Behavioral in the Short-run and Rational in the Long-run? Evidence from S&P 500 options.

J. Driessen, J. Koëter and O. Wilms*

October, 2019

Abstract

We estimate the pricing kernel from options on the S&P 500 index for different horizons and over time. This allows us to compare short- and long-term pricing kernels and analyze their time-series variation. We show that the well documented pricing kernel puzzle—that is, the non-monotonicity of the pricing kernel—only exists for short horizons. For longer horizons the puzzle disappears and the level, shape and time-series variation of the pricing kernel are in line with standard rational expectation asset pricing models. Furthermore, we show that the empirical features of the short-term kernel can be explained by a behavioral asset pricing model.

1 Introduction

Starting with the seminal work of Jackwerth (2000) and Aït-Sahalia and Lo (2000), many studies have used option pricing information to estimate pricing kernels for the U.S. equity market.[1] The main finding in this literature is that the pricing kernel is not strictly downward sloping when it is projected on the equity index return. Depending on the time period, Jackwerth (2000), Aït-Sahalia and Lo (2000) and Rosenberg and Engle (2002) either find the pricing kernel to have a U-shape or an upward sloping region. Both of these findings are in contrast to a monotonically decreasing pricing kernel as predicted by classical finance theory. The result of a non-monotonic pricing kernel is often referred to as the "pricing kernel puzzle". In this paper, we present evidence that the non-monotonicity of the pricing kernel is a unique feature of the short-term one-month

*All authors are at Tilburg University, Finance Department.

[1] See Cuesdeanu and Jackwerth (2018a) for an overview of this literature.

pricing kernel—the maturity that is usually analyzed in the literature. For horizons larger than one month, we show that the puzzle gradually disappears and the level, shape and time-series variation of the pricing kernel are in line with standard rational expectation asset pricing models.

Existing work focuses mainly on the pricing kernel for a horizon of one month. We use options with longer maturities to estimate the pricing kernel for different horizons and analyze its term structure. In line with existing empirical work, we find the one-month pricing kernel to be U-shaped. The U-shape implies that investors are willing to pay more than the expected value for securities paying in bad states and in good states of the economy—a feature that is hard to reconcile with rational expectation asset pricing models, but consistent with a behavorial asset pricing model with probability weighting Baele et al. (forthcoming). We show that standard asset pricing models produce a monotonically decreasing pricing kernel when projecting the pricing kernel onto the market return.[2] To compare short- and long-term pricing kernels, we define a forward kernel which does not depend on any short-term information.[3] In contrast to the results on the short-term kernels found in the literature, we show that for horizons beyond six months, the empirical pricing kernel is monotonically downward sloping. Hence, the pricing kernel puzzle vanishes for longer horizons and we show that the shape and level of the long-term pricing kernel is well in line with the predictions of standard pricing models.

Furthermore, we analyze the time-series variation of the short and long-term pricing kernels to obtain another test for the asset pricing models. For this we calculate the average pricing kernel in good and bad times according to the level of the Chicago Fed National Activity Index (CFNAI). We show that the time-series variation is most pronounced for the short-term pricing kernel, while the time-series variation of the long-term kernel is significantly smaller. Standard rational asset pricing models produce only little time-series variation of the pricing kernel and are therefore in line with the dynamics of the empirical long-term pricing kernel. We show that in terms of the shape and magnitude of the time-series variation, theoretical asset pricing models match the data of the long-term pricing kernel well. In contrast, we find relatively large time-series

[2]In particular, we consider the habit formation model of Campbell and Cochrane (1999), the rare disaster model of Wachter (2013), the long-run risks model of Bansal and Yaron (2004), and the time-varying recovery model of Gabaix (2012).

[3]We refer to the one-month pricing kernel as the short-term pricing kernel and to the expected forward pricing kernel with horizon longer than six months as the long-term pricing kernel.

variation in the short-term pricing kernel and show that the U-shape is stronger in good times than in bad times.[4] The stronger U-shaped pricing kernel indicates that investors are more sensitive towards large negative and positive returns in good times. Hence, we present evidence that the short-term pricing kernel is not only puzzling from a theoretical perspective due to its shape but also due to its dynamics over time. Similar to the average shape of the short-term pricing kernel, also its time-series variation can be explained by a behavioral asset pricing model with probability weighting. In good times volatility is lower, therefore returns smaller in magnitude have lower probabilities and are, due to probability weighting, relatively more overweighted. This leads to a more pronounced U-shape in good times when volatility is low than in bad times when volatility is high.

Summarizing, we show that the pricing kernel puzzle is a unique feature of the short-term pricing kernel and our novel evidence on the shape and time-series variation of the long-term pricing kernel is well in line with the predictions of standard asset pricing models. Furthermore, we argue that the shape and time-series variation of the short-term kernel can be explained by the behavioral asset pricing model of Baele et al. (forthcoming). However, in order to fully explain our empirical results one would need an asset pricing model where our result of the U-shaped short-term pricing kernel and monotonically decreasing long-term pricing kernel arises endogeneously. One direction could be a model with maturity segmentation in the option market and heterogenous investors. Speculative investors with non-standard preferences could mostly trade short-term options whereas investors with standard preferences trade long-term options. We leave it to future research to examine the presence and potential reasons for such segmentation in option markets. Another direction would be to explore the implications of probability weighting with respect to the horizon.

Our empirical procedure to estimate the pricing kernel is conceptually similar to, among others,[5] Aït-Sahalia and Lo (2000). Aït-Sahalia and Lo (2000) exploit the result of Breeden and Litzenberger (1978) to estimate the risk-neutral distribution and model the underlying distribution non-parametrically. Our method uses a discrete-state

[4]Cuesdeanu and Jackwerth (2018b) also document that the pricing kernel has a more pronounced U-shape in times of low volatility, but their focus is mainly to document non-monotonicity of the pricing kernel in different subsamples, and they do not compare their empirical pricing kernels to kernels implied by asset pricing models.

[5]Jackwerth (2000), Rosenberg and Engle (2002), Polkovnichenko and Zhao (2013) and Song and Xiu (2016) all use conceptually the same method as ours to estimate the pricing kernel.

analogue of the result of Breeden and Litzenberger (1978), to capture the risk-neutral distribution of the S&P 500. Using options on the S&P 500 is standard in the literature, as it is a reasonable and commonly-used proxy for the market return and the options are among the most liquid traded options. We estimate the pricing kernel using portfolios of options called butterfly spreads. The prices of butterfly spreads capture the dynamics of the risk-neutral distribution. Roughly speaking, the pricing kernel follows from the quotient of the price of the butterfly spread and the expected payoff, i.e. the expected return of the butterfly spread. In order to calculate the expected payoff, we need to model the returns of the underlying S&P 500 index. We assume a skewed t-distribution with time-varying volatility. An advantage of our method is that the proxy we use for the risk-neutral distribution is an investable portfolio, which facilitates economic interpretation. Given our methodology, we estimate the pricing kernel as a function of the market return. We do this for each day in our sample period, and for different horizons, ranging from one to twelve months.

The remainder of this paper is organized as follows. Section 2 describes the literature to which we contribute into more details. In Section 3 we explain our empirical methodology and define the forward pricing kernel. Section 4 describes the data used to estimate the pricing kernels. In Section 5 we present our findings. The results are compared to the predictions of rational asset pricing models in Section 6 and to a behavioral asset pricing model in Section 7. Section 8 concludes.

2 Related Literature

We contribute to the literature on U-shaped pricing kernels. Jackwerth (2000), Aït-Sahalia and Lo (2000) and Rosenberg and Engle (2002) document that the empirical pricing kernel is not monotonically declining and exhibits a S-shape or U-shape depending on the sample period. Chabi-Yo (2012) and Song and Xiu (2016) confirm the earlier results over longer sample periods. Recently, Sichert (2019) argues that the earlier finding of S-shaped pricing kernels is due to the fact that the nonparametric models of the underlying models are unable to adequately capture stock market volatility. If volatillity is modeled appropriately the pricing kernel is consistently U-shaped. All of the aforemen-

tioned studies use the result of Breeden and Litzenberger (1978) to obtain an estimate of the risk-neutral distribution and model the real-world return distribution using parametric or nonparametric methods.

Chabi-Yo et al. (2007) and Bakshi et al. (2010) try to reconcile the findings of the non-monotonic pricing kernel with either a regime-switching model or a model with disagreeing investors. Christoffersen et al. (2013) show that a reduced form pricing kernel, which allows for a U-shape due to the negative price of variance risk, captures the option pricing data well. Polkovnichenko and Zhao (2013) estimate probability weighting functions from option data and find that the empirical probability weighting functions are such that investors over-weight small probabilities as in prospect theory by Kahneman and Tversky (1979). Baele et al. (forthcoming) show that a model with probability weighting can reconcile a U-shaped pricing kernel and explains the returns on put/call options and the variance premium well.

We also contribute to the literature on term structures of risk premiums. van Binsbergen et al. (2012) analyze the term structure of equity and Dew-Becker et al. (2017) the term structure of variance price risk. Dew-Becker et al. (2017) focus on the pricing of realized volatility and shocks to expected future volatility, and find that investors pay a premium to hedge shocks to realized short-term volatility, whereas hedging shocks to expected long-term volatility cost much less or nothing. We refer to van Binsbergen and Koijen (2017) for a recent overview fo the literature and evidence on term structures for other asset classes. This study is the first that documents stylized facts with respect to the term structure of the pricing kernel.

Linn et al. (2017) argue that the violation of monotonicity documented by earlier work in this literature is spurious. The pricing kernel follows from the real-world and risk-neutral distribution, and, as the risk-neutral distribution is derived from options, it incorporates all information upon time t. In some of the earlier mentioned studies, the real-world probability density is estimated non-parametrically on a sample of past data and therefore unable to capture all information available on time t. Linn et al. (2017) argue that if both the real-world and risk-neutral distribution contain all information on time t, the violation of monotonicity disappears. However, Cuesdeanu and Jackwerth (2018b) show that even when the pricing kernel is estimated using forward-looking data only, it is still U-shaped. In line with Cuesdeanu and Jackwerth (2018b), we find

a U-shaped short-term pricing kernel even when we condition the real-world probability distribution on forward-looking information, the VIX in our case. The level of VIX is forward-looking as it represents the integrated risk-neutral volatility over the following month. We run several specifications of our volatility model, e.g. adding lagged volatility or squared VIX term, however our results are unaffected.

3 Methodology

We estimate the pricing kernel using butterfly spreads, which is a portfolio of options. We use European options on the S&P500 as it is regarded as a reasonable proxy for the market portfolio and the option contracts are among the most traded options. First, we explain how to construct butterfly spreads from options and later on we discuss the details of our estimation procedure. A Butterfly spread can be constructed using three different call (put) options with equal distance in strike. To form a butterfly spread from call (put) options with strike K and spread ΔK, one has to take a long position in call (put) options with strikes $K - \Delta K$ and $K + \Delta K$ and a short position in two calls (puts) with strike K. The payoff of a butterfly is only positive on the interval $[K-\Delta K, K+\Delta K]$, therefore if the spread of the butterfly is small it identifies the pricing kernel well. As we will explain later on, we use out of the money options to construct the butterflies as these contracts are the most liquid.

3.1 Estimation Procedure

The method we use to estimate the pricing kernel exploits the discrete-state analogue of the result by Breeden and Litzenberger (1978). In the paper they prove that the price of an Arrow-Debreu security equals:

$$P(K,T) = \lim_{\Delta K \to 0} \frac{c(K+\Delta K,T) + c(K-\Delta K,T) - 2c(K,T)}{(\Delta K)^2} = \frac{\partial^2 c(K,T)}{\partial K^2}, \qquad (1)$$

where $P(K,T)$ represents the price of an Arrow-Debreu security paying one if the stock price at maturity equals the strike ($S_T = K$). $c(K,T)$ is the price of a call option, therefore

the numerator of equation (1) corresponds to the price of a butterfly spread with strike K, spread ΔK and maturity T. Aït-Sahalia and Lo (2000) utilize the result of Breeden and Litzenberger (1978) to estimate the risk-neutral probability density semiparametrically. It is called semiparametrically as Aït-Sahalia and Lo (2000) assume the option price to follow from Black-Scholes, where volatility is modeled nonparametrically. Together with a non-parametric model of the S&P 500 return distribution, it allows them to tease out the pricing kernel. Conceptually our methodology is similar, but we use butterfly spreads to capture the risk-neutral distribution and model the S&P 500 parameterically. The parametric model we assume for the return distribution of the S&P 500 allows us to calculate expected payoffs of butterfly spreads. From the price of the butterfly spread and the expected payoff, we are able to estimate the pricing kernel. The advantage of our methodology is that the proxy for the risk-neutral distribtution is an investable portfolio which facilitates economic intuition. High levels of the pricing kernel in the states where the butterfly spread has positive payoff, correspond to low expected returns of the butterfly spread.

The pricing kernel follows from the classical asset pricing equation: $p = \mathbb{E}(mX)$, where m is the pricing kernel and X is the payoff. Under the assumption of market completeness, the pricing kernel is unique. In order to solve for the discrete state pricing kernel, we write the price of a butterfly spread as a linear combination of the expected payoffs in each state times the corresponding discount factor. To do so, we discretize the outcome space. In this way, the states correspond to return levels of the S&P 500. Put differently, we estimate the pricing kernel as a function of the return on the S&P 500.

On a certain trading day we observe several butterfly spreads with positive payoffs on different regions of the outcome space. Without loss of generality we assume in this example that each of the observed butterflies has three discrete states in which the payoff is positive. We write the price of the butterfly as a linear combination of the expected payoff in that state times the stochastic discount factor of the particular state. In matrix notation, it is represented as follows, where p_i is the price of butterfly i, m_j the pricing kernel in state j, $x_{i,j}$ is the payoff of butterfly i in state j and b_j, b_{j+1} represent the bounds of the state j:

$$\begin{pmatrix} p_1 \\ \vdots \\ \vdots \\ p_n \end{pmatrix} = \begin{pmatrix} x_{1,1} & x_{1,2} & x_{1,3} & \cdots & 0 \\ 0 & x_{2,2} & x_{2,3} & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & \cdots & x_{n-1,n-2} & x_{n-1,n-1} & 0 \\ 0 & \cdots & x_{n,n-2} & x_{n,n-1} & x_{n,n} \\ b_1 & b_2 & \cdots & \cdots & b_n \quad b_{n+1} \end{pmatrix} \cdot \begin{pmatrix} m_1 \\ \vdots \\ \vdots \\ m_n \end{pmatrix}$$

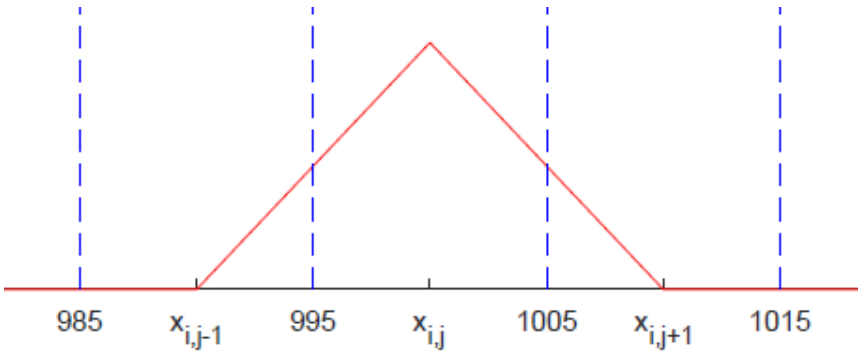
To obtain the expected payoff of the butterfly spreads for each interval, we start with a continuous distribution for the underlying stock price S_T, which we denote $f(S_T)$. The expected payoff of butterfly i in state j, with interval $[b_j, b_{j+1}]$ is then given by:

$$x_{i,j} = \int_{b_j}^{b_{j+1}} X_i(S_T) \cdot f(S_T) dS_T.$$

Whenever the number of states j equals the number of butterflies i, the expected payoff matrix is square and invertible making the pricing kernel m unique.

Using a brief example we illustrate how we discretize the state space. Consider an index with current price $S_0 = 1000$ and a butterfly i with strike $K_i = 1000$ and spread $\Delta K = 10$. Assume for now that the state space is discretized with steps of $1\% \cdot S_0 = 10$, which we graphically illustrate in Figure 1. The state where the return over horizon T is 0% corresponds to the interval $[995; 1005]$.

Figure 1: The figure plots the payoff of a butterfly spread with $K = 1000$ and $\Delta K = 10$.



Three non-zero expected payoffs over the states are important for the valuation of the butterfly spread. Expected payoff $x_{i,j-1}$ calculated over the range $[985; 995]$, expected payoff $x_{i,j}$ calculated over range $[995; 1005]$ and expected payoff $x_{i,j+1}$ calculated over

range $[1005; 1015]$. The expected payoff over the middle state is given by:

$$x_{i,j} = \int_{995}^{1005} X_i(S_T) \cdot f(S_T) dS_T,$$

where $X_i(S_T)$ is the payoff function of a butterfly with strike $K_i = 1000$ and $f(\cdot)$ is the probability density for the index value at time T. To calculate expected payoffs $x_{i,j-1}$ and $x_{i,j+1}$ the intervals are adjusted accordingly. The price of the butterfly spread is calculated by the sum over the states times the corresponding discount factor.

To calculate the expected payoff of a butterfly in a certain state, we have to model the return distribution of the S&P 500. We assume that the return distribution follows a Skewed-t distribution with the volatility being a function of the volatility index (VIX). The details of the model are discussed in Section 3.3.

3.2 The Forward Pricing Kernel

In the following we describe how we derive the forward pricing kernels from the option pricing data. Importantly, these forward kernels do not contain any short-term information and hence, it allows us to separately analyze the kernels for different horizons. We first demonstrate this approach using the example of the two-period forward kernel and then generalize the method to any horizon n. Using the methodology described above, we can estimate the n-period pricing kernel $m_{t,n}$ from an option with maturity n and price $P_{t,n}$ at time t. Given that we have options with different maturities available on a certain day, we are able to estimate a one-period pricing kernel $m_{t,1}$ and a two-period pricing kernel $m_{t,2}$ from option prices $P_{t,1}$ and $P_{t,2}$ from the following equations:

$$P_{t,1} = \mathbb{E}_t\big(m_{t,1} X_{t+1}\big), \qquad P_{t,2} = \mathbb{E}_t\big(m_{t,2} X_{t+2}\big), \tag{2}$$

where X_{t+1} and X_{t+2} are the payoffs at the end of the first and second period, respectively. We estimate the one-period and two-period pricing kernel as a function of the one-period and two-period return. Given that the distributions of the one and two-period return differ, it is difficult to compare these two pricing kernels. In order to do this in an accessible way, we define the forward pricing kernel. Estimating the forward pricing kernel

allows us to compare a one-month pricing kernel dependent on one-month returns with a forward pricing kernel dependent on one-month forward return. As the one-month return distribution and the one-month forward distribution are arguably similar, it allows us to compare the sensitivity of the investor towards one-month risks and one-month forward risks.

We derive the one-period forward pricing kernel from the one- and two-period pricing kernel in the following way. The right equation in (2) can be rewritten in the following way:

$$\begin{aligned} P_{t,2} &= \mathbb{E}_t\big(m_{t,2}X_{t+2}\big) = \mathbb{E}_t\big(m_{t,1}P_{t+1,1}\big) = \mathbb{E}_t\Big(m_{t,1}\cdot\mathbb{E}_{t+1}\big(m_{t+1,1}X_{t+2}\big)\Big) \\ &= \mathbb{E}_t\big(m_{t,1}\cdot m_{t+1,1}X_{t+2}\big), \end{aligned}$$

where $P_{t+1,1}$ is the price of the two-period option in the subsequent period. As we assume complete markets, the pricing kernels are unique. Thus, we can write the two-period pricing kernel as follows:

$$m_{t,2} = m_{t,1}\cdot m_{t+1,1}. \tag{3}$$

Using our methodology, we can estimate the one-period kernel $m_{t,1}$ and two-period pricing kernel $m_{t,2}$ from one-month and two-month options as a function of the one-month and two-month return. To compute the two-period forward pricing kernel $m_{t+1,1}$, we use equation (3) which shows that the forward pricing kernel is dependent on both the first and second-period return:

$$m_{t+1,1}\big(r_{t,t+1}, r_{t+1,t+2}\big) = \frac{m_{t,2}\big(r_{t,t+1}\cdot r_{t+1,t+2}\big)}{m_{t,1}\big(r_{t,t+1}\big)}, \tag{4}$$

where $r_{t,t+1} = \frac{S_{t+1}}{S_t}$ and $r_{t+1,t+2} = \frac{S_{t+2}}{S_{t+1}}$. To obtain the two-period forward pricing kernel as a function of the second-period return only, we calculate the expectation of $m_{t+1,1}$ with respect to the first period return, conditional on the return in the second period which

we call $m^F_{t,2}$:

$$\begin{aligned} m^F_{t,2}\big(r_{t+1,t+2}\big) &= \mathbb{E}_t\Big(m_{t+1,1}\big(r_{t,t+1}, r_{t+1,t+2}\big)\big|r_{t+1,t+2} = r\Big) \\ &= \int_0^\infty \frac{m_{t,2}(r_{t,t+1}\cdot r)}{m_{t,1}(r_{t,t+1})} \cdot f_t\big(r_{t,t+1}\big|r_{t+1,t+2} = r\big) dr_{t,t+1}. \end{aligned} \tag{5}$$

Here $f_t(\cdots|\cdots)$ denotes the one-period return distribution conditional on time t and second-period return $r_{t+1,t+2}$. To be precise, $m^F_{t,2}$ is the expected two-period forward pricing kernel and is estimated on time t. All forward pricing kernels we consider have a horizon of one-period, therefore we omit the horizon subscript.

Equation (5) can be generalized in the following way. Let $m^F_{t,n}$ be the expected n-period forward pricing kernel at time t which is given by:

$$\begin{aligned} m^F_{t,n}\big(r_{t+n-1,t+n}\big) &= \mathbb{E}_t\Big(m_{t+n-1,1}\big(r_{t,t+n-1}, r_{t+n-1,t+n}\big)\big|r_{t+n-1,t+n} = r\Big) \\ &= \int_0^\infty \frac{m_{t,n}(r_{t,t+n-1}\cdot r)}{m_{t,n-1}(r_{t,t+n-1})} f_t\big(r_{t,t+n-1}\big|r_{t+n-1,t+n} = r\big) dr_{t,t+n-1}, \end{aligned} \tag{6}$$

where $f_t(\cdots|\cdots)$ denotes the $(n-1)$-period return distribution conditional on time t and the return in period n, $r_{t+n-1,t+n}$. In the following we refer to $m^F_{t,n}$ as the n-period forward pricing kernel.

3.3 Return Distribution of the S&P 500

In order to estimate the pricing kernel, we need to model the return distribution of the S&P 500. For this we use the skewed t-distribution proposed by Bauwens and Laurent (2002) to model S&P 500 returns. Instead of the constant volatility used in the original paper, we introduce time-varying volatility to adequately capture volatility dynamics of the stock market:

$$r_{t,t+n} = \mu_n + \sigma_{t,t+n}\epsilon_{t,t+n}, \tag{7}$$

$$\sigma_{t,t+n} = \alpha_n + \beta_n VIX_t, \tag{8}$$

where $r_{t,t+n}$ are n-month returns and the standard deviation is a linear function of the VIX at time t. We use this model as our benchmark specification.[6] We assume that the standard deviation of the S&P 500 return distribution is a linear function of the model-free risk-neutral volatility introduced by Britten-Jones and Neuberger (2000). The CBOE introduced the VIX index for which it uses the result of Britten-Jones and Neuberger (2000) to estimate risk-neutral volatilities from monthly options. In the estimation of our model for the S&P 500, we scale the yearly VIX (252 trading days) to the corresponding maturity. Furthermore, the random variable $\epsilon_{t,t+n} \sim SKST(0, 1, \xi_n, v_n)$ follows a standardized skewed t-distribution with parameters $v > 2$ (degrees of freedom) and $\xi > 0$ (skewness parameter). In the following, we drop the subscript n in the definition of the skewed t-distribution for convenience. The density is given by:

$$f(\varsigma_t|\xi, v) = \begin{cases} \frac{2}{\xi+\frac{1}{\xi}} \cdot s \cdot g\left[\xi(s\varsigma_t + m)|v\right] & \text{if} \quad \varsigma_t < -\frac{m}{s} \\ \frac{2}{\xi+\frac{1}{\xi}} \cdot s \cdot g\left[\frac{1}{\xi}(s\varsigma_t + m)|v\right] & \quad\quad \varsigma_t \geq -\frac{m}{s}, \end{cases} \tag{9}$$

where $g(\cdot|v)$ is a symmetric (zero mean and unit variance) Student t-distribution with v degrees of freedom, defined by

$$g(x|v) = \frac{\Gamma\left(\frac{v+1}{2}\right)}{\sqrt{\pi(v-2)}\Gamma\left(\frac{v}{2}\right)}\left[1 + \frac{x^2}{v-2}\right]^{-(v+1)/2}, \tag{10}$$

where $\Gamma(\cdot)$ is the gamma function. The constants $m(\xi, v)$ and $s(\xi, v)$ are the mean and standard deviation of the non-standardized skewed t-distribution, $SKST(m, s^2, \xi, v)$, and given by:

$$m(\xi, v) = \frac{\Gamma\left(\frac{v-1}{2}\right)\sqrt{v-2}}{\sqrt{\pi}\Gamma\left(\frac{v}{2}\right)}\left(\xi - \frac{1}{\xi}\right), \tag{11}$$

$$s^2(\xi, v) = \left(\xi^2 + \frac{1}{\xi^2} - 1\right) - m^2. \tag{12}$$

A skewed t-distribution supports important stylized facts of index returns, namely left-skewed and fat-tailed distributions. The ξ is a skewness related parameter and facilitates economic interpretation. The probability mass right from the mode divided by the probability mass left from the mode equals $\frac{1}{\xi^2}$. Therefore, when $\xi < 1$, the

[6] We also consider other models for volatility, e.g. adding a squared VIX term or lagged realized volatility and find that the model predicts similar volatility levels.

distribution is left skewed and when $\xi > 1$ the distribution is right skewed. Fat tails are related to the parameter of the degrees of freedom v corresponding to the t-distribution. The lower v the fatter the tails and when $v \to \infty$ the distribution converges to a skewed-normal distribution.

We estimate the parameters of equations (7)-(8) using Maximum Likelihood on n-month overlapping returns of the S&P 500 from 1990 to 2016. Our methodology allows us to estimate the pricing kernel for different maturities if we adjust the maturity of the options and the horizon of the return distribution accordingly. As we are interested in the term structure of the pricing kernel, we estimate the pricing kernel for different horizons n, where n ranges from one up to twelve months. The estimates for the one month S&P 500 return distribution are represented in Table 1.

Table 1: The table reports the maximum likelihood estimates of the parameters of the skewed t-distribution with volatility being a linear combination of the VIX. The estimates in the first line correspond to the parameters of equations (7)-(8). The second line adds a squared VIX term (γ_n) to the volatility equation. The parameters are estimated using monthly overlapping simple returns of the S&P 500 from 1990-2016.

μ_1	α_1	β_1	γ_1	ξ_1	v_1
0.0058	-0.0066	0.8488	-	0.7022	14.0910
0.0058	-0.0148	1.1629	-2.6792	0.7028	14.6734

The parameter μ_1 captures the sample average of the one-month returns. Given the estimates of α_1 and β_1 of the first row, we conclude that the overall level of the VIX is larger than the realized volatility of the S&P 500 index. In the literature, this phenomenon is known as the variance premium and first documented by Bollerslev et al. (2009). The estimate of $\xi_1 < 1$ indicates that the return distribution is left-skewed and given the interpretation of ξ_1 it is economically sizable. Furthermore, there is evidence of fat tails given the estimate of the degrees of freedom v_1.

We model the return distribution in order to calculate expected payoffs of the butterfly spreads. To validate the underlying model, we compare expected with realized payoffs of butterflies in our sample from 1996-2016. As we will explain in the Section 4, we get the data on options from OptionMetrics and construct butterflies using raw option data with $\Delta K = 10$. A butterfly is constructed with three options with consecutive difference of 10 index points in strike. The moneyness of the butterfly is calculated

dividing the strike of the butterfly by the current index value (S_0). For the butterflies in the sample we can calculate the expected payoff using our return distribution of the S&P 500. The expected payoff of a butterfly with fixed spread is decreasing in S_0. We account for this by normalizing the realized payoff and the expected payoff with S_0. In order to asses the ability of our return distribution to capture the expected payoff of S&P 500 butterfly spreads, we calculate for each moneyness level the time-series average of the realized payoff divided by S_0 and the time-series average of the expected payoff divided by S_0. The results are shown in Figure 2.

Figure 2: The figure plots time-series averages of realized and expected payoffs of S&P 500 butterfly spreads with maturity of one month and $\Delta K = 10$ as a function of the moneyness. The solid black line shows the time-series average of the realized payoff. The dashed black line shows the expected payoff using equations (7)-(8) for the unterlying return distribution. The dash-dot line adds a squared VIX term to the volatility equation. Results are shown for the full sample ranging from 1996-2016.

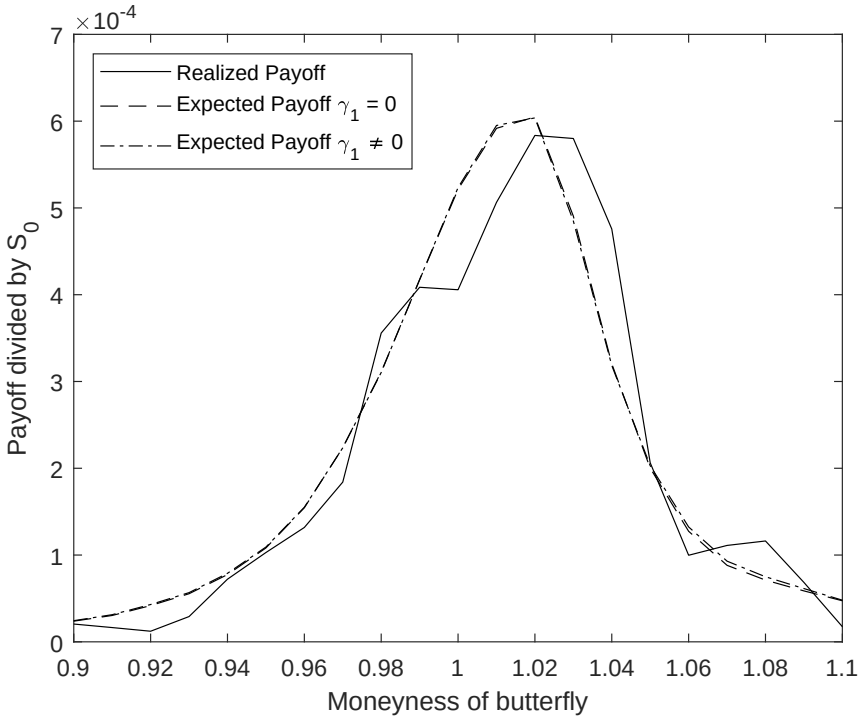


Figure 2 has two important takeaways. First, the expected payoff matches the realized payoff quite well, indicating that we capture the underlying distribution accurately. Second, the expected payoff of the butterflies hardly changes when we add a squared VIX term to the volatility equation. The only difference is that the squared term in the volatility equation makes expected volatility lower for extreme values of the VIX.

As discussed in Section 3.2, we derive the one-month forward pricing kernel from the ratio of the two-month pricing kernel and the one-month pricing kernel and

we proceed correspondingly for longer maturities. We present evidence of the expected forward pricing kernel, by calculating the expectation with respect to the return in the first month, conditional on the return in the second month. The simplest setup for doing this, would be to assume independence of the returns. However, this assumption is quite restrictive and given that variance of the stock market is persistent it does not hold up in the data. This persistence yields the following intuition: if the return in the second period is large in magnitude, it is more likely that the return is drawn from a distribution with large variance, implying that the variance of the first period return is also likely to be larger. For the first n-month return, we use the model given by equations (7)-(8). To model the dependence in variance of the first n-month returns and the one-month return n-periods forward, we assume the following:

$$r_{t,t+n} = \mu_n + \sigma_{t,t+n}\epsilon_{t,t+n}, \tag{13}$$

$$\sigma_{t,t+n} = \alpha_n + \beta_n VIX_t, \tag{14}$$

$$r_{t+n,t+n+1} = \mu_1 + \sigma_{t+n,t+n+1}\epsilon_{t+n,t+n+1}, \tag{15}$$

$$\sigma_{t+n,t+n+1} = \alpha_1 + \beta_1 VIX_{t+n}, \tag{16}$$

$$VIX_{t+n} = \mu_\sigma + \lambda \cdot (VIX_t - \mu_\sigma) + \rho_1\epsilon_{t,t+n} + \rho_2\epsilon^2_{t,t+n} + \sigma_u u_{t,t+n}. \tag{17}$$

In order to calculate the conditional distribution, we assume that $\epsilon_{t,t+n} \sim \text{SKST}(0, 1, \xi_n, v_n)$, $\epsilon_{t+n,t+n+1} \sim \epsilon_{t,t+1} \sim \text{SKST}(0, 1, \xi_1, v_1)$, $u_{t,t+n} \sim N(0, 1)$ and all errors are independently distributed. The intuition of the conditioning works as follows, if $r_{t+n,t+n+1}$ is large in magnitude, it is likely drawn from a distribution with large volatility. In the model volatility is driven by VIX_{t+n}. In order for the VIX_{t+n} to be large, $\epsilon_{t,t+n}$ has to be large implying that the volatility of $r_{t,t+n}$ is large. In this way, the persistence of stock market volatility is taking into account.

Table 2: The table reports in Panel A the Maximum Likelihood Estimates of the parameters of the skewed t-distribution with volatility linear function of the VIX, equations (7)-(8). These parameters are estimated using n-month overlapping returns of the S&P 500 from 1990-2016. In Panel B the estimates for the VIX process with n periods forward, equations (15)-(17).

Panel A

n	μ_n	α_n	β_n	ξ_n	v_n
1	0.0058	−0.0066	0.8488	0.7022	14.0911
5	0.0337	−0.0090	0.8055	0.7751	7.6392
6	0.0400	−0.0060	0.8054	0.7568	6.1238
11	0.0790	0.0296	0.6483	0.8026	6.5666
12	0.0867	0.0434	0.5977	0.8027	6.3923

Panel B

n	μ_σ	λ	ρ_1	ρ_2	σ_u
5	0.0517	0.5105	−0.0094	0.0049	0.0120
11	0.0505	0.3522	−0.0084	0.0060	0.0157

The results of maximum likelihood estimation of model (13)-(17) are reported in Table 2 for different maturities n. Using the results of Panel A, we can estimate with our methodology the n-month pricing kernel. From the 6-month and 5-month pricing kernel, we determine the 6-month forward pricing kernel $\left(m^F_{t,6}\right)$ for which we need the conditional distribution with estimates in Panel B of Table 2. In a similar way we estimate the 12-month forward pricing kernel $\left(m^F_{t,12}\right)$.

As expected, the average return estimates (μ_n) of Panel A in Table 2 increases with maturity. Even though the VIX equals the model-free one-month risk neutral volatility, it does capture time-variation of the volatility for return distributions with maturities beyond one month, indicated by the postive estimate of β_n. The skewness related parameter goes up when the maturity is longer, showing that the left-skew of the distribution becomes less pronounced for longer maturity. Lastly, for each of the maturities we find evidence of fat tails given that the degrees of freedom are rather small.

In Panel B of Table 2 we estimate the VIX process in order to derive the conditional distribution. μ_σ is approximately the average monthly VIX.[7] As λ is positive it indicates some persistence in the VIX process, and as expected the λ is lower when

[7]To be precise: $\mathbb{E}\big(VIX_t\big) = \mu_\sigma + \frac{\rho_2}{1-\lambda}$.

predicting VIX_{t+n} for the more distant future. ρ_1 indicates that, if returns over n periods are negative (positive), the VIX n months from now increases (decreases)—a phenomenon also known as the leverage effect. ρ_2 indicates that, if the volatility of the return of the first n-months is large, the VIX n-months from now increases. In this way, λ, ρ_1 and ρ_2 make volatility of the stock market persistent as the VIX n-months from now increases when the VIX is currently high or for large negative n-month ahead returns.

4 Data

We use European options on the S&P 500 available on OptionMetrics to construct the volatility surface from 1996-2016. One of the reasons to make use of the surface is to address noise in the data and make our methodology less vulnerable to outliers. Furthermore, the volatility surface allows us to obtain a daily estimate of the n-month pricing kernel. The volatility surface of OptionMetrics represents options with a delta of 20% to 80% for call and -20% to -80% for put options. One of our goals is to explore the behavior of the pricing kernel in the tails of the distribution, therefore we use the same kernel smoothing algorithm as used by OptionMetrics to obtain estimates up to a delta of ($\pm$) 0.1%. Furthermore, we use out of the money options for the construction of the butterfly spreads as these are usually more liquid than in the money options.

On every trading day for which option prices are observed we estimate the pricing kernel over the following interval $\left[K_{t,\min}; K_{t,\max}\right]$, where $K_{t,\min}$ and $K_{t,\max}$ are the smallest and largest strike of the butterfly spreads on day t. We choose the amount of states on which the pricing kernel is estimated to be stable at 20 estimates each day, which at the same time pins down the distance between two consecutive states. In order for the pricing kernel to be unique we use 20 different butterfly spreads. The spread of the butterfly (ΔK) is set to be equal to the difference in strike of a call options with delta 0.35 and 0.50. In this way the spread scales with maturity, volatility and the level of the index. The states are distributed over the interval as follows: relatively twice as many states lay in the interval of one-standard deviation around the mode of the skewed-t

distribution. The confidence interval is calculated in the following way:

$$\left[m_{t,n} - \frac{c_n \cdot \sigma_{t,n}}{1+\xi_n^2}; m_{t,n} + \frac{c_n \cdot \sigma_{t,n}}{1+\frac{1}{\xi_n^2}} \right],$$

where $m_{t,n}$ is the mode of the n-month return distribution at time t, $c_n = 2 \cdot t^{-1}(0.84, v_n)$ comes from a t-distribution with v_n degrees of freedom and $\sigma_{t,n}$ is the predicted volatility from our model. The described interval is the skewed-t equivalent of the interval $[\mu - \sigma; \mu + \sigma]$ of a normal distribution. Outside the interval with the largest probability mass, the relative amount of states goes down by a factor two, or equivalently, the step-size between intermediate states goes up by a factor of two. In order to keep the number of states for which a butterfly spread has a positive payout stable, we scale the spread of the butterfly by a factor of two outside the abovementioned interval. We provide further details regarding the empirical strategy in Appendix A.1.

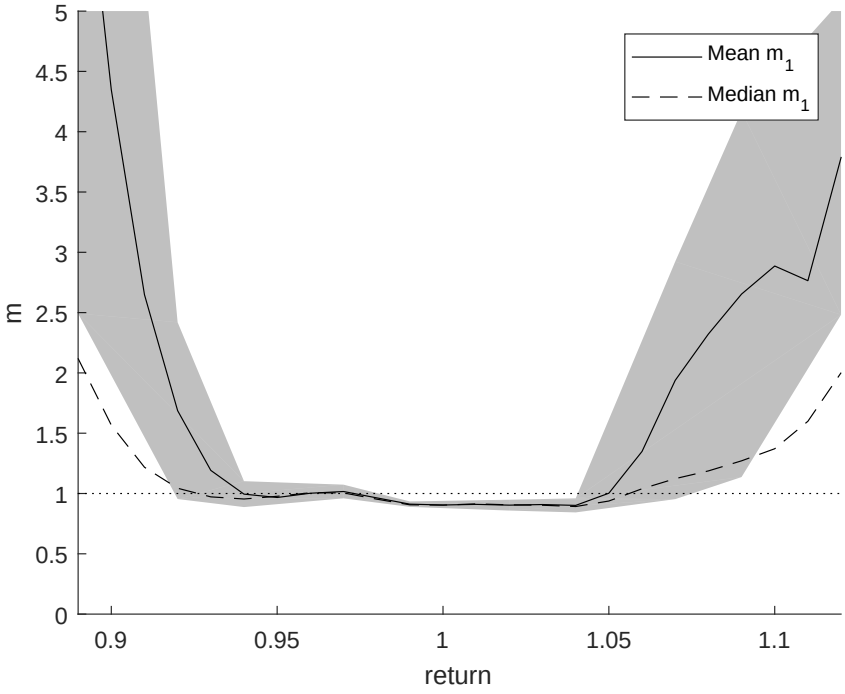
5 Results

We begin our analysis by estimating the one-month pricing kernel and we show in Section 5.1 that, in line with previous research, the short-term pricing kernel is strongly U-shaped. In Section 5.2 we analyze the term structure of the pricing kernel and we show that the pricing kernel puzzle disappears for longer horizons. While we still find a small violation of monotonicity for the six-month forward pricing kernel (but no U-shape), the twelve-month forward pricing kernel is strictly downward sloping. In Section 5.3 we report the time-series variation of the pricing kernels. We find that the short-term pricing kernel exhibits significant time-series variation and the U-shape is more pronounced in good times. The six and twelve-month forward pricing kernel on the other hand exhibit relatively little time-series variation. In Section we compare these findings to the predictions of several asset pricing models.

5.1 The Pricing Kernel Puzzle

First, we estimate the one-month kernel as it is commonly done in the literature. Figure 3 plots the time-series mean and median of the one-month pricing kernel as a function of the return.[8]

Figure 3: The figure plots the time-series average (solid) and median (dashed) of the one-month pricing kernel. The shaded area in the figure indicate 95% confidence interval using Newey-West standard errors. Note, for each moneyness level 1% highest and lowest estimates are discarded.



In line with previous studies, we find that the one-month pricing kernel is U-shaped and hence violates monotonicity. Researchers have called this phenomenon the pricing kernel puzzle (see for example Jackwerth (2000), Aït-Sahalia and Lo (2000), Rosenberg and Engle (2002), Chabi-Yo (2012) and Song and Xiu (2016)). The economic implication of the U-shape is that investors are willing to pay more for instruments which pay out either in bad or good states of the economy. Bad states are characterized by monthly returns being negative and good states vice versa. Economically, an investor is

[8]The range of return levels for which we report the pricing kernel corresponds to a 99% confidence interval based on a normal distribution with mean $\mu_1 = 0.58\%$ and average volatility $\bar{\sigma}_1 = 4.5\%$. The estimates for μ_1 and $\bar{\sigma}_1$ follow from our model of the underlying distribution, equations (7)-(8). We choose a 99% confidence interval in order to account for the fact that the empirical return distribution is skewed and heavy-tailed.

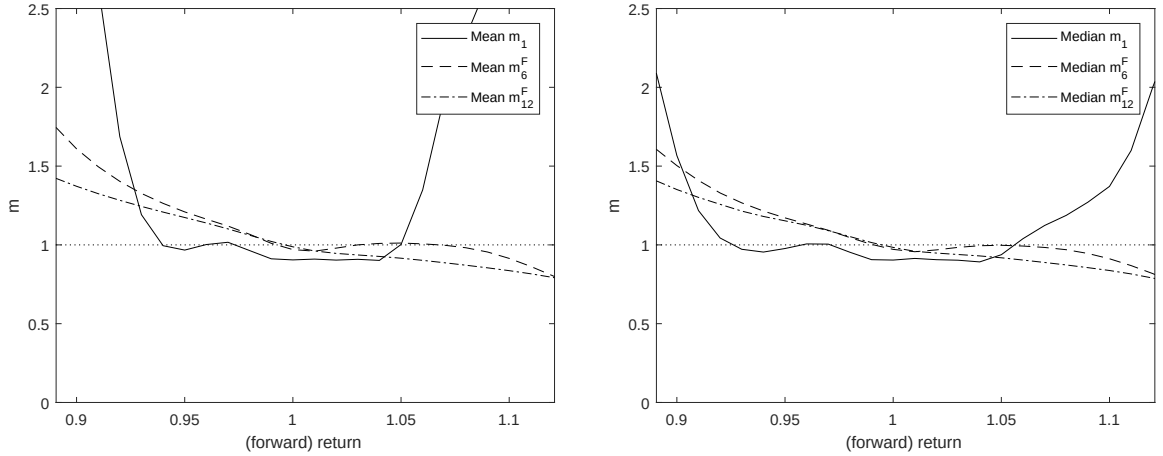
willing to pay about 1.5 times (median) the expected value for an Arrow-Debreu security paying when the stock market looses 10% (0.90 in the figure), which is equivalent to accepting an expected return of -33%. In the good states of the economy, the pricing kernel slopes upward for monthly stock returns greater than 5%. This result can be characterized as a preference of investors for betting on (extremely) good states of the economy and willing to accept a negative expected return. The difference between the mean and median increases more towards the tails of the return distribution, indicating that the time-series distribution of the pricing kernel for a given return level in either of the tails is right-skewed. This suggests that there is time-series variation of the pricing kernel which we analyze in detail in Section 5.3.

5.2 The Term Structure of the Pricing Kernel

Using the methodology described in Section 3.2, we estimate forward pricing kernels. We first compare the results for the one-month pricing kernel, the six-month forward pricing kernel $\left(m_{t,6}^{F}\right)$ and the twelve-month forward pricing kernel $\left(m_{t,12}^{F}\right)$. The latter two are derived from the five- and six-month pricing kernel and the eleven- and twelve-month pricing kernel, respectively. To compute the forward kernels, we use the estimates for the return distribution reported in Table 2. Later on in this section we show that the results are robust when we use different maturities. Figure 4 shows the means and medians of the one-month, the six-month forward and twelve-month forward pricing kernels.

Figure 4: The figure plots on the left (right) the time-series mean (median) of the one, six and twelve-month (forward) pricing kernel. The solid, dashed and dot-dashed line represent the one, six and twelve month forward pricing kernel, respectively. Forward pricing kernels are estimated using the methodology described in section 3.2.

(a) Mean Pricing Kernel. (b) Median Pricing Kernel.



As reported in the previous section, the one-month kernel is strongly U-shaped. This U-shape disappears for longer horizons and the pricing kernels approach to be monotonically decreasing. For the six-month forward kernel we find a small violation of monotonicity, whereas the twelve-month forward kernel slopes downward monotonically. Investors are sensitive towards extreme bad and good returns for one-month horizon, whereas for longer horizons the investors are only sensitive towards large negative returns. Hence, we show that the pricing kernel puzzle, which has been discussed by Jackwerth (2000), Aït-Sahalia and Lo (2000), Rosenberg and Engle (2002), Chabi-Yo (2012) and Song and Xiu (2016), vanishes for longer horizons. Note, in Appendix A.2 we show that the differences between one, six and twelve-month (forward) pricing kernel are statistically significant.

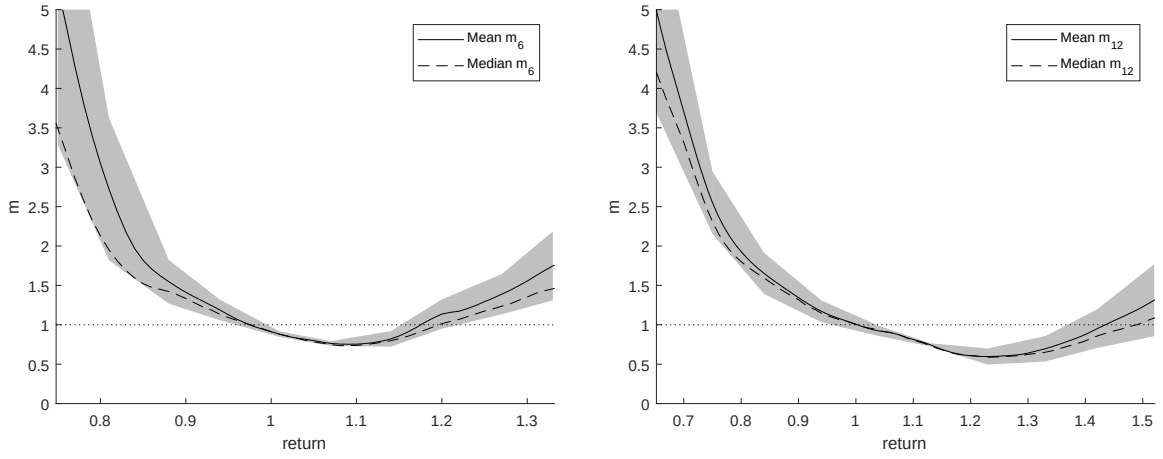
The difference between the one-month pricing kernel and the six and twelve-month forward pricing kernel implies the following: realized one-month risk is priced differently from forward six-month and twelve-month risk. Especially in the tails, the difference between the one-month pricing kernel and the forward pricing kernels is large. The evidence suggests that investors are highly sensitive towards large good or bad returns for short-horizons, whereas for longer horizons the investor is only sensitive towards large negative forward returns. The main difference in shape between the six-month and twelve-month forward pricing kernel is the relatively small violation of monotonicity we document for the six-month forward pricing kernel. As we will discuss in details later on in this section, the violation of monotonicity disappears gradually when the maturity of the

forward pricing kernel increases.

To illustrate in more detail how the shapes of the pricing kernel lead to the presented forward kernels, we also present the pricing kernels over the total six-month and twelve-month periods. We estimate these pricing kernels as a function of the six-month and twelve-month return, respectively.

Figure 5: The figure on the left (right) plots the time-series mean and median of the six (twelve)-month pricing kernel. The solid (dashed) line represent the time-series mean (median). The shaded area in the figure indicate 95% confidence interval using Newey-West standard errors. Note, for each moneyness level 1% highest and lowest estimates are discarded.

(a) Six-month Pricing Kernel. (b) Twelve-month Pricing Kernel.



The results for the six and twelve-month pricing kernel are shown in Figure 5.[9] We observe that both kernels are U-shaped. However, the U-shapes are less pronounced compared to the one-month pricing kernel (see Figure 3). The long-term pricing kernel is the product of the short-term pricing kernel and the forward kernels. As the forward pricing kernels decrease almost monotonically, we can conclude that the violation of monotonicity found in the long-term pricing kernels is mostly driven by the shape of the short-term pricing kernel. Furthermore, the difference between the mean and median is much smaller for the six and twelve-month pricing kernel than it is for the one-month pricing kernel. This indicates that the time-series variation of the short-term pricing kernel is much larger than it is for the long-term pricing kernel. We confirm this finding and discuss it in more detail in Section 5.3.

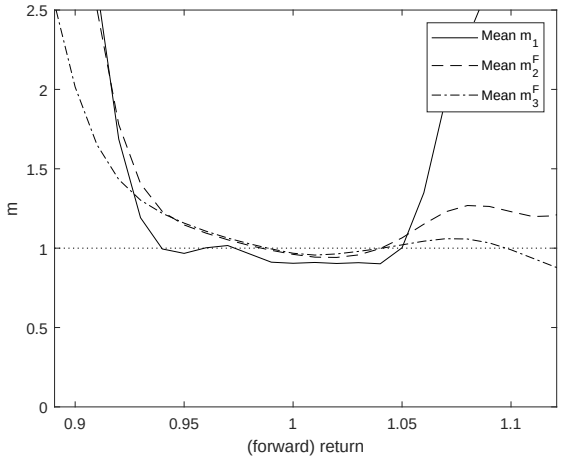
To dissect our results with respect to the horizon in more detail, we consider the full term structure of the pricing kernel up to twelve months and show that the U-shape disappears gradually with the maturity. First, we compare the short-end of the pricing

[9]The range on which we show the pricing kernels is computed in the same way as for the one-month pricing kernel.

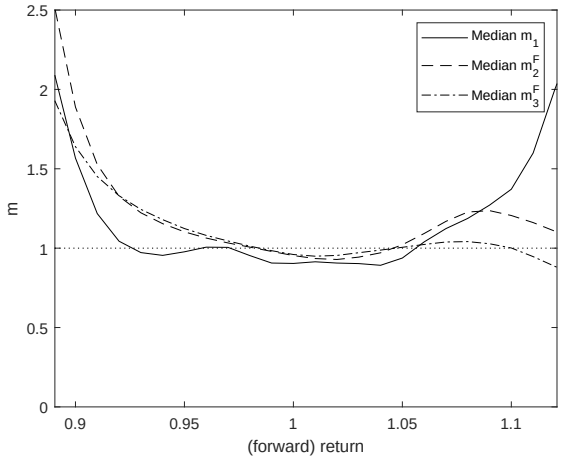
kernel for horizons up to three months.

Figure 6: The figure plots on the left (right) the time-series mean (median) of the one, two and three-month (forward) pricing kernel. The solid, dashed and dot-dashed line represent the one, two and three-month forward pricing kernel, respectively. Forward pricing kernels are estimated using the methodology described in section 3.2.

(a) Mean Pricing Kernel.



(b) Median Pricing Kernel.

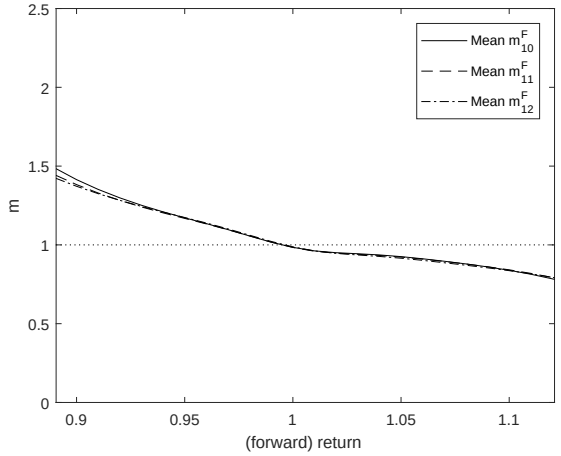


The left graph of Figure 6 indicates that the violation of monotonicity is strongest for the one-month pricing kernel. The two- and three-month forward pricing kernel violate monotonicity, but the strong U-shape has disappeared. Furthermore, the level of the forward pricing kernel for low returns goes down quite substantially for the two- compared to the three-month forward pricing kernel. This indicates that investors become less willing to accept negative returns to hedge bad states of the economy for more distant forward risks.

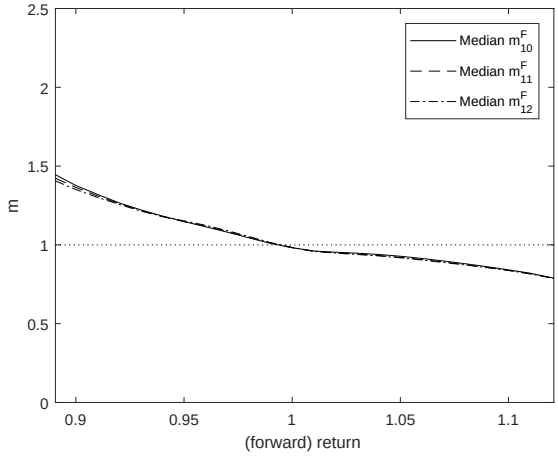
Accordingly, we also compare the long maturity forward pricing kernels using the ten-, eleven- and twelve-month forward pricing kernel which are depicted in Figure 7.

Figure 7: The figure plots on the left (right) the time-series mean (median) of the ten, eleven and twelve-month (forward) pricing kernel. The solid, dashed and dot-dashed line represent the ten, eleven and twelve-month forward pricing kernel, respectively. Forward pricing kernels are estimated using the methodology described in section 3.2.

(a) Mean Pricing Kernel.



(b) Median Pricing Kernel.



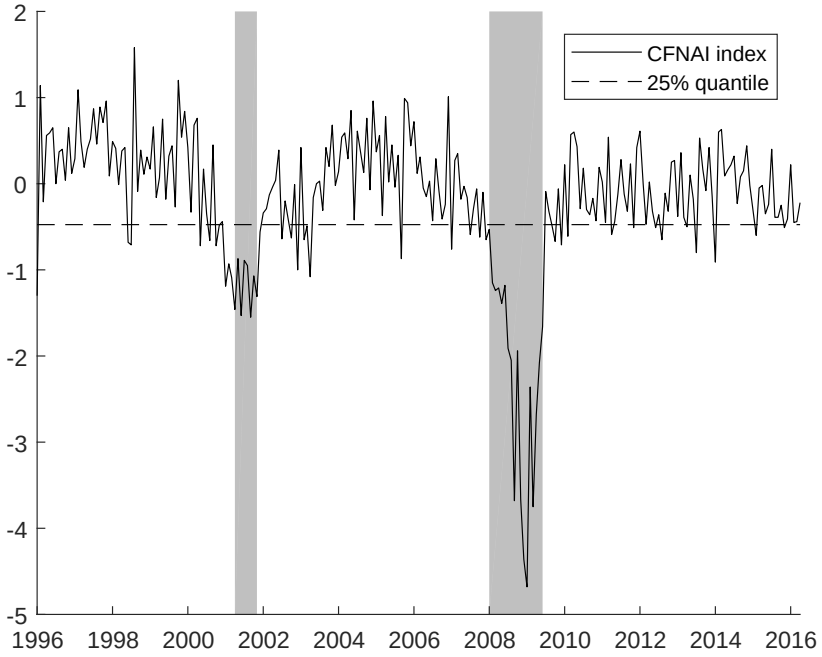
The long-term forward pricing kernels are very similar and all are monotonically decreasing in the market return. We conclude that the strong U-shape and hence the pricing kernel puzzle is a unique feature of the short-term pricing kernel. For longer horizons the puzzle gradually disappears and the pricing kernel becomes monotonically decreasing. We argue in Section 6 that the empirical properties of the long-term forward pricing kernels are well in line with standard asset pricing models. See Appendix A.2 for the other forward pricing kernels up to twelve months.

5.3 Time-Series Variation of the Pricing Kernel

In this section we provide results on the time-series variation of the pricing kernel which can be used as an additional test for asset pricing models. The previous section reports a significant difference between the mean and median of the one-month pricing kernel which indicates possible time-series variation. In the following we first analyze the empirical features of the time-series variation of the pricing kernel and later compare it to the predictions of prominent asset pricing models. A simple way to do this is to compare the absolute size of the time-series variation found in the data with the predictions of the models. For this, we define subsamples and compare the means of the pricing kernel in each of the subsamples. We use the CFNAI, which is a business cycle indicator issued by the Chicago Fed at the beginning of each month, to divide the sample. We define bad

(good) times as periods when the CFNAI is lower (larger) than the 25% quantile over the period of 1996-2016. We choose the 25% cut-off in order for the periods to be sufficiently bad. Figure 8 shows the time-series of the CFNAI of our sample period.

Figure 8: The figure plots the level of the CFNAI index at the beginning of every month. The dashed line corresponds to the 25% quantile over our sample period. The shaded area corresponds to the NBER-recessions.

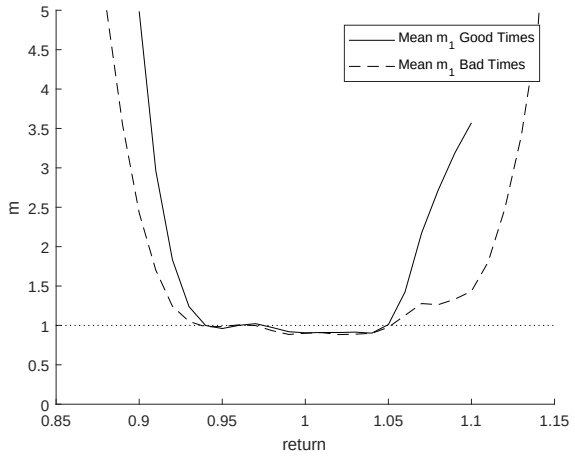


The shaded area in Figure 8 correspond to the NBER recessions. We observe that the two long recessions indicated by the CFNAI coincide with the NBER recessions which suggests that our methodology is appropriate. The advantage of using the CFNAI over NBER recessions is however, that the CFNAI is available at the beginning of each month while the NBER recessions are specified ex post.

We calculate the mean (median) pricing kernel for each of the maturities in good and bad times. As before, we plot the mean and median of the pricing kernel over the range corresponding to a 99% confidence interval if the return distribution would have been normally distributed. The only difference is that we calculate the average volatility according to our return distribution in good and bad times separately. The volatility in good times is lower than in bad times, therefore the average and median pricing kernel in good times is represented on a smaller range.

Figure 9: The left (right) figure plots the time-series mean (median) of the one-month pricing kernel conditional on being in good and bad times. The solid and dashed line represent the good and bad times, respectively.

(a) Time-series variation of the mean.



(b) Time-series variation of the median.

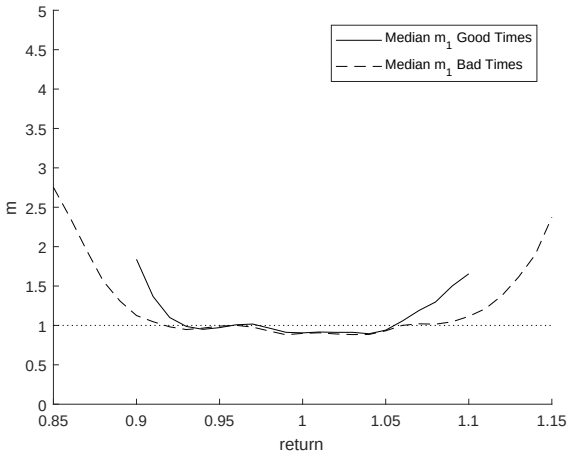
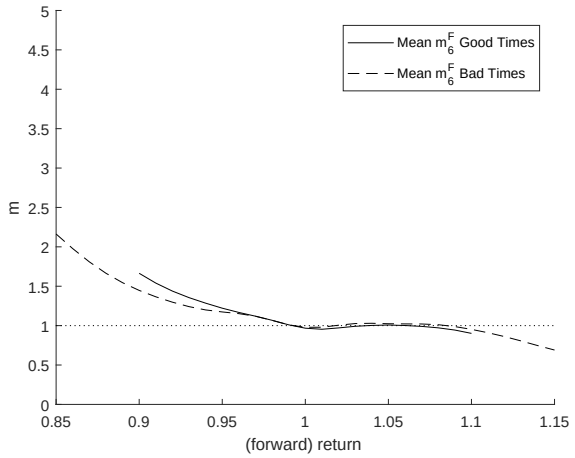


Figure 9 plots the mean and median of the one-month pricing kernel over the different subsamples. We find that the U-shape for both, the mean and median, is stronger in good times than in bad times. For returns larger than 5% in absolute value, the investor is willing to pay significantly more to hedge against (or bet on) extreme events in good than in bad times. As we show in Section 6, capturing the magnitude of the time-series variation provides a new challenge (beyond the U-shape) to standard asset pricing models which predict only very little time-series variation. In the following we repeat the analysis for the six and twelve month forward pricing kernels.

Figure 10: The left (right) figure plots the time-series mean (median) of the six-month forward pricing kernel conditional on being in good and bad times. The solid and dashed line represent the good and bad times, respectively.

(a) Time-series variation of the mean.



(b) Time-series variation of the median.

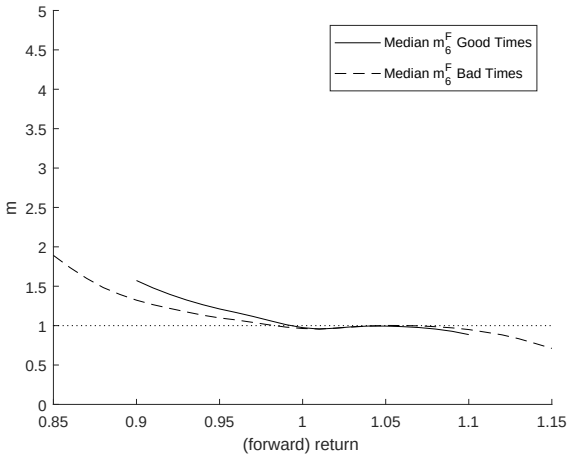
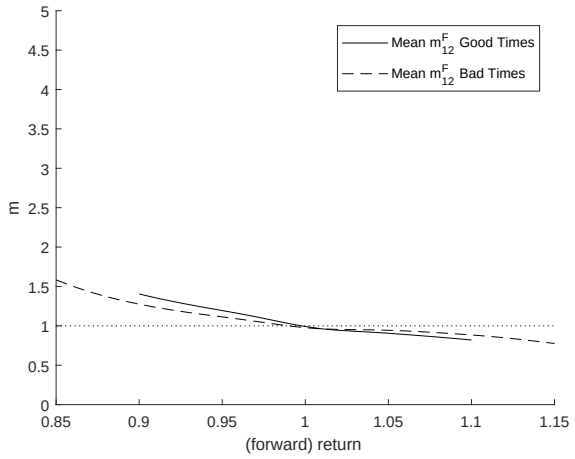


Figure 10 plots the mean and median of the six-month forward pricing kernel in good

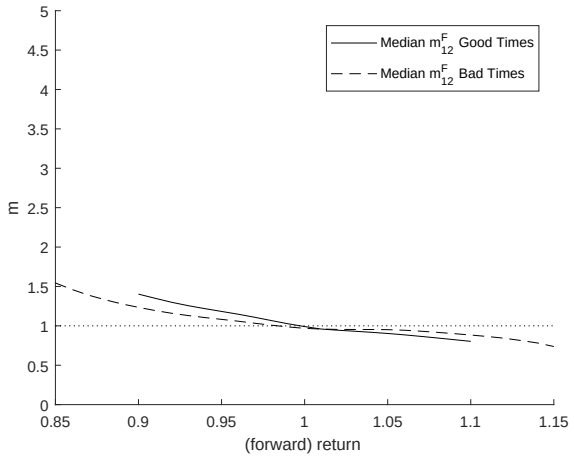
and bad times. The level of the pricing kernel increases only marginally in good times compared to bad times for low returns. Given the relative size of the time-series variation of the six-month forward pricing kernel compared to the one-month pricing kernel, the results suggest that the long-term pricing kernel exhibits only little time-series variation.

Figure 11: The left (right) figure plots the time-series mean (median) of the twelve-month forward pricing kernel conditional on being in good and bad times. The solid and dashed line represent the good and bad times, respectively.

(a) Time-series variation of the mean.



(b) Time-series variation of the median.



The results for the twelve-month kernel support this claim (see Figure 11). In good times the level of the forward pricing kernel is slightly higher compared to bad times. But the overall magnitude of the time-series variation further decreases compared to the variation of the six-month kernel. Hence, the strong time-series variation of the pricing kernel seems again to be a unique feature of the short-term kernel while the long-term kernels show relatively little variation.

In order to test the significance of the change of curvature, we do a regression of the slope of the pricing kernel in a certain region on a dummy variable equaling one in good times. We analyze four different slopes which are defined in the following way:

$$\begin{aligned}
\mathrm{s}_1(T) &= m_T\big(1.00\big) - m_T\big(0.90\big),\\
\mathrm{s}_2(T) &= m_T\big(1.00\big) - m_T\big(0.95\big),\\
\mathrm{s}_3(T) &= m_T\big(1.05\big) - m_T\big(1.00\big),\\
\mathrm{s}_4(T) &= m_T\big(1.10\big) - m_T\big(1.00\big).
\end{aligned}$$

The slope is dependent on the horizon of the pricing kernel (T) and again we distinguish bewteen a horizon of one, six and twelve months.

Table 3: The table reports regressions of the slope of the kernel on the time-series dummy indicating one in good times. Time-series regression: $\text{slope}_i(T) = \beta_0 + \beta_1 \cdot \text{good} + \epsilon$. In brackets the hac t-statistics are given, using the appropiate amount of lags according to the ACF of the residuals.

		1-month		6-month		12-month	
$s_1(T)$	β_0	-3.49 (-2.55)	-1.53 (-2.04)	-0.64 (-6.40)	-0.47 (-6.94)	-0.38 (-8.13)	-0.30 (-8.73)
	β_1	–	-2.61 (-2.06)	–	-0.23 (-2.22)	–	-0.11 (-2.89)
$s_2(T)$	β_0	-0.06 (-3.10)	-0.09 (-5.40)	-0.24 (-14.88)	-0.19 (-7.30)	-0.19 (-6.04)	-0.14 (-5.91)
	β_1	–	0.03 (1.25)	–	-0.06 (-1.79)	–	-0.07 (-3.11)
$s_3(T)$	β_0	0.10 (1.48)	0.08 (1.62)	0.04 (3.41)	0.05 (2.97)	-0.07 (-1.93)	-0.03 (-1.50)
	β_1	–	0.02 (0.49)	–	-0.00 (-0.25)	–	-0.05 (-2.14)
$s_4(T)$	β_0	1.99 (2.42)	0.53 (3.40)	-0.06 (-3.66)	-0.03 (-2.03)	-0.15 (-2.94)	-0.09 (-3.02)
	β_1	–	2.17 (1.94)	–	-0.04 (-1.68)	–	-0.08 (-2.20)

Table 3 shows the results. The first column of each T-month slope corresponds to the total sample average. The second column adds a good times dummy equaling one when the CFNAI is larger than the 25% quantile over our sample. The results show the statistical significance of the change in slope over time. For the one-month kernel, we observe that the slope for negative market returns, $s_1(1)$, is driven by the highly negative slope in good times. Furthermore, the U-shape, or positive slope for positive market returns, $s_4(1)$, is also driven by the large slope in good times. The pattern for the slope of negative market returns of the one-month kernel is similar for the pricing kernels with longer horizons. However, the U-shape disappears for long horizons, as we only find a positive slope for the six-month kernel in $s_3(6)$. For the twelve-month kernel, the U-shape disappears. For the six- and twelve-month forward kernel we still find that the level for bad returns increases in good times. However, the overall magnitude is much smaller, especially for the twelve-month forward pricing kernel. So we conclude from Table 3 that the short-term pricing kernel exhibits large and significant time-series variation, whereas the long-term pricing

kernel varies only slightly.

6 The Pricing Kernel in Rational Expectation Asset Pricing Models

In this section we discuss the predictions of several famous asset pricing models with respect to the average pricing kernel, the time-series variation and the term structure. We consider the following models: the habit model by Campbell and Cochrane (1999), the time-varying rare disaster model by Wachter (2013),[10] the long run risk model by Bansal and Yaron (2004) and the time-varying recovery by Gabaix (2012). We use standard calibrations of the models and analyze their predictions for the pricing kernel. We show that all models produce a downward sloping pricing kernel as a function of the market return and the models predict relatively little time-series variation. While the models can not match the dynamics of the short-term pricing kernel (as has been noted by previous research), we argue that they match the empirical dynamics of the long-term kernels surprisingly well.

To compare the predictions of the models to the data, we need to copmute the pricing kernel of each model as a function of the market return. In the models the market return is a function of stochastic variation in the consumption/dividend growth and stochastic variation of the state variables. Similarly, the pricing kernel is a function of these stochastic variables as well. Given that the probability distribution of these variables are specified in the models, we can calculate the average level of the pricing kernel for each return level. To calculate the expected pricing kernel at a given return level in each of the models, we map all stochastic variation of consumption, dividends and state-variables to the market return and calculate the average in order to obtain the pricing kernel as a function of the market return. This approach is the theoretical equivalent of our empirical procedure.

In the following we discuss the shape and time-series variation of the pricing kernel in the different models. For a detailed description of the models as well as the

[10] As all other models considered in this study are in discrete time, we also use the discrete-time equivalent of the Wachter (2013) model which has also been used in Dew-Becker et al. (2017).

calibrations that we use, we refer to Appendices A.3-A.6. The pricing kernel projections for the habit model, the rare-disaster model, the long-run risk model and the time-varying recovery model together with the empirical one-month and the twelve-month forward kernels are shown in Figures 12, 13, 14 and 15 respectively.

Figure 12: The figure plots the mean one-month pricing kernel of the Habit model (solid line), the mean empirical one-month pricing kernel (dot-dashed line) and the mean empirical twelve-month forward pricing kernel (dashed line).

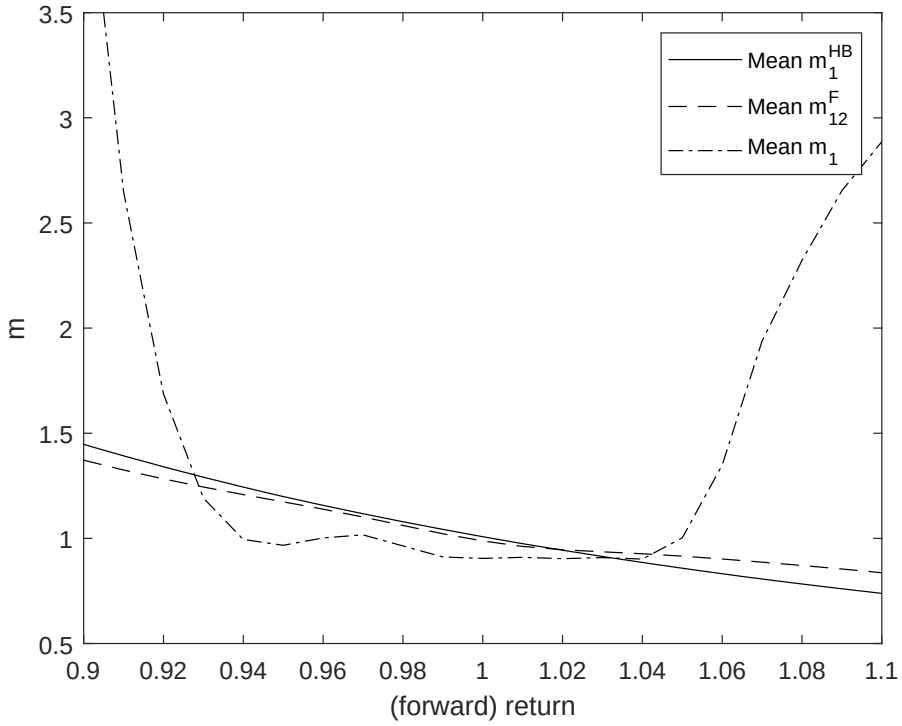


Figure 13: The figure plots the mean one-month pricing kernel of the Rare Disaster model (solid line), the mean empirical one-month pricing kernel (dot-dashed line) and the mean empirical twelve-month forward pricing kernel (dashed line).

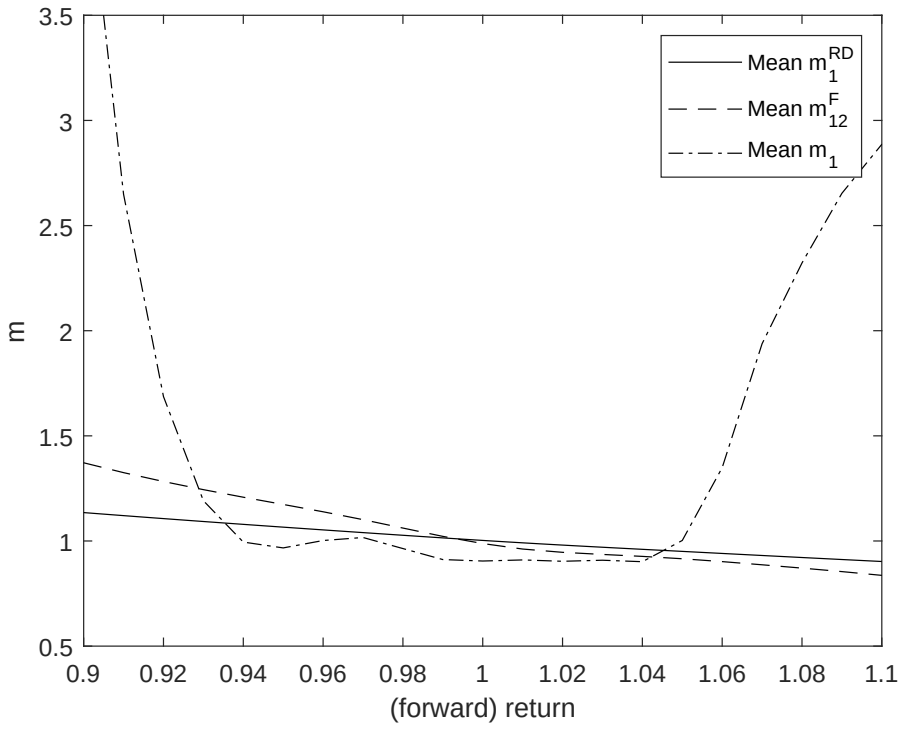


Figure 14: The figure plots the mean one-month pricing kernel of the Long Run Risk model (solid line), the mean empirical one-month pricing kernel (dot-dashed line) and the mean empirical twelve-month forward pricing kernel (dashed line).

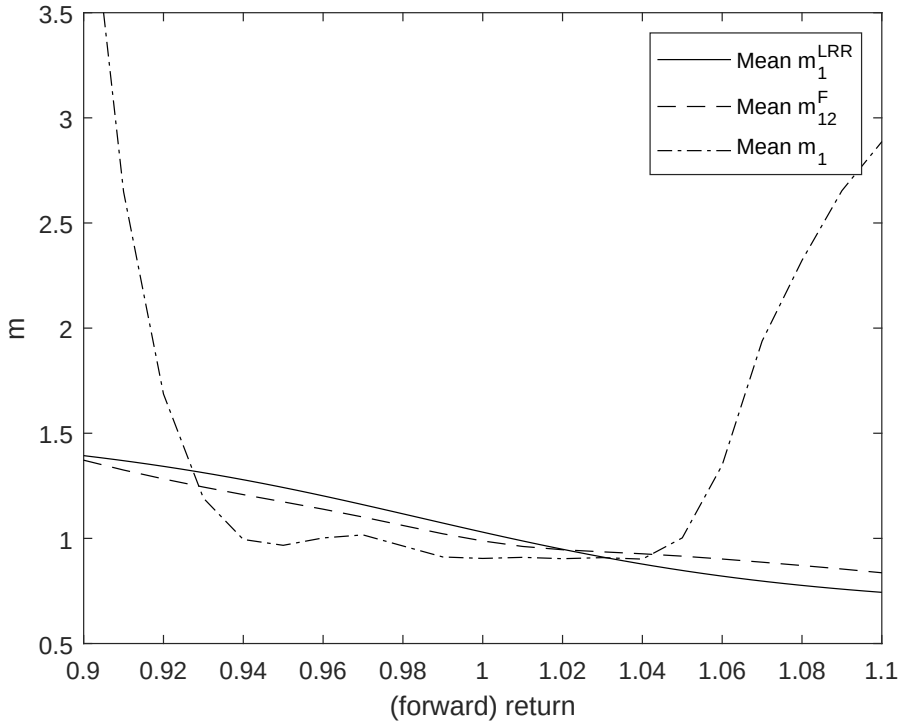
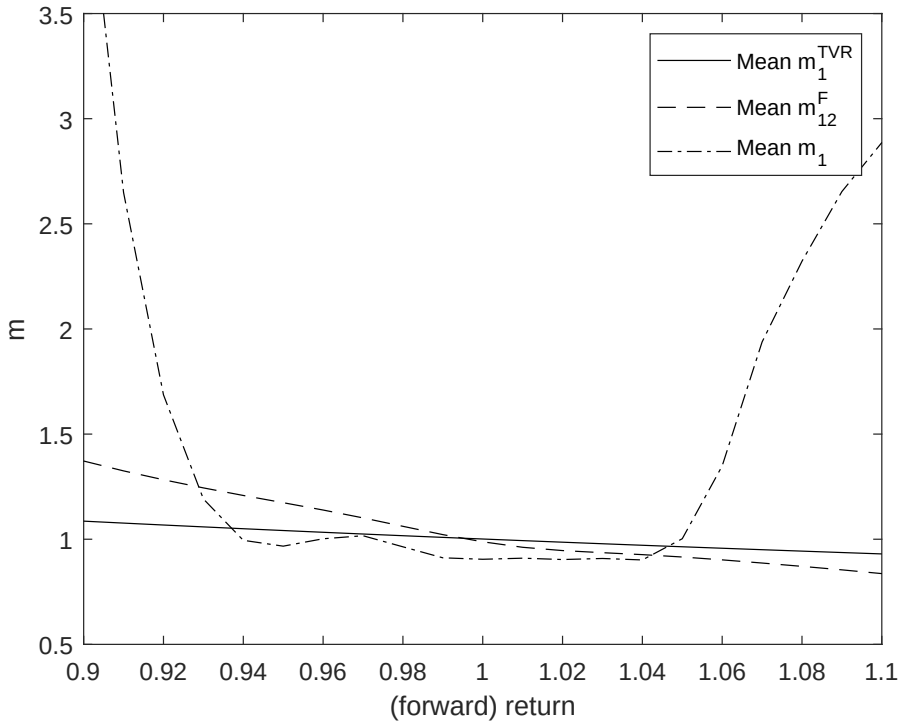


Figure 15: The figure plots the mean one-month pricing kernel of the Time-Varying Recovery model (solid line), the mean empirical one-month pricing kernel (dot-dashed line) and the mean empirical twelve-month forward pricing kernel (dashed line).



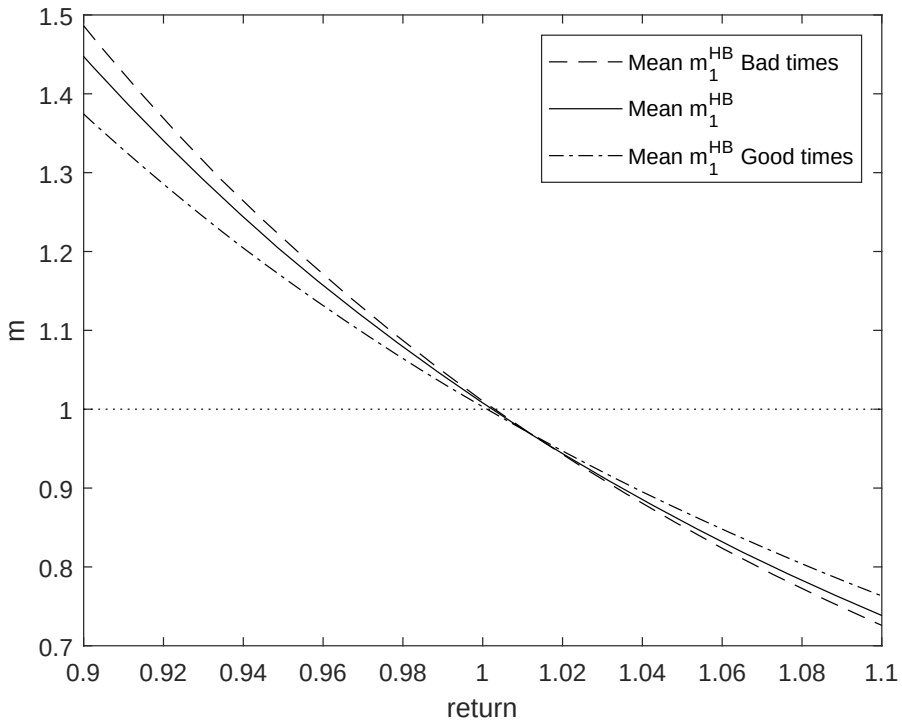
For each of the models the average pricing kernel is monotonically decreasing in the market return. Hence, the shape of the empirical one-month pricing kernel cannot be matched by any of the rational expectation models. The fact that the models fail to generate a U-shaped pricing kernel is well known. However, the figures also demonstrate that the models can not match the short term pricing kernel in terms of levels for negative returns. In particular the empirical pricing kernel for negative returns is significantly larger than the kernel implied by the models which suggest that investors in the model require a much higher return for holding assets that pay off in the bad state compared to the investors in the real world. So both, the upward slope for positive returns and the level of the short-term pricing kernel for large negative returns is at odds with the predictions of the models.

As we have argued in Section 5.2, the violation of monotonicity of the pricing kernel disappears gradually when the maturity of the forward pricing kernel increases. Therefore, the figures also show the empirical twelve-month forward pricing kernel $\left(m^F_{t,12}\right)$ which is strictly downward sloping. We observe that the shape of the twelve-month

forward kernel is in line with the kernels predicted by the models and also in terms of levels, the models do a surprisingly good job in matching the data.[11] While we find that the pricing kernels in the rare disaster model and the time-varying recovery model are slightly less curved than the empirical long-term forward pricing kernel, it is important to note that none of the calibrations used in this paper was altered in order to match the empirical properties of the pricing kernel. Instead we used standard calibrations from the literature and even with these calibrations we get a good fit of the forward pricing kernel.

As an additional test for the asset pricing models, we compare the time-series variation of the pricing kernels in the models with the time-series variation observed in the data. Figure 16 shows the time-series variation of the pricing kernel in the habit model.[12]

Figure 16: The figure plots the average pricing one-month kernel (solid line), as well as the kernel in good (dot-dashed line) and in bad times (dashed line) for the Habit model.



[11] We compare the expected forward pricing kernel estimated from the data with one-month pricing kernels from the models. The models exhibit little time-series variation and have state variables that are mean-reverting. Therefore, we know that if the current one-month pricing kernel is different from the long-run mean the forward pricing kernel is even closer towards the long-run mean, due to the fact that the state variables in the model revert back to the long-run mean in expectation.

[12] The good and bad times are defined as the 75% and 25% quantile of s_t in a simulation study.

We observe that there is only little time-series variation in the pricing kernel of the Habit model and we find similar magnitudes for the time-series variation for the other asset pricing models.[13] The results in Section 5.3 suggest that the time-series variation of the short-term pricing kernel is relatively large. The time-series variation of the long-term pricing kernel on the other hand is significantly lower. As the models predict only little time-series variation, they can not match the properties of the short term kernel, but the time-series variation is well in line with the variation of the long-term forward kernel.

To summarize, we show that rational expectation asset pricing model can adequately capture the shape, level and time-series variation of the long-term forward pricing kernel. We argue that the well documented pricing kernel puzzle is a unique feature of the one-month pricing kernel. For longer maturities the puzzle disappears and the predictions of the standard models are well in line with the data.

7 The Pricing Kernel in a Behavioral Asset Pricing Model

In the previous section we show that the rational expectation asset pricing models can explain the empirical evidence of the long-term pricing kernel. However, the evidence for the short-term pricing kernel is puzzling in terms of its U-shape and its time-series variation. In this section we argue that a model with an investor with cumulative prospect theory (CPT) preferences from Tversky and Kahneman (1992) can potentially explain the empirical evidence of the short-term pricing kernel. The key feature is probability weighting of the agent. Probability weighting is the tendency of individuals to overweight small probabilities of extreme outcomes due to which payoffs in unlikely states of the world, both negative *and* positive, are more attractive to CPT investors. The formulas of the CPT-framework in Tversky and Kahneman (1992) are given in Appendix A.7.

We follow the model of Baele et al. (forthcoming), in which they assume a representative investor with CRRA-utility over terminal wealth and CPT-utility over gains

[13]The corresponding results for the time-varying disaster, long-run risk and time-varying recovery model are shown in Figures 23, 24 and 25 in the appendix.

and losses. Baele et al. (forthcoming) show that such a model does a good job in explaining the low returns on both OTM put and call options and, therefore, the variance premium.

The ability of the pricing kernel implicitly defined in (20) of Appendix A.7 to match the empirical U-shaped short-term pricing and its time-series variation comes from probability weighting. A CPT-investor evaluates its utility using decision weights π_i instead of objective probabilities p_i which an investor under expected utility uses. Intuitively, under the parameters in Tversky and Kahneman (1992), the probability weighting leads to an overweighting of unlikely and extreme states, that is, the tails of the distribution. Due to this overweighting, the extreme negative and positive states become more attractive and can make the pricing kernel U-shaped for a scaling term $\hat{b}$ sufficiently large. Where $\hat{b}$ governs the relative importance of the CRRA and CPT-utility in the investor's preference specification.

Using equation (20) in Appendix A.7, we can calculate the pricing kernel by plugging in a distribution for the market returns (R_i^E). We use the same model for the return distribution of the S&P 500 specified in Section 3.3. The next step is evaluate how well the model does compared to the empirical results we obtained for the one-month pricing kernel. Similar to what we did in the data, we calculate the model implied pricing kernel on every day we estimate the empirical pricing kernel.

Figure 17: The figure plots on the left (right) the empirical mean (median) one-month pricing kernel and the mean (median) pricing kernel from the CPT model. The solid line represents the empirical one-month pricing kernel, whereas the dashed line represents the CPT-model implied pricing kernel.

(a) Mean Pricing Kernel.

(b) Median Pricing Kernel.

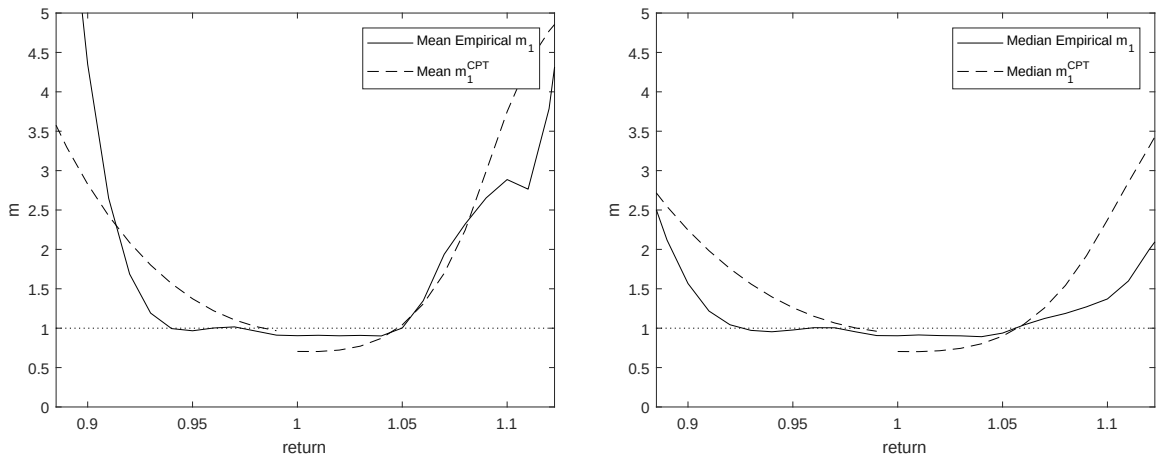


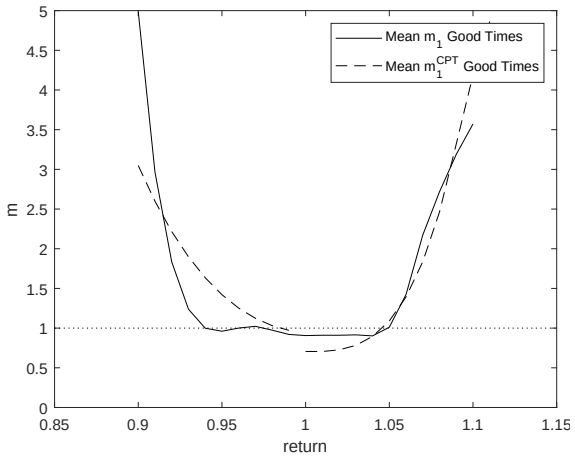
Figure 17 shows that using the parameters from Baele et al. (forthcoming), the

model is able to match the shape of the empirical pricing kernel quite well. A second feature of the model-implied pricing kernel is the discontinuity around return of 1.00, which is the loss-aversion from equation (20) and only has an effect when the return on the index is lower than the risk-free rate. Furthermore, for both, the empirical result and the model, the mean of the pricing kernel is above the median, indicating time-series variation.

In the following we analyze the time-series variation in the pricing kernel of the CPT-model. Time-series variation of the pricing kernel in the CPT-model is driven by the time-variation of the volatility of our return distribution. Therefore, we can define good and bad times in the same way as we have done in our empirical study.

Figure 18: The figure plots on the left (right) the empirical mean (median) one-month pricing kernel and the mean (median) pricing kernel from the CPT model in good times. The solid and dashed line represent the empirical and CPT pricing kernel, respectively.

(a) Mean Pricing Kernel in Good Times.



(b) Median Pricing Kernel in Good Times.

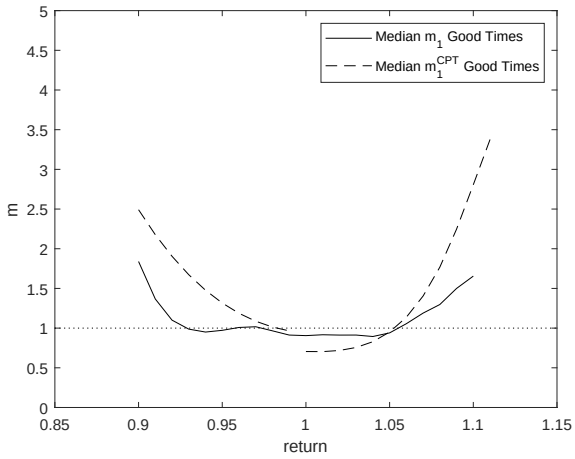
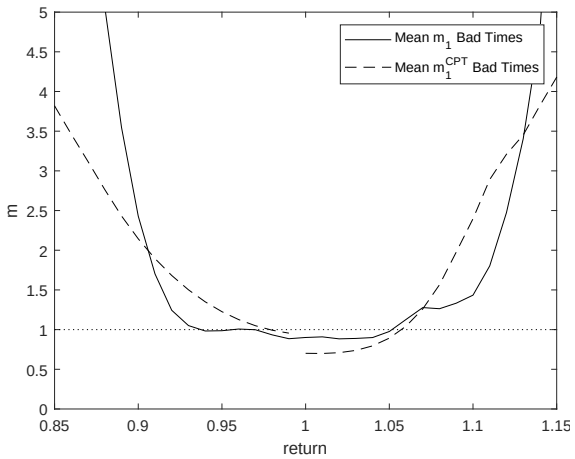
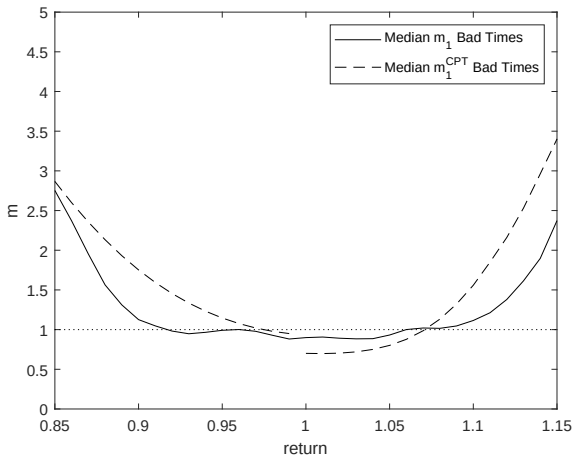


Figure 19: The figure plots on the left (right) the empirical mean (median) one-month pricing kernel and the mean (median) pricing kernel from the CPT model in bad times. The solid and dashed line represent the empirical and CPT pricing kernel, respectively.

(a) Mean Pricing Kernel in Bad Times.



(b) Median Pricing Kernel in Bad Times.



The CPT-model can produce substantial time-series variation of the pricing kernel and the variation goes in the same direction as empirically documented. Both in the model and empirically the U-shape of the one-month pricing kernel increases in good times. The intuition from the CPT-model is as follows, lower probabilities of extreme events are relatively more overweighted. In good times, when the standard deviation of the return distribution is low, the probability of a 5% increase or decrease in stock prices is very low. Whereas in bad times, when the standard deviation of the return distribution is high, the likelihood of a 5% increase or decrease in stock prices is larger. Given that in CPT, lower probabilities of extreme events are relatively more overweighted, the U-shape is stronger in good times when the probability of a 5% increase or decrease is very low.

8 Conclusion

This paper shows that the pricing kernel puzzle, that is the non-monotonicity of the pricing kernel, is a unique feature of short-term pricing kernel and it disappears at longer horizons. More precisely, for maturities longer than one month we show that the U-shape of the forward pricing kernels disappears and for horizons beyond six months, the pricing kernel is strictly monotonically decreasing. While standard asset pricing models have difficulties to explain the empirical properties of the short-term kernels, we demonstrate that they match the shape and level of the long-term kernels surprisingly well. Furthermore we show that an asset pricing model with behavioral preferences is able to reconcile the evidence of the short-term pricing kernel. We provide further support to our hypothesis by looking at the time-series variation of the pricing kernels. In line with the previous results we find that standard pricing models can match the time-series variation of long-term kernels while the behavioral model can explain the time-series variation of the short-term kernel. It would be an interesting direction for future research to combine the findings of our work and develop a pricing model that can jointly explain the dynamics of the short and long-term kernels. Furthermore our result can also be explained by maturity segmentation in the option market and our result could be seen as first evidence of this. Potentially, investors with non-standard preferences trade mostly short-term options and investor with more standard preferences trade long-term options.

However, whether this result when combining these types of investors arises endogenously would be an interesting avenue for future research. Especially given that the implications of CPT-preferences with the respect to the horizon are unknown.

9 Bibliography

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

Baele, L., Driessen, J., Ebert, S., Londono, J. and Spalt, O. (forthcoming). Cumulative prospect theory, option prices, and the variance premium. *Review of Financial Studies*.

Bakshi, G., Madan, D. and 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.

Bansal, R. and Yaron, A. (2004). Risks for the long run: A potential resolution of asset pricing puzzles. *The Journal of Finance*, 59(4):1481 – 1509.

Bansal, R., Kiku, D. and Yaron, A. (2012). An empirical evaluation of the long-run risks model for asset prices. *Critical Finance Review*, 1(1):183–221.

Bauwens, L. and Laurent, S. (2002). A new class of multivariate skew densities, with applications to garch models. *Working paper*.

Bollerslev, T., Tauchen, G. and Zhou, H. (2009). Expected stock returns and variance risk premia. *The Review of Financial Studies*, 22(11):4463–4492.

Breeden, D. and Litzenberger, R. (1978). Prices of state-contingent claims implicit in option prices. *Journal of Business*, 51(4):621–651.

Britten-Jones, M. and Neuberger, A. (2000). Option prices, implied price processes, and stochastic volatility. *The Journal of Finance*, 55(2):839–866.

Campbell, J. and Cochrane, J. (1999). By force of habit: A consumption-based explanation of aggregate stock market behavior. *Journal of Political Economy*, 107(2):205 – 251.

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

Chabi-Yo, F., Garcia, R. and Renault, E. (2007). State dependence can explain the risk aversion puzzle. *The review of Financial studies*, 21(2):973–1011.

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

Cuesdeanu, H. and Jackwerth, J. (2018a). The pricing kernel puzzle: Survey and outlook. *Annals of Finance*, 14(3):1–41.

Cuesdeanu, H. and Jackwerth, J. (2018b). The pricing kernel puzzle in forward looking data. *Review of Derivatives Research*, 21(3):1–24.

Dew-Becker, I., Giglio, S., Le, A. and Rodriguez, M. (2017). The price of variance risk. *Journal of Financial Economics*, 123(2):225–250.

Epstein, L. and Zin, S. (1989). Substitution, risk aversion, and the temporal behavior of consumption and asset returns: A theoretical framework. *Econometrica*, 57(4):937–969.

Gabaix, X. (2012). Variable rare disasters: An exactly solved framework for ten puzzles in macro-finance. *The Quarterly Journal of Economics*, 127(2):645–700.

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

Kahneman, D. and Tversky, A. (1979). Prospect theory: An analysis of decisions under risk. *Econometrica*, 47:263–292.

Linn, M., Shive, S. and Shumway, T. (2017). Pricing kernel monotonicity and conditional information. *The Review of Financial Studies*, 31(2):493–531.

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

Rosenberg, J. and Engle, R. (2002). Empirical pricing kernels. *Journal of Financial Economics*, 64(3):341–372.

Sichert, T. (2019). The pricing kernel is u-shaped. *Working paper*.

Song, Z. and Xiu, D. (2016). A tale of two option markets: Pricing kernels and volatility risk. *Journal of Econometrics*, 190(1):176–196.

Tversky, A. and Kahneman, D. (1992). Advances in prospect theory: Cumulative representation of uncertainty. *Journal of Risk and Uncertainty*, 5(4):297–323.

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

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

Wachter, J. (2013). Can time varying risk of rare disasters explain aggregate stock market volatility? *The Journal of Finance*, 68(3):987 – 1035.

A Appendix

A.1 Details empirical analysis

In this section we discuss the details of our empirical analysis and consider a trading day on which we estimate the pricing kernel. We denote the spread of the butterfly by ΔK, which captures the distance between the strike K and the point where the payoff of the butterfly is zero. We make ΔK increasing in S_0, T and implied volatility by choosing ΔK equal to the difference in strike of call options with deltas 0.35 and 0.50 on a certain trading day. Approximately, we keep the expected payoff of a butterfly given a moneyness level constant, as the expected value would decrease in any of the aforementioned variables. To make sure that the prices of butterflies with moneyness more towards the tails of the distribution are significantly larger than zero, we scale ΔK with moneyness outside the region $\left[m_t - \frac{c\cdot\sigma_t}{1+\xi^2}; m_t + \frac{c\cdot\sigma_t}{1+\frac{1}{\xi^2}}\right]$. By doing so, we can still exploit the differences in price for butterflies with extreme moneyness levels.

The amount of states we consider is constant and equal to 20 for every day and maturity, therefore in order for the pricing kernel to be unique we consider 20 butterfly spreads of the corresponding maturity. Lets consider an example of the estimation on how we set up a grid and compute the needed butterfly spreads in order to estimate the pricing kernel. First, we use the kernel smoothing algorithm of OptionMetrics to construct the volatility surface for each option with delta $(\pm)0.001$ up to delta $(\pm)0.999$. On the first day in our sample January 4th 1996, the strike of a call option with maturity 30 days and delta 0.50 is given by 619.57 and the strike for option with delta 0.35 equals 627.03 yielding $\Delta K = 7.46$. The total interval over which we estimate the pricing kernel follows from $\left[K(-0.001) + 2\cdot\Delta K; K(0.001) - 2\cdot\Delta K\right]$, where $K(\cdot)$ is the strike of an option with delta equal to $(\pm)0.001$. On January 4th the total interval equals $\left[534.15; 660.91\right]$ over which there will be 20 states. Over the interval $\left[m_t - \frac{c\cdot\sigma_t}{1+\xi^2}; m_t + \frac{c\cdot\sigma_t}{1+\frac{1}{\xi^2}}\right]\cdot S_0$, the amount of states will be double as large as outside this region. The region on this day $\left[610.23; 636.52\right]$, the amount of states equals to following $20\cdot 2(636.52-610.23)/(660.91-534.15+636.52-610.23) \approx 7$. Given that we know the amount of states on this interval, we can calculate the distance between two consecutive states. The remaining 13 states are divided accordingly over the rest of the total grid, making sure that the total amount

of states sums to 20. In the end we know that we have 7 states, and therefore butterflies, over the interval $\left[610.23; 636.52\right]$ with $\Delta K = 7.46$ and 13 states, and butterflies, over the intervals $\left[534.15; 610.23\right]$ and $\left[636.52; 660.91\right]$ with $\Delta K = 14.92$.

A.2 The Term Structure of the Pricing Kernel

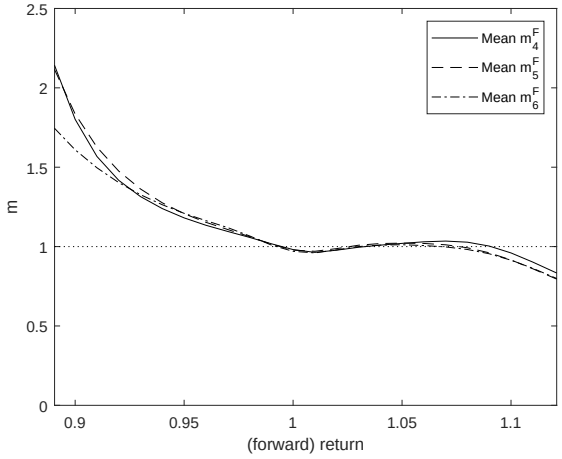
Test of differences one-month, six-month forward and twelve-month forward pricing kernel.

Table 4: The mean differences between seven points of the (forward) pricing kernel are tested. $m_1(r)$ corresponds to the one-month pricing kernel, $m_6(r)$ corresponds to six months forward kernel and $m_{12}(r)$ corresponds to the twelve months forward kernel. In brackets the hac t-statistics are given, with appropriate amount of lags according to the ACF of the residuals.

r	0.90	0.95	0.98	1.00	1.02	1.05	1.10
$m_1(r) - m_6(r)$							
Mean	2.75 (2.10)	−0.24 (−7.40)	−0.10 (−4.75)	−0.07 (−3.74)	−0.08 (−4.56)	−0.01 (−0.18)	1.91 (2.35)
$m_1(r) - m_{12}(r)$							
Mean	2.98 (2.20)	−0.21 (−5.08)	−0.10 (−5.03)	−0.08 (−8.31)	−0.04 (−1.23)	0.09 (1.50)	1.98 (2.43)
$m_6(r) - m_{12}(r)$							
Mean	0.24 (6.31)	0.04 (1.14)	0.01 (0.28)	−0.02 (−1.06)	0.03 (2.13)	0.10 (7.90)	0.08 (9.13)

Figure 20: The figure plots on the left (right) the time-series mean (median) of the four, five and six-month forward pricing kernel. The solid, dashed and dot-dashed line represent the four, five and six-month forward pricing kernel, respectively. Forward pricing kernels are estimated using the methodology described in section 3.2.

(a) Mean Pricing Kernel.



(b) Median Pricing Kernel.

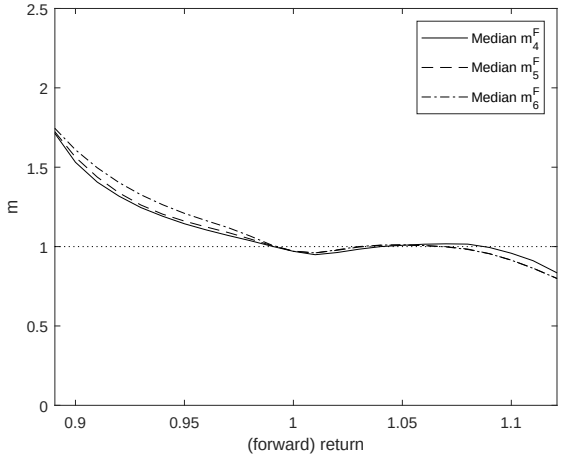
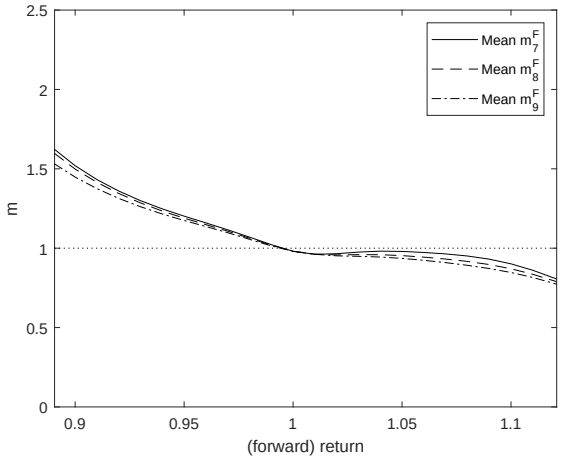
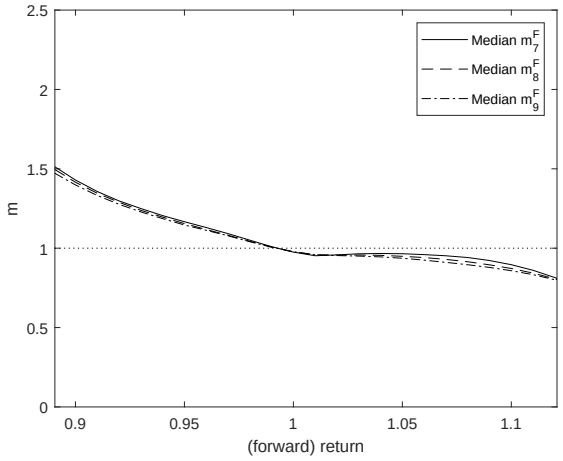


Figure 21: The figure plots on the left (right) the time-series mean (median) of the seven, eight and nine-month forward pricing kernel. The solid, dashed and dot-dashed line represent the seven, eight and nine-month forward pricing kernel, respectively. Forward pricing kernels are estimated using the methodology described in section 3.2.

(a) Mean Pricing Kernel.



(b) Median Pricing Kernel.



A.3 The Habit Model of Campbell and Cochrane (1999)

In the habit model of Campbell and Cochrane (1999), log consumption is given by:

$$\Delta c_{t+1} = g + v_{t+1},$$

where $v_{t+1} \sim N\big(0, \sigma_v^2\big)$. Different from the standard CRRA case, the agent derives utility over consumption in excess of the habit level $\big(X_t\big)$, yielding the following pricing kernel specification:

$$m_{t,1}^{HB} = e^{-\delta}\Big(\frac{C_{t+1}}{C_t}\Big)^{-\gamma}\Big(\frac{S_{t+1}}{S_t}\Big)^{-\gamma},$$

where C_t and $S_t := (C_t - X_t)/C_t$ are the consumption and surplus ratio at time t, respectively. The surplus ratio evolves according to:

$$\log\Big(\frac{C_{t+1}}{C_t}\Big) = g + v_{t+1},$$
$$s_{t+1} = (1-\phi)\overline{s} + \phi s_t + \lambda(s_t)v_{t+1}$$

where $s_t = \log(S_t)$ and $\lambda(s_t)$ is given by:

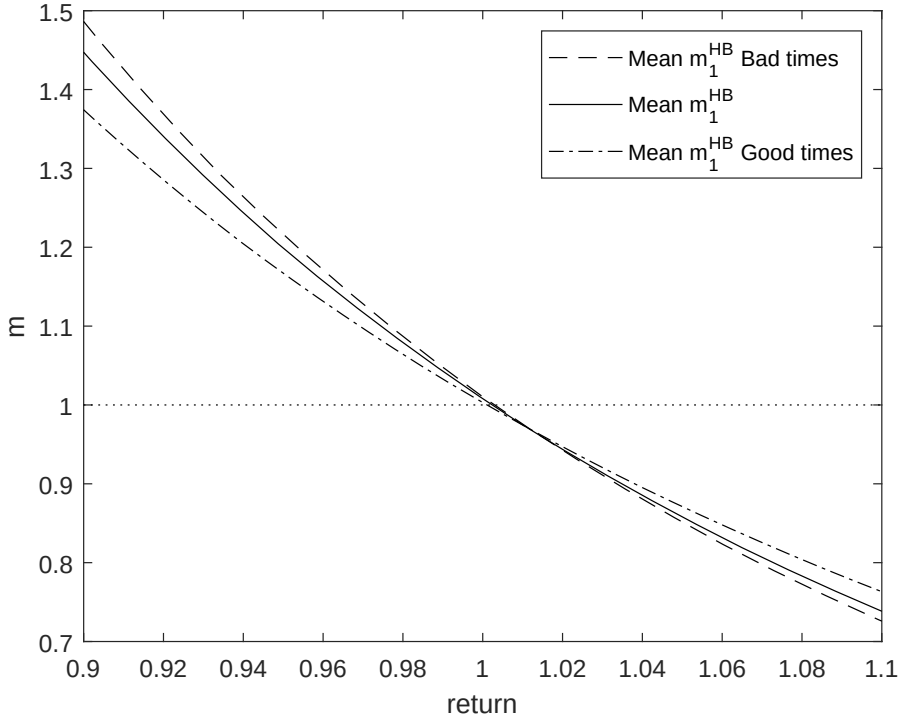
$$\lambda(s_t) = \frac{1}{\overline{S}}\sqrt{1 - 2(s_t - \overline{s})} - 1,$$
$$\overline{S} = \sigma_v \sqrt{\frac{\gamma}{1 - \phi - b/\gamma}}.$$

Note, $\lambda(s_t)$ is set to zero when $s_t > s_{max}$:

$$s_{max} = \overline{s} + \frac{1}{2}\big(1 - \overline{S}^2\big).$$

Given the specified pricing kernel in the model we can use the standard asset pricing equation to price the market portfolio and compute its returns. We take the calibration from Campbell and Cochrane (1999) which is given in Table 5. Figure 22 plots the average pricing kernel of the model as well as its time-series variation. The good and bad times are defined as the 25% and 75% quantile of s_t in a simulation study. In bad times the surplus ratio is low and in good times high.

Figure 22: The figure plots the average pricing kernel, as well as the kernel in good and in bad time for the Habit model of Campbell and Cochrane (1999).



From Figure 22 we conclude that the time-series variation is comparatively very small to the time-series variation of the empirical short-term pricing kernel.

Table 5: Calibration of the Campbell and Cochrane (1999) model.

Parameter	Value	Parameter	Value
μ_c	$0.0189/12$	σ_c	$0.015/\sqrt{12}$
ϕ	$0.87^{\frac{1}{12}}$	δ	$0.89^{\frac{1}{12}}$
r^f	$0.0094/12$	γ	2

A.4 The Time-Varying Rare Disasters Model of Wachter (2013).

In this section we consider the time-varying disaster model of Wachter (2013). We use the discrete time equivalent of the model as used in Dew-Becker et al. (2017). Log consumption and dividend growth in the model are given by:

$$\Delta c_t = \mu_c + \sigma_c \epsilon_{c,t} + J_t,$$
$$\Delta d_t = \eta \Delta c_t,$$

where $\epsilon_c \sim N(0,1)$ and J_t is the jump process (rare disaster). The process is modeled as a poisson mixture of normal distributions and is given by:

$$J_t = \sum_{i=1}^{N_t} \xi_{i,t},$$
$$\xi_{i,t} \sim N\big(\mu_d, \sigma_d\big),$$
$$N_t \sim Poisson(\lambda_t).$$

The intensity of the jump process is time-varying and follows:

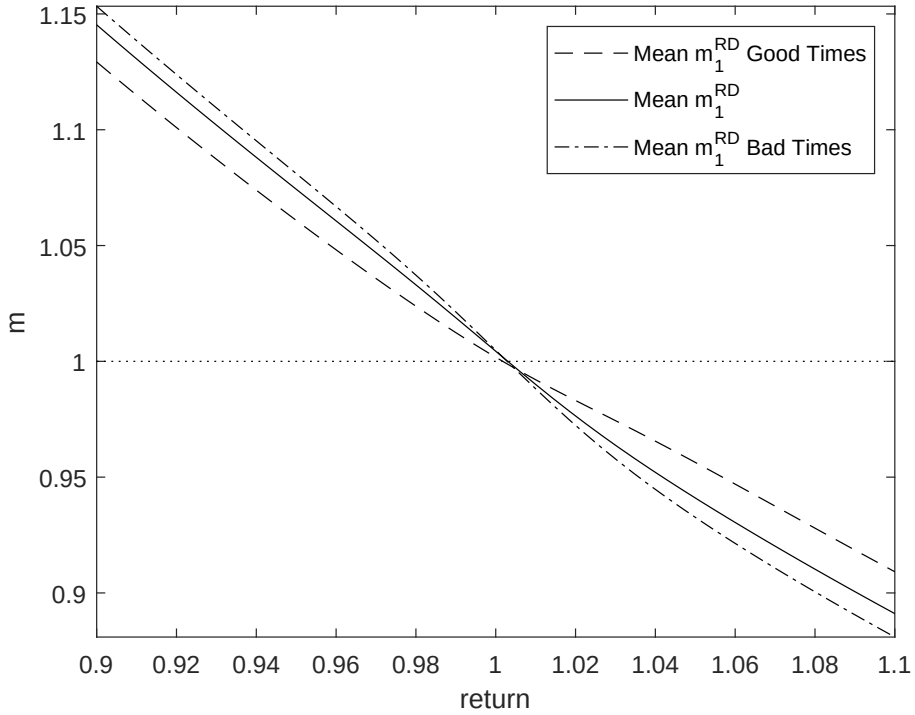
$$\lambda_{t+1} = \phi\lambda_t + (1-\phi)\mu_\lambda + \sigma_\lambda \sqrt{\lambda_t}\epsilon_{\lambda,t+1},$$

where $\epsilon_{\lambda,t} \sim N(0,1)$. The investor has Epstein-Zin utility with Elasticity of Interemporal Substitution (EIS) equal to one so lifetime utility is given by:

$$v_t = (1-\beta)c_t + \frac{\beta}{1-\alpha}\log \mathbb{E}_t \exp\big(v_{t+1}(1-\alpha)\big),$$
$$m_{t,1}^{RD} = \beta \exp\big(-\Delta c_{t+1}\big)\frac{\exp\big((1-\alpha)v_{t+1}\big)}{\mathbb{E}_t \exp\big((1-\alpha)v_{t+1}\big)}.$$

Given the specified pricing kernel in the model we can use the standard asset pricing equation to price the market portfolio and compute its returns. We take the calibration from Wachter (2013) which is given in Table 6. Figure 23 plots the average pricing kernel of the model as well as its time-series variation. The good and bad times are defined as the 25% and 75% quantile of λ_t in a simulation study. In bad times the jump intensity is high and in good times low.

Figure 23: The figure plots the average pricing kernel and its time-series variation from the disaster model of Wachter (2013) with the calibration from Table 6.



Similar to the results of the Habit model, we find that the time-series variation of the rare disaster model by Wachter (2013) is comparatively small. In the Table 6 the calibration is represented.

Table 6: Calibration of the Wachter (2013) model.

Parameter	Value	Parameter	Value
μ_c	$0.0252/12$	σ_c	$0.02/\sqrt{12}$
μ_d	-0.15	σ_d	0.10
μ_λ	$0.0355/12$	σ_λ	$0.067/12$
ϕ	$\exp(-0.08/12)$	β	$\exp(-0.012/12)$
η	2.6	γ	$3.0 = 1 - \alpha$

A.5 Long-Run Risk Model of Bansal and Yaron (2004)

In this section we consider is the long-run risk model from Bansal and Yaron (2004) and we use the specification of Bansal et al. (2012) which is given by:

$$\begin{aligned}
\Delta c_{t+1} &= \mu_c + x_t + \sigma_t \eta_{t+1}, \\
x_{t+1} &= \rho x_t + \varphi_e \sigma_t e_{t+1}, \\
\sigma^2_{t+1} &= \bar{\sigma}^2 + \nu\big(\sigma^2_t - \bar{\sigma}^2\big) + \sigma_w w_{t+1}, \\
\Delta d_{t+1} &= \mu_d + \phi x_t + \pi \sigma_t \eta_{t+1} + \varphi_u \sigma_t u_{t+1},
\end{aligned}$$

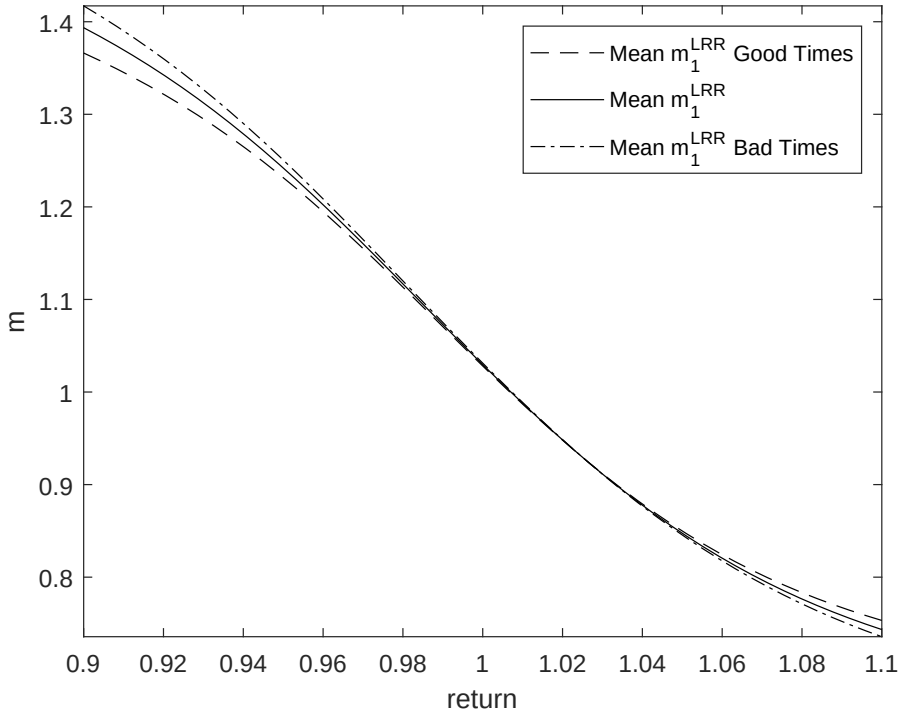
where $\eta_{t+1}, e_{t+1}, w_{t+1}, u_{t+1} \sim N(0,1)$. The representative agent is assumed to have Epstein and Zin (1989) recursive preferences and maximizes lifetime utility and implies the following pricing kernel:

$$\begin{aligned}
V_t &= \Big[(1-\delta) C_t^{\frac{1-\gamma}{\theta}} + \delta\big(\mathbb{E}_t[V_{t+1}^{1-\gamma}]\big)^{\frac{1}{\theta}}\Big]^{\frac{\theta}{1-\gamma}}, \\
m_{t,1}^{LRR} &= \exp\Big(\theta \log\delta - \frac{\theta}{\psi}\Delta c_{t+1} + (\theta - 1) r_{a,t+1}\Big),
\end{aligned}$$

where $r_{a,t+1}$ is the return on the aggregate claim to consumption. Given the specified pricing kernel in the model we can use the standard asset pricing equation to price the market portfolio and compute its returns. We take the calibration from Bansal and Yaron (2004) which is given in Table 7. Figure 24 plots the average pricing kernel of the model as well as its time-series variation. In the long-run risk model, there is one state variable more compared to the previous models. Good times are defined a period where x_t is

one standard deviation larger than the long-run average and σ_t is one standard deviation smaller than the long-run average. The opposite holds for bad times, low consumption growth and high volatility.

Figure 24: The figure plots the average pricing kernel and its time-series variation from the long-run risk model of Bansal and Yaron (2004) with the calibration from Table 7.



The results are similar to before.

Table 7: Calibration of the long run risk model.

Parameter	Value	Parameter	Value
μ_c	0.0015	μ_d	0.0015
$\bar{\sigma}$	0.0078	σ_w	0.0000023
ρ	0.979	φ_e	0.044
ϕ	3	π	0
φ	4.5	δ	0.998
γ	10	ψ	1.5
ν	0.987		

A.6 The Time-Varying Recovery Model of Gabaix (2012)

The next model we consider is the time-varying recovery model of Gabaix (2012), we use the same specification as in Dew-Becker et al. (2017) and is specified as follows:

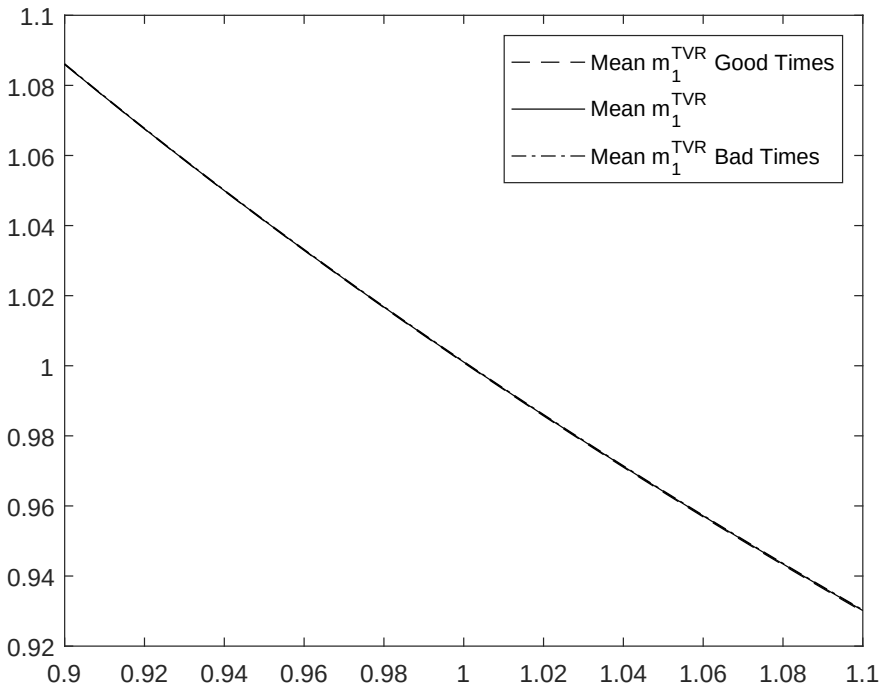
$$\begin{aligned}
\Delta c_{t+1} &= \mu_c + \sigma_c \epsilon_{c,t+1} + J_{c,t+1}, \\
L_{t+1} &= \big(1 - \rho_L\big)\bar{L} + \rho_L L_t + \sigma_L \epsilon_{L,t+1}, \\
\Delta d_{t+1} &= \lambda \sigma_c \epsilon_{c,t+1} - L_t \cdot \mathbb{1}_{J_{c,t} \neq 0},
\end{aligned}$$

where $\epsilon_{c,t+1}, \epsilon_{L,t+1} \sim N(0,1)$ and $J_{c,t+1}$ follows the same distribution as in the rare-disaster model defined in Section A.4. In the model the representative agent has power utility preferences. Corresponding to a pricing kernel with the following specification:

$$m_{t,1}^{TVR} = \beta \cdot \Big(\frac{C_{t+1}}{C_t}\Big)^{-\gamma}.$$

As one can see from the formulas, the pricing kernel is only driven by shocks to consumption, ϵ_c or disaster J_c. Therefore, it is not dependent on the level of the state-variable in the model L_t and there is no time-series variation. Given the specified pricing kernel in the model we can use the standard asset pricing equation to price the market portfolio and compute its returns. We take the calibration from Dew-Becker et al. (2017) which is given in Table 8. Figure 25 plots the average pricing kernel of the model as well as its time-series variation. Good times are defined as the mean recovery rate plus twice the standard deviation and bad times vice versa.

Figure 25: The figure plots the average pricing kernel and its time-series variation of the time-varying recovery model using the parameters of table 8.



Indeed, the model does not exhibit any time-series variation as the pricing kernel is i.i.d. due to the power utility assumption. The term structure in this model will be flat as well due to the power utility preferences.

Table 8: Calibration of the model by Gabaix (2012).

Parameter	Value	Parameter	Value
μ_c	$0.01/12$	σ_c	$0.02/\sqrt{12}$
μ_d	-0.3	σ_d	0.15
$\bar{L}$	$-\log(0.5)$	σ_L	0.04
ρ_L	$0.87^{1/12}$	λ	5
β	$0.96^{1/12}$	γ	7

A.7 Cumulative Prospect Theory

In this section the CPT framework of Tversky and Kahneman (1992) is introduced formally. The utility in CPT is defined over gains or losses as opposed to wealth in expected utility. Lets say the wealth at time T is equal to W_T, then the gain or loss at time is defined as $X_T = W_T - W_{Ref}$, where W_{Ref} denotes the reference point, which determines which wealth level is a gain or loss. In all of our analysis we use that the reference point of the investor is the risk-free interest rate, i.e. $W_{Ref} = W_0 \cdot r^f$, where r^f is the T period interest rate. The investor in Baele et al. (forthcoming) includes the CPT utility over the gain or loss X_T in its objective function, $CPT\big(X_T\big)$. To define $CPT(\cdot)$, let denote $v(\cdot)$ denote the value function, which is defined over gains and losses X_T. The following functional form of $v(\cdot)$ is assumed in Tversky and Kahneman (1992):

$$v\big(X_T\big) = \begin{cases} \big(X_T\big)^{\alpha} & \text{if} \quad X_T > 0 \\ -\lambda\big(-X_T\big)^{\beta} & \quad\quad X_T \leq 0 \end{cases}. \tag{18}$$

Similar to Baele et al. (forthcoming) we assume a piecewise linear function, i.e. $\alpha = \beta = 1$. The assumption is needed to solve for the equilibrium, for more on this see Baele et al. (forthcoming). In equation (18) λ is called the loss aversion parameter. For example, if $\lambda = 2$, then a \$1 loss resonates twice as much as a \$1 gain. Besides the value function, also probabilities may be processed nonlinearly in CPT. We define probability weighting functions $w^+(\cdot)$ and $w^-(\cdot)$ for gains and losses, respectively, by:

$$w^-(p) = \frac{p^{c_1}}{\big[p^{c_1} + (1-p)^{c_1}\big]^{1/c_1}}, \quad w^+(p) = \frac{p^{c_2}}{\big[p^{c_2} + (1-p)^{c_2}\big]^{1/c_2}}, \tag{19}$$

where, $c_1, c_2 \in [0.28; 1]$ control the curvature of each weighting function. For parameters of Tversky and Kahneman (1992) the weighting functions are inverse S shaped and distorts cumulative probabilities, which means that probabilities close to zero of extreme states (negative and positive), i.e. tails of the distribution, are overweighted $\big(w(p) > p\big)$.
The decision weights a CPT agent uses are dependent on the ranking of the outcomes. Lets assume that there are N states of the world at time T, each occuring with objective probability p_i and each outcome is associated with wealth level $W_{T,i}$. All states are ranked from worst to best, and depening on the reference point W_{Ref} in the gain or loss domain:

$W_1 \leq \cdots \leq W_{k-1} \leq W_{Ref} \leq W_k \leq \cdots \leq W_N$. Then decision weight π_i for state i is given by:

$$\pi_i = \begin{cases} w^-(p_1 + \cdots + p_i) - w^-(p_1 + \cdots + p_{i-1}) & \text{for} \quad 2 \leq i < k \\ w^+(p_i + \cdots + p_N) - w^-(p_{i+1} + \cdots + p_N) & \qquad\quad k \leq i < N \end{cases},$$

where $\pi_1 = w^-(p_1)$ and $\pi_N = w^+(p_N)$. The CPT value of gain or loss X_T is then computed as:

$$CPT\big(X_T\big) = \sum_{i=1}^{N} \pi_i \cdot v\big(X_{T,i}\big).$$

In the model by Baele et al. (forthcoming), the investor does not just experience CPT utility over gains and losses but also CRRA utility over terminal wealth, she seeks to maximize:

$$\mathbb{E}u\big(W_T\big) + b_0 CPT\big(X_T\big),$$

where b_0 is the scaling term that governs relative importance of the CRRA and CPT part. The equilibrium condition from which the equity premium follows, equation (10) in Baele et al. (forthcoming), is given by:

$$0 = \sum_i p_i (R_i^E - R^f) \left[\big(R_i^E\big)^{-\gamma} + \hat{b} \frac{\pi_i}{p_i} \big(1 + (\lambda - 1) 1_{R_i^E < R^f}\big) \right], \tag{20}$$

where R_i^E is the return on the market in state i with probability p_i, γ is the risk aversion, $\hat{b}$ is the scaling term that governs the relative importance of the CPT- and CRRA-part, π_i the decision weight of state i and λ is loss-aversion.

Table 9: Calibration CPT model by Baele et al. (forthcoming).

Parameter	Value	Parameter	Value
α	1.00	β	1.00
r^f	0.00	γ	1.00
λ	2.25	$\hat{b}$	0.65
c_1	0.65	c_2	0.65