kernels.

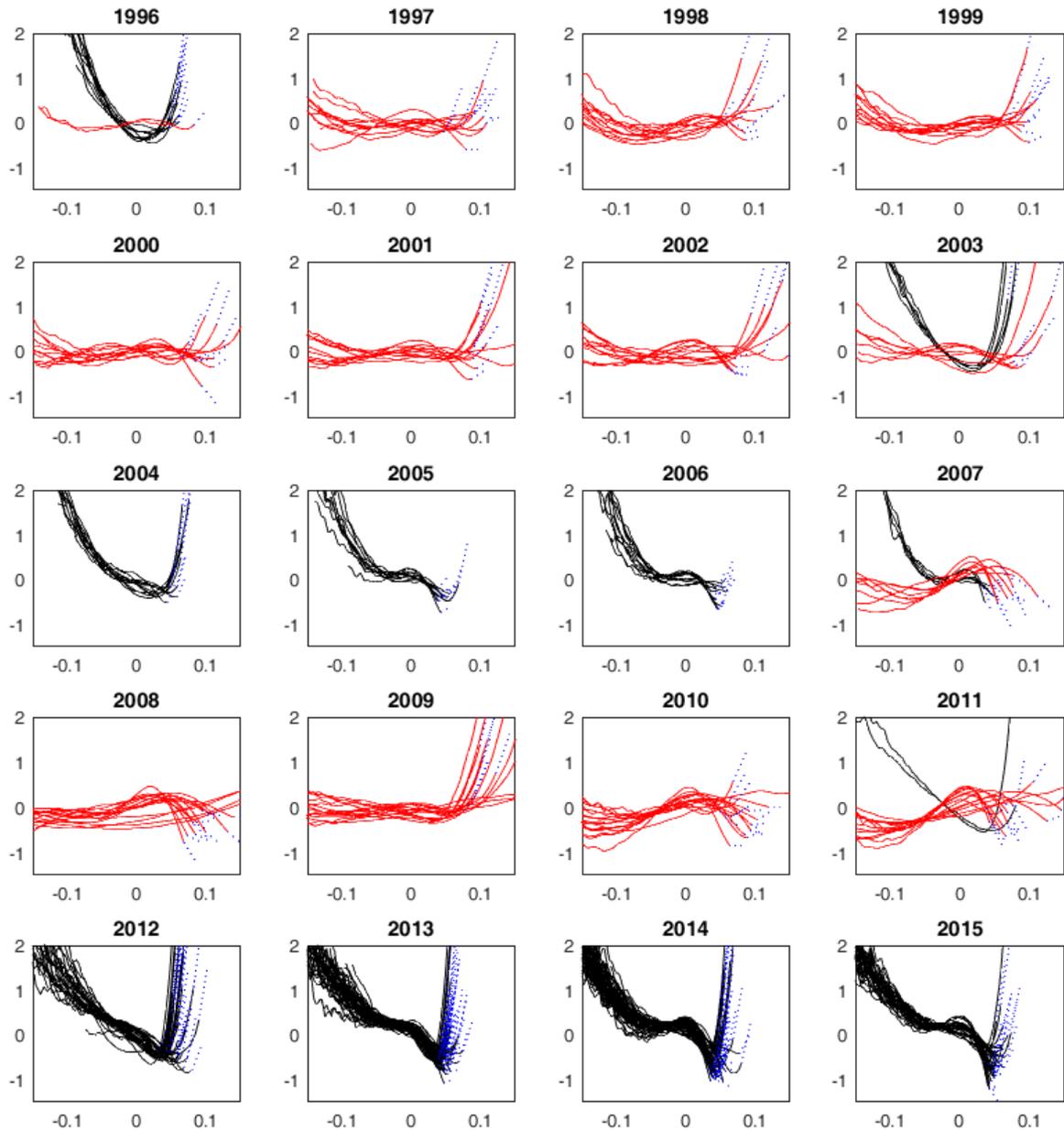


Figure 13: Empirical pricing kernels with changing parameters in the HN model with simulated densities

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

## 6.2 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 use it. 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 (16) and (17). This means that the level of the VIX is used both for normalizing the monthly returns and for rescaling. By first dividing and then multiplying any forecasting error due to risk premia at least partly cancels out, as long as the relationship is monotonic. If the VIX is a linear function of the physical volatility expectation everything cancels out and no error is introduced.

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 (see e.g. Figlewski & Malik (2014)). Figure 14 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.

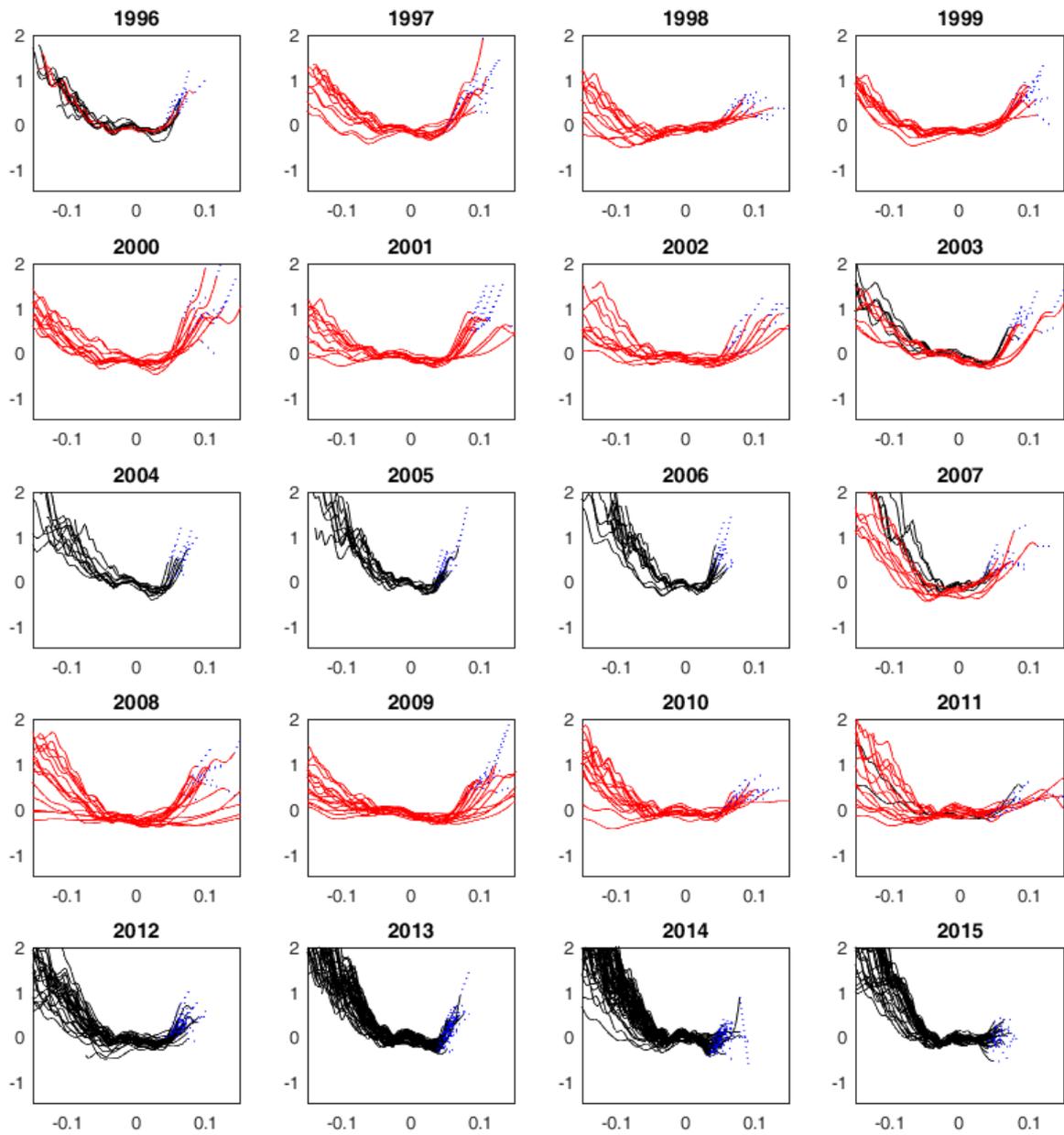


Figure 14: Empirical pricing kernels using the VIX as volatility forecast

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. Log-returns are on the horizontal axis. The horizon is one month. The blue line connects the points, which depict the ratio of the CDFs of the tail, with the corresponding pricing kernels.

### 6.3 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 5 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. The heterogeneous autoregressive (HAR) model of Corsi (2009) uses volatility components constructed over different time horizons and is given by:

$$\sqrt{RV_t} = \alpha_0 + \alpha_d \sqrt{RV_{t-1}} + \alpha_w (\sqrt{RV})_{t-5:t-1} + \alpha_m (\sqrt{RV})_{t-22:t-1} + u_t, \quad (28)$$

where:

$$(\sqrt{RV})_{t+1-k:t} = \frac{1}{k} \sum_{j=1}^k \sqrt{RV_{t-j}}, \quad (29)$$

$$RV_t = \sum_{j=1}^M r_{t,j}^2, \quad (30)$$

where  $RV_t$  is the realized variance over day  $t$ ,  $M$  is the sampling frequency and  $(\sqrt{RV})_{t+1-k:t}$  is the  $k$  period realized volatility. For brevity, I refer the reader to the original paper for further details of the model. I use realized volatility data based on 5 minute returns of the S&P 500 obtained from the website of the Oxford-Man Institute, which starts at 1.1.2000. The estimated parameters are displayed in Table 7. The parameter estimates are very similar to those of Corsi (2009) for all parameters, besides  $\alpha_0$ . Analogously to Table 2 above, Table 8 presents the volatility forecasts of the model, together with the forecasts of the CP-GARCH over the same time period. In times of low volatility, the HAR forecasts are slightly below the realization, but by a very similar magnitude by which the CP-GARCH forecasts are too high. In times of high volatility, the HAR model is significantly below the realized volatility on average. The RMSEs of the HAR model are always larger than their CP-GARCH counterparts. In sum, this suggests that the HAR model produces generally downward bias multi-period volatility forecasts. The reason could be that the model is specified in terms of volatility, not variance. This removes the bias towards the rather infrequent high volatility levels that other models have. But at the same time, the model appears to be biased towards the more fre-

quent low volatility periods.<sup>29</sup> Introducing structural breaks into the HAR model would probably improve its performance in that regard. I leave this for future research.

Table 7: Parameter estimates for the HAR model

	$\alpha_0$	$\alpha_d$	$\alpha_w$	$\alpha_m$
Estimate	.0005	.3632	.3876	.1965

The table shows the parameter estimates for the heterogeneous autoregressive (HAR) model of Corsi (2009) with data from 1.1.2000-31.8.2015.

Table 8: Predicted vs. realized 21 day volatility

	'00-'03	'03-'07	'07-'11	'11-'15
Average realized 21d volatility	.626	.0300	.0683	.0334
HAR Avg. predicted 21d volatility	.0508	.0291	.0523	.0302
CP Avg. predicted 21d volatility	.0619	.0308	.0701	.0354
HAR RMSE predicted 21d volatility	.0204	.0069	.0323	.0102
CP RMSE predicted 21d volatility	.0160	.0066	.0302	.0085

The table shows the average realized 21 day volatility across the different regimes, as well as the average predicted volatility by both the HAR and CP-GARCH model and the root-mean-square error (RMSE) of the predictions.

<sup>29</sup>The performance of the HAR model increases when the variance is calculated based on 5 minute data. However, the overall picture that the HAR model underpredicts volatility in times of high variance remains unchanged.

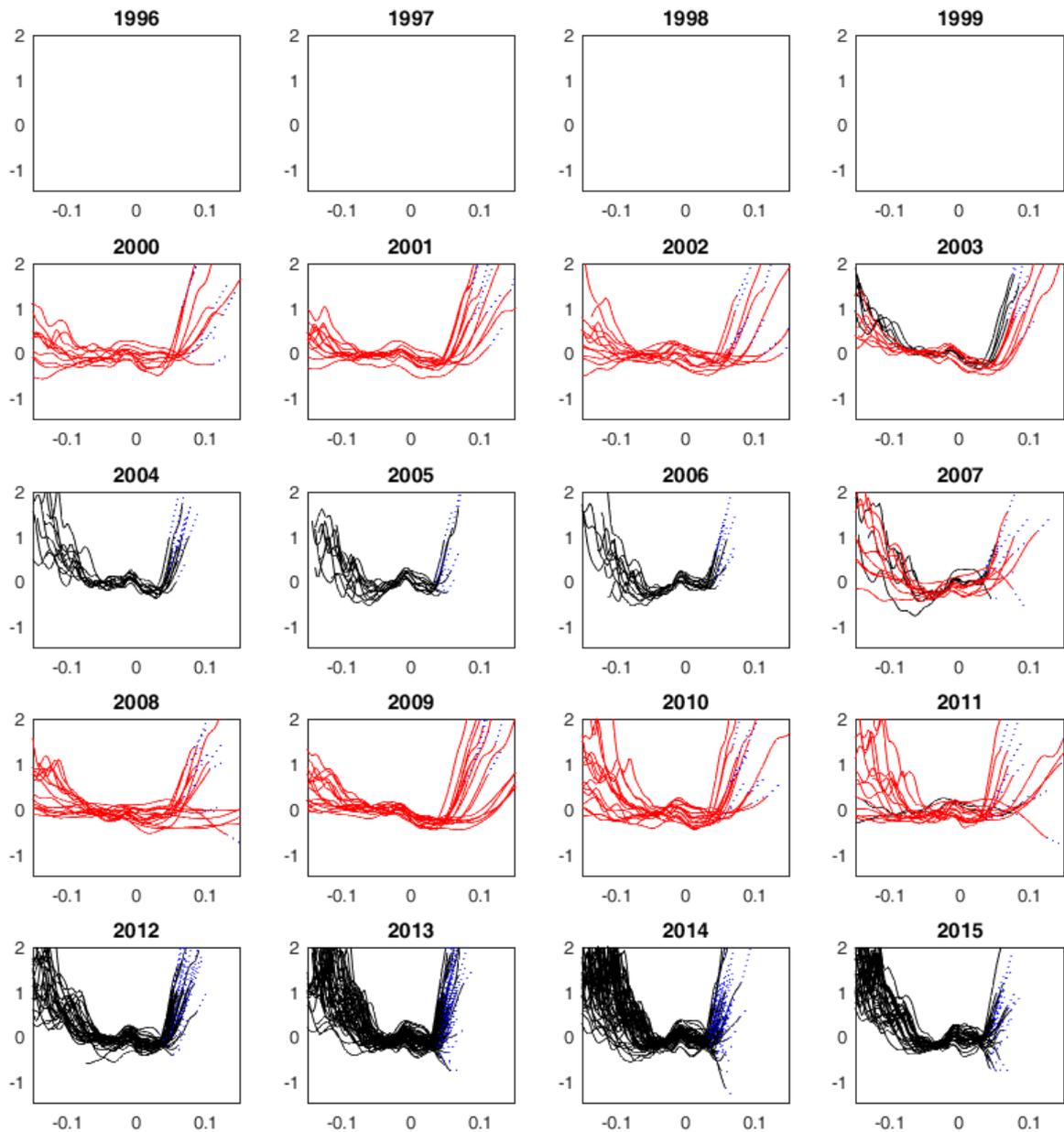


Figure 15: Empirical pricing kernels using the volatility forecast of the Corsi (2009) HAR model

The figure shows the natural logarithm of estimated pricing kernels obtained by using the volatility forecast from the Corsi (2009) HAR model. Data only exists from 1.1.2000 onwards. Red (black) depicts times with high (low) variance. Log-returns are on the horizontal axis. The horizon is one month. The blue line connects the points, which depict the ratio of the CDFs of the tail, with the corresponding pricing kernels.

I incorporate this volatility model into my pricing kernel estimation methodology, by employing its volatility forecast in (16) and (17), while everything else remains unchanged. The results, which are displayed in Figure 15 are interesting in several regards. First, 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 Chapter 4.5 fully support this conjecture. Lastly, the wiggles around zero are the result of an unsmooth empirical return shock density. Figure 18 in Appendix G illustrates this. Overall, using the HAR model can at least partly confirm the u-shape of the pricing kernels.

## 6.4 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 6.1 are reported. The graphs show that the results are impressively robust against changes in the analyzed horizon. All the observations above can be found for the other maturities and are equally strong. To save space, not all the results are shown. Figures 19 - 22 in the Appendix G contain the results for the two weeks and six weeks horizon with the benchmark method.

## 6.5 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. Due to space limitations, the results are not displayed, but are available from the author upon request.

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. This comes at the cost that the approach is more involved and Appendix E contains the technical details. Again, the results show that the impact is only marginal and the omission of the switching probability is a reasonable simplification.

## 6.6 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 dynamics of the NGARCH model as well as the corresponding parameter estimates are in Appendix F. The likelihood functions and algorithms are analogous to those presented in Chapter 3.

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. Due to space limitations, the results are not displayed, but are available from the author upon request.

The major difference in the dynamics of the two GARCH models are the drift term in the return equation and the concatenation of the variance in 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 modelled. Although the dynamics are different, there are no differences in the results. Furthermore, the above analysis shows that the key driver of the results are the biased volatility forecast of the fixed parameter GARCH. This effect most likely carries over to any GARCH specifications and the results can also be obtained using other GARCH models.

## 6.7 Second order polynomial

As a minor robustness check the fourth order polynomial, that was used to interpolate implied volatilities as described in 4.2, is replaced with the also popular second order

polynomial. The estimated pricing kernels do barely change and the results stay unaltered. The fourth order polynomial is preferred since untabulated analysis shows that it fits the data better.

## 6.8 Generalizability to other equity indices

A relevant question is whether the obtained results are generalizable to other equity indices. Due to a lack of option data, I cannot extend the analysis to other indices. It is promising, however, that Grith et al. (2013), who use DAX 30 data from June 2003 to June 2006, conclude that the hump is more pronounced in calm periods.

## 6.9 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. However, some approaches are not covered. This paragraph discusses how the presented results relate to these approaches.

First, there are several studies that assume a specific distribution for the physical return density, as variance-gamma, normal-inverse-Gaussian, or lognormal. This is potentially restrictive and it is not clear how the chosen parametric distribution influences the results and what empirical PK shapes are actually possible. Furthermore, some studies do not condition their return density forecasts on current market volatility, which most of the literature agrees on is important. Those who do condition on volatility typically do this using a standard GARCH model with fixed parameters and are therefore most likely prone to the bias documented above. If the method is altered to condition on volatility correctly, the results are expected to be similar, at least qualitatively, since also both the GARCH kernel method and the empirical shock method lead to qualitatively similar results, despite different forecasted return distributions.

Second, some studies make assumptions that restrict the functional form of the pricing kernel.<sup>30</sup> However, this by assumption typically makes certain shapes hard to find or even rules them out a priori.

Third, there are several studies that estimate an unconditional pricing kernel, which

---

<sup>30</sup>Cuesdeanu & Jackwerth (2016) provide a good and comprehensive overview of these studies in their Ch. 2.2.

allows some of them to be more or less non-parametric. This approach is fundamentally different from the approach taken here. Section 5.1 below discusses this approach in the light of the presented results. The main insight is that this approach might deliver wrong results.

Fourth, some studies use a rolling window of pure historic returns and kernel estimation methods to obtain the (conditional) physical returns density. This method does not condition the estimate on current market volatility. Since market volatility is time-varying and clustered, the past few years are often not a good proxy for the future expected returns. This is in particular the case at the beginning of a calm periods, where the past mostly consists of a crisis. Since the risk-neutral distribution is purely forward looking, this approach leads to findings of pronounced humps especially at the beginning of a calm period.

## 7 Conclusion

This paper identifies structural breaks in the volatility process of the S&P 500 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. This is in contrast to a standard GARCH model that, when estimated over a long time series, leads to volatility forecasts that systematically and significantly biased, especially in times of low volatility. Using a standard GARCH model is a popular approach in the literature on empirical pricing kernels. The results brought forward show that the biased volatility forecasts are the reason why many researchers find S-shaped pricing kernels, especially in times of low volatility. When replacing the standard GARCH by the CP-GARCH in the otherwise identical methodology, the S-shaped pricing kernels disappear and U-shaped pricing kernels are observed over the entire time series. It is very encouraging that this observation is robust to a number of changes in the method, model and data. Furthermore, 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 modelling the risk-neutral GARCH dynamics, that are otherwise also biased. Overall, the results are very helpful for asset pricing models as they show which kind of risk-factors and pricing kernel patterns are to be explained. Most obviously, they rule out the theoretically very challenging need to explain the coexistence of U-shaped and S-shaped pricing kernels.

The results and empirical exercises in this paper can be extended and generalized in a number of ways. First, the proposed methodology can be applied to other major stock indices. The finding of Grith et al. (2013) for DAX 30 data that the S-shape is more pronounced in calm periods, suggests that the conclusions extend to other indices. 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 suggests that a model should generate a U-shaped projection of the stochastic discount factor on returns. Identifying risk factors which make states with high returns risky and expensive for investors is economically very interesting.

# Appendix

## A Relation to Markov Switching Model

A close relative to a change-point model is the probably more popular Markov-switching (MS) model, also called regime-switching GARCH. The above model could become a Markov-switching model by changing the transition matrix  $\mathbf{P}$  from absorbing to recurring. But there are several reasons why the CP model is preferred here over the MS model. First of all, the estimation of a MS-GARCH model is even more complicated than that of a CP-GARCH model. One way to illustrate this is: for a time series of length  $T$ , the MS model has  $(K + 1)^T$  possible paths, while in the CP model the number of paths is at least linear in  $T$ , but not exponential. Augustyniak (2014) is the first and so far only one to provide an algorithm for and to conduct a maximum-likelihood estimation of a MS-GARCH model. However, preliminary analysis as well as correspondence with the author revealed that (at least a two state) MS-GARCH model is not well specified. Without any restrictions on the parameters, the optimal variance parameters are explosive in the (infrequently observed) very high volatility regime. Furthermore, the model seems to overfit the data, because paths generated from estimated parameters exhibit unreasonable dynamics. The main reason are the extreme positive and negative daily returns that occur from time to time. The estimation fits especially those in the (very) high variance regime. Therefore, one would probably need three or even four states to produce reasonable dynamics. The estimation of this model would be very difficult. Furthermore, the model then would be not so different from the CP model used here that has five states over the full time series.

In addition, the solution of the variance depended pricing kernel model, if one correctly includes the switching probabilities, is even more complicated for the MS version than for the CP. Lastly, the CP might have the better economic interpretation. The idea behind a model with structural breaks is that some fundamentals in the economy change. The regimes of a CP model, which usually last a few years, can roughly be seen as business cycles. Even though the MS model roughly identifies the same regimes, if one simulates the model it switches at random points in time, which can easily be implausible close together to represent a business cycle. On the other hand, the CP model amplifies a problem that also the MS shares: The problem of how an individual could know the parameters of the next state (and maybe even this state) if that state is different to all others and has never occurred before. This concern is addressed in Section 6.2 of the

robustness section by using a purely forward looking volatility forecast.

Overall, for the outlined reasons the less common CP model is preferred here over the MS model.

## B Data Cleaning

Here is a detailed list of the option data cleaning steps applied.

### B.1 General Option Data

1. All S&P 500 Options in the period from 1.1.1996-31.8.2015 with a time to maturity  $>14$  days and  $<366$  days are downloaded from OptionMetrics. This amounts to roughly 6 million option quotes.
2. Remove all quotes that have:
  - zero trading volume on the day of the price quote
  - best bid is below \$0.50
  - that are more than 20 points in-the-money
  - violate
3. For each week in the sample, use only Wednesday data. If the Wednesday of a week is not a trading day the Tuesday of that week is used. If this is also not a trading day the week is excluded. This happened a few times, (e.g. mostly around Christmas and New Year, e.g. at and after 11.09.2001).
4. Of the cleaned data, identify the six most actively traded contracts (by total volume). If some contracts have the same volume, the ones with the larger number of outstanding options are used. If this would also tie, the selection is randomized (i.e. the given random ordering is used).
5. For the sub-sample with only close to 30 day maturity: in each week, the option series which has a remaining time to maturity closest to 30 days is selected. If 30 days are not matched exactly, the next closest maturity is selected, starting with 31 days, next 29, then 32, then 28 and so on.

## B.2 Pricing Kernel Option Data

1. All S&P 500 Options in the period from 1.1.1996-31.8.2015 with a time to maturity  $>27$  days and  $<34$  days (monthly horizon) are downloaded from OptionMetrics. (For the other horizons, the used thresholds are:  $>11$  and  $<20$ ,  $< 39$  and  $> 48$ ,  $< 55$  and  $> 65$ )
2. Remove all quotes that:
  - have zero trading volume on the day of the price quote
  - have best bid is below \$0.50
  - are more than 20 points in-the-money.
3. 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 16 days (two weeks), 44 days (six weeks) and 60 days (two months).

## C Estimation of Risk-Neutral Density

1. Clean the data as described above.
2. Get risk-free rate from OptionsMetrics, and interpolate linearly for the correct maturity.
3. Calculate implied dividend yield from at the money call and put pair.  
First, the implied forwards from put-call parity:

$$C + Ke^{-r\tau} = P + Fe^{-r\tau}, \quad (31)$$

and then dividend yield via the spot-forward parity:

$$F = Se^{(r-DivYield)\tau}. \quad (32)$$

4. Transform mid-prices into implied volatilities using Black and Scholes (1973). In the region of  $\pm 20$  points from at-the-money, take a weighted average (by volume) of put and call implied volatilities.

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.
6. Numerically differentiate the call prices using (14) to recover the risk-neutral return distribution.

## D Particle-MCEM-MCML algorithm

This section repeats the main steps in the Particle-MCEM-MCML algorithm suggested by Sichert (2017) to estimate the CP-GACH model. This hybrid algorithm, called Particle-MCEM-MCML, is based on the algorithms proposed by Augustyniak (2014) and Bauwens et al. (2014). For further details and empirical applications see Sichert (2017) .

Let  $\theta$  denote the parameter vector,  $S_t$  the stock price,  $y_t$  the latent state variable, and  $S = S_{1:T} = \{s_1, \dots, s_T\}'$  and  $Y = Y_{1:T} = \{y_1, \dots, y_T\}'$  the corresponding vectors.<sup>31</sup> The algorithm starts with an initial guess of the parameters  $\theta^{(0)}$  and  $r$  counts the number of steps, starting from 1.<sup>32</sup>

**Algorithm 1** (Particle-MCEM-MCML Algorithm).

1. Simulate  $m_r$  samples of the state vector  $Y_{1:T}$  from  $p(Y_{1:T}|S_{1:T}, \theta^{(r-1)}, P)$  using the particle Gibbs sampler. The sampler is described next.

Particle Gibbs sampler:

Let  $w_t^i$  denote the normalized weights that are associated to  $N$  particles  $\{y_t^1, \dots, y_t^N\}$  which represent possible realizations of  $y_t$ . These weights serve to approximate the probability  $p(y_t|s_t, \theta, P)$ . More specifically  $p(y_t = j|s_t, \theta, P) \approx \sum_{i=1}^N w_t^i \mathbb{1}_{\{y_t^i=j\}}$  with  $\mathbb{1}_{\{\cdot\}}$  being the indicator function.

Define the ancestor variable  $A_t^k$  as the particle from which the particle  $k$  at time  $t$  is sampled, and the lineage variable  $b_t^k$  as the particle belonging to the path of the particle  $k$  at time  $t$ . Set  $b_T^k := k$  so that we have the backward recursion

---

<sup>31</sup>The increase readability, here  $s_t$  denotes the stock price at time  $t$ , which was  $S_t$  in the main text above.

<sup>32</sup>The main steps are mostly based on Augustyniak (2014), while the particle sampler is mostly based on Bauwens et al. (2014), both with several adjustments. For a detailed discussion see the respective papers.

$b_t^k = A_t^{b_t^k}$ . The  $b_t^k$  variable represents next the lineage of the previous  $y_t$  draw. The conditional SMC can be computed for  $p(y_t|s_t, \theta, P)$  for  $t = 1, \dots, T$ , assuming we have  $\{y_1^{b_1^k}, \dots, y_T^{b_T^k}\}$  and given uniform initial weights  $w_0^i = 1/N$  and initial particles  $s_0^i = 1$  as:

- (a)  $\forall i \in [1, N]$ , compute  $g_t^i = w_{t-1}^i \sum_{j=1}^K p(y_t = j | y_{t-1}^i, P) f(s_t | F_{t-1}, \theta, P, y_t = j)$ ,  $F_{t-1}$  denoting the data and particles until  $t - 1$ , and the normalized weights  $\tilde{\omega}_t^i = g_t^i / \sum_{j=1}^N g_t^j$ .
- (b)  $\forall i \in [1, N] \setminus b_t^k$  sample independently a label variable  $A_{t-1}^i \sim \tilde{\omega}_t$  such that  $A_{t-1}^i \in [1, N]$ .
- (c)  $\forall i \in [1, N] \setminus b_t^k$ , sample a particle  $y_t^i \sim p(y_t | y_{t-1}^{A_{t-1}^i})$ .
- (d)  $\forall i \in [1, N]$  compute  $\hat{w}_t^i = \frac{f(s_t | F_{t-1}, y_t^i, \theta, P)}{\sum_{j=1}^K f(s_t | F_{t-1}, \theta, P, y_t = j) p(y_t = j | s_{t-1}^{A_{t-1}^i})}$  and the normalized.

After reaching  $T$ :

- (e)  $\forall i \in [1, N] \setminus b_T^k$  recursively calculate the ancestral lineage and path for each  $Y_{1:T}^i$ .
  - (f)  $\forall i \in [1, m_r]$  sample  $k^i \sim w_T$ . The  $Y_{1:T}^{k^i}$  are the new state vectors.
2. Monte Carlo E-step: Calculate  $\hat{Q}(\theta | \theta^{(r-1)})$ , an approximation of the conventional E-step  $Q(\theta | \theta^{(r-1)})$ , where

$$\begin{aligned}
\hat{Q}(\theta | \theta^{(r-1)}) &= \frac{1}{m_r} \sum_{i=1}^{m_r} \log[f(S, Y^{(i)} | \theta)] \\
&= -\frac{T \log(2\pi)}{2} - \frac{1}{2m_r} \sum_{t=1}^T \sum_{i=1}^{m_r} [\log(h_t^{(i)})] + \frac{(S_t - (r_t + (\mu_{y_t^{(i)}} - \frac{1}{2})h_t^{(i)}))^2}{(h_t^{(i)})} \\
&\quad + \frac{1}{m_r} \sum_{t=1}^T \sum_{i=1}^{m_r} \log(p_{y_{t-1}^{(i)}, y_t^{(i)}}) \\
&= \text{term 1} + \text{term 2}.
\end{aligned} \tag{33}$$

In the previous expressions,  $h_t^{(i)}$  is shorthand for  $h_t(Y_{1:t}^{(i)})$ .

3. M-step: Perform the following maximization:

$$\theta^{(r)} = \arg \max_{\theta} \hat{Q}(\theta | \theta^{(r-1)}) \quad (34)$$

This optimization can again be split into two independent steps since terms 1 and term 2 of Eq. (33) involve different subsets of the parameters. Term 1 includes the mean and GARCH parameters while term 2 only contains transition probabilities. Maximization of term 1 must be performed numerically and is similar to a standard GARCH optimization to calculate the MLE. To improve the performance of that optimization, the gradient of term 1 with respect to the mean and GARCH parameters should be provided to the optimization routine (see Appendix C of Augustyniak (2014)). Maximization of term 2 can be done analytically. Term 2 is at its maximum when the transition probabilities take the values

$$p_{jk} = \frac{f_{jk}}{\sum_{l=1}^N f_{jl}}, \quad j, k = 1, \dots, K, \quad (35)$$

where  $f_{jk}$  denotes the total number of transitions from state  $j$  to state  $k$  in all of the  $m_r$  simulated state vectors. A proof of this result is in Appendix B of Augustyniak (2014).

4. Apply a decision rule to determine whether to switch to the MCML algorithm. If the decision is to switch, go to step 5 and set  $\theta^* = \theta^{(r)}$ . Otherwise, add 1 to  $r$  and go to step 1.
5. Simulate  $m^*$  samples of the state vector  $Y$  from  $p(Y|S, \theta^*, P^*)$  using the particle Gibbs sampler described in step 1 of the algorithm to obtain the importance sample  $\{Y^{(i)}\}_{i=1}^{m^*}$ .
6. MCML-step: Perform the following maximization to obtain the MLE:

$$\hat{\theta} = \arg \max_{\theta} \left[ \log \sum_{i=1}^{m^*} \omega_{\theta|\theta^*}^{(i)} \right], \quad (36)$$

where  $\omega_{\theta|\theta^*}^{(i)} = \log f(S, Y^{(i)}|\theta, P) - \log f(S, Y^{(i)}|\theta^*, P^*)$ .

## E Formulas with switching probabilities

This section describes the calculation of the expected variance that includes the switching probability. When the current regime is  $j$ , the formula for the expected variance is:

$$\begin{aligned}
VAR_t(T) &= E_t\left(\sum_{\tau=t+1}^T h_\tau\right) = \left[p_{j,j}^T \cdot VAR_{j,t}(T) + p_{j,j}^{T-1} p_{j,j+1} \cdot VAR_{j+1,t+\tau}(T)[h_{j+1,t+1}] \right. \\
&\quad \left. + p_{j,j}^{T-1} p_{j,j+1} \cdot \sum_{\tau=1}^{T-1} \left(VAR_{j,t}(t+\tau) + VAR_{j+1,t+\tau}(T)[E_t(h_{j+1,t+\tau+1})]\right)\right] \\
&\quad / (p_{j,j}^T + p_{j,j}^{T-1} p_{j,j+1} T) = \\
&\quad = \frac{1}{p_{j,j} + p_{j,j+1} T} \left[ p_{j,j} VAR_{j,t}(T) + p_{j,j+1} VAR_{j+1,t+\tau}(T)[h_{j+1,t+1}] + \right. \\
&\quad \left. p_{j,j+1} \sum_{\tau=1}^{T-1} \left(VAR_{j,t}(t+\tau) + VAR_{j+1,t+\tau}(T)[E_t(h_{j+1,t+\tau+1})]\right) \right], \tag{37}
\end{aligned}$$

where:

- $p_{j,j}$  is the probability to stay in regime  $j$ ,
- $p_{j,j+1}$  is the probability to switch from regime  $j$  to regime  $j+1$ ,
- $VAR_{j,t}(T)$  is the at time  $t$  expected variance from  $t+1$  to  $T$ , conditional on staying in regime  $j$  from  $t+1$  to  $T$  and calculated using Equ. 5 with the parameters of the respective regime,
- $h_{j+1,t+1} = \omega_{j+1} + \alpha_{j+1}(z_{t-1} - \gamma_{j+1}\sqrt{h_{t-1}})^2 + \beta_{j+1}h_{t-1}$
- $E_t(h_{j+1,t+\tau+1})$  is the expected variance in  $t+\tau+1$  conditional on staying in regime  $j$  from  $t$  to  $t+\tau$  and switching regime in  $t+\tau+1$ ;  
 $E_t(h_{j+1,t+\tau+1}) = \omega_{j+1} + (\beta_{j+1} + \alpha_{j+1} \cdot \gamma_{j+1}^2)E_t(j, h_{t+\tau})$ , and  
 $E_t(j, h_{t+\tau}) = (\beta_j + \alpha_j \cdot \gamma_j^2)^{\tau-1} h_{j,t+1} + \frac{\omega_j \alpha_j}{\beta_j + \alpha_j \gamma_j^2} [1 - (\beta_j + \alpha_j \cdot \gamma_j^2)^{\tau-1}]$
- and  $VAR_{j+1,t+\tau}(T)[\cdot]$  is the expected variance from  $t+\tau+1$  to  $T$ , when the process switches to regime  $j+1$  in  $t+\tau+1$ . It is calculated using Equ. 5 where  $h_1$  is replaced with  $[\cdot]$ .

To understand the formula, it might be useful to further discuss the dynamics in 6 and 7: On (the morning of) day  $t$ , first nature draws the latent state variable  $y_t$ , then the

variance process is updated, and then the return is realized during the day, using the new variance. The variance process is updated using the variance from  $t - 1$  and the parameters of the new regime.

The formula gives the expected variance over all possible path with zero switches (first term) or one switch (second and third term). The second term is the case when the regime switches on the next day, while the third term contains all other possible paths with one regime switch. The denominator normalizes the probabilities.

The presented formula omits the probability of more than one regime switch, to keep the analysis tractable. This is both economically and statistically reasonable. First, the idea of the breaks are that they represent long swings like business cycles and a regime that lasts less than 21 days seems unreasonable. Second, the statistic probability of observing more than one switch within 21 days is very low (e.g. 0.02% when using a relatively high switching probability of 0.001). Since multiple switches are ignored, the probabilities would not exactly add up to one and hence they are normalized to one by the denominator.

Lastly, for the last regime no  $j + 1$  exists. I chose to use the second to last regime as a proxy for the unknown regime. Using the parameters of regime  $j + 1$  in regime  $j$  does contain some forward looking bias. However, this can be seen as a proxy for using an alternating pattern of high an low volatility regimes, since the long-run volatility across the high an low volatility regimes respectively is very similar.

## F NGARCH formulas and parameter estimates

### F.1 Dynamics

The dynamics of the standard NGARCH model are are:

$$\ln\left(\frac{S_t}{S_{t-1}}\right) = r_t + \mu\sqrt{h_t} - \frac{1}{2}h_t + \sqrt{h_t}z_t, \quad (38)$$

$$h_t = \omega + \alpha h_{t-1}(z_{t-1} - \gamma)^2 + \beta h_{t-1} \quad (39)$$

$$z_t \sim N(0, 1).$$

Furthermore the long run variance of the NGARCH model is:

$$E[h_t] = \frac{\omega}{1 - \beta - \alpha(1 + \gamma^2)}, \quad (40)$$

and the expected variance over any period is:

$$E_0 \left[ \sum_{t=1}^T h_t \right] = T \cdot E[h_t] + (h_1 - E[h_t]) \frac{1 - (\beta + \alpha(1 + \gamma^2))^T}{1 - (\beta + \alpha(1 + \gamma^2))}. \quad (41)$$

Finally, the dynamics of the NGARCH model with structural breaks are:

$$\ln\left(\frac{S_t}{S_{t-1}}\right) = r_t + \mu_{y_t} \sqrt{h_t} - \frac{1}{2} h_t + \sqrt{h_t} z_t, \quad (42)$$

$$h_t = \omega_{y_t} + \beta_{y_t} h_{t-1} + \alpha_{y_t} h_{t-1} (z_{t-1} - \gamma_{y_t})^2 \quad (43)$$

$$z_t \sim N(0, 1),$$

where  $y_t$  and  $\mathbf{P}$  are as defined in Section 2.2.

## F.2 Parameter estimates

Table 9: Estimation Results of the NGARCH model

Parameters	FP NGARCH		CP-NGARCH			
	'92-15	'92-'97	'97-'03	'03-'07	'07-'12	'12-'15
$\omega$	1.84E-06	2.86E-06	6.55E-06	4.73E-06	4.05E-06	3.88E-06
$\alpha$	7.44E-02	4.19E-02	5.07E-02	2.14E-02	7.31E-02	6.25E-02
$\beta$	0.830	0.791	0.739	0.394	0.793	0.487
$\gamma$	1.065	1.470	1.878	4.763	1.317	2.547
$\mu$	0.021	0.050	0.001	0.046	-0.013	0.062
$p_{jj}$		0.99918	0.99941	0.99898	0.99911	1
Properties						
$\beta + \alpha(1 + \gamma^2)$	0.9884	0.924	0.9679	0.9017	0.9934	0.9545
Long-run volatility	0.200	0.097	0.227	0.110	0.392	0.147
Log-likelihood						
Total	19555.5	19653.1				
AIC	-39101.0	-39248.3				
BIC	-39067.5	-39054.2				

Parameter estimates are obtained by optimising the likelihood on returns. Parameters are daily, long run volatility is calculated as  $\sqrt{\text{long} - \text{run variance} \cdot 252}$ . For each model, the total likelihood value at the optimum is reported. The volatility parameters are constrained such that the variance is positive ( $0 \leq \alpha < 1$ ,  $0 \leq \beta < 1$ ,  $\alpha(1 + \gamma^2) + \beta < 1$ ,  $0 < \omega$ ). 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.

## G Additional Plots

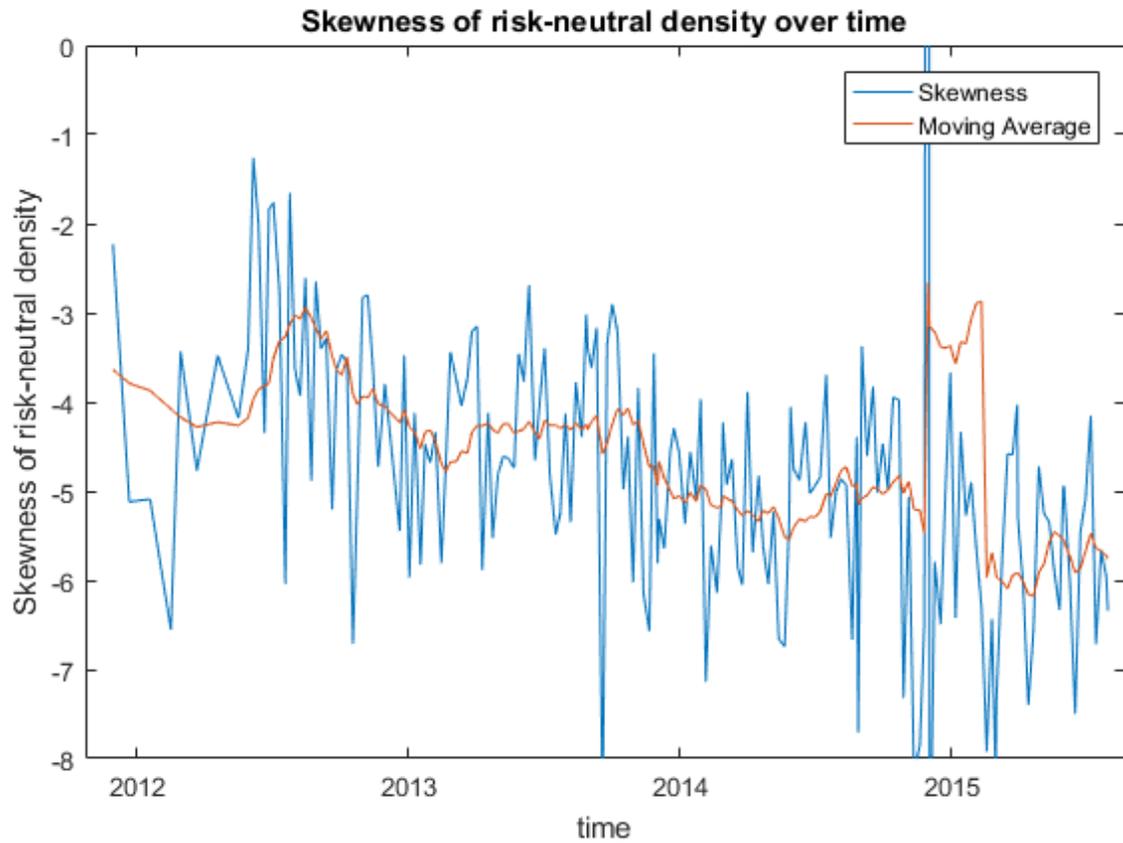


Figure 16: Skewness of risk-neutral return density and moving average

The figure shows the skewness of the estimated risk-neutral density during the last low volatility regime (29.11.2011-31.08.2015). The blue line shows the estimate for each date, and the red line shows the moving average of the past 12 observations. The data and method to obtain the option implied distribution is as in Section 4. The tails of the distribution are not completed.

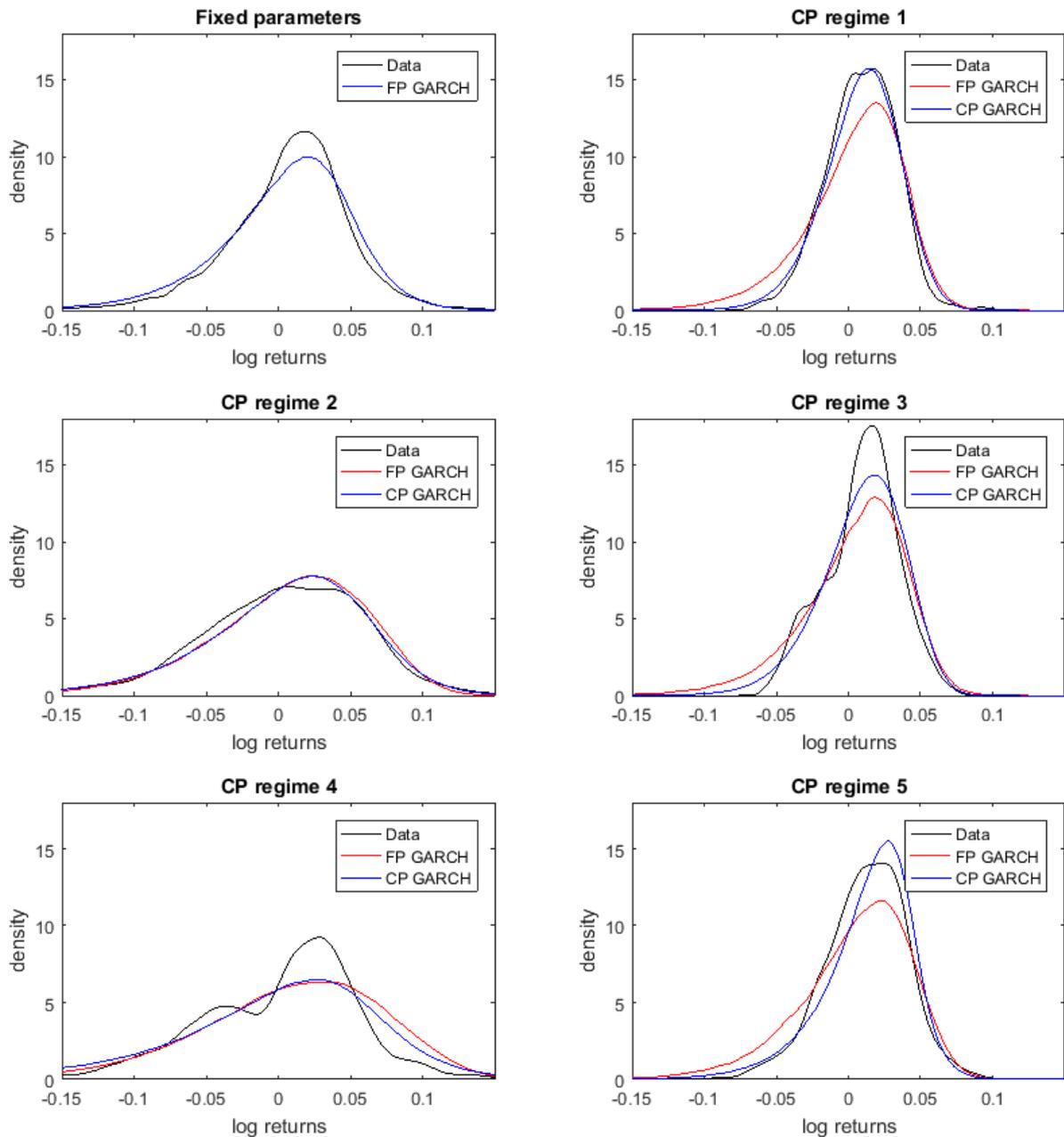


Figure 17: Monthly return density for the CP-GARCH model, FP GARCH model and empirical returns

The figure shows the estimated monthly return densities (21 days) for the CP-GARCH, the FP GARCH and the empirical returns. The first FP GARCH densities and all CP-GARCH densities are obtained from a long simulation (1 million days). The FP GARCH density in the regimes is obtained from 100,000 simulations, that are started from the average variance of the regime. The first sub-plot contains the full time series and the FP GARCH model, while the remaining ones only contain the distributions of the respective regimes in timely order.

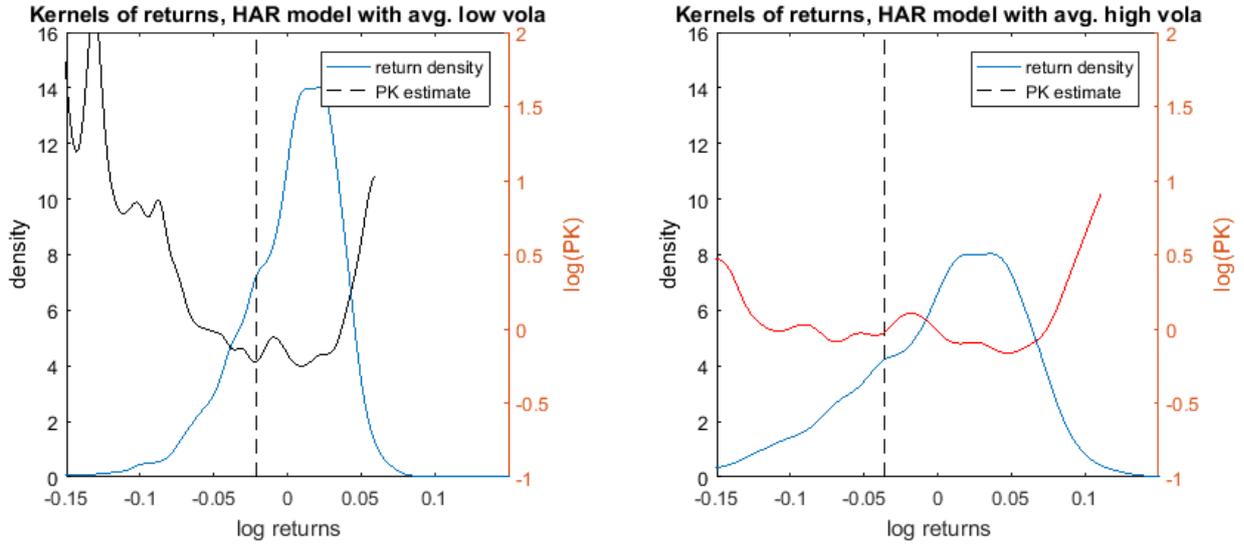


Figure 18: Monthly return density for the CP-GARCH model, FP GARCH model and empirical returns

The figure shows the estimated monthly return densities (21 days) for the HAR model of Corsi (2009) (blue line), and the corresponding pricing kernel estimates (black and red, respectively). The left plot shows data of the date that is closest to the average low volatility, while the right plot shows data of the date that is closest to the average high volatility. The vertical line (dashed) shows where the kink in the return density forecasts translates into a hump in the pricing kernel estimate.

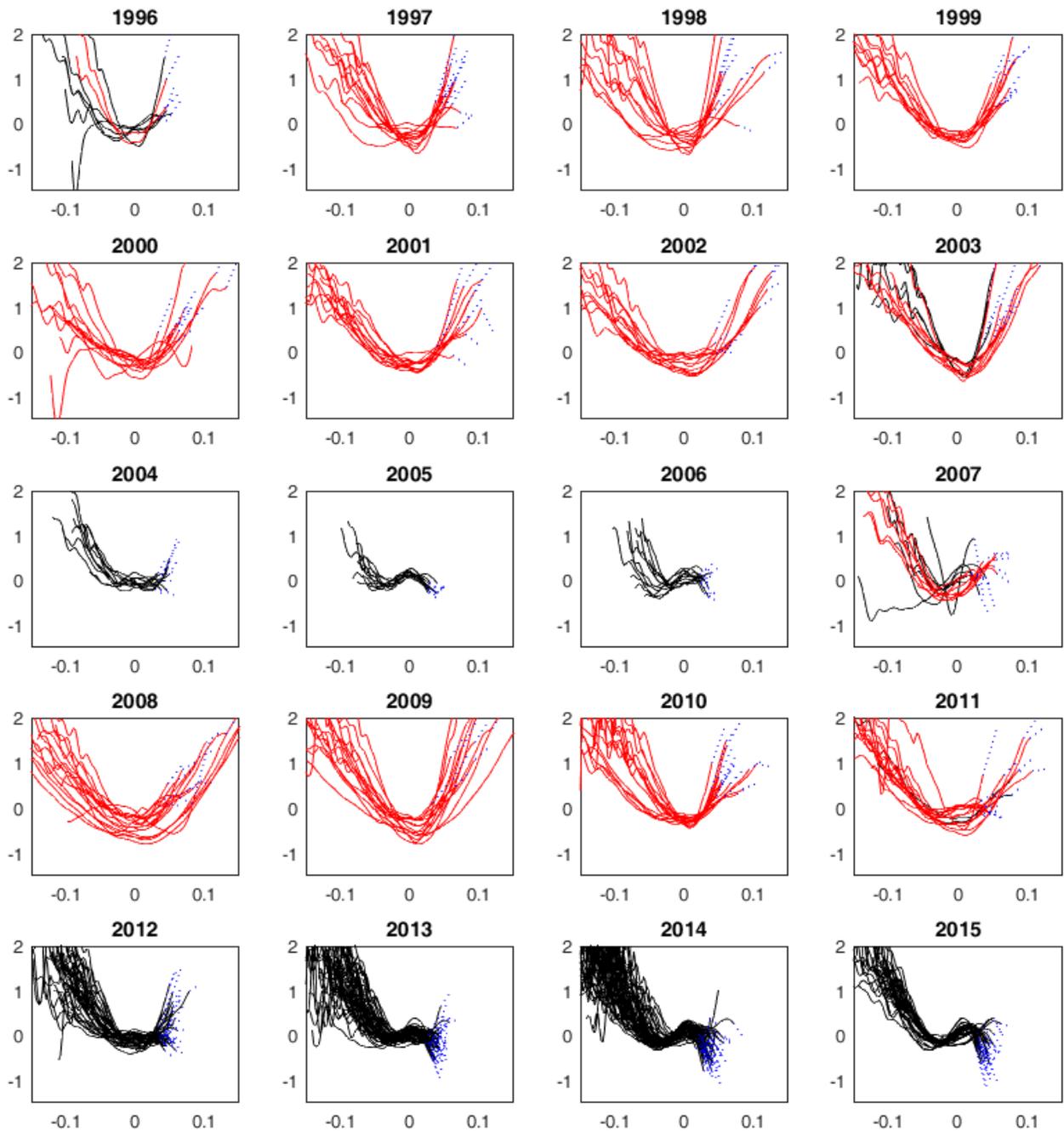


Figure 19: Empirical pricing kernels with fixed parameters in the HN model and two weeks horizon

The figure shows the natural logarithm of estimated pricing kernels obtained from using the Heston-Nandi model with fixed parameters. Red (black) depicts times with high (low) variance. Log-returns are on the horizontal axis. The horizon is two weeks. The blue line connects the points <sup>74</sup> which depict the ratio of the CDFs of the tail, with the corresponding pricing kernels.

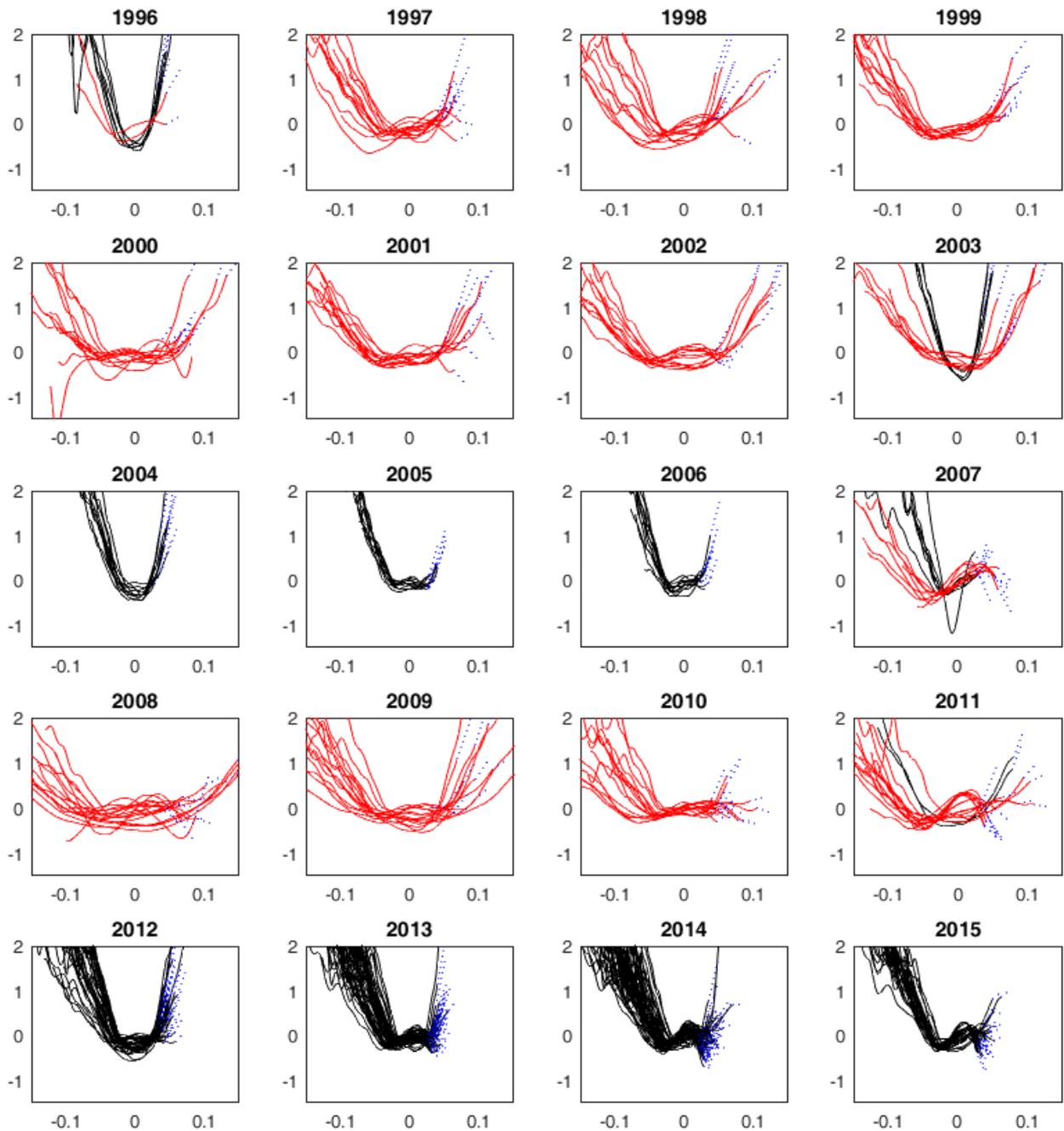


Figure 20: Empirical pricing kernels with CP parameters in the HN model and two weeks horizon

The figure shows the natural logarithm of estimated pricing kernels obtained from using the Heston-Nandi model with CP parameters. Red (black) depicts times with high (low) variance. Log-returns are on the horizontal axis. The horizon is two weeks. The blue line connects the points <sup>75</sup>, which depict the ratio of the CDFs of the tail, with the corresponding pricing kernels.

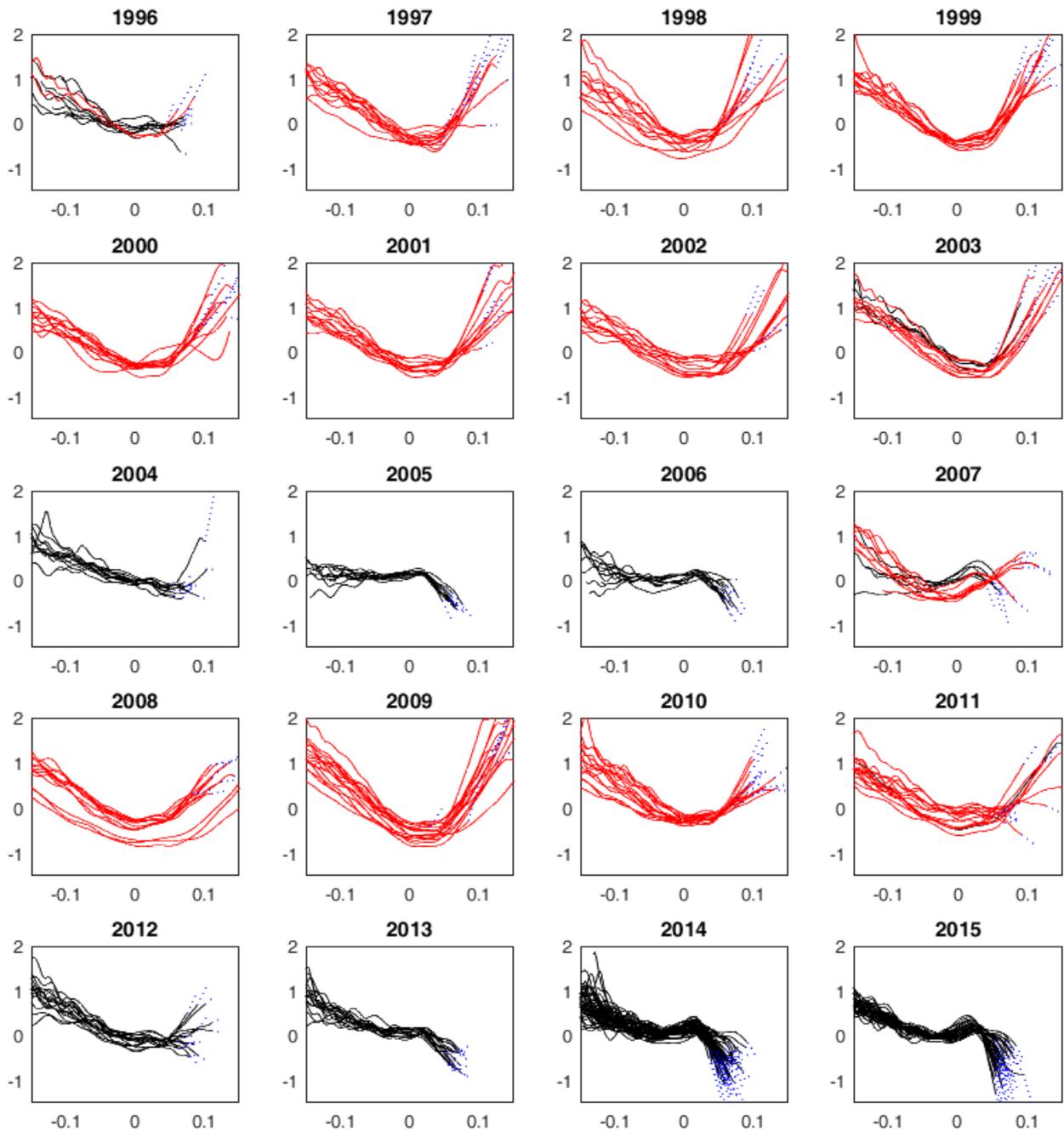


Figure 21: Empirical pricing kernels with fixed parameters in the HN model and six weeks horizon

The figure shows the natural logarithm of estimated pricing kernels obtained from using the Heston-Nandi model with fixed parameters. Red (black) depicts times with high (low) variance. Log-returns are on the horizontal axis. The horizon is six weeks. The blue line connects the points <sup>76</sup>, which depict the ratio of the CDFs of the tail, with the corresponding pricing kernels.

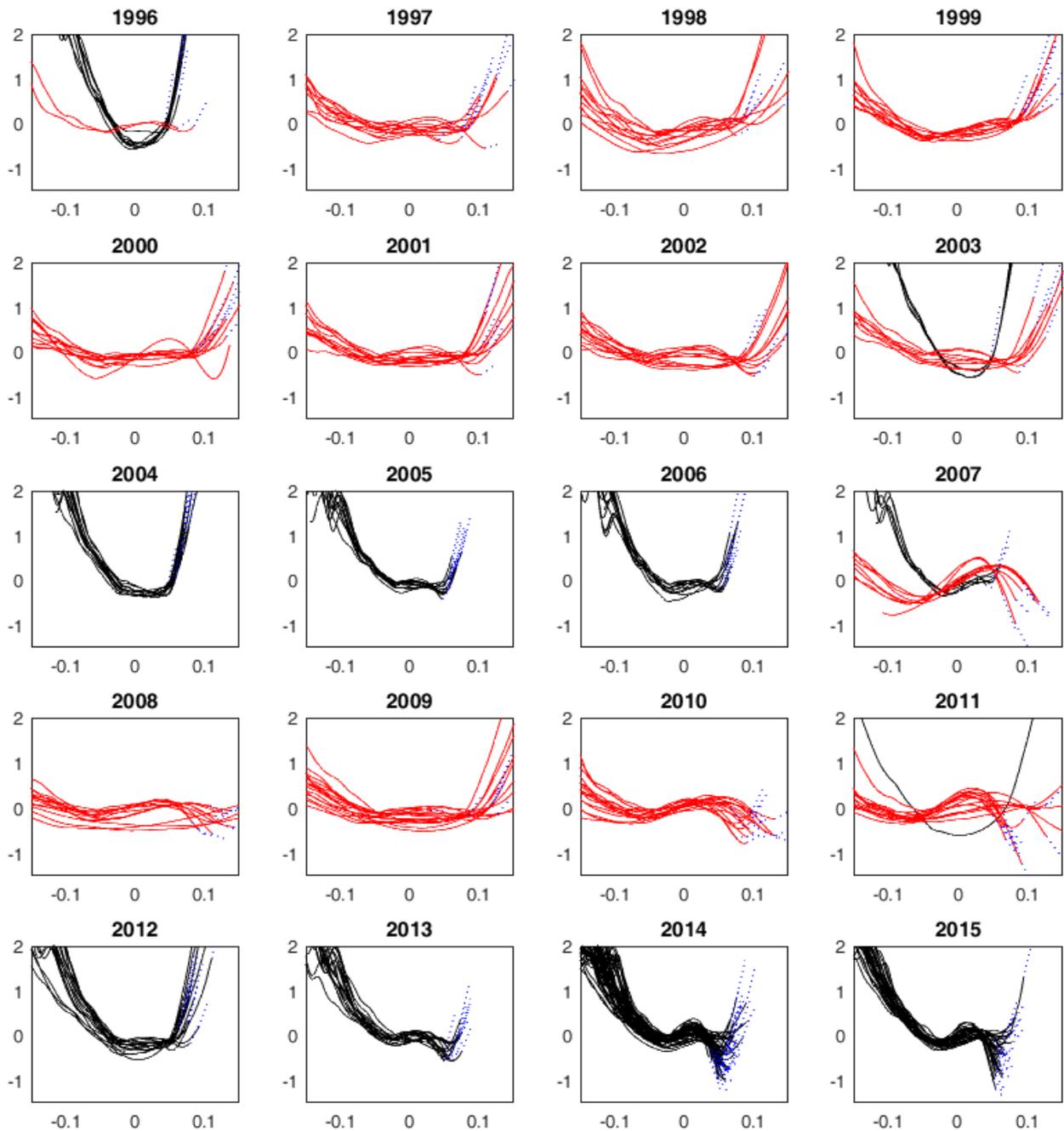


Figure 22: Empirical pricing kernels with CP parameters in the HN model and six weeks horizon

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

## References

- 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.
- 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.
- Bauwens, L., Preminger, A. & Rombouts, J. V. (2010), ‘Theory and inference for a markov switching garch model’, *The Econometrics Journal* **13**(2), 218–244.
- 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*.
- Billio, M., Casarin, R. & Osuntuyi, A. (2016), ‘Efficient gibbs sampling for markov switching garch models’, *Computational Statistics & Data Analysis* **100**, 37–57.

- Branger, N., Schlag, C. & Zaharia, S. (2011), ‘An equilibrium foundation for the heston stochastic volatility model and u-shaped pricing kernels’, *Working paper, University Münster* .
- Breedon, D. T. & Litzenberger, R. H. (1978), ‘Prices of state-contingent claims implicit in option prices’, *Journal of Business* pp. 621–651.
- 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.
- 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. (2016), ‘The pricing kernel puzzle: Survey and outlook’, *Available at SSRN* .
- 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. (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. & 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*.
- Gray, S. F. (1996), ‘Modeling the conditional distribution of interest rates as a regime-switching process’, *Journal of Financial Economics* **42**(1), 27–62.
- 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., Hardle, W. & Park, J. (2013), ‘Shape invariant modeling of pricing kernels and risk aversion.’, *Journal of Financial Econometrics* **11**(2), 370 – 399.
- Härdle, W. K., Okhrin, Y. & Wang, W. (2014), ‘Uniform confidence bands for pricing kernels’, *Journal of Financial Econometrics* **13**(2), 376–413.
- 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.
- Hsieh, K. & Ritchken, P. (2005), ‘An empirical comparison of garch option pricing models’, *Review of Derivatives Research* **8**(3), 129–150.
- 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.
- 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. (2017), ‘Maximum likelihood estimation of change-point garch models’, *Working Paper, Goethe University Frankfurt* .
- Ziegler, A. (2007), ‘Why does implied risk aversion smile?’, *Review of Financial Studies* **20**(3), 859–904.