


# Conditional risk and the pricing kernel<sup>☆</sup>

David Schreindorfer<sup>a,\*</sup>, Tobias Sichert<sup>b</sup> 

<sup>a</sup> Eli Broad College of Business, Michigan State University, 632 Bogue St, 48824 MI, East Lansing, USA

<sup>b</sup> Stockholm School of Economics, Bertil Ohlins gata 4, Stockholm, 11350, Sweden

## ARTICLE INFO

### JEL classification:

G12  
G13  
G33

### Keywords:

Pricing kernel  
Risk–return trade-off  
Volatility  
Equity index options  
Variance premium  
Conditional density estimation

## ABSTRACT

We propose a statistical methodology for jointly estimating the pricing kernel and conditional physical return densities from option prices. Pricing kernel estimates show that negative stock market returns are significantly more painful to investors in low-volatility periods. Density estimates reflect a significantly positive risk–return trade-off, suggest that Martin’s (2017) lower bound on the equity premium is violated in high-volatility periods, and provide new evidence on the variance premium’s predictive power for excess returns as well as the co-movement between higher return moments. Lastly, we show that leading macrofinance models are at odds with basic features of conditional stock market risks and risk pricing.

Stock market risks and their pricing are important for investors and policymakers alike. Unfortunately, conditional risks beyond volatility are challenging to estimate and expected returns are even more challenging to estimate. Our understanding of conditional risk and its relation to expected returns is therefore far from complete. In this paper, we propose an empirical framework for jointly estimating conditional expected returns, conditional risk, and conditional risk prices based on index options and return data.

Our estimator is based on the fact that the pricing kernel’s projection onto returns equals the ratio of risk-neutral and physical return densities, scaled by the risk-free rate (see, e.g., [Cochrane 2005](#)),

$$E_t[M_{t+1}|R_{t+1}] = \frac{1}{R_t^f} \frac{f_t^*(R_{t+1})}{f_t^f(R_{t+1})}. \quad (1)$$

We extract risk-neutral return densities,  $f_t^*(R_{t+1})$ , from index option prices based on the seminal result of [Breedon and Litzenberger \(1978\)](#). Next, we map them to physical return densities,  $f_t(R_{t+1})$ , based on Eq. (1) and a polynomial approximation of the projected pricing kernel,  $E_t[M_{t+1}|R_{t+1}]$ . Finally, we estimate polynomial parameters by maximizing the likelihood of realized returns, which are drawn from each  $f_t(R_{t+1})$ . To account for the possibility that risk and risk prices co-move, we allow the polynomial’s shape to vary with volatility. Our approach yields maximum likelihood estimates of the projected pricing kernel and conditional return density on different dates.

The pricing kernel’s projection onto stock market returns reveals how investors’ marginal utility varies with returns. [Fig. 1](#) plots our

<sup>☆</sup> Dimitris Papanikolaou was the editor for this article. This paper largely subsumes our 2021 working paper “Volatility and the Pricing Kernel”. We thank an anonymous referee, our discussants Ian Dew-Becker, Jens Christensen, Kris Jacobs, Christian Skov Jensen, Mete Kilic, Dmitriy Muravyev, Sean Myers, and Paola Pederzoli, as well as Caio Almeida, Adrien d’Avernas, Tyler Beason, Oliver Boguth, Niels Gormsen, Christian Julliard, Lars-Alexander Kuehn, Ian Martin, Jonathan Payne, Seth Pruitt, Robert Ready, Guillaume Roussellet, Riccardo Sabbatucci, Petra Sinagl, Ivan Shaliastovich, Per Strömberg, and seminar participants at Arizona State, Stockholm School of Economics, McGill, Princeton, Carnegie Mellon, University of Amsterdam, University of Arizona, University of Oregon, Goethe University, ITAM, University of Georgia, University of Virginia, University of Iowa, Michigan State, Vanderbilt University, the Virtual Derivatives Workshop, JEF seminar, BI-SHoF Conference 2022, CICF 2022, SoFiE Conference 2022, EFA 2022, BMRG Conference 2023, FMA CBOE 2023, MFA 2023, SGF Conference 2023, AFA 2024 and ESSFM 2024 for helpful comments and suggestions. Tobias Sichert gratefully acknowledges support from the Jan Wallanders and Tom Hedelius foundation, Tore Browaldhs foundation, grant no. Fh21-0026 and from the Swedish House of Finance.

Corresponding author.

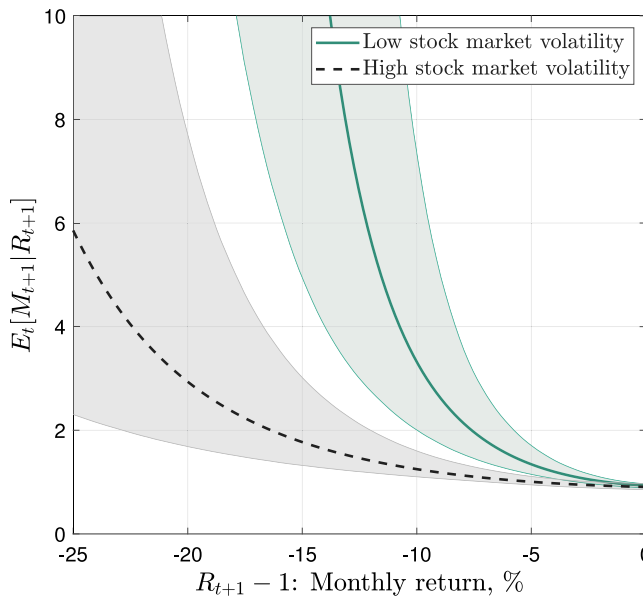


Fig. 1. Volatility and the projected pricing kernel. We plot the projected pricing kernel for the 10th and 90th percentile of conditional stock market volatility. The pricing kernel is measured at a monthly horizon, parameterized by Eqs. (4) and (5) with a polynomial order of  $N = 2$ , and estimated over the 1990–2023 sample. Shaded areas represent pointwise 90% confidence bounds, computed based on a block bootstrap with a block length of 21 trading days.

estimate for the 10th and 90th percentile of return volatility.<sup>1</sup> The steeper curve in periods of low volatility shows that negative returns are significantly more painful to investors when they occur in calm markets. For example, monthly returns of  $-10\%$  coincide with an average marginal utility of 3.68 when they occur in periods of low volatility, but only a value of 1.32 when they occur in periods of high volatility. A battery of robustness tests illustrates that this finding is not sensitive to our parametric assumptions.

Time-variation in the projected pricing kernel drives a wedge between the dynamics of conditional return moments under the physical and risk-neutral probability measures. In particular, the flattening projection in times of high volatility lowers risk premia and thereby reduces the difference between physical and risk-neutral moments. It therefore plays a key role in our estimation of conditional return densities from option prices. We explore the economic implications of our estimates based on three applications.

Our first application examines the risk–return relation in the time series. A large literature has found mixed and statistically weak results (e.g., French et al. 1987 and Glosten et al. 1993) and has failed to reach a consensus on the existence of a risk–return trade-off. However, an important limitation of most prior studies is that they rely on realized returns as a proxy for expected returns. In his presidential address, Elton (1999) points out that realized returns reflect large and potentially serially correlated innovations that are likely to bias inference about expected returns. A few prior papers have assessed the risk–return relation based on an explicit estimate of expected returns

(Guo and Whitelaw 2006, Ludvigson and Ng 2007, Pástor et al. 2008), and consistently found a significantly positive relationship. Our analysis complements this literature by examining the risk–return trade-off based on option prices. An advantage of this approach is that it allows us relate the risk–return relation to properties of the pricing kernel.

Our estimates imply that rising volatility affects expected returns through two opposing channels. It increases expected returns due to higher risk, as suggested by economic intuition, and it decreases expected returns due to lower risk prices, as shown in Fig. 1. Moreira and Muir (2017) previously argued for a countercyclical price of risk as an explanation for the missing risk–return trade-off. An important contribution of our paper is to provide an explicit estimate of risk prices (the projected pricing kernel) and the risk–return relation it implies. To that end, we consider different pricing kernel specifications and evaluate their fit to the data.

We show that, when return densities are estimated based on a log-quadratic (or higher order) polynomial for the pricing kernel, the risk effect dominates the price effect and expected returns are positively related to conditional volatility. Statistically, the estimated regression slope is significant at the 5% or 1% level, depending on the assumed polynomial order. Economically, a one standard deviation increase in volatility raises expected returns by a sizable 3.8 percentage points per year. Hence, our estimates imply a strong risk–return trade-off despite countercyclical risk prices.

Two alternative specifications illustrate key drivers of this finding. First, when we estimate return moments based on a (time-varying) log-linear pricing kernel, we fail to find evidence of a significant risk–return trade-off. Second, when we estimate return moments based on a time-invariant (log-quadratic) pricing kernel, which implies that increasing volatility is not accompanied by decreasing risk prices, we unsurprisingly find a much stronger risk–return trade-off. However, both alternatives provide a poor fit to the data and are statistically rejected at the 1% level in favor of the benchmark specification. Convexity and time-variation in the projected pricing kernel are therefore both critically important for identifying the empirical risk–return relation based on option prices.

Pricing kernel convexity has the additional implication that systematic skewness is a priced risk factor, as shown theoretically by Harvey and Siddique (2000). Their study documents a sizable premium for systematic skewness risk in the cross section of stock returns. We provide corroborating time series evidence based on bivariate risk–return regressions for the market. Specifically, we show that, when return moments are estimated based on a log-quadratic (or higher order) polynomial for the pricing kernel, expected excess returns are significantly related to conditional skewness after controlling for volatility. A one standard deviation decrease in skewness (which makes skewness more negative) increases expected returns by about 1 percentage point per year. However, skewness is not significantly related to expected returns when return moments are estimated based on a log-linear pricing kernel, in line with Harvey and Siddique’s (2000) theoretical results. To our knowledge, these findings represent the first time-series evidence on the relation between systematic skewness risk and expected returns.

As a second application, we use our estimates to provide a critical assessment of prior findings about conditional risk and risk premium in Martin (2017), Johnson (2019), and Gormsen and Jensen (2023).

Martin (2017) shows that, under the “negative correlation condition” (NCC)  $\text{Cov}_t[M_{t+1}R_{t+1}, R_{t+1}] \leq 0$ , the market’s risk-neutral variance serves as a lower bound for its expected excess return. If the NCC is violated, risk neutral variance instead provides an upper bound. Martin does not test the NCC empirically, but verifies that it holds in a variety of asset pricing models. We test the condition empirically by comparing our estimate of expected excess returns to Martin’s bound. We find that, as a result of time-variation in the projected pricing kernel, the NCC is violated in periods with large volatility spikes. Specifically, Martin (2017) shows that the condition holds for an investor who is fully invested in the market, as long as the investor’s risk aversion exceeds

<sup>1</sup> A number of influential papers have documented the puzzling fact that the projected pricing kernel is a non-monotonic function of returns; see, Aït-Sahalia and Lo (2000), Jackwerth (2000), and Rosenberg and Engle (2002). Our estimates imply the same non-monotonicity in the positive return region, as we show in the online appendix. The main text focuses on the negative return region to draw attention to the novel fact we document – covariation with volatility – and away from the existing fact. Martin and Papadimitriou (2022) show that pricing kernel non-monotonicity arises as an equilibrium outcome in a model with heterogeneous beliefs.

one at all times. This situation is equivalent to the projected pricing kernel maintaining a sufficiently steep slope. However, as Fig. 1 shows, the kernel flattens in times of high volatility, which induces large NCC violations during the financial crisis of 2008 and the Covid-19 pandemic of 2020. Our estimates therefore show that Martin's bound overstates the extent to which risk premia spike during economic crises. Simulation evidence suggests that the detected NCC violations cannot be explained by estimation noise.

Bollerslev et al. (2009) show based on OLS regressions that the variance premium is a significant predictor of excess stock market returns. Johnson (2019) challenges this seminal finding by showing that the predictive relation is not robust to WLS estimation, which is statistically more efficient than OLS because it under-weights noisy return realizations in periods of high volatility. Based on Johnson's evidence, the variance premium's predictive power appears spurious. However, the findings in Bollerslev et al. (2009) and Johnson (2019) are both subject to the aforementioned (Elton, 1999) critique, because they rely on realized returns as a proxy of expected returns. Our estimates allow us to revisit this issue by regressing expected (rather than realized) excess returns on the variance premium. We find a highly significant relation, in line with Bollerslev et al.'s original OLS finding.

Gormsen and Jensen (2023) study time-variation in higher moments. One of their key findings is that stock market returns become more left-skewed in times of low volatility, which has important implications for regulatory value-at-risk measures. A caveat is that Gormsen and Jensen's conclusions about (physical) returns are based on risk-neutral volatility and skewness estimates, whose dynamics may differ from those of their physical counterparts. Our estimates show that rising volatility affects physical skewness via two channels. First, it is associated with an increase in risk-neutral skewness, as documented by Gormsen and Jensen (2023). This channel increases physical skewness. Second, it is associated with a flattening pricing kernel, which reduces the difference between risk-neutral and physical moments. This channel decreases physical skewness. We find that the two channels quantitatively offset each other. As a result, physical skewness is not significantly correlated with volatility. However, in line with Kozhan et al.'s (2013) observation that the returns of variance and skewness swaps are strongly correlated, we find significant co-movement between ex ante measures of variance and skewness premia. This co-movement explains why Gormsen and Jensen (2023) find a strong correlation between risk-neutral variance and skewness measures.

As a third application, we investigate the ability of macrofinance models to capture basic features of conditional stock market risk and risk prices, as reflected in  $f_t(R_{t+1})$  and  $E_t[M_{t+1}|R_{t+1}]$ . Our analysis considers eleven different models with a variety of economic mechanism, including external habits (Campbell and Cochrane 1999, Bekaert and Engstrom 2017), long-run risks (Bansal and Yaron 2004, Drechsler and Yaron 2011, Schorfheide et al. 2018), rare disasters (Barro 2009, Wachter 2013, Gabaix 2012), incomplete markets (Constantinides and Ghosh 2017), disappointment aversion (Schreindorfer 2020), and slow-moving beliefs about volatility (Lochstoer and Muir 2022). We summarize the typical shape and the amount of time-variation in  $f_t(R_{t+1})$  and  $E_t[M_{t+1}|R_{t+1}]$  with a small set of moments and examine the models' ability to replicate them in finite sample simulations. The results are disillusioning. None of the models come close to capturing basic properties of  $f_t(R_{t+1})$  or  $E_t[M_{t+1}|R_{t+1}]$ . The models fail to capture asymmetries in the conditional return distribution and the projection's steep slope. Additionally, they are at odds with the amount of cyclical variation in both functions. These findings are troubling because it is arguably the models' main objective to explain the nature and pricing of stock market risks.

**Related Estimation Approaches.** Our estimation approach builds on several prior estimators of conditional return densities and the pricing kernel.

The most common approach to conditional density estimation assumes a parametric density and models its parameters as functions of

covariates (Hansen 1994). In a similar spirit, we model parameters of the projected pricing kernel as functions of conditional volatility. However, our methodology has several advantages. First, we specify conditional return densities as transformation of risk-neutral densities, which can be extracted non-parametrically from option prices. As a result, our density estimates can take on a much broader set of shapes than commonly used parametric families, such as the Normal, Student- $t$ , or skewed Student- $t$  distribution. Second, it leverages the rich conditioning information in risk neutral densities. This makes it easier to identify time-variation in higher return moments. Third, our methodology estimates conditional return densities jointly with the projected pricing kernel. This enables us to quantify the relative importance of risk and risk prices for fluctuations in risk premia.

In modeling the logarithm of the projected pricing kernel as a polynomial, we build on Rosenberg and Engle (2002). Our study differs from theirs along two key dimensions. First, Rosenberg and Engle model the physical return density as a GARCH process, whereas our methodology requires no parametric assumptions about the physical probabilities. Instead, we extract risk-neutral densities from option prices and transform them to conditional physical densities based on a parametric pricing kernel. Second, Rosenberg and Engle estimate the parameters of their pricing kernel polynomial separately for every day in their sample. Instead, we model them as (time-invariant) functions of volatility. Doing so allows us to formally test for time-variation in  $E_t[M_{t+1}|R_{t+1}]$  by evaluating the hypothesis that the pricing kernel's volatility-dependence parameter equals zero.

We also build on estimation approaches that take the risk-neutral distribution as given, parameterize the projected pricing kernel, and estimate parameters via a criterion function based on realized returns. Bliss and Panigirtzoglou (2004) maximize the  $p$ -value of a Berkowitz test for uniformity and independence of returns, whereas Linn et al. (2018) minimize a generalized method of moments criterion for moments of the inverse conditional CDF of returns. We follow the same general idea, but recognize that estimation can be performed based on a more conventional maximum likelihood approach. In particular, once a functional form has been specified for  $E[M|R]$ , it can be used to map the option-implied risk-neutral distribution to an analogous physical distribution, based upon which the likelihood function of returns can be computed. Compared to Bliss and Panigirtzoglou (2004) and Linn et al. (2018), an important advantage of our approach is that it implies conditional return densities that integrate to a probability mass of one.<sup>2</sup>

A contemporaneous and independently developed paper by Kim (2022) also studies time-variation in the projected pricing kernel. Kim builds on the GMM approach of Linn et al. (2018) and shows how  $E_t[M_{t+1}|R_{t+1}]$  varies with different macroeconomic covariates, including volatility.<sup>3</sup> Apart from proposing a different estimation approach, our study differs from Kim's in its applications. We evaluate what our estimates imply about the risk-return trade-off, prior findings on conditional risk and risk premia, and macrofinance models, whereas Kim (2022) studies a conditional version of the equity premium decomposition in Beason and Schreindorfer (2022) and uses his approach for out-of-sample prediction of market returns. Lastly, we show that, despite being more parsimonious, our approach for modeling the kernel's volatility dependence provides a significantly better fit to the data than Kim's.

<sup>2</sup> Cuesdeanu and Jackwerth (2018) point out that Linn et al.'s (2018) estimation approach implies return density whose mass deviates substantially from one.

<sup>3</sup> Similarly, Driessen et al. (2020) examine whether the projected pricing kernel varies with the Chicago Fed National Activity Index, but find relatively small effects and do not evaluate their statistical significance.

## 1. Estimation

This section explains our estimation approach, discusses data sources, and illustrates the robustness and statistical significance of our estimates. Throughout, the pricing kernel in period  $(t+1)$  is denoted by  $M_{t+1}$ , the risk-free rate by  $R_t^f$ , the ex-dividend market return by  $R_{t+1}$ , risk-neutral objects by a “\*”-superscript, and objects that condition on investors’ information set at time- $t$  by a “t”-subscript. We later drop “t”-subscripts for readability when not needed for clarity.

### 1.1. Estimation approach

The absence of arbitrage opportunities implies Eq. (1) in the introduction: The pricing kernel’s projection onto stock market returns equals the ratio of the conditional risk neutral and physical return densities, scaled by the risk-free rate. The projected pricing kernel  $E_t[M_{t+1}|R_{t+1}]$  measures the mean of  $M_{t+1}$  conditional on investors’ information set at time- $t$  and conditional on a (potential) return outcome at time- $(t+1)$ . To estimate it empirically, we extract  $f_t^*$  from option prices for each day of the sample based on the classic result of Breeden and Litzenberger (1978). Our implementation is standard and follows Beason and Schreindorfer (2022). Next, we model the projection with a flexible parametric function of returns and conditional volatility,  $M(R_{t+1}, \sigma_t; \theta)$ , and combine it with Eq. (1) to express the conditional physical density as

$$f_t(R_{t+1}; \theta) = \frac{f_t^*(R_{t+1})}{R_t^f \times M(R_{t+1}, \sigma_t; \theta)}. \quad (2)$$

$\sigma_t \equiv \text{Std}_t[\ln R_{t+1}]$  denotes the conditional volatility of log returns and is part of investors’ time- $t$  information set. Its estimation is detailed in Appendix A and discussed below. Given a functional form for  $M(R_{t+1}, \sigma_t; \theta)$ , the unknown parameter vector  $\theta$  can be estimated by maximizing the log-likelihood of realized returns,

$$LL(\theta) = \sum_{t=1}^T \ln f_t(R_{t+1}; \theta). \quad (3)$$

A few comments may be useful. First, our notation highlights  $f_t$ ’s dependence on the parameter vector  $\theta$ , but it is important to emphasize that the density does not belong to a known parametric family of distributions. Rather, it results from applying a (parametric) change-of-measure to the (nonparametric) risk-neutral distribution  $f_t^*$ , whose shape is implied by the market prices of equity index options. Compared to parametric models of  $f_t$ , our density estimates can therefore take on a wider variety of shapes. Obviously, it is unusual and perhaps even unique for a maximum likelihood estimator not to require any distributional assumptions. Second, with a slight abuse of notation, we use “ $R_{t+1}$ ” to denote both a potential return outcome (of which there are many each period) and an actual return realization (of which there is only one each period). Third, Eqs. (2) and (3) show that our estimation of the projected pricing kernel can equivalently be interpreted as an estimation of the conditional return density. In particular,  $\theta$  is chosen to find the densities  $f_t$  that maximize the likelihood of realized returns, given the information in  $f_t^*$  and an assumed functional form for  $E_t[M_{t+1}|R_{t+1}]$ . Many prior estimators of the pricing kernel do not allow for this dual interpretation because they are inconsistent with a return density that integrates to one; e.g., Bliss and Panigirtzoglou (2004) and Linn et al. (2018). Fourth, the term “projected” pricing kernel is taken from prior literature and should not be interpreted as a linear projection. Rather,  $E_t[M_{t+1}|R_{t+1}]$  is generally a nonlinear conditional expectation function of  $R_{t+1}$  for any time- $t$  information set. Apart from the market return, it averages over all shocks that affect the pricing kernel at  $(t+1)$ . Fifth, the maximum likelihood methodology makes our estimator statistically efficient and it incorporates conditioning information from the entire risk-neutral distribution. Both features represent important advantages over moment-based estimation approaches of the pricing kernel. Sixth, the log-likelihood in Eq. (3) has to be maximized

numerically, which we implement based on Matlab’s `fmincon` routine. The optimization executes most efficiently if one provides `fmincon` with an analytical expression of the score vector (the log-likelihood’s gradient), which we derive in Section II of the online appendix.

### 1.2. Parameterization

To ensure positivity, we model the projection as an exponential polynomial,

$$M(R_{t+1}, \sigma_t; \theta) = \exp \left\{ \delta_t + \sum_{i=1}^N c_{it} \times (\ln R_{t+1})^i \right\}, \quad (4)$$

where the polynomial coefficients  $c_{it}$  vary with volatility according to

$$c_{it} = \frac{c_i}{\sigma_t^{b \times i}}, \quad (5)$$

and  $\theta = (b, c_1, \dots, c_N)$  is a vector of unknown parameters. The time-varying intercept,  $\delta_t$ , is calculated for each day of the sample to satisfy the theoretical restriction that  $f_t(R_{t+1}; \theta)$  integrates to one, i.e.,  $\delta_t$  does not represent a free parameter.<sup>4,5</sup> This restriction is crucial because the likelihood function is not well-defined for “densities” whose probability masses differ from one another (and from one).

We experimented with different functional forms for the time-varying polynomial coefficients  $c_{it}$ , and found that (5) provides a very good fit (in terms of log-likelihood) despite its parsimony. Additionally, we found that the relationship between  $c_{it}$ ’s and  $\sigma_t$  looks very similar when estimated based on the more flexible polynomial specification  $c_{it} = b_0 + b_1 \sigma_t + b_2 \sigma_t^2 + \dots$  (see Online Appendix I.B). This alternative is used to illustrate the robustness of our results in Section 1.6.

Lastly, it is worth noting that (5) nests two interesting special cases. For  $b = 0$ , the projected pricing kernel equals a time-invariant function of returns,

$$M(R_{t+1}; \theta) = \exp \left\{ \delta_t + \sum_{i=0}^N c_i \times (\ln R_{t+1})^i \right\}, \quad (6)$$

i.e., the graph of  $E[M|R]$  does not vary with volatility, apart from a small shift induced by  $\delta_t$ . For  $b = 1$ , the projected pricing kernel equals a time-invariant function of *standardized returns* (up to a small shift due to  $\delta_t$ ),

$$M(R_{t+1}, \sigma_t; \theta) = \exp \left\{ \delta_t + \sum_{i=0}^N c_i \times \left( \frac{\ln R_{t+1}}{\sigma_t} \right)^i \right\}. \quad (7)$$

In this case, the graph of  $E[M|R]$  scales horizontally and proportionally with volatility. Intermediate values of  $b$  allow  $E[M|R]$  to change with volatility to varying degrees. To formally test whether  $E[M|R]$  varies with volatility, we evaluate the hypothesis  $H_0 : b = 0$ . To test whether the  $E[M|R]$  is well-described by a time-invariant function of standardized returns, as in Eq. (7), we evaluate the hypothesis  $H_0 : b = 1$ .

<sup>4</sup> The intercept equals  $\delta_t = -\ln R_t^f + \ln \left( \int_0^\infty f^* \times \exp \left\{ -\sum_{i=1}^N c_{it} \times (\ln R_{t+1})^i \right\} dR_{t+1} \right)$ , i.e., its value is implied by  $R_t^f$ ,  $f_t^*$ , and the polynomial coefficients  $(b, c_1, \dots, c_N)$ . We find  $\delta_t$  for each date by evaluating this integral numerically.

<sup>5</sup> Instead of computing  $\delta_t$  based on the theoretical restriction  $\int f = 1$ , one could add a time-varying intercept  $c_{0t}$  to polynomial (4) and model  $c_{0t}$  as a function of volatility. Since this approach does not guarantee  $\int f = 1$ , however, it becomes necessary to add a penalty for violations of the restriction to the objective function. In turn, doing so requires the researcher to make a subjective choice on the relative importance of the restriction and the fit to realized returns. In the context of a moment-based estimation of the pricing kernel, a contemporaneous paper by Kim (2022) adds such a penalty term to his objective function, whereas Linn et al. (2018) simply ignore the theoretical restriction that probabilities add up to one.



**Table 1**  
Sensitivity of moments to parameters.

This table shows the sensitivity of conditional return moments (in rows) with respect to pricing kernel parameters (in columns). The sensitivity of moment  $i$  with respect to parameter  $j$  equals  $\frac{d m_i}{d \theta_j} \times \left| \frac{\theta_j}{m_i} \right|$  and is evaluated at the point estimates for the  $N = 2$  specification from Table 2.

	$b$	$c_1$	$c_2$
Average $E_t[R_{t+1}]$	1.85	-0.34	0.14
Average $\text{Var}_t[R_{t+1}]$	-1.41	0.04	-0.23
Average $\text{Skew}_t[R_{t+1}]$	2.45	-0.08	0.35
Average $\text{Kurt}_t[R_{t+1}]$	-1.44	0.01	-0.22
$\text{corr}(E_t[R_{t+1}] - E_t^*[R_{t+1}], \text{Var}_t[R_{t+1}])$	-0.25	-0.07	0.01

### 1.3. Parameter identification

The pricing kernel controls the extent to which conditional real world probabilities differ from their risk-neutral counterparts. Specifically, (2) shows that  $f_t(R)$  takes on smaller values than  $f_t^*(R)$  for return regions where  $M(R_{t+1}, \sigma_t; \theta) > 1/R_t^f$  and higher values where  $M(R_{t+1}, \sigma_t; \theta) < 1/R_t^f$ . Individual elements of  $\theta = \{c_1, \dots, c_N, b\}$  are therefore identified if they alter the shape of  $E[M|R]$  in such a way that it better explains the relative likelihood of different return realizations. Since  $f_t^*$  does not vary with  $\theta$ , one can equivalently think of parameters as being identified by risk premia: An increase in the mean of  $f_t$  is equivalent to a higher equity premium, an increase in the variance of  $f_t$  is equivalent to a higher (less negative) variance premium, etc.

Most elements of  $\theta$  alter the shape of  $f_t$  in multiple ways relative to that of  $f_t^*$ . This is illustrated quantitatively in Table 1, which shows the sensitivity of different  $f_t$  moments to individual parameters. In what follows, we discuss the main sources of parameter identification.  $c_1$ , the slope of  $E[M|R]$ , controls the relative probabilities of negative and positive returns. If the slope is negative, for example, the left tail of  $f_t^*$  gets downweighted in computing  $f_t$ , whereas the right tail gets upweighted.  $c_1$  is therefore identified by the mean of  $f_t$  and the relative likelihood of negative returns.  $c_2$ , the curvature of  $E[M|R]$ , controls the relative probabilities of small and large absolute returns. If the curvature is positive, both extreme tails of  $f_t$  get downweighted relative to the tails of  $f_t^*$ , whereas the center of the distribution gets upweighted. Hence,  $c_2$  is identified by the variance of  $f_t$  and the relative likelihood of extreme returns. The mean of  $f_t$  (equivalently, the equity premium) also contributes to the identification of  $c_2$ . Specifically,  $E_t[R_{t+1}]$  is increasing in  $c_2$  because  $f_t^*$  is left-skewed. Similarly,  $c_3, c_4$ , etc. are mainly identified by higher order moments of  $f_t$ . The scaling parameter  $b$  controls how parameters of  $E[M|R]$  vary with volatility, and therefore the amount of time-variation in the probabilities of different returns. For  $b = 0$ , the shape of  $E[M|R]$  is time-invariant and an increase in volatility is accompanied by an increase in the equity premium, which implies that the expected excess returns and volatility are positively correlated. For  $b > 0$ , an increase in volatility makes the slope of  $E[M|R]$  less negative and its curvature less positive. Both effects lower expected returns and thereby reduce the positive correlation between the expected excess returns and volatility.  $b$  is therefore identified by the strength of the risk–return trade-off.

### 1.4. Data

Our empirical analysis is based on the S&P 500 index and a return horizon of one month (30 calendar days). Return data comes from the Center for Research in Security Prices (CRSP). Price quotes on SPX options for the estimation of  $f_t^*$  come from the Chicago Board Options Exchange (CBOE). Because this data limits our sample to the 34-year period from 1990 to 2023, we sample daily to maximize the efficiency of our estimates, i.e., we work with a daily sample of  $T = 8,553$  overlapping monthly returns. The conditional volatility of log returns,  $\sigma_t$ , is estimated with the heterogeneous autoregressive (HAR) model

of Corsi (2009), as detailed in Appendix A.<sup>6</sup> The estimation relies on intra-daily price quotes for S&P 500 futures, which were purchased from TickData. We use quotes for the large futures contract (ticker “SP”) from 1990 to 2002, and for the E-Mini Futures contract (ticker “ES”) from 2003 to 2023, i.e., we use data for the more actively traded futures contract in each part of the sample.

The COVID-19 episode represents an important challenge for our estimation. The S&P 500 dropped -31.9% in the 30-day period following 2020/02/19. This event was so extreme based on the economic environment at the time that our estimator assigns it an ex ante probability of only 0.000017% (1 event every 500,253 years),<sup>7</sup> despite being designed to maximize the likelihood of returns. Clearly, this observation makes the overall sample non-representative. We therefore drop eight overlapping 30-day returns that follow the surrounding days between 2020/02/11 (whose conditional probability maps to 1 event every 8432 years) and 2020/02/21 (whose conditional probability maps to 1 event every 64,059 years) from the sample.

### 1.5. Estimation results

Table 2 shows estimates of the parameterized pricing kernel in Eqs. (4) and (5) and polynomial orders between  $N = 1$  and  $N = 5$ . To account for autocorrelation that results from our use of overlapping return data, we rely on a block bootstrap with a block length of 21 trading days for all statistical tests.

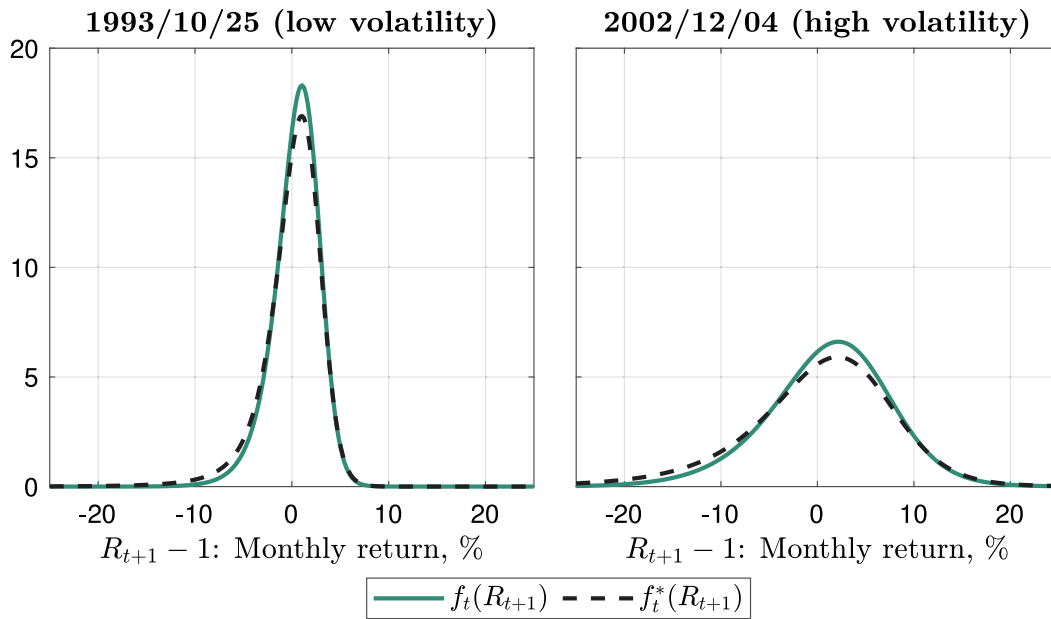
The first important feature of our estimates lies in time-variation of the projected pricing kernel, which is governed by the volatility-scaling parameter  $b$ . When we model the projected pricing kernel as a log-linear ( $N = 1$ ) polynomial, we fail to reject the hypothesis  $H_0 : b = 0$  at common significance levels. Log-linearity therefore suggests that the projection is time-invariant. However, the same hypothesis is rejected at the 1% level when we model the projection with a log-quadratic ( $N = 2$ ) or higher-order ( $N \geq 3$ ) polynomial. To conclude that the shape of  $E[M|R]$  varies with volatility, it is therefore necessary to capture the projection’s convexity. We examine evidence for convexity below.

We also find that  $E[M|R]$  is well-described as scaling proportionally with volatility, as the volatility-scaling parameter  $b$  is estimated to be close to (and insignificantly different from) one for all polynomial orders. As explained above, this implies that  $E[M|R]$  is approximately a time-invariant function of standardized returns.

The second important feature of our estimates lies in the fit of different polynomial orders. Based on the log-linear ( $N = 1$ ) projection, we reject the hypothesis  $H_0 : c_1 = 0$  at the 1% level. This rejection is unsurprising because  $c_1 = 0$  implies a horizontal pricing kernel and zero risk premia – a poor description of the data. For the log-quadratic ( $N = 2$ ) specification, we similarly reject the hypothesis  $H_0 : c_2 = 0$  at the 1% level, i.e., the  $N = 2$  specification is statistically preferred to the  $N = 1$  specification. The projected pricing kernel therefore features significant convexity. This finding mirrors (Harvey and Siddique, 2000), who find significant evidence for pricing kernel convexity in the cross-section of stock returns. Further increasing the polynomial order shows that the  $N = 2$  specification is not rejected in favor of  $N = 3$ ,  $N = 3$  is rejected in favor of  $N = 4$ , and  $N = 4$  is not rejected in favor of  $N = 5$ . To evaluate formally whether  $N = 4$  provides a better fit than  $N = 2$ , one needs to test the joint hypothesis  $H_0 : c_3 = c_4 = 0$ . Unfortunately, the bootstrap is unsuitable for

<sup>6</sup> Figure IA.III of the online appendix shows that we find similar time-variation in  $E[M|R]$  when conditional volatility is measured by the VIX index instead. This approach has the advantage of not relying on a parametric time series model for volatility, but the disadvantage that the VIX index embeds a risk premium and therefore systematically overestimates volatility.

<sup>7</sup> We compute the probability as  $\int_{-100\%}^{-31.9\%} f_t(R) dR$  based on the conditional density  $f_t$  on 2020/02/19.



**Fig. 2.** Conditional density estimates for select days. We plot the estimated physical and risk-neutral return density for the days on which conditional volatility is closest to its 10th (left panel) or 90th (right panel) percentile. Estimates are based on the  $E[M|R]$  specification in Eqs. (4) and (5) and a polynomial order of  $N = 2$ .

**Table 2**

**Estimation results.**

We estimate the projected pricing kernel in Eqs. (4) and (5) for different polynomial orders  $N$  by maximizing the log-likelihood of realized returns, Eq. (3). Bootstrap tests account for autocorrelation based on a block bootstrap with a block length of 21 trading days.

$N$	1	2	3	4	5
Log-likelihood	15,974	16,067	16,067	16,090	16,090
<b>Parameter Estimates</b>					
$\hat{c}_1$	-0.071	-0.072	-0.074	-0.117	-0.117
$\hat{c}_2$	—	0.087	0.090	0.015	0.015
$\hat{c}_3$	—	—	0.000	0.016	0.016
$\hat{c}_4$	—	—	—	0.004	0.004
$\hat{c}_5$	—	—	—	—	0.000
$\hat{b}$	1.168	0.980	0.976	1.080	1.079
<b>Bootstrap Tests (p-values in %)</b>					
$H_0 : b = 0$	17.2	0.7	0.7	0.2	0.2
$H_0 : b = 1$	90.5	92.0	91.3	80.7	81.1
$H_0 : c_N = 0$	0.0	0.0	52.1	0.2	27.9

joint hypothesis tests and we are unaware of established methods for adjusting likelihood ratio tests for autocorrelation. We therefore rely on an ad hoc approach that adjusts the likelihood ratios in Table 2 for our effective sample size. In Appendix B.1.4, we estimate that the effective sample size equals 804, i.e., approximately twice the number of months in our sample. Based on this effective sample size and our nominal sample size of  $T = 8,553$ , the adjusted likelihood ratio test statistic equals  $LR = -2 \left( 16,067 \times \frac{804}{8,553} - 16,090 \times \frac{804}{8,553} \right) = 4.1836$ , which has a  $p$ -value of 12.35% (untabulated) under the asymptotic chi-squared distribution with two degrees of freedom. Because we fail to reject  $N = 2$  in favor of  $N = 4$ , we rely on the more parsimonious  $N = 2$  specification as our benchmark case. All subsequent results are based on this estimate, unless otherwise mentioned.

Fig. 1 in the introduction illustrates graphically how  $E[M|R]$  varies with volatility by plotting it for the 10th and 90th percentile of  $\sigma_t$  ( $p_{10}$  and  $p_{90}$ ). The figure shows that the pricing kernel is considerably steeper when volatility is low. For example, for a monthly return of  $-10\%$ , the projected pricing kernel equals  $M(R_{t+1} = -0.1, \sigma_t = p_{10}; \theta) = 3.32$  when volatility is low and  $M(R_{t+1} = -0.1, \sigma_t = p_{90}; \theta) = 1.25$  when volatility is high.

Fig. 2 shows conditional return densities for two dates. For comparability with Fig. 1, we select days on which volatility is closest to its

10th or 90th percentile. Because our  $E[M|R]$ -parameterization implies a smooth change-of-measure,  $f_t$  inherits many of  $f_t^*$ 's properties. It is unimodal, roughly bell-shaped, and its conditional volatility moves with that of  $f_t^*$ . Relative to  $f_t^*$ , however,  $f_t$  has more probability mass in the center and less mass in the left tail. As a result, the physical density is less left-skewed and leptokurtic than its risk-neutral counterpart, the equity premium is positive, and the variance premium is negative.

Fig. 3 shows the time series of expected excess returns and conditional higher moments. Mean and volatility are visibly countercyclical and positively correlated with one another. In contrast, conditional skewness and kurtosis appear roughly acyclical. Across the 8553 trading days in our sample, the conditional physical (risk-neutral) density has an average mean of 9.34% (0.92%) p.a., standard deviation of 14.22% (18.34%) p.a., skewness of  $-0.63$  ( $-1.54$ ), and kurtosis of 4.49 (10.92).

It is important to ask whether our estimation methodology reliably identifies expected returns. We address this question by regressing realized excess returns onto our estimate of expected excess returns,

$$R_{t+1} - E_t^*[R_{t+1}] = \beta_0 + \beta_1 (E_t[R_{t+1}] - E_t^*[R_{t+1}]) + \varepsilon_{t+1}. \quad (8)$$

Table IA.I of the online appendix shows that the regression slope  $\beta_1$  is significant at the 5% level for  $N \geq 2$ , but insignificant at common significance levels for  $N = 1$ . For polynomial orders  $N \geq 2$ , the regression yields  $R^2$  values between 1.29% and 1.30% and a Wald test fails to reject the joint hypothesis  $H_0 : \beta_0 = 0, \beta_1 = 1$  at the 5% significance level. These results suggest that our estimator reliably identifies expected returns for log-quadratic and higher order specifications of the projected pricing kernel. Building on this observation, we use our expected return estimates to assess the risk–return trade-off in Section 2.

Another noteworthy feature of our estimates is the variance premium they imply.  $\text{Var}_t[R_{t+1}] - \text{Var}_t^*[R_{t+1}]$  averages  $-5.3\%^2$  per month for the  $N = 1$  specification,  $-12.6\%^2$  for  $N = 2$ , and between  $-12.6\%^2$  and  $-12.3\%^2$  for higher order polynomials. Which magnitude is more plausible? A simple estimate of the premium that does not depend on our estimate of the pricing kernel is  $\sigma_t^2 - \left( \frac{VIX_t}{100} \right)^2$ , which averages  $-13.6\%^2$  per month over our sample period. This comparison suggests that the log-linear ( $N = 1$ ) specification is misspecified, in the sense that it generates an implausibly small variance premium.

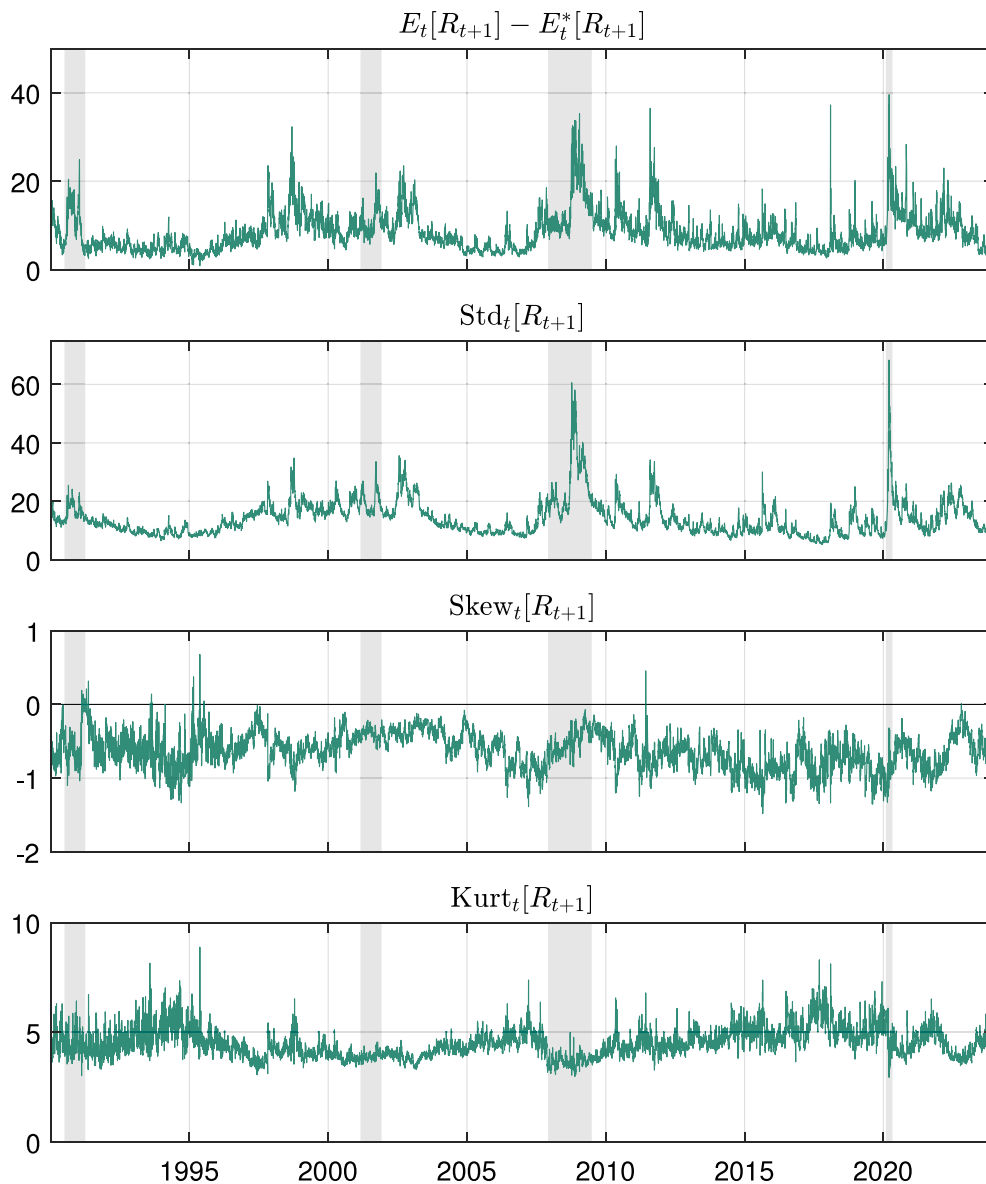


Fig. 3. Conditional return moments. This figure shows conditional return moments based on the  $E[M|R]$  specification in Eqs. (4) and (5) and a polynomial order of  $N = 2$ .

### 1.6. Robustness

We perform four robustness tests. Parameter estimates for these tests are reported in Section III of the online appendix.

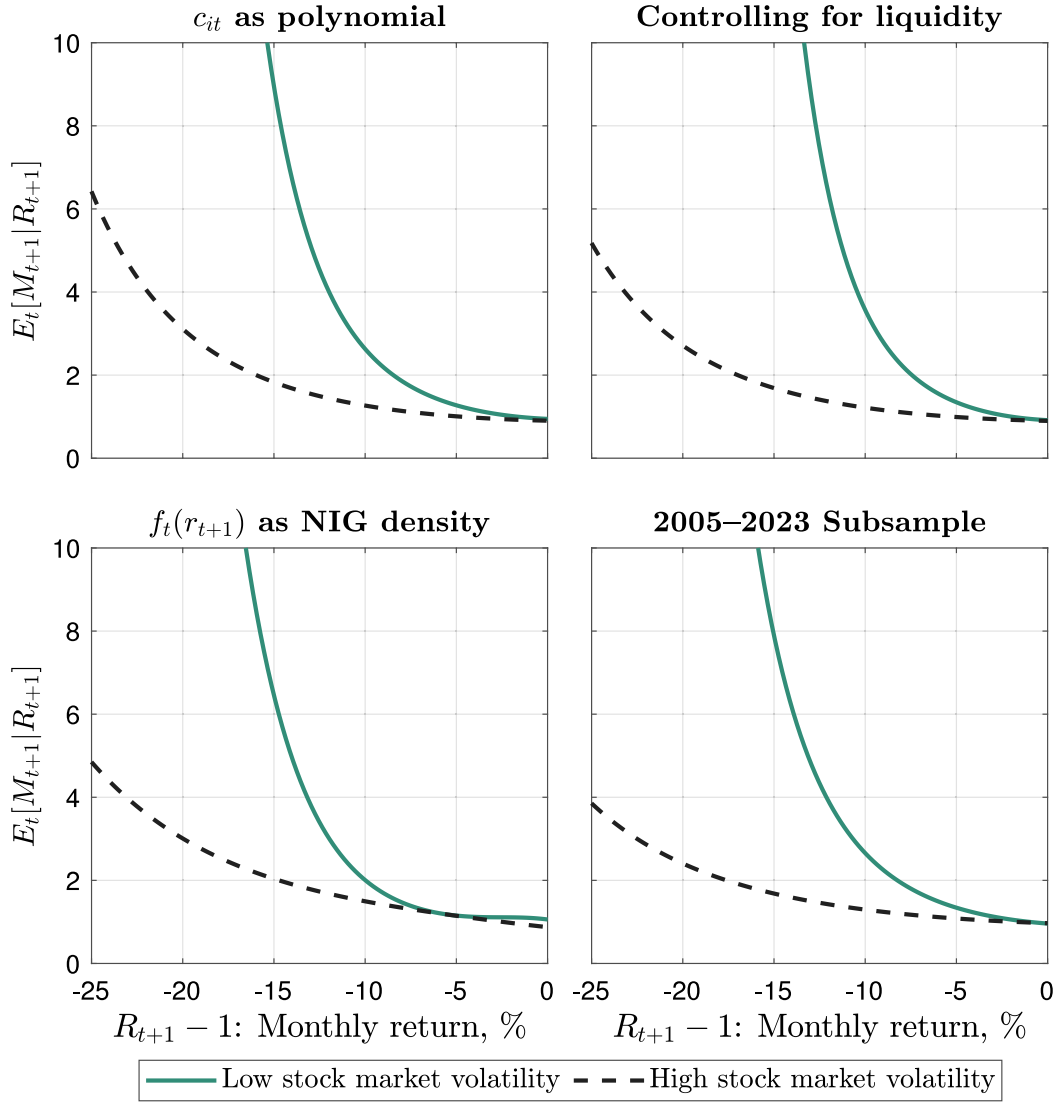
First, we model the projected pricing kernel's volatility-dependence with the alternative specification

$$c_{it} = \sum_{k=0}^K c_{ik} \times \sigma_t^k, \quad (9)$$

which assumes that coefficients of the  $E[M|R]$ -polynomial are themselves polynomials of volatility. The combination of (4) and (9) is equivalent to a bivariate polynomial in  $\ln R_{t+1}$  and  $\sigma_t$  with a tensor product base.<sup>8</sup> Table 3 reports log-likelihoods for  $N = 2$  and different polynomial orders  $K$ . Likelihoods fall below the one from our  $N = 2$  benchmark specification for  $K \in \{1, 2\}$  and above it for  $K \geq 3$ . To formally compare the alternative specification to our benchmark, we

<sup>8</sup>  $K = 1$  corresponds to the specification of Kim (2022), whose study is discussed under “Related Estimation Approaches” on page 8.

rely on Vuong's (1989) likelihood-ratio test for non-nested models to test the null hypothesis that our benchmark Eq. (5) and the alternative Eq. (9) are equally close to the true data generating process against the alternative hypothesis that our benchmark specification is closer. The Vuong test penalizes models with more parameters analogous to the BIC and is commonly-used as a model-selection tool. Table 3 shows that the Null is rejected at the 1% level for any  $K$ , which implies that our benchmark specification is statistically preferred to the alternative. However, we illustrate in Section I.B of the online appendix that the functional relationship between  $\sigma_t$  and the polynomial coefficients  $c_{it}$  converges to a shape that closely resembles the one implied by our benchmark as  $K \rightarrow \infty$ . This observation is noteworthy because our benchmark specification relies on a single parameter ( $b$ ) to control time-variation in  $c_{it}$ 's, whereas the alternative specification relies on  $N \times K$  parameters to do so. Of course, the fact that time-variation in polynomial coefficients  $c_{it}$  looks similar to our benchmark for high  $K$  implies that time-variation in the projected pricing kernel looks similar as well. We illustrate this fact for  $K = 5$  in the top-left panel of Fig. 4. Despite its parsimony, our benchmark specification therefore does not appear to restrict the projected pricing kernel's volatility-dependence in an unrealistic manner.



**Fig. 4. Robustness.** We plot the projected pricing kernel for the 10th and 90th percentile of conditional stock market volatility. Top-left:  $E[M|R]$  is a  $N = 2$  polynomial with coefficients that depend on volatility via (9) with  $K = 5$ . Top-right:  $E[M|R]$  is equivalent to the benchmark specification, but  $\sigma_t$  is orthogonalized with respect to different liquidity measures. Bottom-left: We model the distribution of standardized log returns  $\ln R_{t+1}/\sigma_t$  with a Normal Inverse Gaussian distribution, compute  $f_t(r_{t+1})$  via a change-of-variables, and obtain  $E[M|R]$  from (1). Bottom-right: The benchmark  $E[M|R]$  specification is estimated over the 2005–2023 subsample.

**Table 3**

**Bivariate polynomial — specification test.**

We estimate the projected pricing kernel based on the bivariate polynomial specification in Eqs. (4) and (9) for  $N = 2$  and different polynomial orders  $K$  by maximizing the log-likelihood of realized returns, Eq. (3). The Vuong (1989) test for non-nested models tests the null hypothesis that our benchmark model “b” (Eqs. (4) and (5) for  $N = 2$ ) and this alternative model “a” are equally close to the true data generating process, against the alternative that the benchmark model is closer. To adjusted for autocorrelation due to overlapping data, we multiply log-likelihoods by the ratio of our effective and nominal sample sizes (804/8,553) in computing the test statistic  $Z = \frac{(LL^b - LL^a) \times \frac{804}{8553} - \frac{\text{para a} \cdot \# \text{ para b}}{2} \ln(804)}{\sqrt{804 \times \omega}}$ . The effective sample size is computed in Appendix B.1.4. “ $LL^x$ ” denotes the log-likelihood of the respective models  $x \in \{a, b\}$  and  $\omega$  is defined by setting  $\omega^2$  equal to the mean of the squares of the pointwise log-likelihood ratios  $\ln \frac{f_t(R_{t+1})}{f_t^b(R_{t+1})}$ .  $p$ -values are based on the test statistic’s asymptotic  $N(0, 1)$  distribution.

$K$	1	2	3	4	5
Log-likelihood	16,043	16,063	16,070	16,071	16,072
Vuong test statistic	2.83	10.82	17.40	23.52	29.94
Vuong $p$ -value, %	0.23	0.00	0.00	0.00	0.00

Second, we orthogonalize  $\sigma_t$  with respect to proxies of option market liquidity. This test addresses the potential concern that movements in  $E[M|R]$  reflect frictions in the options market that are positively correlated with stock market volatility. We proxy option market liquidity by (1) daily option volume (across all maturities and strike prices), normalized by the average daily volume over the prior three month to remove the time trend in volume (following Chen et al. 2019), (2) daily open interest (across all maturities and strike prices), similarly normalized by the trailing three months average and (3) the bid–ask spread of the put option that is closest to being at-the-money and a maturity of 30 calendar days, normalized by the option’s midquote. We then regress  $\sigma_t$  on the three liquidity variables, scale the resulting regression residuals so that they have the same mean and variance as  $\sigma_t$ , and re-estimate our benchmark specification of  $E[M|R]$  based on this liquidity-adjusted volatility measure. The estimated volatility-scaling parameter of  $\hat{b} = 1.06$  is very close to the benchmark estimate of 0.98 and it remains highly statistically significant with a  $p$ -value of 0.7%. The top-left panel of Fig. 4 shows that, as a result, the projected pricing kernel’s volatility-dependence closely resembles the one without liquidity controls.



Third, instead of modeling  $E[M|R]$  as a polynomial, we model  $f_t(R_{t+1}; \theta)$  as a parametric density and obtain  $E[M|R]$  from (1) as the ratio of risk-neutral and physical densities, scaled by the risk-free rate. Specifically, we parameterize the density of standardized log returns,  $g_t\left(\frac{\ln R_{t+1}}{\sigma_t}; \theta\right)$ , with a normal inverse Gaussian (NIG) distribution and compute the distribution of simple returns via a change of variables as  $f_t(R_{t+1}; \theta) = g_t\left(\frac{\ln R_{t+1}}{\sigma_t}; \theta\right) / (\sigma_t \times R_{t+1})$ . The NIG distribution is unimodal, bell-shaped, allows for nonzero skewness and excess kurtosis, and depends on four parameters, which we estimate via maximum likelihood. This method for estimating the conditional distribution resembles the popular approach of scaling historical return innovations from a GARCH model with a current estimate of conditional volatility – see, e.g., [Rosenberg and Engle \(2002\)](#), [Barone-Adesi et al. \(2008\)](#) and [Christoffersen et al. \(2013\)](#) — and shares its limitation that higher conditional moments (beyond volatility) are time-invariant by construction. In contrast, the parameterized pricing kernel in our benchmark specification allows all return moments to vary over time. The bottom-left panel of [Fig. 4](#) shows that projected pricing kernel's volatility-dependence for this alternative specification nevertheless closely resembles the one in our benchmark. The parametric assumptions we make about the projected pricing kernel therefore do not appear to be overly restrictive.

Fourth, we re-estimate the benchmark specification in the second part of the sample (2005–2023) to address concerns about a possible segmentation between index option and equity markets. In particular, [Dew-Becker and Giglio \(2022\)](#) argue that the two markets have historically been segmented, but also provide evidence suggesting that they have become well-integrated since about the mid 2000's. If time-variation in the estimated projected pricing kernel was a result of market segmentation, one would expect it to be substantially weaker in more recent data. The bottom-right panel of [Fig. 4](#) shows that this is not the case. The projected pricing kernel's volatility-dependence closely its full sample equivalent.

Overall, the robustness tests show that our main result is not sensitive to the way  $E[M|R]$  is parameterized, the volatility-dependence of  $E[M|R]$  does not reflect co-movement between volatility and liquidity in the options market, and it also does not appear to arise from market segmentation.

## 2. Implications for the risk–return trade-off

This section uses our conditional moment estimates to examine the risk–return trade-off in the time series of stock market returns.

### 2.1. Background

Economic intuition and mainstream asset pricing theories suggest that periods of elevated risk should coincide with higher risk premia, i.e., regressions such as

$$E_t[R_{t+1}] - E_t^*[R_{t+1}] = \beta_1 + \beta_2 \text{Std}_t[R_{t+1}] + \varepsilon_t, \quad (10)$$

should yield a positive slope coefficient. Yet a large literature fails to find evidence of a significant risk–return trade-off in the data when using realized returns  $R_{t+1}$  as a proxy for unobserved expected returns  $E_t[R_{t+1}]$  (e.g., [Glosten et al. 1993](#)). An important shortcoming of this proxy, highlighted by [Elton's \(1999\)](#) presidential address, is that  $R_{t+1}$  features large and potentially serially correlated innovations that are likely to bias inference about expected returns. Our estimation approach allows us to address [Elton's](#) critique because it equips us with an estimate of  $E_t[R_{t+1}]$ .

Before doing so, we confirm that one fails to detect a risk–return trade-off in our sample when relying on realized returns as a proxy. Specifically, we estimate

$$R_{t+1} - E_t^*[R_{t+1}] = \beta_1 + \beta_2 \text{Std}_t[R_{t+1}] + \varepsilon_t, \quad (11)$$

where  $\text{Std}_t[R_{t+1}]$  is the conditional volatility of ex-dividend returns that is implied by our estimate of  $f_t(R_{t+1})$ .<sup>9</sup> For consistency with our conditional moment estimates, we compute realized excess returns as the difference between the ex-dividend return  $R_{t+1}$  and its conditional risk-neutral expectation  $E_t^*[R_{t+1}]$ , rather than the difference between the cum-dividend return and its conditional risk-neutral expectation (which equals  $R_t^f$ ).<sup>10</sup> Because realized returns are heteroscedastic, we estimate the regression via weighted least squares for efficiency by scaling regression residuals by  $1/\text{Std}_t[R_{t+1}]$ . For ease of interpretation, we standardize the regressor by subtracting its unconditional mean and dividing by its unconditional standard deviation. Panel A of [Table 4](#) shows that the regression has little explanatory power ( $R^2 < 0.4\%$ ) and features an insignificant slope coefficient, regardless of the polynomial order  $N$  we use for  $E[M|R]$  in the estimation. The economic magnitude of the estimated risk–return relation is relatively small as well. A one standard deviation increase in conditional volatility is estimated to raise excess returns by 1.2 to 1.3 percentage points per year, depending on the polynomial order. Hence, the classic finding of [Glosten et al. \(1993\)](#) continues to hold in our sample.

One interpretation of the [Glosten et al.](#) evidence is that the risk–return trade-off is weak or even non-existent (e.g., [Moreira and Muir 2017](#) and [Lochstoer and Muir 2022](#)). From this perspective, our finding on time variation in the projected pricing kernel and the missing risk–return trade-off may appear like two sides of the same coin: If a rise in volatility does not lead to higher expected returns (a non-existent risk–return trade-off), it must be accompanied by lower risk prices (a flattening pricing kernel). Another interpretation is that the risk–return trade-off is masked by the (possibly non-IID) noise in realized returns. It is possible that expected returns, which are free from such noise, co-move significantly with conditional volatility despite the negative correlation between risk prices and volatility. Put differently, it is a quantitative question which of the two opposing channels dominates.

### 2.2. New estimates

We now explore this question by estimating risk–return trade-off regressions based on expected, rather than realized returns. The conditional moments  $E_t[R_{t+1}]$  and  $\text{Std}_t[R_{t+1}]$  are taken from estimations with a range of different polynomial orders  $N$  for the projected pricing kernel. Statistical tests are bootstrapped to account for first-stage estimation error in conditional moments.<sup>11</sup> [Appendix B.2](#) address additional econometric concerns about the regression slope  $\hat{\beta}_2$  with an extensive simulation study. This simulation evidence suggests that  $\hat{\beta}_2$  is consistent and essentially unbiased in finite samples. It also shows

<sup>9</sup> Prior literature on the risk–return trade-off has used both volatility and variance as regressors. In unreported results, we estimated analogous regressions based on variance and found nearly identical results, in terms of both magnitude and significance of the regression slope (using standardized variance as a regressor).

<sup>10</sup> Our measure of excess returns differs from the traditional measure by the risk premium for the dividend yield. Because S&P 500 companies announce dividends on average three weeks in advance of the ex dividend date ([Table 1](#) in [Schulz 2016](#)), however, this difference is negligible at the monthly horizon we focus on. Quantitatively,  $R_{t+1} - E_t^*[R_{t+1}]$  averages 8.33% p.a. over our 1990–2023 sample, whereas cum-dividend returns on the S&P 500 in excess of the one-month Treasury rate (from Ken French's data library) average 8.49% p.a.

<sup>11</sup> Specifically, we create 10,000 artificial samples based on a bootstrap with block length 21 (the overlap of our daily estimates). In each sample, we estimate the projected pricing kernel, compute the mean and volatility of the conditional return density for each day, and estimate risk–return trade-off regressions based on the resulting moment estimates.  $p$ -values of the regressions' slope coefficients in [Table 4](#) are based on these 10,000 bootstrap estimates.

Table 4

## Risk–return trade-off regressions.

We estimate risk–return trade-off regressions for realized returns (Panel A) via WLS, with residuals scaled by  $1/\text{Std}_t[R_{t+1}]$ , and regressions for expected returns (Panels B–D) via OLS. Conditional moments are taken from estimations with various polynomial orders ( $N$ ) for the log projected pricing kernel and either allow the projection's shape to vary with volatility ( $b \neq 0$ ) or not ( $b = 0$ ). Returns are measured in annualized percent. Regressors are standardized by subtracting their unconditional mean and then dividing by their unconditional standard deviation. Hypothesis tests are based on a block bootstrap with a block length of 21 trading days. The sample spans 1990–2023.

$N$	$b \neq 0$ (estimated)					$b = 0$
	1	2	3	4	5	2
Panel A: $R_{t+1} - E_t^*[R_{t+1}] = \beta_1 + \beta_2 \text{Std}_t[R_{t+1}] + \varepsilon_{t+1}$						
$\hat{\beta}_2$	1.30	1.21	1.21	1.26	1.26	1.37
$p$ -value, $H_0 : \beta_2 = 0$ , %	53.72	58.74	58.36	57.64	57.30	49.22
$R^2$ , %	0.38	0.33	0.33	0.34	0.34	0.33
Panel B : $E_t[R_{t+1}] - E_t^*[R_{t+1}] = \beta_1 + \beta_2 \text{Std}_t[R_{t+1}] + \varepsilon_t$						
$\hat{\beta}_2$	3.18	3.77	3.78	3.42	3.42	6.32
$p$ -value, $H_0 : \beta_2 = 0$ , %	30.68	1.16	1.02	0.36	0.08	0.00
$R^2$ , %	79.68	75.77	75.98	71.88	71.90	83.04
Panel C : $E_t[R_{t+1}] - E_t^*[R_{t+1}] = \beta_1 + \beta_2 \text{Std}_t[R_{t+1}] + \beta_3 \text{Skew}_t[R_{t+1}] + \varepsilon_t$						
$\hat{\beta}_2$	3.18	3.97	3.99	3.60	3.61	7.68
$p$ -value, $H_0 : \beta_2 = 0$ , %	29.46	1.28	1.06	0.38	0.08	0.00
$\hat{\beta}_3$	0.00	−0.90	−0.90	−0.99	−0.99	−2.10
$p$ -value, $H_0 : \beta_3 = 0$ , %	94.92	1.08	0.92	0.68	0.08	0.00
$R^2$ , %	79.68	79.85	80.08	77.69	77.73	88.40
Panel D : $E_t[R_{t+1}] - E_t^*[R_{t+1}] = \beta_1 + \beta_2 \text{Std}_t[R_{t+1}] + \beta_3 \text{Skew}_t[R_{t+1}] + \beta_4 \text{Kurt}_t[R_{t+1}] + \varepsilon_t$						
$\hat{\beta}_2$	3.16	3.79	3.97	3.26	3.27	8.17
$p$ -value, $H_0 : \beta_2 = 0$ , %	32.30	1.28	1.12	0.20	0.08	0.00
$\hat{\beta}_3$	0.03	−1.13	−0.96	−1.84	−1.85	−1.59
$p$ -value, $H_0 : \beta_3 = 0$ , %	99.18	0.46	0.62	0.04	0.00	0.00
$\hat{\beta}_4$	−0.05	−0.45	−0.18	−1.19	−1.19	1.15
$p$ -value, $H_0 : \beta_4 = 0$ , %	98.66	59.88	47.44	7.88	18.74	0.10
$R^2$ , %	79.69	80.35	80.23	80.58	80.60	89.41

that our estimation and bootstrap testing methodology reliably detects a risk–return trade-off when one exists in population and reliably fails to detect it if the true  $\beta_2$  equals zero.

Panel B of Table 4 shows OLS estimates of regression (10). The hypothesis  $H_0 : \beta_2 = 0$  is associated with a  $p$ -value of 30.96% when conditional moments are estimated with a log-linear ( $N = 1$ ) projection, i.e., this case yields no evidence of a significant risk–return trade-off. However, it is important to recall from Section 1.5 that the log-linear specification of  $E[M|R]$  is rejected (at the 1% level) in favor of a log-quadratic specification, and the latter is therefore likely to yield more reliable conditional moment estimates. Indeed, Table 4 shows that, for  $N = 2$  and higher polynomial orders, the risk–return trade-off is statistically significant. The hypothesis  $H_0 : \beta_2 = 0$  is rejected at the 5% significance level for all  $N \geq 2$  and at the 1% level for  $N \geq 4$ . The implied risk–return trade-off is economically significant as well. Based on the  $N = 2$  estimates, a one standard deviation increase in conditional volatility raises expected excess returns by 3.77 percentage points per year.

The time-variation we document in the projected pricing kernel is therefore not equivalent to the absence of a risk–return trade-off. To the contrary, convex specifications of the projection allow us to establish that a significantly positive risk–return trade-off exists in the data. This relation is undetectable in short samples of realized returns due to their large amount of (possibly non-IID) noise. The relation can also not be established based on a log-linear specification of the projected kernel. The reason is that “ $b$ ”, the parameter governing time variation in the projection and the strength of the risk–return trade-off, is poorly identified in the  $N = 1$  case, as shown in Table 2.<sup>12</sup>

<sup>12</sup> It is worth noting that, more generally, the risk–return slope estimate  $\hat{\beta}_2$  (shown in panel B of Table 4) is closely related to the estimate of “ $b$ ” (shown in Table 2), which governs time-variation in the projected pricing kernel. First,  $\hat{b}$  displays little variation across different polynomial orders (it is always close to 1) and  $\hat{\beta}_2$  therefore displays little variation across polynomial orders as

Section 1 showed that the shape of the projected pricing kernel varies significantly over time. To quantify the extent to which this time-variation dampens the risk–return relation, we report risk–return regressions for return moments from a time-invariant specification of the projection (which imposes  $b = 0$ ) in the last column of Table 4. For the expected excess return regression in Panel B, the slope coefficient increases from 3.77 with time-variation to 6.32 without time-variation, whereas the regression's  $R^2$  increases from 75.77% to 83.04%. Time-variation in the projected pricing kernel therefore dampens the risk–return relation substantially, but does not render it statistically or economically insignificant.

### 2.3. The role of higher moments

Up to this point, we have followed the prior risk–return trade-off literature in measuring risk by the second moment. This approach is rooted in a large literature on measuring and modeling conditional volatility, and a lack of comparable tools for conditional higher moments. Our methodology allows us to break with the tradition of equating risk and volatility because it equips us with an estimate of the entire conditional return density  $f_t(R_{t+1})$ . We do so by adding the conditional skewness and kurtosis of returns as explanatory variables to the risk–return regression,

$$E_t[R_{t+1}] - E_t^*[R_{t+1}] = \beta_1 + \beta_2 \text{Std}_t[R_{t+1}] + \beta_3 \text{Skew}_t[R_{t+1}] + \beta_4 \text{Kurt}_t[R_{t+1}] + \varepsilon_t. \quad (12)$$

As before, we estimate the regression via OLS and standardize all regressors for ease of interpretation. Panel C of Table 4 shows results

well (it always lies between 3 and 4). Second, the precision of  $\hat{b}$  increases monotonically in the polynomial order  $N$ , and the precision of  $\hat{\beta}_2$  therefore does so as well. For  $N = 1$ , both  $b$  and  $\beta_2$  are poorly identified and therefore insignificantly different from 0.

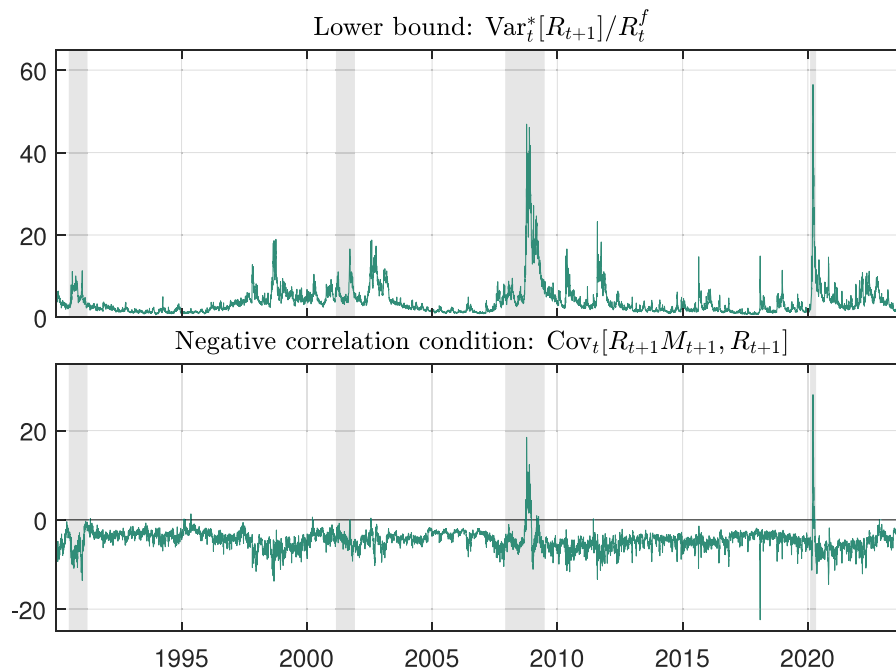


Fig. 5. The Martin (2017) lower bound. We plot Martin's (2017) lower bound (top panel) and the difference between expected excess returns and the lower bound (bottom panel) in percent p.a. Both objects are based on the estimation with a log-quadratic ( $N = 2$ ) projected pricing kernel.

for a regression based on volatility and skewness, whereas Panel D adds kurtosis as a third regressor as in (12). Mimicking the findings for univariate regressions, estimates based on the log-linear ( $N = 1$ ) specification of the projected kernel suggest that neither volatility nor higher moments are significantly related to expected excess returns. For higher order polynomials, however, conditional skewness is statistically significant at the 5% level for  $N \geq 2$  and at the 1% level for  $N \geq 5$ , controlling for volatility. Compared to the univariate specification, adding skewness as a second regressor increases  $R^2$  by 4–6 percentage points. A one standard deviation decrease in conditional skewness (which makes returns more left-skewed) increases expected excess returns by about 1 percentage point per year, holding volatility constant. In contrast, Panel D shows that conditional kurtosis is insignificant for all polynomial orders with the exception of  $N = 4$ , where it is marginally significant with a  $p$ -value of 7.88%. It is possible and perhaps plausible that conditional kurtosis is estimated with less precision than conditional skewness, which provides a potential explanation for the insignificant relation between kurtosis and expected returns.

Our finding that conditional skewness is significantly related to expected returns in the time series is consistent with Harvey and Siddique's (2000) finding that conditional skewness commands a significant risk premium in the cross-section of stock returns. It is also supports the view that left tail risks are a key determinant of the equity premium (Beason and Schreindorfer 2022) and cross-sectional differences in expected returns (Ang et al. 2006).

### 3. Implications for prior work

In this section, we use our estimates to provide a critical assessment of prior findings about conditional risk and risk premia. Specifically, we re-examine Martin's (2017) lower bound on the equity premium, Bollerlev et al.'s (2009) and Johnson's (2019) evidence on the variance premium's predictive power for excess returns, and Gormsen and Jensen's (2023) finding on co-movement between higher return moments.

#### 3.1. Martin (2017)

Martin (2017) shows that the absence of arbitrage opportunities implies

$$E_t[R_{t+1}] - E_t^*[R_{t+1}] = \frac{\text{Var}_t^*[R_{t+1}]}{R_t^f} - \text{Cov}_t[M_{t+1}R_{t+1}, R_{t+1}]. \quad (13)$$

If the second term on the RHS is non-positive,

$$\text{Cov}_t[M_{t+1}R_{t+1}, R_{t+1}] \leq 0, \quad (14)$$

which Martin (2017) refers to as the “negative correlation condition” (NCC), then  $\frac{\text{Var}_t^*[R_{t+1}]}{R_t^f}$  represents a lower bound on the expected excess return. Otherwise it represents an upper bound. Martin does not evaluate the NCC empirically, but confirms that it holds in a variety of asset pricing models. The resulting lower bound, which Martin argues to be approximately tight empirically, suggests that the equity premium is more volatile, more right-skewed, and fluctuates at a higher frequency than suggested by traditional estimates based on valuation ratios.

Relative to our in-sample estimate of expected excess returns, Martin's estimate has the advantage of being available in real time. It only requires an estimate of the risk-neutral return variance from option prices, but no additional parameter estimates. Its disadvantage is that the lower bound is only valid if the NCC holds. Furthermore, the bound only reveals the equity premium if the NCC holds with equality, i.e., if the bound is tight. Our estimates make it feasible to evaluate the NCC empirically and thereby re-examine Martin's bound. Fig. 5 shows our estimate of Martin's lower bound,  $\frac{\text{Var}_t^*[R_{t+1}]}{R_t^f}$ , in the top panel and the difference between our estimate of expected excess returns and the lower bound,  $\text{Cov}_t[M_{t+1}R_{t+1}, R_{t+1}]$ , in the bottom panel. The NCC states that  $\text{Cov}_t[M_{t+1}R_{t+1}, R_{t+1}]$  is negative, which is the case whenever expected excess returns (shown in the top panel of Fig. 3) exceed the lower bound. A number of observations are noteworthy.

First, expected excess returns exceed the lower bound on 8465 (or 99.0%) of the 8553 days in our sample. Our estimates therefore suggest that the NCC nearly always holds.

Table 5

## Predicting excess returns with the variance premium.

We use the variance premium to predict either realized excess returns (Panel A) or expected excess returns (Panels B and C). Conditional moments are taken from estimations with various polynomial orders ( $N$ ) for the log projected pricing kernel and either allow the projection's shape to vary with volatility ( $b \neq 0$ ) or not ( $b = 0$ ). Realized return regressions are estimated based on either OLS or WLS, as indicated in column one, whereas expected return regressions are estimated via OLS only.

		$b \neq 0$ (estimated)					$b = 0$
$N$		1	2	3	4	5	2
Panel A : $R_{t+1} - E_t^*[R_{t+1}] = \lambda_1 + \lambda_2 \left( \text{Var}_t[R_{t+1}] - \text{Var}_t^*[R_{t+1}] \right) + \varepsilon_{t+1}$							
OLS	$\hat{\lambda}_2$	-5.25	-6.15	-6.15	-6.09	-6.09	-5.81
	$p$ -value, $H_0 : \lambda_2 = 0$ , %	14.20	3.18	3.06	2.66	2.80	12.18
	$R^2$ , %	0.97	1.33	1.33	1.30	1.30	1.18
WLS	$\hat{\lambda}_2$	-2.68	-3.72	-3.73	-3.23	-3.23	-5.33
	$p$ -value, $H_0 : \lambda_2 = 0$ , %	24.70	11.34	11.38	12.28	11.90	9.22
	$R^2$ , %	0.73	1.12	1.12	1.01	1.02	1.18
Panel B : $E_t[R_{t+1}] - E_t^*[R_{t+1}] = \lambda_1 + \lambda_2 \left( \text{Var}_t[R_{t+1}] - \text{Var}_t^*[R_{t+1}] \right) + \varepsilon_t$							
OLS	$\hat{\lambda}_2$	-2.83	-4.16	-4.17	-3.81	-3.82	-6.50
	$p$ -value, $H_0 : \lambda_2 = 0$ , %	27.92	0.00	0.04	0.00	0.00	0.00
	$R^2$ , %	63.48	92.28	92.29	89.40	89.38	87.81
Panel C : $E_t[R_{t+1}] - E_t^*[R_{t+1}] = \lambda_1 + \lambda_2 \left( \text{Var}_t[R_{t+1}] - \text{Var}_t^*[R_{t+1}] \right) + \lambda_3 \text{Std}_t[R_{t+1}] + \varepsilon_t$							
OLS	$\hat{\lambda}_2$	-0.98	-3.92	-3.91	-3.31	-3.31	-4.01
	$p$ -value, $H_0 : \lambda_2 = 0$ , %	28.16	0.00	0.04	0.04	0.04	0.00
	$\hat{\lambda}_3$	2.43	0.27	0.28	0.59	0.59	3.14
	$p$ -value, $H_0 : \lambda_3 = 0$ , %	41.40	79.32	78.20	26.58	25.66	0.12
	$R^2$ , %	82.83	92.36	92.38	89.98	89.96	95.43

Second, the correlation between Martin's lower bound and our estimate of the market's risk premium equals 87.6% and the two series share similar fluctuations at higher frequencies. This co-movement is not mechanical despite the fact that both measures rely on information from index options. In particular, Martin connects the equity premium to options data via theory, whereas we do so based on a purely statistical criterion function. Our estimates therefore provide independent support for Martin's claim that the equity premium fluctuates at higher frequencies than implied by traditional estimates based on valuation ratios.

Third, our estimate of expected excess returns has a time series average 8.43% p.a., compared to only 3.90% for the lower bound, which suggests that the bound is far from tight. Indeed, Martin's bound falls below 2.78% p.a. on half of the days in our sample, whereas the median of our expected excess return estimate equals 7.40% p.a. This evidence is only suggestive about the bound's tightness, but it confirms the result of a formal statistical test based on realized returns in Back et al. (2022). A potential reason for this finding is that the bound, if interpreted as being tight, implies that variance is the only dimension of risk investors care about. This condition holds for log utility, but not for more general utility functions. Theoretically, it is well-known that log utility is quantitatively inconsistent with the equity premium and a range of other asset market puzzles. Empirically, there is strong evidence that market skewness is a key driver of cross-sectional differences in expected returns (Harvey and Siddique 2000) and, as we saw in the previous section, time-variation in the equity premium. It is therefore perhaps not surprising that we find the bound not to be tight empirically.

Fourth, our estimates show that days with NCC violations are not randomly distributed. Instead, almost all of them occur either during the 2007–2008 financial crisis or during the COVID-19 pandemic in 2020. These are precisely the periods for which Martin's bound implies the largest spikes in the market risk premium. At its peak, the bound exceeds our estimate of the equity premium by 18.5% p.a. during the financial crisis (on October 10th, 2008) and by 28.0% p.a. during the COVID-19 pandemic (on March 16th, 2020). It therefore appears that Martin's bound considerably overstates the extent to which the equity premium spikes during economic crises.<sup>13</sup>

The NCC violations we document are a direct consequence of time-variation in the projected pricing kernel. In particular, Martin (2017) shows that the condition holds for an investor who is fully invested in the market, as long as the investor's risk aversion exceeds one at all times. This situation is equivalent to the projected pricing kernel remaining sufficiently steep at all times. However, our estimates imply that the pricing kernel becomes very flat during times of extreme volatility.

Of course, it is possible that our methodology delivers imprecise estimates during times of extreme volatility. The NCC violations we detect may therefore simply reflect estimation error. We evaluate this possibility in two ways. The first approach relies on the Monte Carlo experiment in Appendix B.2, which was previously used to examine the statistical properties of the risk–return slope  $\hat{\beta}_2$ . Specifically, we simulate 5000 finite samples and use them to compute the probability that the NCC holds (for true  $f_t$  moments, which are known in the simulation), conditional on our estimate of  $\text{Cov}_t[M_{t+1}R_{t+1}, R_{t+1}]$  exceeding thresholds of 18.5% p.a. (as in 2008) or 28.0% p.a. (as in 2020). The conditional probabilities of these “false positives” equal 5.1% and 3.2%, respectively. The second approach relies on the same bootstrap methodology that we previously used to compute standard errors. Specifically, we simulate 10,000 (block) bootstrap samples, estimate pricing kernel parameters in each, and use the resulting conditional return moments to check for NCC violations in each sample. This experiment shows that 92.6% of bootstrap samples contain NCC violations.<sup>14</sup> Both approaches suggest that it is unlikely that the NCC violations we detect empirically are the result of estimation error.

Overall, our estimates confirm Martin's (2017) finding that the equity premium is more volatile and fluctuates at higher frequencies than implied by traditional estimates. However, they also suggest that the equity premium is considerably less right-skewed than implied by his lower bound, because we find the bound not to be tight during times of low volatility and to be violated during times of very high volatility.

<sup>14</sup> Note that this approach faces a joint hypothesis problem: We are evaluating whether the projection varies sufficiently with volatility to admit NCC violations in high volatility states, and whether volatility reaches sufficiently high levels to trigger those violations. Theoretically, any positive value of the volatility-scaling parameter  $b$  (which 99.3% of bootstrap samples feature) implies that the NCC is violated in states of (perhaps unrealistically) high volatility.

<sup>13</sup> See Bakshi et al. (2024) for related empirical evidence on NCC violations.



Table 6

## Co-movement of higher moments.

This table examines co-movement between conditional variance and conditional skewness. Conditional moments are taken from estimations with various polynomial orders ( $N$ ) for the log projected pricing kernel and either allow the projection's shape to vary with volatility ( $b \neq 0$ ) or not ( $b = 0$ ). We regress risk-neutral (physical) skewness on risk-neutral (physical) variance in Panel A (Panel B), and regress the skewness premium on the variance premium in Panel C. The specification in Panel A is not affected by our estimates and therefore identical across columns. All regressions are estimated via OLS. For comparability with Gormsen and Jensen (2023), variance is expressed in annualized percent.

$N$	$b \neq 0$ (estimated)					$b = 0$
	1	2	3	4	5	2
Panel A : $\text{Skew}_t^*[R_{t+1}] = \gamma_1 + \gamma_2 \text{Var}_t^*[R_{t+1}] + \varepsilon_t$						
$\hat{\gamma}_2$				0.060		
$p$ -value, $H_0 : \gamma_2 = 0$ , %				0.00		
$R^2$ , %				9.71		
Panel B : $\text{Skew}_t[R_{t+1}] = \gamma_1 + \gamma_2 \text{Var}_t[R_{t+1}] + \varepsilon_t$						
$\hat{\gamma}_2$	0.024	0.015	0.016	0.009	0.009	0.133
$p$ -value, $H_0 : \gamma_2 = 0$ , %	56.64	31.12	32.32	25.20	13.48	0.00
$R^2$ , %	1.13	3.00	3.07	1.97	1.99	34.32
Panel C : $\text{Skew}_t^*[R_{t+1}] - \text{Skew}_t[R_{t+1}] = \gamma_1 + \gamma_2 (\text{Var}_t^*[R_{t+1}] - \text{Var}_t[R_{t+1}]) + \varepsilon_t$						
$\hat{\gamma}_2$	0.337	0.104	0.103	0.118	0.118	0.010
$p$ -value, $H_0 : \gamma_2 = 0$ , %	10.48	1.18	1.02	1.62	0.38	12.68
$R^2$ , %	8.23	4.44	4.44	4.23	4.23	0.22

## 3.2. Bollerslev et al. (2009) and Johnson (2019)

In a seminal paper, Bollerslev et al. (2009) show that the variance premium  $\text{Var}_t[R_{t+1}] - \text{Var}_t^*[R_{t+1}]$  is a significant predictor of realized excess stock market returns  $R_{t+1} - E_t^*[R_{t+1}]$  in OLS regressions. Subsequent work has confirmed this finding in other asset markets and used it as an empirical target for equilibrium asset pricing models. In an important challenge to this literature, Johnson (2019) shows that “regardless of variable construction, forecast horizon, sampling frequency, sample period, or country, [weighted least squares] estimates of the relation between [the variance premium] and future market returns are not statistically significant” (p. 3), which indicates that “the significant OLS estimates may be false positives driven by a few periods with high expected volatility.” (p. 1). Panel A of Table 5 replicates both the original Bollerslev et al. and Johnson findings in our 1990–2023 sample. In OLS regressions, the slope coefficient on the variance premium is significant at the 5% level for  $N \geq 2$ . The  $N = 1$  estimator produces insignificant results but, as we saw in Section 1.5, it produces misspecified variance premium estimates and is statistically rejected in favor of higher-order polynomials. In WLS regressions that scale residuals by  $1/\text{Std}_t[R_{t+1}]$ , the slope coefficient is insignificant at common significance levels and its point estimate is noticeably smaller than the corresponding OLS estimate. As shown by Johnson (2019), statistically efficient estimates based on realized returns therefore suggest that expected excess returns are not (linearly) related to the variance premium.

Of course, the Elton (1999) critique on the use of realized returns as a proxy of expected returns applies here as well. We therefore rerun the regressions with our estimates of expected excess returns on the left hand side. Panel B shows that, in this case, the variance premium is significant at the 1% level (with a  $p$ -value of 0.0%) for  $N \geq 2$ . This finding suggests that WLS regressions for realized returns produce misleading conclusions about expected returns.

It is worth noting that the variance premium is a substantially better predictor of expected excess returns than conditional volatility, with  $R^2$  values around 90% for the regressions in Panel B, compared to  $R^2$  values between 70% and 80% for the risk–return trade-off regressions in Table 4. In fact, Panel C of Table 5 shows that conditional volatility becomes insignificant when the variance premium is added as a second regressor to the regression. This finding mimics similar OLS evidence based on realized excess returns in Bollerslev et al. (2009). The intuitive reason is that risk premia reflect fluctuations in both the amount and price of risk. As a result, the equity premium and the variance premium co-move more closely with each other than each one of them co-moves with conditional volatility. Overall, our findings confirm (contradict) prior OLS (WLS) evidence on the predictive power of the variance premium.

## 3.3. Gormsen and Jensen (2023)

Gormsen and Jensen (2023) study time-variation in the shape of the distribution of stock market returns. Using both ex ante risk-neutral moments extracted from option prices and ex post moments based on realized high-frequency returns, they show that stock market returns become more left-skewed in times of low volatility. The authors explore the implications of this finding for asset pricing models with rare disasters and note that it has important implications for regulators, because it “suggest[s] that [value-at-risk] measures only based on variance are likely to understate the true risk of portfolios that contain equities. More importantly, this mistake is going to be largest during good times with low variance and high prices, which are potentially periods where regulators should worry about overaccumulation of risk in the economy.” (pp. 4–5). Panel A of Table 6 replicates Gormsen and Jensen’s finding in our sample by showing that risk-neutral variance is a positive and highly significant predictor of risk-neutral skewness. Because risk-neutral moments do not depend on our estimate of the pricing kernel, this result does not vary with the polynomial order  $N$ .

Of course, what matters for risk management applications is co-movement between conditional physical moments. Evidence based on ex post realized moments or risk-neutral moments is only suggestive in this context. In particular, while it is well-known that realized variance is a good proxy for conditional variance, the literature has not established an analogous result for realized skewness. Inference about conditional skewness based on realized skewness can therefore be misleading.<sup>15</sup> Additionally, risk-neutral moments reflect both physical moments and moment risk premia and co-movement between risk-neutral moments could reflect time-variation in risk premia. It is therefore important to examine co-movement between conditional physical variance and skewness based on explicit estimates of these moments. We do so in Panel B of Table 6. In line with Gormsen and Jensen’s finding on risk-neutral moments, we find that the relation between conditional physical variance and skewness is positive. Different from their results, however, we find the relationship to be insignificant, regardless of the polynomial order  $N$  we use in the estimation. More importantly, point estimates of the regression slope are substantially smaller than those for risk-neutral moments.<sup>16</sup> Hence, our estimates do

<sup>15</sup> A useful example of this issue occurs in the rare disaster model of Wachter (2013). In that model, the realized variance and skewness of daily returns within a month are negatively correlated, as shown by Gormsen and Jensen (2023), whereas the conditional variance and skewness of monthly returns are positively correlated, as we show in Section 4.

Table 7

## Conditional risk and risk pricing in macrofinance models.

This table shows small sample moments of log returns (subscript “ $r$ ”) and the log projected pricing kernel (subscript “ $m$ ”).  $\sigma_{t,x}$  and  $skew_{t,x}$  denote the conditional volatility and skewness of  $x \in \{r, m\}$ . Data moments are taken from the  $N = 2$  estimation. Model moments are shown for the models of Campbell and Cochrane (1999), Bekaert and Engstrom (2017), Bansal and Yaron (2004), Drechsler and Yaron (2011), Schorfheide et al. (2018), Barro (2009), Wachter (2013), Gabaix (2012), Constantinides and Ghosh (2017), Schreindorfer (2020), and Lochstoer and Muir (2022). Model moments represent medians of each statistic across 100,000 simulated samples of length  $T = 804$  months, the effective size of our empirical sample. See Appendix B.1.4 for a discussion on the effective sample size of our sub-sampled data. NA indicates moments that are undefined due to division by zero. Asterisks indicate 1-sided  $p$ -values of the data moments under the null of the model, computed based on the same finite sample simulation.

	Data	Habits		Long-Run Risks			Rare Disasters			Other		
		CC99	BE17	BY04	DY11	SSY18	B09	G12	W13	CG17	S20	LM22
Risk												
$E\left[\sigma_{t,r}\right]$	14.4	16.7***	16.5***	16.5***	17.7***	14.0	12.5***	12.9***	20.8**	9.0	13.6***	20.8***
$E\left[\text{Skew}_{t,r}\right]$	−0.8	0.0***	−1.7***	0.0***	−0.6***	0.0***	−38.1***	−2.7***	−13.7***	−2.1***	−0.8***	−0.3***
$\text{Std}\left[\sigma_{t,r}\right]$	6.3	1.6***	3.5***	1.7***	5.4	2.4***	0.0***	0.7***	6.6	12.2*	0.0***	5.6***
$\text{corr}\left[\text{Skew}_{t,r}, \sigma_{t,r}\right]$	−0.05	NA	0.97***	NA	−0.60***	NA	NA	−0.95***	0.97***	0.51**	NA	0.95***
Risk Pricing												
$E\left[\frac{\sigma_{t,m}}{\sigma_{t,r}}\right]$	4.1	2.1***	2.7***	2.0***	1.7***	3.1***	1.5***	2.2***	1.2***	0.9***	4.1***	4.6***
$E\left[\frac{\text{Skew}_{t,m}}{\text{Skew}_{t,r}}\right]$	−8.5	NA	−0.9***	NA	−7.7	NA	−1.0***	−30.1**	−0.9***	−6.2***	−11.5***	0.8***
$\text{Std}\left[\frac{\sigma_{t,m}}{\sigma_{t,r}}\right]$	1.2	0.9	0.5***	0.0***	0.5***	0.6***	0.0***	0.1***	0.0***	0.3***	0.0***	2.2***
$\text{corr}\left[\frac{\sigma_{t,m}}{\sigma_{t,r}}, \sigma_{t,r}\right]$	−0.84	0.88***	−0.75**	NA	0.86***	−0.58***	NA	−0.76	NA	0.78***	NA	−0.50***

\*  $p < 0.1$ .\*\*  $p < 0.05$ .\*\*\*  $p < 0.01$ .

not support the idea that stock market returns become more left skewed in good times.

To reconcile the difference between Gormsen and Jensen’s finding on risk-neutral moments and our finding on physical moments, we examine moment risk premia. Specifically, Panel C shows that the variance premium  $\text{Var}_t^*[R_{t+1}] - \text{Var}_t[R_{t+1}]$  is a positive and statistically significant predictor of the skewness premium  $\text{Skew}_t^*[R_{t+1}] - \text{Skew}_t[R_{t+1}]$  for  $N \geq 2$ . This finding confirms similar evidence in Kozhan et al. (2013), who show that returns on variance and skewness swaps (ex-post measures of variance and skewness premia) are significantly positively correlated.<sup>17</sup> The positive relation between risk neutral variance and skewness documented by Gormsen and Jensen (2023) therefore appears to be driven by co-movement in moment risk premia, rather than co-movement in conditional physical moments. Overall, our findings suggest that co-movement between variance and skewness does not present a challenge for regulatory value-at-risk measures.

The last column of Table 6 illustrates that time-variation in the projected pricing kernel is critical for this conclusion. Specifically, when we estimate return moments based on the pricing kernel specification without time-variation ( $N = 2$ ,  $b = 0$ ), we fail to replicate the evidence in Kozhan et al. (2013) and, as a result, find significant co-movement between the physical variance and skewness of returns.

<sup>16</sup> Our analysis is based on estimates of conditional skewness. A potential concern is therefore that the smaller regression slope estimate for physical moments results from attenuation bias. In particular, Fig. 3 shows that skewness estimates are noticeably more noisy in the pre-1996 part of our sample. We therefore re-ran the regression in Panel B of Table 6 in the 1996–2023 subsample. For  $N = 2$ , the estimated regression slope equals 0.016, which is quite similar to the 0.015 estimate in the full sample. Attenuation bias therefore does not appear to explain our finding.

<sup>17</sup> Another way to examine differences between the dynamics of physical and risk-neutral return moments is to test the hypothesis that the regression slope for risk-neutral moments (in Panel A of 6) equals the regression slope for physical moments (in Panel B of 6). In untabulated results, we find that a bootstrap test (using a bootstrap block length of 21) rejects this hypothesis at the 1% significance level for all polynomial orders  $N$ . This finding supports the view that co-movement between risk-neutral moments results primarily from co-movement between risk premia.

However, it is important to recall that this specification is statistically rejected at the 1% level in favor of the time-varying benchmark, and that its implications should therefore not be interpreted as a reasonable description of the data.

#### 4. Implications for macrofinance models

The primary objective of most macrofinance models is to explain the characteristics and pricing of stock market risks. The two functions  $f_t(R_{t+1})$  and  $E_t[M_{t+1}|R_{t+1}]$  represent an almost ideal diagnostic tool for such models because they provide a complete description of conditional risks in returns and investors’ attitude towards them. The goal of this section is to provide a broad overview of existing models’ ability to match basic facts about stock market risks and risk prices, as reflected in  $f_t(R_{t+1})$  and  $E_t[M_{t+1}|R_{t+1}]$ .

This analysis complements (Moreira and Muir, 2017; Martin, 2017), who test macrofinance models based on their ability to match the empirical risk–return relation and risk-neutral return moments, respectively. We add value by considering a broader set of models. Additionally, it is worth noting that many existing facts about stock market returns can be viewed as summary statistics of  $f_t(R_{t+1})$  and  $E_t[M_{t+1}|R_{t+1}]$ . For example, the same risk-neutral return moments can arise from different combinations of risk (physical densities) and risk pricing (projected pricing kernels). Similarly, the same risk–return relationship can arise from different combinations of time-varying risk and time-varying risk pricing. Additionally, our estimates show that non-normalities in  $f_t(R_{t+1})$  and non-linearities in  $E_t[M_{t+1}|R_{t+1}]$  both play important roles in shaping the risk–return trade-off. By separating risks from risk prices on a state-by-state basis,  $f_t(R_{t+1})$  and  $E_t[M_{t+1}|R_{t+1}]$  provide a more complete characterization of stock market premia than commonly reported summary statistics.

##### 4.1. Conditional risk and risk pricing in macrofinance models

We consider eleven models with various economic mechanisms, including external habits (Campbell and Cochrane 1999, Bekaert and Engstrom 2017), long-run risks (Bansal and Yaron 2004, Drechsler and Yaron 2011, Schorfheide et al. 2018), rare disasters (Barro 2009, Wachter 2013, Gabaix 2012), incomplete markets

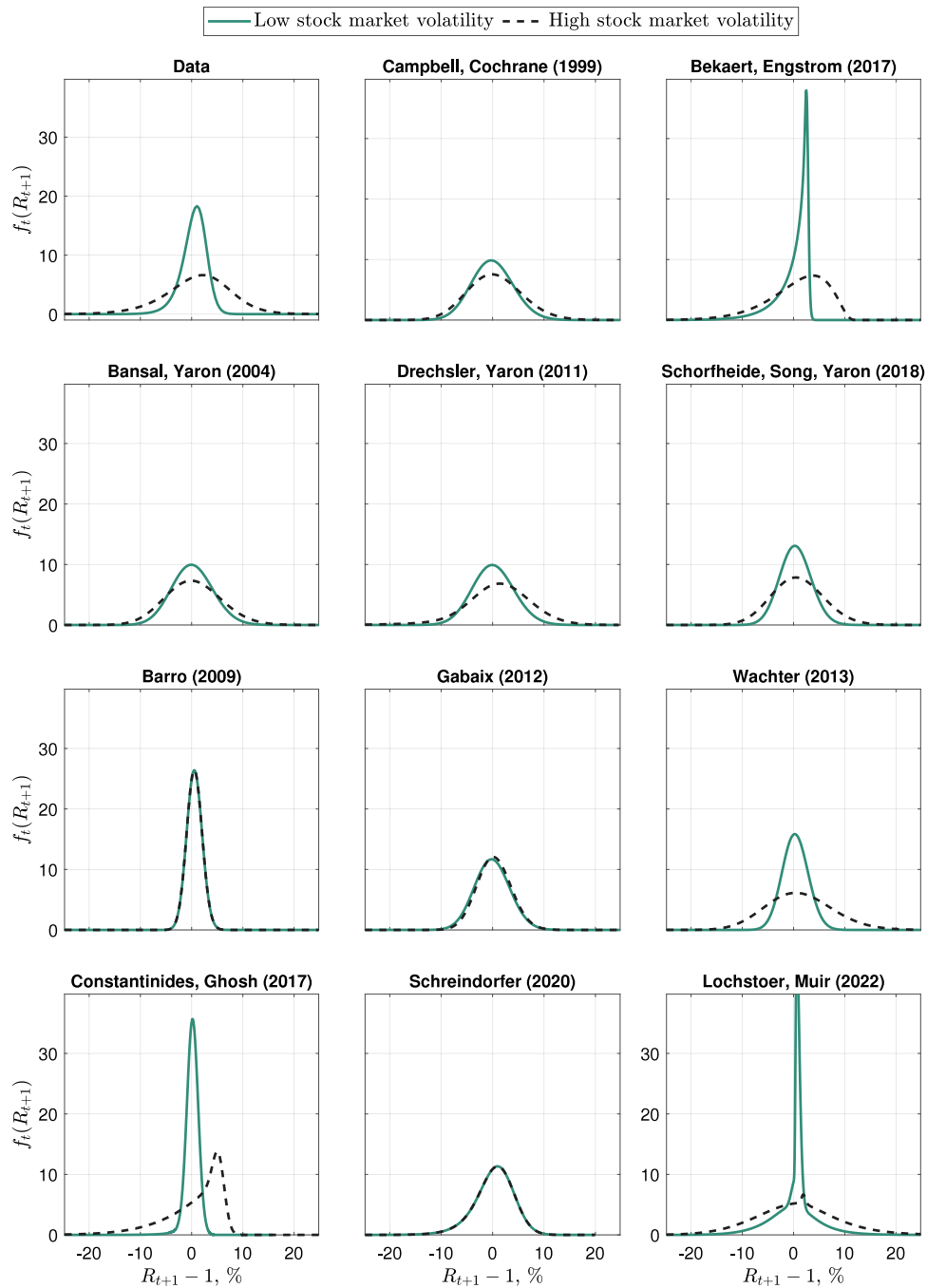


Fig. 6. The conditional distribution of returns in macrofinance models. We plot the conditional distribution of returns for the 10th and 90th percentile of conditional volatility.

(Constantinides and Ghosh 2017), disappointment aversion (Schreindorfer 2020), and slow-moving beliefs about volatility (Lochstoer and Muir 2022). To save space, we refer to the models by the first initials of the authors' last names and the publication year; e.g., the Campbell and Cochrane (1999) model is abbreviated "CC99", the Wachter (2013) model is abbreviated "W13", etc. We rely on Dew-Becker et al.'s (2017) calibration of the G12 model and Beason and Schreindorfer's (2022) calibration of the CG17 model.<sup>18</sup> All other models are calibrated as in

<sup>18</sup> Dew-Becker et al.'s version of the G12 model adds Gaussian innovations to consumption growth (which is constant absent disasters in the model's original version), it assumes a normally distributed (rather than constant)

the original studies. The calculation of  $f_t(R_{t+1})$  and  $E_t[M_{t+1}|R_{t+1}]$  in the models are detailed in Section IV of the online appendix.

Figs. 6 and 7 show  $f_t(R_{t+1})$  and  $E_t[M_{t+1}|R_{t+1}]$  for the 10th and 90th percentile of volatility and Table 7 summarizes key properties of the two functions via small sample statistics. To generate these statistics,

disaster size, and it specifies the recovery rate of dividends as an autoregressive (rather than linearity generating) process. These modifications do not alter the model's basic economics, but result in a pricing kernel that is a continuous (rather than discontinuous) function of the model's state. Beason and Schreindorfer (2022) provide a monthly calibration of the CG17 model, which is calibrated at the quarterly frequency in the original study.

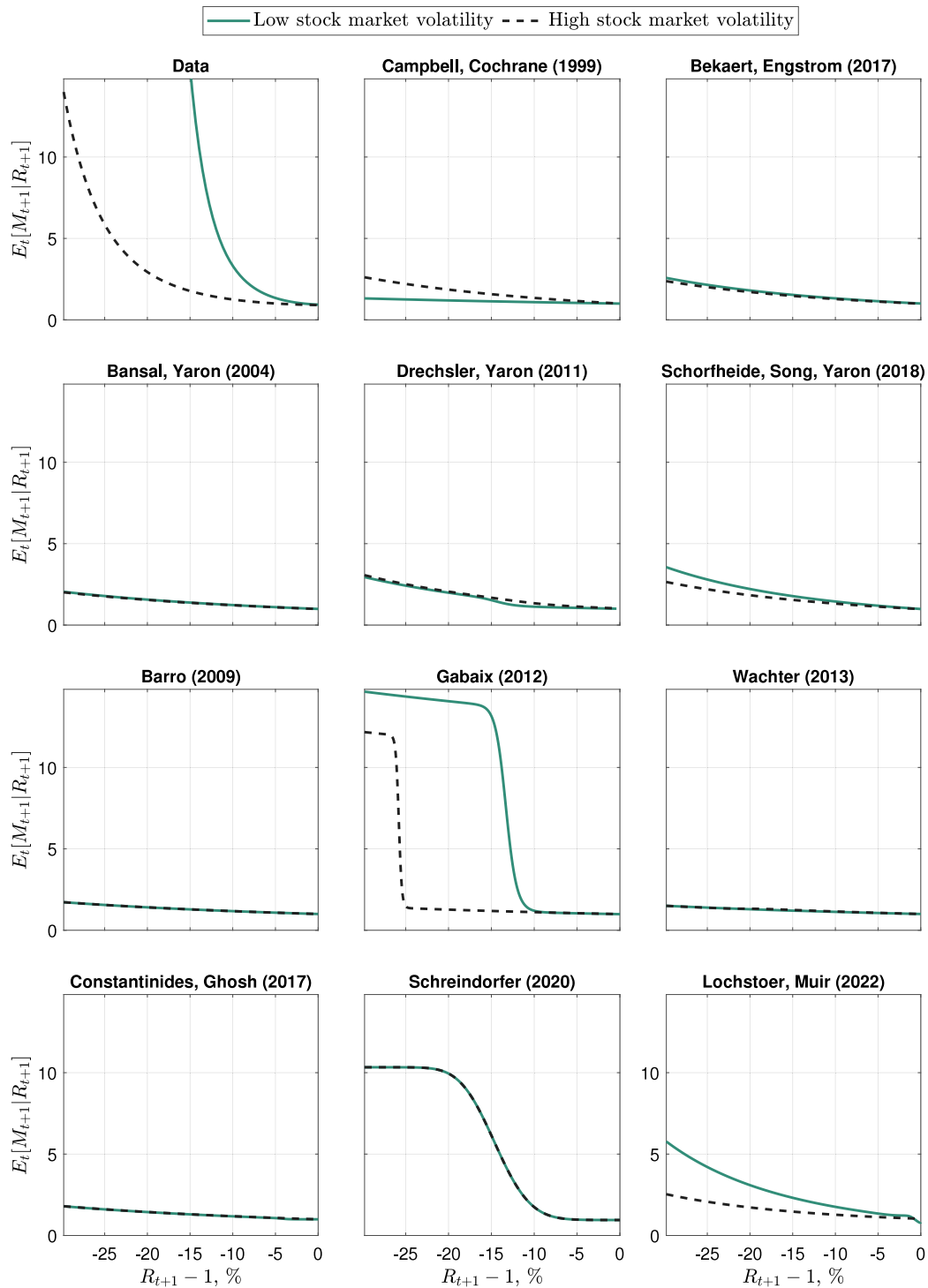


Fig. 7. The projected pricing kernel in macrofinance models. We plot the projected pricing kernel for the 10th and 90th percentile of conditional volatility.

we simulate 100,000 finite samples from each model, compute the statistic of interest in each sample, and report the median value across samples. We also generate an empirical  $p$ -value for each statistic: this represents the proportion of the 100,000 samples that generate values that are as or more extreme as observed in the data. In this context, it is worth noting that the models of Barro (2009) and Schreindorfer (2020) are IID and therefore feature no time-variation in conditional moments. This implies that even small differences between data and model moments result in  $p$ -values of zero.

The table shows moments of log returns and the log projected pricing kernel, which allows us to use log-normality and log-linearity as quantitative benchmarks. The conditional volatility and skewness of log returns and the log projected pricing kernel are denoted by  $(\sigma_{t,r}, \text{Skew}_{t,r})$  and  $(\sigma_{t,m}, \text{Skew}_{t,m})$ , respectively. We summarize the typical shape of the conditional return distribution by the average conditional moments  $E[\sigma_{t,r}]$  and  $E[\text{Skew}_{t,r}]$  and time-variation in its shape by  $\text{Std}[\sigma_{t,r}]$  and  $\text{corr}[\sigma_{t,r}, \text{Skew}_{t,r}]$ . Log-normality implies that  $\text{Skew}_{t,r}$  is equal to zero in every period and that  $\text{corr}[\sigma_{t,r}, \text{Skew}_{t,r}]$  is undefined. Typical deviations from log-normality can therefore be quantified by  $E[\text{Skew}_{t,r}]$



and fluctuations and the cyclical nature of such deviations by the combination of  $\text{Std}[\sigma_{t,r}]$  and  $\text{corr}[\sigma_{t,r}, \text{Skew}_{t,r}]$ . We summarize the projected pricing kernel's typical shape by  $E[\sigma_{t,m}/\sigma_{t,r}]$  and  $E[\text{Skew}_{t,m}/\text{Skew}_{t,r}]$  and time-variation in its shape by  $\text{Std}[\sigma_{t,m}/\sigma_{t,r}]$  and  $\text{corr}[\sigma_{t,m}/\sigma_{t,r}, \sigma_{t,r}]$ . Log-linearity, i.e., the log projection being linear in log returns, implies that  $\sigma_{t,m}/\sigma_{t,r}$  equals the absolute value of the projection's slope and that  $\text{Skew}_{t,m}/\text{Skew}_{t,r}$  is time-invariant and equal to  $-1$ . We can therefore quantify typical deviations from log-linearity (i.e. convexity) by  $E[\text{Skew}_{t,m}/\text{Skew}_{t,r}]$  and the projection's steepness by  $E[\sigma_{t,m}/\sigma_{t,r}]$ . Additionally,  $\text{Std}[\sigma_{t,m}/\sigma_{t,r}]$  quantifies the amount of time-variation in the projection's slope and  $\text{corr}[\sigma_{t,m}/\sigma_{t,r}, \sigma_{t,r}]$  the degree of co-movement between the slope and volatility. Our empirical estimates show that the projection becomes flatter when volatility increases, i.e., they imply that this correlation is negative.

The results are easily summarized. None of the models come close to matching the typical shapes of  $f_t(R_{t+1})$  or  $E_t[M_{t+1}|R_{t+1}]$ , and none come close to matching the amount of time-variation in either function.

Among the risk metrics summarizing  $f_t(R_{t+1})$ , the conditional skewness of returns is particularly problematic. For all eleven models, the empirical value of  $E[\text{Skew}_{t,r}]$  has a  $p$ -value below 1%. For the models with rare disasters (B09, W13, G12) or gamma-distributed shocks to state variables (BE17, CG17), conditional skewness is significantly too negative. The discrepancy is very visible in Fig. 6 for the latter group, but less visible for the former due to the low probability and extreme magnitude of disasters. For models that are conditionally log-normal (BY04, SSY18) or approximately so (CC99), conditional skewness equals zero. It also remains too close to zero for the modified long-run risks model of DY11, who add jumps to the model's state variables. S20 matches the amount of conditional skewness up to the reported precision, but is nevertheless statistically rejected due to the combination differences in trailing decimals and being IID. Many models are also inconsistent with average volatility,  $E[\sigma_{t,r}]$ , but this discrepancy is less concerning because one could re-calibrate the amount of "leverage" in dividends to match this moment.

Most models struggle with the amount of time-variation in condition risk, with  $\text{Std}[\sigma_{t,r}]$  values that fall significantly below the empirical estimate. The only models that come statistically close to matching the volatility-of-volatility (DY11, W13, CG17) do so via time-variation in the probability of jumps. This mechanism results in a counterfactually high correlation between conditional skewness and volatility, which lies outside of the 99% confidence interval for all three models.<sup>19</sup> In sum, the models we consider all fail to explain the magnitude and time-variation of conditional return risks.

Among the risk price metrics summarizing  $E_t[M_{t+1}|R_{t+1}]$ , the typical slope of the projected pricing kernel is most problematic. This shortcoming is easily visible in Fig. 7, which shows that  $E_t[M_{t+1}|R_{t+1}]$  is noticeably too flat in almost every model. Table 7 shows that the empirical estimate of  $E[\sigma_{t,m}/\sigma_{t,r}] = 4.1$  has a probability of less than 1% in all eleven models. S20 matches the slope's magnitude but is once again statistically rejected due to being IID. An important reason for most models' inability to generate a sufficiently steep projection is that they imply almost no convexity, as apparent from the counterfactually low values of  $E[\text{Skew}_{t,m}/\text{Skew}_{t,r}]$ . Two exceptions are G12, which implies a counterfactually large amount of convexity, and S20, which generates an economically plausible amount of convexity but is once again statistically rejected due to its IID nature.

<sup>19</sup> Some readers may notice that we report a positive value for  $\text{corr}[\text{Skew}_{t,r}, \sigma_{t,r}]$  in the W13 model, while Gormsen and Jensen (2023) report a negative value. This difference results from the fact that we compute the conditional skewness of monthly returns, whereas Gormsen and Jensen (2023) compute the realized skewness of daily returns within months. The positive correlation we find arises from the fact that an increase in the disaster probability (i) increases conditional skewness (making it less negative) because the disaster outcome becomes less unusual, i.e., less of an outlier and (ii) conditional variance rises because extreme outcomes become more likely.

The models also struggle with the amount of variation in condition risk prices, with  $\text{Std}[\sigma_{t,m}/\sigma_{t,r}]$  values that are either zero or fall significantly below the empirical estimate. The single exception is CC99, which generates a realistic amount of variation in the slope of the projection but a cyclical pattern that contradicts the data. Like DY11, CC99 implies that the projected pricing kernel becomes steeper, rather than flatter, when volatility rises. Hence, the models we consider all fail to explain the magnitude and time-variation of conditional risk prices.

## 5. Conclusion

Option markets provide us with invaluable information about conditional stock market risks and the pricing of such risks. Our paper proposes an empirical framework for jointly estimating conditional expected returns, conditional risks, and conditional risk prices based on index options and return data. We find that negative returns are substantially more painful to investors when they occur during low-volatility periods, which plays a key role in inferring conditional return moments from option prices.

We explore the economic implications of our moment estimates through three applications. First, we decompose the risk–return trade-off into two opposing channels: as risk increases, the price of risk decreases. Our results show that the risk effect dominates the price effect when the estimation is based on higher-order polynomials for the pricing kernel, leading to a positive and statistically significant risk–return trade-off. Second, we revisit a number of key findings about conditional return moments and risk premia in prior work. Specifically, we show that (i) Martin's (2017) lower bound on the equity premium is violated during high-volatility periods, (ii) contrary to WLS evidence in Johnson (2019), the variance premium is significantly related to expected returns and, (iii) co-movement between conditional risk-neutral volatility and skewness, as documented by Gormsen and Jensen (2023), is by driven by co-movement in risk premia rather than co-movement in underlying risks. Third, we find that eleven prominent macrofinance models fail to capture basic features of conditional stock market risks and risk prices, such as the conditional skewness of returns, the steepness of the projected pricing kernel, and the amount of cyclical variation in both metrics.

Our approach can be extended and generalized in a number of ways. First, conditional return distributions are paramount for risk management purposes. Our methodology offers a way to quantify the probability of tail events from option prices, which could be valuable in value-at-risk estimations. Our paper has deliberately focused on in-sample estimations to maximize power in detecting true predictive relationships (Inoue and Kilian, 2005; Campbell and Thompson, 2008; Cochrane, 2008; Hansen and Timmermann, 2015). In the context of risk management applications, however, it would be useful to explore out-of-sample implementations of our approach. Second, it would be interesting to apply our estimation methodology to other asset classes with active option markets, such as bonds, currencies, commodities, and international stock market indices. In ongoing research, Almeida, Freire, and Sichert (2025) are exploring an application to single stocks to better understand the dynamics of their expected returns and conditional higher moments. Lastly, while our third application illustrated several shortcomings of existing macrofinance models, it also revealed which existing mechanisms are most promising for generating realistic risk and risk price dynamics. In our view, it would be most useful to explore models that allow stock market volatility to evolve independently from the volatility of macroeconomic fundamentals (as, e.g., in Gabaix 2012), as well as models that are able to generate high risk prices for tail events, either through frictions or tail-sensitive preferences.

## CRediT authorship contribution statement

**David Schreindorfer:** Writing – review & editing, Writing – original draft, Visualization, Validation, Supervision, Software, Resources, Project administration, Methodology, Investigation, Funding acquisition, Formal analysis, Data curation, Conceptualization. **Tobias Sichert:** Writing – review & editing, Writing – original draft, Visualization, Validation, Supervision, Software, Resources, Project administration, Methodology, Investigation, Funding acquisition, Formal analysis, Data curation, Conceptualization.

## Declaration of competing interest

The authors have no other potential conflicts to disclose.

## Acknowledgments

Tobias Sichert acknowledge support from the Jan Wallanders and Tom Hedelius Foundation and the Tore Browaldhs Foundation, grant no. Fh21-0026 and from the Swedish House of Finance.

## Appendix

This appendix details our time series model for conditional volatility and results from a simulation study of our estimator.

### Appendix A. Conditional volatility estimation

We estimate the conditional volatility of log returns based on the heterogeneous autoregressive (HAR) model of [Corsi \(2009\)](#):

$$RV_t^{(21)} = \alpha + \beta^m RV_{t-21}^{(21)} + \beta^w RV_{t-21}^{(5)} + \beta^d RV_{t-21}^{(1)} + \epsilon_t, \quad (\text{A.1})$$

where the realized volatility  $RV_t^{(1)} = (\sum_{i=1}^{N_t} r_{it}^2)^{0.5}$  equals the square root of the sum of  $N_t$  squared intra-daily log returns on day  $t$ , and  $RV_t^{(h)} = (\frac{1}{h} \sum_{j=0}^h RV_{t-j}^{(1)})^{0.5}$ . Weekly ( $RV_t^{(5)}$ ) and monthly ( $RV_t^{(21)}$ ) volatility measures allow the model to capture the long-memory feature of stock market volatility. Returns are sampled every five minutes to compute  $RV$ , which is commonly viewed as a good trade-off between sampling noise (which increases at low sampling frequencies) and microstructure noise (which decreases at low sampling frequencies). We sub-sample our estimator every minute, which reduces the noise without affecting the estimator's bias, and add the squared log overnight return to each intra-daily variance estimate.

Intra-daily returns are computed from S&P 500 future prices, which we obtained from Tick Data Inc. In 1997, the Chicago Mercantile Exchange (CME) introduced the E-mini future (symbol: ES). Over time, the standard “large” futures contract (symbol: SP) lost market share to the E-mini, and eventually was discontinued in 2021. Since the dollar trading volume of the E-mini overtook that of the large contract during 2002, we switch our  $RV$  calculation from the large contract to the mini in 2003.

Our volatility forecasts are out-of-sample and implemented based on an expanding estimation window. We start the sample in 1988, so that our first forecast on Jan 02, 1990 is based on a two year estimation window. The model forecasts volatility very well with an out-of-sample  $R_{OOS}^2$  of 60.4%.

### Appendix B. A simulation study of our estimator

The applications in Sections II–IV rely on estimates of the projected pricing kernel. Naturally, these estimates are affected by estimation noise and possible misspecification of the parametric function for  $E_t[M_{t+1}|R_{t+1}]$ . In this appendix, we conduct an extensive Monte Carlo simulation study to assess the importance of these issues for our economic conclusions. Specifically, we examine statistical properties of the risk–return slope estimate  $\beta_2$  and the reliability of our detection of NCC violations.

### B.1. Simulation design

Our simulation results are all based on 5000 finite samples.

#### B.1.1. Data generating process

The data generating process is designed to capture the non-linearity of  $E_t[M_{t+1}|R_{t+1}]$ , the non-normalities in  $f_t(R_{t+1})$ , and the amount of time-variation in both functions. We estimate and a first-order vector autoregressive (VAR) model for the first four conditional moments of  $\ln R_{t+1}$ . We then simulate this model by drawing innovations from their joint empirical distribution. To ensure positivity of variance and kurtosis, we specify the VAR model for the natural logarithms of these moments, but include the mean and skewness without a log transformation. In each period of the simulation, the four moments are mapped to a density  $\tilde{f}_t(\ln R_{t+1})$  based on the Normal Inverse Gaussian (NIG) distribution of [Barndorff-Nielsen \(1977, 1978, 1997\)](#). The NIG distribution is well-suited for our purposes because it provides a unique mapping from the first four moments to its four parameters; see, e.g., Theorem 2.2 in [Eriksson et al. \(2009\)](#). The conditional density of simple returns, obtained via the change-of-measure  $f_t(R_{t+1}) = \tilde{f}_t(\ln R_{t+1})/R_{t+1}$ , is used to generate return realizations for each simulation period. By characterizing the return density based on four moments, our data generating process provides a realistic depiction of the shape of  $f_t(R_{t+1})$  at different points in time. By capturing the dynamics of these moments via a VAR model, it also provides a realistic depiction of how the shape of  $f_t(R_{t+1})$  evolves over time.

To generate the conditional risk-neutral distribution, we map  $f_t(R_{t+1})$  to  $f_t^*(R_{t+1})$  based on the parametric  $N = 2$  estimate of  $E_t[M_{t+1}|R_{t+1}]$ , the simulated volatility series from the VAR model, and Eqs. (1), (4), and (5). The polynomial intercept  $\delta_t$  in Eq. (4) is set to ensure that  $f_t^*(R_{t+1})$  integrates to one in each simulation period, which implies that we do not need to specify  $R_t^f$ .

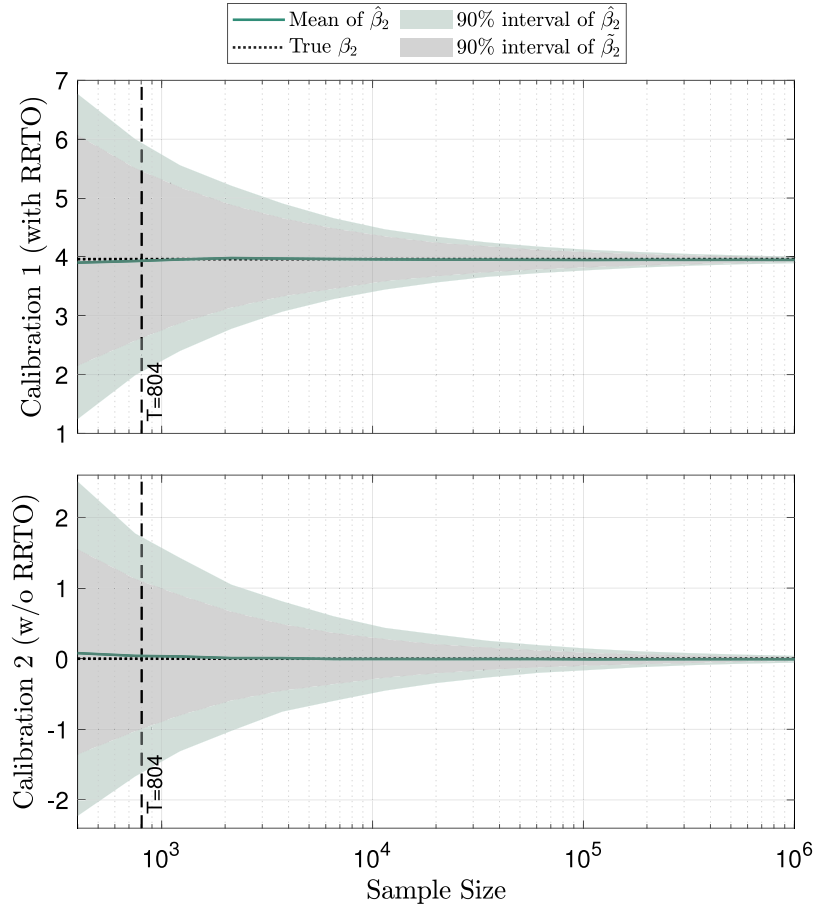
#### B.1.2. Estimated versus true return densities

In the empirical data, only the risk-neutral density  $f_t^*(R_{t+1})$  and return realization  $R_{t+1}$  are observed. The conditional physical density  $f_t(R_{t+1})$  and associated return moments have to be estimated. In the simulation, we observe the true return density  $f_t(R_{t+1})$ , in addition to  $f_t^*(R_{t+1})$  and  $R_{t+1}$ . This allows us to compare properties of the true and estimated return densities. To do so, we follow our empirical estimation approach, i.e., we estimate the parameters of  $E[M|R]$  based on simulated data for  $R_{t+1}$  and  $f_t^*(R_{t+1})$  and use them to compute an estimate of  $f_t(R_{t+1})$ .

#### B.1.3. Calibrations

We consider two calibrations. Calibration 1 is based on our empirical parameter estimates of the projected pricing kernel and therefore features a positive risk–return trade-off. Calibration 2 increases the amount of time-variation in  $E[M|R]$  relative to the empirical estimates, such that there is no risk–return trade-off and the true  $\beta_2$  equals zero.<sup>20</sup> We compute the true value of the risk–return slope coefficient  $\beta_2$  by simulating the data generating process for 10 million periods and estimating regression (10) based on true rather than estimated  $f_t(R_{t+1})$ -moments. The true value of  $\beta_2$  equals 3.96 in calibration 1 and, by construction, equals zero in calibration 2.

<sup>20</sup> We do so by increasing the pricing kernel parameter “ $b$ ” in Eq. (5) relative to its empirical estimate. The remaining pricing kernel parameters are estimated by maximizing the log-likelihood of realized returns, given the restricted value of  $b$ . To ensure internal consistency, we re-estimate the VAR based on the return moments that are implied by this alternative estimate of the projected pricing kernel.



**Fig. 8. Sampling Distribution of the Risk-Return Slope Coefficient.** This figure shows the sampling distribution of the slope coefficient  $\hat{\beta}_2$  in risk–return trade-off regression (10). The sampling distribution is generated based on a Monte Carlo simulation that closely mimics our empirical setting; see the main text for additional details. The true value of  $\beta_2$  is indicated by the dotted line, the mean of the sampling distribution by the solid line, and a symmetric 90% interval of the sampling distribution by the wider shaded region. We also plot the 90% interval of an alternative estimator  $\tilde{\beta}_2$  based on the true (but empirically unobservable) conditional return distribution  $f_i(R_{i+1})$ . The calibration in the top panel is based on our empirical estimates and reflects a positive risk–return trade-off. The calibration in the bottom panel implies the absence of a risk–return trade-off, i.e., a true value of  $\beta_2 = 0$ .

#### B.1.4. Effective sample size

A noteworthy shortcoming of our simulation design is that it does not allow us to mimic the daily sub-sampling approach that was used in the empirical data. Doing so would necessitate the simulation of daily returns, but our empirical estimates contain no information about return moments at this frequency. To simulate finite samples that can be meaningfully compared to the empirical data, it is therefore necessary to determine our effective sample size. The effective sample size equals the number of non-overlapping monthly observations that results in the same estimation precision as our daily sample of 8553 overlapping monthly observations. We compute the effective sample size based on the precision of the risk–return trade-off regression slope  $\beta_2$ . Specifically, we first compute the standard error of  $\hat{\beta}_2$  in overlapping data based on a block bootstrap with a block length of 21 (the average number of trading days per month). This estimate of the standard error, which we denote by  $SE(\hat{\beta}_2)$ , accounts for autocorrelation due to overlapping data. Next, we compute the standard error of  $\hat{\beta}_2$  in non-overlapping bootstrap samples of different sizes, i.e., samples that are generated with a bootstrap block length of one. Denote these estimates by  $SE_T(\hat{\beta}_2)$ . The effective sample size equals the number of non-overlapping observations  $\hat{T}$  that produces the same standard error as  $T = 8,553$  overlapping observations, i.e.,  $\hat{T}$  is defined by  $SE_{\hat{T}}(\hat{\beta}_2) = SE(\hat{\beta}_2)$ . Unless otherwise noted, all simulation results are based on this “effective sample size” of  $\hat{T} = 804$  non-overlapping monthly observations.

#### B.2. Statistical properties of the risk–return slope $\hat{\beta}_2$

A number of plausible econometric concerns can be brought forth about our estimation of the risk–return trade-off. First, the independent variable of regression (10) is estimated, which may render our estimator of the regression slope  $\beta_2$  inconsistent due to an errors-in-variables problem. Second, the dependent and independent variable are determined jointly based on our estimation of  $E[M|R]$ . It is therefore conceivable that a misspecified pricing kernel induces covariation between conditional return moments that does not really exist, i.e., our non-zero estimate of the risk–return slope coefficient  $\beta_2$  may reflect a bias. One piece of evidence against the latter concern is that we find a very similar risk–return relationship based on different polynomial orders for  $E[M|R]$ , including very flexible higher order polynomials. Nevertheless, it appears important to examine the statistical properties of  $\hat{\beta}_2$  in more detail.

Fig. 8 illustrates the sampling distribution of  $\hat{\beta}_2$  for different sample sizes. The top panel shows calibration 1 and the bottom panel calibration 2. The dotted line represents the true value of  $\beta_2$ . Estimates of  $\beta_2$  are computed based on estimated return densities, i.e., exactly as in the empirical data. Sampling variation in  $\hat{\beta}_2$  therefore reflects both small sample variation in true  $f_i(R_{i+1})$  moments and estimation error in these moments. Fig. 8 depicts the mean of  $\hat{\beta}_2$  with a solid line and values between the 5th and 95th percentiles of  $\hat{\beta}_2$  with the wider of the two shaded areas. For both calibrations,  $\hat{\beta}_2$  converges to  $\beta_2$  as the sample size approaches infinity. Hence,  $\hat{\beta}_2$  is a consistent estimator.

Table 8

**Bootstrap Test of the Risk-Return Trade-Off in Simulated Data.**

We simulate 5,000 finite samples with a length of  $\hat{T} = 804$  months from the data generating process described in the text. For each simulated sample, we perform the same bootstrap test of  $H_0 : \beta_2 = 0$  in regression (10) as for the empirical sample. The order of the pricing kernel polynomial is set to  $N = 2$ . This table reports the fraction of simulated samples for which  $H_0$  is rejected at different significance levels. The left panel reports results for the calibration with a positive risk–return trade-off (calibration 1), whereas the right panel reports results for calibration without a risk–return trade-off (calibration 2).

	Calibration 1 (with RTO)			Calibration 2 (without RTO)		
Significance level	1%	5%	10%	1%	5%	10%
Rejection rate, %	64.9	85.2	91.3	1.5	7.0	12.6

The figure shows that  $\hat{\beta}_2$  is nearly unbiased in both calibrations. At the effective sample size of  $\hat{T} = 804$ , the bias equals  $-0.03$  (compared to a true  $\beta_2$  value of  $3.96$ ) for calibration 1 and  $0.04$  (compared to a true  $\beta_2$  value of  $0$ ) for calibration 2. It also features high precision. At the effective sample size, calibration 1 (with a positive risk–return trade-off) implies a only  $0.25\%$  probability of obtaining a negative  $\hat{\beta}_2$ -estimate. At the same sample size, calibration 2 (without a risk–return trade-off) implies a  $0.59\%$  probability of obtaining a  $\hat{\beta}_2$ -estimate that exceeds our empirical point estimate of  $3.77$ . It is therefore highly unlikely that the true data generating process does not feature a positive risk–return trade-off.

To understand how estimation error in  $f_t(R_{t+1})$  affects  $\hat{\beta}_2$ , we also compute an alternative (but empirically infeasible) estimator  $\tilde{\beta}_2$  based on moments of the true  $f_t(R_{t+1})$ . Fig. 8 depicts values between the 5th and 95th percentiles of  $\tilde{\beta}_1$  with the narrower of the two shaded areas. This analysis shows that most variability in  $\hat{\beta}_2$  reflects small sample variation in true  $f_t(R_{t+1})$  moments, rather than estimation error in these moments. For example, for a sample size of  $\hat{T} = 804$ , calibration 1 implies that the  $90\%$  interval has a width of  $2.86$  for  $\tilde{\beta}_2$ , compared  $3.92$  for  $\hat{\beta}_2$ . This suggests that conditional return moments (and the projected pricing kernel) are estimated with relatively high precision in our empirical sample.

**B.2.1. Bootstrap test of  $\beta_2 = 0$** 

We next examine the reliability of our bootstrap test of  $H_0 : \beta_2 = 0$ . Like the empirical sample, each simulated sample is bootstrapped 10,000 times (see Footnote 11 for details). Because the simulated monthly data is non-overlapping, we use a bootstrap block length of 1, rather than the block length of 21 that was used in the overlapping empirical data. The null hypothesis is rejected in any given simulation if the bootstrapped  $p$ -value falls below the chosen significance level. Table 8 reports the rejection rates across the 5,000 simulated samples for both calibrations and a range of common significance levels.

When a risk–return trade-off exists in population (calibration 1), our estimation and testing methodology is likely to detect it. For example, the null hypothesis  $H_0 : \beta_2 = 0$  is rejected in  $85.2\%$  of simulated samples when testing at the  $5\%$  significance level. Hence, the test has fairly high power. When no risk–return trade-off exists in population (calibration 2), the bootstrap test is unlikely to incorrectly detect one. For example, the null hypothesis  $H_0 : \beta_2 = 0$  is incorrectly rejected in only  $7.0\%$  of simulated samples (a Type 1 error) when testing at the  $5\%$  significance level. Furthermore, a look across the different significance levels in Table 8 shows that the test is approximately correctly sized for the effective sample size of our data. This finding alleviates the biggest potential concern about our results: Our estimation of the projected pricing kernel (and conditional return moments) is quite unlikely to detect a risk–return trade-off if none exists in the true data generating process. In sum, our estimation and testing methodology provides a novel way to establish that a significantly positive risk–return trade-off exists in the data.

**B.3. Detection of NCC violations**

Lastly, we assess whether our detected NCC violations are genuine or the result of estimation noise. To that end, note the covariance term

in the NCC equals

$$\begin{aligned} \text{Cov}_t[M_{t+1}R_{t+1}, R_{t+1}] &= E_t[M_{t+1}R_{t+1}^2] - E_t[M_{t+1}R_{t+1}]E_t[R_{t+1}] \\ &= E_t[E_t[M_{t+1}|R_{t+1}]R_{t+1}^2] - E_t[E_t[M_{t+1}|R_{t+1}]R_{t+1}]E_t[R_{t+1}], \end{aligned}$$

where the last equality uses the law of iterated expectations. In each of the 5000 simulated samples, we use this expression to compute  $\text{Cov}_t[M_{t+1}R_{t+1}, R_{t+1}]$  based on the true versions of  $f_t(R_{t+1})$  and  $E_t[M_{t+1}|R_{t+1}]$ , and also based on their estimated counterparts. Next, we compute the probability of the true  $\text{Cov}_t[M_{t+1}R_{t+1}, R_{t+1}]$  being non-positive (the NCC holding), conditional on its finite sample estimate exceeding thresholds of  $18.5\%$  p.a. or  $28.0\%$  p.a. These thresholds reflect the magnitude of the NCC violations we detect in the empirical sample for the financial crisis and Covid-19 episode. The conditional probabilities equal  $5.1\%$  and  $3.2\%$ , respectively.

**References**

- Ait-Sahalia, Y., Lo, A.W., 2000. Nonparametric risk management and implied risk aversion. *J. Econometrics* 94 (1), 9–51.
- Ang, A., Chen, J., Xing, Y., 2006. Downside risk. *Rev. Financ. Stud.* 19 (4), 1191–1239.
- Back, K., Crotty, K., Kazempour, S.M., 2022. Validity, tightness, and forecasting power of risk premium bounds. *J. Financ. Econ.* 144 (3), 732–760.
- Bakshi, G., Crosby, J., Gao, X., Xue, J., Zhou, W., 2024. The options-inferred equity premium and the slippery slope of the negative correlation condition. *J. Invest. Manag.*
- Bansal, R., Yaron, A., 2004. Risks for the long run: A potential resolution of asset pricing puzzles. *J. Financ.* 59 (4), 1481–1509.
- Barndorff-Nielsen, O., 1977. Exponentially decreasing distributions for the logarithm of particle size. *Proc. R. Soc. A* 353 (1674), 401–419.
- Barndorff-Nielsen, O., 1978. Hyperbolic distributions and distributions on hyperbolae. *Scand. J. Stat.* 151–157.
- Barndorff-Nielsen, O., 1997. Normal inverse Gaussian distributions and stochastic volatility modelling. *Scand. J. Stat.* 24 (1), 1–13.
- Barone-Adesi, G., Engle, R.F., Mancini, L., 2008. A GARCH option pricing model with filtered historical simulation. *Rev. Financ. Stud.* 21 (3), 1223–1258.
- Barro, R.J., 2009. Rare disasters, asset prices, and welfare costs. *Am. Econ. Rev.* 99 (1), 243–264.
- Beason, T., Schreindorfer, D., 2022. Dissecting the equity premium. *J. Polit. Econ.* 130 (8), 2203–2222.
- Bekaert, G., Engstrom, E., 2017. Asset return dynamics under habits and bad environment—good environment fundamentals. *J. Polit. Econ.* 125 (3), 713–760.
- Bliss, R.R., Panigirtzoglou, N., 2004. Option-implied risk aversion estimates. *J. Financ.* 59 (1), 407–446.
- Bollerslev, T., Tauchen, G., Zhou, H., 2009. Expected stock returns and variance risk premia. *Rev. Financ. Stud.* 22 (11), 4463–4492.
- Breeden, D.T., Litzenberger, R.H., 1978. Prices of state-contingent claims implicit in option prices. *J. Bus.* 51 (4), 621–651.
- Campbell, J.Y., Cochrane, J.H., 1999. By force of habit: A consumption-based explanation of aggregate stock market behavior. *J. Polit. Econ.* 107 (2), 205–251.



- Campbell, J.Y., Thompson, S.B., 2008. Predicting excess stock returns out of sample: Can anything beat the historical average? *Rev. Financ. Stud.* 21 (4), 1509–1531.
- Chen, H., Joslin, S., Ni, S.X., 2019. Demand for crash insurance, intermediary constraints, and risk premia in financial markets. *Rev. Financ. Stud.* 32 (1), 228–265.
- Christoffersen, P., Heston, S., Jacobs, K., 2013. Capturing option anomalies with a variance-dependent pricing kernel. *Rev. Financ. Stud.* 26 (8), 1963–2006.
- Cochrane, J.H., 2005. *Asset Pricing: Revised Edition*. Princeton University Press.
- Cochrane, J.H., 2008. The dog that did not bark: A defense of return predictability. *Rev. Financ. Stud.* 21 (4), 1533–1575.
- Constantinides, G.M., Ghosh, A., 2017. Asset pricing with countercyclical household consumption risk. *J. Financ.* 72 (1), 415–460.
- Corsi, F., 2009. A simple approximate long-memory model of realized volatility. *J. Financ. Econ.* 7 (2), 174–196.
- Cuesdeanu, H., Jackwerth, J.C., 2018. The pricing kernel puzzle in forward looking data. *Rev. Deriv. Res.* 21 (3), 253–276.
- Dew-Becker, I., Giglio, S., 2022. Risk preferences implied by synthetic options. Working Paper.
- Dew-Becker, I., Giglio, S., Le, A., Rodriguez, M., 2017. The price of variance risk. *J. Financ. Econ.* 123 (2), 225–250.
- Drechsler, I., Yaron, A., 2011. What's vol got to do with it. *Rev. Financ. Stud.* 24 (1), 1–45.
- Driessen, J., Koëter, J., Wilms, O., 2020. Behavioral in the short-run and rational in the long-run? Evidence from S&P 500 options. Working Paper, Tilburg University.
- Elton, E.J., 1999. Presidential address: expected return, realized return, and asset pricing tests. *J. Financ.* 54 (4), 1199–1220.
- Eriksson, A., Ghysels, E., Wang, F., 2009. The normal inverse Gaussian distribution and the pricing of derivatives. *J. Deriv.* 16 (3), 23.
- French, K.R., Schwert, G.W., Stambaugh, R.F., 1987. Expected stock returns and volatility. *J. Financ. Econ.* 19 (1), 3–29.
- Gabaix, X., 2012. Variable rare disasters: An exactly solved framework for ten puzzles in macro-finance. *Q. J. Econ.* 127 (2), 645–700.
- Glosten, L.R., Jagannathan, R., Runkle, D.E., 1993. On the relation between the expected value and the volatility of the nominal excess return on stocks. *J. Financ.* 48 (5), 1779–1801.
- Gormsen, N., Jensen, C., 2023. Higher-moment risk. *J. Financ.* (forthcoming).
- Guo, H., Whitelaw, R.F., 2006. Uncovering the risk–return relation in the stock market. *J. Financ.* 61 (3), 1433–1463.
- Hansen, B.E., 1994. Autoregressive conditional density estimation. *Internat. Econom. Rev.* 705–730.
- Hansen, P.R., Timmermann, A., 2015. Equivalence between out-of-sample forecast comparisons and wald statistics. *Econometrica* 83 (6), 2485–2505.
- Harvey, C.R., Siddique, A., 2000. Conditional skewness in asset pricing tests. *J. Financ.* 55 (3), 1263–1295.
- Inoue, A., Kilian, L., 2005. In-sample or out-of-sample tests of predictability: Which one should we use? *Econometric Rev.* 23 (4), 371–402.
- Jackwerth, J.C., 2000. Recovering risk aversion from option prices and realized returns. *Rev. Financ. Stud.* 13 (2), 433–451.
- Johnson, T.L., 2019. A fresh look at return predictability using a more efficient estimator. *Rev. Asset Pricing Stud.* 9 (1), 1–46.
- Kim, H.J., 2022. Characterizing the conditional pricing kernel: A new approach. Working Paper, University of Houston.
- Kozhan, R., Neuberger, A., Schneider, P., 2013. The skew risk premium in the equity index market. *Rev. Financ. Stud.* 26 (9), 2174–2203.
- Linn, M., Shive, S., Shumway, T., 2018. Pricing kernel monotonicity and conditional information. *Rev. Financ. Stud.* 31 (2), 493–531.
- Lochstoer, L.A., Muir, T., 2022. Volatility expectations and returns. *J. Financ.* 77 (2), 1055–1096.
- Ludvigson, S.C., Ng, S., 2007. The empirical risk–return relation: A factor analysis approach. *J. Financ. Econ.* 83 (1), 171–222.
- Martin, I., 2017. What is the expected return on the market? *Q. J. Econ.* 132 (1), 367–433.
- Martin, I., Papadimitriou, D., 2022. Sentiment and speculation in a market with heterogeneous beliefs. *Am. Econ. Rev.* 112 (8), 2465–2517.
- Moreira, A., Muir, T., 2017. Volatility-managed portfolios. *J. Financ.* 72 (4), 1611–1644.
- Pástor, L., Sinha, M., Swaminathan, B., 2008. Estimating the intertemporal risk–return tradeoff using the implied cost of capital. *J. Financ.* 63 (6), 2859–2897.
- Rosenberg, J.V., Engle, R.F., 2002. Empirical pricing kernels. *J. Financ. Econ.* 64 (3), 341–372.
- Schorffheide, F., Song, D., Yaron, A., 2018. Identifying long-run risks: A Bayesian mixed-frequency approach. *Econometrica* 86 (2), 617–654.
- Schreindorfer, D., 2020. Macroeconomic tail risks and asset prices. *Rev. Financ. Stud.* 33 (8), 3541–3582.
- Schulz, F., 2016. On the timing and pricing of dividends: Comment. *Am. Econ. Rev.* 106 (10), 3185–3223.
- Vuong, Q.H., 1989. Likelihood ratio tests for model selection and non-nested hypotheses. *Econometrica* 307–333.
- Wachter, J.A., 2013. Can time-varying risk of rare disasters explain aggregate stock market volatility? *J. Financ.* 68 (3), 987–1035.