GENERALIZED RISK PREMIA – THE ECONOMIC VALUE OF PREDICTABILITY

PAUL SCHNEIDER

ABSTRACT. Using a new measure for predictability combining economic and statistical criteria I find that in the S&P 500 market neither stochastic volatility, nor valuation-based information are advantageous over a homoscedastic return model. The testing framework is based on a benchmark trading strategy with optimal Sharpe ratio. The strategy’s expected excess returns naturally accommodate compensation for higher-order moment risk, to first order variance risk and the equity premium. Further analysis of the price of moment risk reveals a connection between flight-to-quality in crisis periods and the compensation for variance, skew and kurtosis risk.

1. INTRODUCTION

When designing, selecting, and estimating models for asset allocation, trading signals or risk management a number of simple-to-ask, but hard-to-answer questions arise: Which dynamic trading strategy gives me the highest Sharpe ratio? Would I trade in less disperse forecast errors for less skewed ones? I can predict well according to standard criteria, why am I still losing money when I use my model’s signals in a trading strategy?

In this paper I show that these questions are very much related and that they pose an additional layer of complication on top of model selection based on statistical criteria,
a hard topic its own right. In the equity index market, for instance, Ang and Bekeart (2007) and Welch and Goyal (2008) cast doubt on the abilities of elaborate models to beat the historical average in predicting stock returns. Campbell and Thompson (2008), Cochrane (2008), and Rapach et al. (2010) have more encouraging results, but in both cases it is not clear how the predictability, if there was any, would best be translated into a trading strategy and how it would map into gains and losses.

Risk premia can be interpreted as expected profits from trading strategies and there is a growing literature on equity, variance and skew risk premia in the S&P 500 market (Bondarenko, 2003b; Bakshi et al., 2003; Eraker, 2008; Bollerslev et al., 2009; Carr and Wu, 2009; Bondarenko, 2010; Neuberger, 2012). From economic theory, these risk premia are all functions of the volatility of the pricing kernel and therefore one might ex-ante expect them to co-move. While variance and skew risk premia have shown remarkable similarities in Kozhan et al. (2013), they both seem to be unconnected to the equity premium. This example highlights that assets can be exposed to multiple risk factors and carry multiple risk premia.

To investigate the connection between the different risk compensations I introduce the likelihood ratio swap. Similar to how moments can be computed by taking derivatives of the moment-generating function, compensation for moment risk appears order by order in a series representation of the risk premium on this likelihood ratio swap. Analogous to the variance swap which trades implied variance for realized variance, the likelihood ratio swap trades \textit{implied pricing kernel variance} for \textit{realized pricing kernel variance}.

I show how to make the likelihood ratio swap fully tradeable using a model and a panel of options. The resulting trading strategy prescribes portfolio weights depending directly on the model parameters and the option-implied forward-neutral density, thereby relating the predictive density of a model to how much money could be made or lost with
it in the market. It thus combines statistical and economic information, addressing the concerns raised in Leitch and Tanner (1991) about the possible divergence of statistical and economic predictability criteria. Being benchmarked with respect to its Sharpe ratio, the strategy can also be extended to subjective (representative agent) loss functions similar to Aït-Sahalia and Brandt (2001).

My construction extends the pricing kernel misspecification framework from Hansen and Jagannathan (1997) and Li et al. (2010) to time-series-only models while still using all the information contained in option prices in the spirit of Aït-Sahalia and Brandt (2007), in addition to full pricing kernel specifications. It furthermore makes the projection of the pricing kernel onto an asset fully tradeable and yields a criterion for model risk based on the entire distribution of forecast errors, taking into account that risk factors driving the economy are not independent of each other. This is identified as a severe problem in Harvey and Siddique (2000), Aït-Sahalia and Brandt (2001) and more recently in Chabi-Yo (2012). The Sharpe ratio of this trading strategy is used as the test statistic for predictability. This is an advantage over Diebold et al. (1998), whose criterion is based on the independence of draws from a uniform distribution, which is very difficult to assess. Furthermore, it is an out-of-sample measure, to alleviate concerns raised in Thornton and Valente (2012) about the relevance of in-sample economic measures. The technology I employ is similar to Chapman (1997), who approximates the pricing kernel using an expansion with orthonormal polynomials, with the difference that I make full use of conditional information. In Bekaert and Liu (2004) and Chabi-Yo et al. (2008) it is argued that using conditioning information is important to improve inference and to alleviate pathologies in asset pricing. The polynomial expansion I employ to make the
strategy tradeable has an advantage over the expansion in terms of cumulants in Backus et al. (2011) and Backus et al. (2012) in the interpretation of risk premia.\(^1\)

In the S&P 500 index market I investigate whether a conditional volatility specification or valuation-based information improves the predictive power of a model. For this purpose I perform an out-of-sample analysis with the Heston (1993) stochastic volatility model and the homoscedastic Bilateral Gamma model from Küchler and Tappe (2008a) for S&P 500 forward returns. There is no clear distinction between them according to standard statistical criteria and economic value measured by comparing gains from simple moment trading strategies where the predictions are used as signals. Specifically, the Heston model produces higher bias and RMSE in forecasting returns, but is at the same time able to outperform the Bilateral Gamma model in betting on log return predictions. The criterion developed in this paper is able to rationalize this ambiguity and gives preference to the homoscedastic model.

A subsequent study of the conditional second moment of the pricing kernel through the lens of the Bilateral Gamma model reveals interesting patterns and relations between risk premia associated with different return moments. Compensation for moment risk is highly correlated across moments and a factor analysis reveals that it seems to be driven by not more than 2 factors, with far over 90% of the variation explained by the first principal component. There is an indication of flight to quality to the bond markets, which is intriguing, because it is obtained from the S&P 500 index and S&P 500 index options, without even specifying a term structure model.

The paper is organized as follows. I start from the definition of the likelihood ratio swap in Section 2. Section 3 then explains how it can be replicated and traded using a model, and how excess returns from this swap are a measure of predictability. Section 4

\(^1\)The implications of using raw moments over central moments for risk premia are detailed in Section 3.1, Footnote 6.
tests the economic value that can be generated through knowledge of conditional volatility and valuation-based information, and Section 5 concludes. Appendix A contains proofs for the claims made in the main text, B and C develop supplementary technical tools for Section 3.

2. Risk Premia and the Likelihood Ratio Swap

This section introduces a benchmark claim which forms the basis for assessing the value of predictability in terms of economic value. It also establishes the connection between trading strategies and risk premia.

Risk premia, whether in equity, bond, currency or other asset markets, have largely been concerned with the first moment of returns with only a few exceptions investigating variance and skewness risk premia (Bakshi et al., 2003; Carr and Wu, 2009; Kozhan et al., 2013, for instance). Researchers examine the predictability of excess returns to estimate the asset risk premium. An excess return can be interpreted as the profit from a trading strategy, which in the simplest case is to enter into a forward contract on the asset at time 0, and hold it until expiry at time $T$. Denoting the forward price of an asset contracted at time 0 with maturity $T$ by $F_{0,T}$ and the forward pricing probability measure parameterized by a maturity $T$ by $\mathbb{Q}_T$, the profit is therefore\(^2\)

$$F_{T,T} - \mathbb{E}_0^{\mathbb{Q}_T} [F_{T,T}] = F_{T,T} - F_{0,T}, \quad (1)$$

and the corresponding (conditional) risk premium is the conditional expectation thereof under the historical, or time-series probability measure $\mathbb{P}$

$$\mathbb{E}_0^{\mathbb{P}} [F_{T,T}] - \mathbb{E}_0^{\mathbb{Q}_T} [F_{T,T}]. \quad (2)$$

\(^2\)I work with forwards under the $T$ forward measure for ease of exposition, to avoid notational complications through stochastic dividends and interest rates.
In a market free from frictions and other imperfections, the expected excess return is interpreted as a premium for risk bearing. Introducing the likelihood ratio, state price density, or pricing kernel
\[ \mathcal{L} := \frac{dQ_T}{dP}, \]
the risk premium can be expressed as minus the conditional covariance of the forward price with the pricing kernel: 
\[ -\text{Cov}^P_0 (\mathcal{L}, F_{T,T}). \]
Yet another interpretation of the risk premium is as the expected payoff from a swap contract and this is the notion which I will adopt in this paper.

The exposure obtainable through the “first moment swap” in (2) is rather limited, so it is necessary to consider other payoff functions. Let
\[ R_{t,T} := \log \frac{F_{T,T}}{F_{t,T}}, \]
keeping in mind that absence of arbitrage the distribution of the forward price \( F_{T,T} \) is the same as the distribution of the underlying asset. Assume that \( \mathcal{L} \in L^2_P. \) This assumption states that the conditional variance of the pricing kernel is finite. Now choose the payoff function\(^4\)
\[ \mathcal{L}(R_{0,T}) := \mathbb{E}^{Q_T} \left[ \mathcal{L} \mid R_{0,T} \right]. \]

\(^3\)Define the weighted Hilbert space \( L^2_w \) as the set of (equivalence classes of) measurable functions \( f \) on \( \mathbb{R} \) with finite \( L^2_w \)-norm defined by
\[ \|f\|^2_{L^2_w} = \int_{\mathbb{R}} |f(\xi)|^2 \, dw(\xi) < \infty. \]
Accordingly, the scalar product on \( L^2_w \) is denoted by
\[ \langle f, h \rangle_{L^2_w} = \int_{\mathbb{R}} f(\xi) \, h(\xi) \, dw(\xi). \]
\(^4\)A similar payoff has been used before in Bakshi et al. (2010), with the difference that they take the conditional expectation under \( P \) rather than \( Q_T. \) In their context they use it within a “kernel call”, a theoretical instrument which is designed to optimally detect U-shaped pricing kernels. As we will see, the kernel call and the likelihood ratio swap introduced below are fundamentally different in that the one minimizes conditional expected returns in the presence of a U-shaped pricing kernel and the other minimizes the conditional Sharpe ratio. More examples of a projection of the pricing kernel under the \( P \) measure onto the variable of interest are the papers by Bondarenko (2003b) and Bondarenko (2003a).
This definition creates exposure to the variance of the pricing kernel and projects it onto log forward returns. The conditioning list in (4) could well be extended with forward returns other than \( R_{0,T} \), provided there are options written on the joint distribution of the chosen variables for empirical purposes. The properties of the profit from a swap contract written on the projected likelihood ratio are collected in

**Result 2.1 (Likelihood Ratio Contract).** Consider the likelihood ratio swap contract with net payoff \( \mathcal{L}(R_{0,T}) - \mathbb{E}^{Q_T}_0 [\mathcal{L}(R_{0,T})] \) and assume that \( \mathcal{L} \) is \( R_{0,T} \)-measurable

(i) In an economy with no risk premia, \( \mathbb{P} = Q_T \), the likelihood ratio swap pays nothing. Otherwise it is guaranteed to have, both, strictly positive and strictly negative payoff regions as a function of \( R_{0,T} \).

(ii) The expected payoff from a likelihood ratio swap, the likelihood ratio risk premium

\[
1 - \mathbb{E}^{Q_T}_0 [\mathcal{L}],
\]

involves only a conditional \( Q_T \) expectation, corresponds to minus the conditional variance of the pricing kernel, and is therefore non-positive and bounded.

(iii) The likelihood ratio contract attains the Hansen-Jagannathan bound.

**Proof.** In Appendix A.1. □

Result 2.1 outlines properties of a benchmark claim. Most models in finance, for example stochastic volatility models, do not satisfy the measurability requirement in Result 2.1. The likelihood ratio swap is designed such that it measures exposure to pricing kernel risk precisely under idealized assumptions. This is similar to variance swaps, which measure the premium on quadratic variation precisely only in a diffusion world (Bondarenko, 2010; Neuberger, 2012). Deviations from the idealized world make the measurement imperfect, but a variance swap will still pick up variance exposure in the presence of price
discontinuities. Analogously, the likelihood ratio swap will reflect pricing kernel risk also in the presence of path dependencies. The likelihood ratio swap creates exposure to the variance of the pricing kernel, projected onto an asset return. It ensures that there will be states of the world with positive payoffs for insurance sellers, as well as insurance buyers. The risk premium on the likelihood ratio contract is negative. This reminds of the empirical observation that the variance risk premium is large and negative. This is no coincidence, however, since there is a strong connection between the two (Result 3.1 below). To develop intuition about the likelihood ratio claim consider

Example 2.1 (Likelihood Ratio Swap Under Geometric Brownian Motion). Consider a world where the forward price is driven by \( \mathbb{P} \) Brownian motion \( W^p \). The \( \mathbb{P} \) distribution is \( N(\mu - \frac{1}{2}\sigma^2 T, \sigma^2 T) \), and the \( \mathbb{Q}_T \) distribution of log returns is \( N(-\frac{1}{2}\sigma^2 T, \sigma^2 T) \), because the forward price is a \( \mathbb{Q}_T \)-martingale. The likelihood ratio is

\[
\mathcal{L} = \exp \left( -\frac{\mu}{\sigma} W^p_T - \frac{\mu^2}{2\sigma^2} T \right),
\]

(6)

and the projection onto \( R_{0,T} \) is

\[
\mathcal{L}(R_{0,T}) = \exp \left( \frac{\mu T (\mu - \sigma^2) - 2R_{0,T}\mu}{2\sigma^2} \right).
\]

(7)

Computing the risk premium in this world yields \( 1 - e^{\frac{\mu^2}{2\sigma^2} T} \). It corresponds to one minus the exponential of the Sharpe ratio squared. Figure 1a shows that the payoff from a likelihood ratio swap in a Black-Scholes world reminds of a put option payoff. For U-shaped kernels, and recent empirical evidence points towards such a shape, the payoff need not be monotonic. This can be seen in Figure 1b. In both cases the risk premium associated with the likelihood ratio swap is negative, in line with Result 2.1 above.
The likelihood ratio swap introduced in Result 2.1 is only a theoretical benchmark claim, since the likelihood ratio is unattainable through a trading strategy, but the next section shows how it can nevertheless be used as a tool to measure the predictability of a model in terms of economic value.

3. The Economic Value of Predictability

To assess the predictive abilities of a model in economic terms ideally one could write financial contracts on the forecast errors across the entire distribution. In practice, usually only forecasts of levels are evaluated. A trading strategy based on such a signal would suggest entering into a long forward position on the stock today for an expected profit if the level prediction was higher than the current forward price, and going short otherwise. The net payoff from the corresponding strategy, a first-moment swap, can be seen in equation (1).

There are many reasons why such an approach can yield surprisingly bad economic results, even with unbiased forecast errors: Excess returns could be very small on average,
with high variation. In the S&P 500 market this is the case for first-moment swaps (cf. Figure 4a below). Furthermore, neglecting higher-order moments can lead to large unexpected losses. Even if higher-order moments were taken into account, for instance by simultaneously trading first-moment swaps and variance swaps, it is not clear how one would simultaneously hedge the joint exposure. The reason is that first-moment swaps and second-moment swaps do not move independently, because they are exposed to moments of the same distribution. This can already be seen from the decomposition of the pricing kernel in terms of its cumulants (Backus et al., 2011). To meet these concerns, a trading strategy based on the conditional likelihood ratio in (4) relates the predictive $P$ density to the market-implied $Q_T$ density and thereby captures risk premia simultaneously across the entire distribution of returns.

To operationalize this strategy, let $P_M(\theta)$, denote the probability measure induced by a model $M(\theta)$ with parameter $\theta$ and introduce\textsuperscript{5}

$$L_M := \frac{dQ_T}{dP_M(\theta)}, \quad \text{and} \quad L_M(R_{0,T}) := \mathbb{E}^{Q_T}[L_M \mid R_{0,T}]. \quad (8)$$

Provided that sufficiently many options are written on $R_{0,T}$, the $Q_T$ distribution can be computed through the results of Breeden and Litzenberger (1978) and Carr and Madan (2001), and provided the density function of $P_M$ is known, the likelihood ratio $L_M$ above is fully tractable.

Writing a contract on $L_M(R_{0,T})$ can then be interpreted as a trade on the model $M(\theta)$. As the risk return ratio of the likelihood ratio contract is the most extreme possible in the market, it can be thought of as a well-specified insurance contract written directly on the

\textsuperscript{5}Analogous to the payoff in (4) the list of conditioning arguments could be extended with other asset returns. The above payoff is conditional only on $R_{0,T}$ for ease of exposition and to reflect the availability of option prices. A higher-order expansion in the multivariate case requires options written on joint payoffs. To the best of my knowledge this is only the case in the foreign exchange market, where claims on products of exchange rates can be engineered through the no-arbitrage cross rate condition, and through spread options which are used in the commodity market.
model. Intuitively, if $P_M$ was close to $P$, then the payoffs from writing swap contracts on $L_M(R_{0,T})$ should be close to the payoffs from the unattainable benchmark likelihood ratio swap from Result 2.1. For a model $P_M$ which approximates $P$ badly, the risk premium will move away from the Hansen-Jagannathan bound.

In Section 3.1 below I develop a trading strategy replicating the model-implied likelihood ratio swap contract, and also introduce a representation of $L_M(R_{0,T})$ in terms of orthonormal polynomials yielding an equivalent portfolio of moment swaps for improved economic interpretation.

3.1. Likelihood Ratio Expansion. Provided that both $P_M$ and $Q_T$ have exponential tails and $L_M \in L^2_{P_M}$, Lemmas 3.1 and 3.3 in Filipović et al. (2013) guarantee that the set of polynomials is dense, and that an orthonormal basis of polynomials exists in $L^2_{P_M}$. This admits a polynomial representation

$$L_M(R_{0,T}) = 1 + \sum_{j=1}^{\infty} c_j(\theta, Q_T) H_j(R_{0,T} | \theta),$$

(11)

$I$ choose a polynomial expansion of the pricing kernel over an expansion in terms of cumulants as it is done in Backus et al. (2011) and Backus et al. (2012), because the risk premia on powers of $R_{0,T}$ are much easier estimated and interpreted than the risk premia on centralized powers of $R_{0,T}$. To see this consider the case of variance. Trading strategies can only attain $g(R_{0,T}) - \mathbb{E}_0^{Q_T}[g(R_{0,T})]$ rather than $g(R_{0,T}) - \mathbb{E}_0^P[g(R_{0,T})]$, such that the excess return on central variance attainable through a trading strategy is

$$\left( R_{0,T} - \mathbb{E}_0^{Q_T}[R_{0,T}] \right)^2 - \mathbb{E}_0^{Q_T}\left[ \left( R_{0,T} - \mathbb{E}_0^{Q_T}[R_{0,T}] \right)^2 \right],$$

(9)

rather than the desired

$$\left( R_{0,T} - \mathbb{E}_0^P[R_{0,T}] \right)^2 - \mathbb{E}_0^{Q_T}\left[ \left( R_{0,T} - \mathbb{E}_0^{Q_T}[R_{0,T}] \right)^2 \right].$$

(10)
which I truncate for practical purposes

\[
\mathcal{L}^{(j)}(R_{0,T}) := 1 + \sum_{j=1}^{J} c_j(\theta, Q_T) H_j(R_{0,T} | \theta) \quad (12)
\]

\[
= \sum_{j=0}^{J} a_j(\theta, Q_T) R_{0,T}^j. \quad (13)
\]

The orthonormal polynomials \( H_j(R_{0,T} | \theta) \) are specific to \( \mathbb{P}_M \) and the algorithm for developing them order by order from the canonical basis can be found in Appendix B. The coefficients \( c_j(\theta, Q_T) \) jointly depend on \( \mathbb{P}_M \) and \( Q_T \) and they are obtained from computing \( Q_T \) expectations of \( H_j^2(R_{0,T} | \theta) \) by applying the Carr and Madan (2001) and Bakshi and Madan (2000) formula repeatedly (in Appendix C). The coefficients \( a_j(\theta, Q_T) \) are defined implicitly by collecting terms of order \( R_{0,T}^j \).

The polynomial representation in (13) establishes a link between the likelihood ratio projection and linear CAPM-type pricing kernels (Nagel and Singleton, 2011), as well as nonlinear polynomial pricing kernel models (Harvey and Siddique, 2000). Martellini and Ziemann (2010) show that higher-order moments play an important role in investment decisions involving non-normal returns. The subsequent example is nevertheless done in a Black-Scholes setting to deepen intuition about the construction in terms of orthonormal polynomials without technicalities.

**Example 3.1** (Black Scholes Example 2.1 continued). Take the model \( \mathcal{M} = BS \) with \( R_{0,T} \sim N((\eta - 1/2\sigma^2)T, \sigma^2T) \). This corresponds to a world where an investor knows \( \sigma \), but has to estimate the growth rate \( \mu \). In that case we can compute in closed form

\[
\mathcal{L}_{BS}(R_{0,T}) = \exp \left( \frac{\eta(T(\eta - \sigma^2) - 2R_{0,T})}{2\sigma^2} \right). \quad (14)
\]

An order \( J = 1 \) projection onto multiple assets is feasible using regular options on the constituents and would lead to a CAPM-type model with time-varying coefficients as in Ang and Kristensen (2012) and also has a correspondence to the linear framework from Kan and Robotti (2009).
To illustrate the polynomial expansion (12), a second-order expansion of the above is
\[ \mathcal{L}_{BS}^{(2)}(R_0,T) = \sum_{i=0}^{2} c_{BS} H_{BS}^i(R_0,T), \]
where
\[ H_0^{BS}(x) = \tilde{H}_0^{BS}(x) = 1, \quad HO_0^{BS} = 1, \quad c_0^{BS} = 1 \]
\[ H_1^{BS}(x) = x + \frac{1}{2} T(\sigma^2 - 2\eta), \quad HO_1^{BS} = T\sigma^2, \quad c_1^{BS} = -T\eta \]
\[ H_2^{BS}(x) = x^2 + Tx(\sigma^2 - 2\eta) + \frac{1}{4} T^2(\sigma^2 - 2\eta)^2 - T\sigma^2, \quad HO_2^{BS} = 2T^2\sigma^4, \quad c_2^{BS} = T^2\eta^2. \]

Computing the risk premium associated with a first-order likelihood ratio swap
\[ \mathcal{L}_{BS}^{(1)}(R_0,T) = 1 - \frac{\eta(R_0,T + \frac{1}{2} T(\sigma^2 - 2\eta))}{\sigma^2} \]
gives
\[ \mathbb{E}_0^P \left[ \mathcal{L}_{BS}^{(1)}(R_0,T) \right] - \mathbb{E}_0^{QT} \left[ \mathcal{L}_{BS}^{(1)}(R_0,T) \right] = -T \cdot \frac{\eta\mu}{\sigma^2}, \]
which is, with \( \eta = \mu \), to first order the risk premium we had computed in Example 2.1.

The expression is minus the Sharpe ratio squared, consolidating the \( QT \) and \( P \) information into one number. Figure 2 shows how well we can approximate the likelihood ratio with a few orders, provided the model is correctly specified. The risk premium only depends on the true, unobserved mean \( \mu \) under the \( P \) measure, and the estimated mean \( \eta \), since there is no variance risk in the economy.

Comparing panels 2b and 2a shows that a misspecified model or an imprecisely estimated parameter diminishes the quality of the likelihood ratio approximation, as can be expected. This in turn moves the likelihood ratio swap excess return away from the Hansen-Jagannathan bound. Whichever cause, a misspecified model, a low-order truncation of the likelihood ratio approximation error, or parameter estimation error, will result in a lower absolute value of the realized Sharpe ratio.
Moving away from the stylized Black Scholes economy back to the unattainable $L_M(R_{0,T})$ swap, a first-order expansion with respect to the true, unknown measure $\mathbb{P}$ reveals an interesting insight relating the risk premium on the likelihood ratio swap to variance risk and the equity premium. This is empirically relevant, since both the variance premium and the equity premium can be estimated non-parametrically (model-free).

**Result 3.1** (Equity Premium, Variance Premium and Likelihood Ratio Risk). *The risk premium on $L(R_{0,T})$ defined in Result 2.1 (ii) depends to first order on the variance risk premium, the difference between expected realized variance and $VIX^2$, and the equity risk premium.*

*Proof.* In Appendix A.2

Having obtained a representation of the likelihood ratio in terms of polynomials in $R_{0,T}$, the next section discusses how to compute a “realized likelihood ratio” and an “implied
3.2. Trading the Likelihood Ratio Swap. The truncated representation (12) in terms of polynomials in $R_{0,T}$ grants an important advantage for trading the $L_M$ swap over the original form (11): It can be directly replicated by dynamically trading options and the underlying forward contract using the methodology from Kozhan et al. (2013, KNS). In the present case this will yield a dynamically hedged trading strategy where the weights on the option portfolios and the forward contract will depend on the model parameters $\theta$ and the option prices themselves. This is beneficial over a buy-and-hold strategy using a static options position.\footnote{In principle one could compute the likelihood ratio as a function of the model parameters $\theta$ and options prices, and then approximate the payoff with an option portfolio.} The KNS strategy ensures that the risk characteristics of the exposure stay the same over the trading interval. As a consequence, the excess return from a likelihood ratio swap engineered as a superposition of KNS moment swaps maintains exposure to the model’s original predictive density $\mathbb{P}_M$. With a static buy-and-hold strategy, the risk characteristics would change over time, and the portfolio would pick up other, unwanted exposures on the way. As a result buy-and-hold strategies deliver noisier estimates than their dynamically hedged counterparts.

The KNS strategy is model-independent. In terms of swap contracts we have from KNS, to trade the $(n)$-th moment, the floating leg

$$r^{(n)}_{M}(0, T) := \sum_{i=1}^{M} G^{(n)}_{t_i, t_{i-1}, T} - G^{(n)}_{t_{i-1}} - \frac{G^{(n)'}_{t_{i-1}}}{F_{t_{i-1}, T}} (F_{t_i, T} - F_{t_{i-1}, T})$$

(17)
at observation frequencies $0 = t_0, \ldots, t_M = T$, where

$$G_{t,T}^{(n)} := \mathbb{E}_t^{Q_T} \left[ r_M^{(n)}(t, T) \right] = \mathbb{E}_t^{Q_T} \left[ R_t^{n} \right],$$

$$G_{t,t,T}^{(n)} := \mathbb{E}_t^{Q_T} \left[ r_M^{(n)}(t, T) \right] = \mathbb{E}_t^{Q_T} \left[ R_t^{n} \right],$$

$$G_{t,T}^{(n)'} := n \mathbb{E}_t^{Q_T} \left[ R_t^{n-1} \right],$$

where the representation of the conditional $Q_T$ expectations in terms of option prices is given in Appendix C. The fixed (implied) leg is then

$$I_{0,T}^{(n)} := \mathbb{E}_0^{Q_T} \left[ r_M^{(n)}(0, T) \right] = \mathbb{E}_0^{Q_T} \left[ R_0^{n} \right],$$

independently of the hedging frequency $M$. Finally, I define the realized risk premium on the $(n)$-th moment swap over the period $[0, T]$ as the excess return

$$RP_{0,T}^{(n)} := r_M^{(n)}(0, T) - I_{0,T}^{(n)}. \quad (22)$$

Analogous to variance swaps measuring variance risk premia and skew swaps measuring skew risk premia, the likelihood ratio excess return is then obtained from (13) by trading weighted moment swaps

$$RP_{0,T}^{L,M(J)} := \sum_{n=1}^{J} a_n(\theta, Q_T) \cdot RP_{0,T}^{(n)}. \quad (23)$$

The likelihood ratio swap payoff gives the portfolio weights on the moment swaps as a function of the parameters of the model $\mathcal{M}$ and current option prices. By construction this portfolio exposes the model to the distribution of forecast errors through the conditional

---

9 The idea behind the KNS trading strategy originates from the fact that in a diffusion world the $G^{(n)}$ trading strategy is a perfect hedge for $R^n$ risk, in the sense that the floating leg of a $G^{(n)}$ swap is predictable. Note that this is just a motivation, the actual trading strategy does not make any assumptions on the forward price process, in particular it is jump-robust.
\( \mathbb{P}_M \) moments, and measures this exposure in a model-free way, since \( RF_{0,T}^{(n)} \) are model-free. The terminology \textit{generalized risk premium} is motivated by the series representation in (23) (which unconditionally also holds for the unobservable \( \mathbb{P} \) measure), as the risk premium on the likelihood ratio swap is seen to be a weighted sum of risk premia on moment swaps.

3.3. \textbf{Models for} \( \mathbb{P} \) \textit{and} \( \mathbb{Q}_T \). The recent literature on empirical asset pricing is focused on models which jointly model the \( \mathbb{Q}_T \) as well as the \( \mathbb{P} \) distribution. To accommodate this notion define

\[
\mathcal{L}_{\mathcal{M}^*} := \frac{dQ_{T,M}}{dP_M},
\]

and

\[
\mathcal{L}_{\mathcal{M}^*}(R_{0,T}) := \mathbb{E}^{Q_T}[\mathcal{L}_{\mathcal{M}^*} \mid R_{0,T}].
\]

The formula in Appendix B for developing the expansion stays the same, only \( \mathbb{Q}_T \) is swapped for \( \mathbb{Q}_{T,M} \).

\textbf{Example 3.2} (Power Utility). Consider the power utility pricing kernel

\[
\mathcal{L}_{\mathcal{M}^*}(R_{0,T}; \gamma) := C \cdot e^{-\gamma R_{0,T}},
\]

where \( C \) is the normalization constant, and suppose the investor uses a lognormal model for the log asset return \( R_{0,T} \sim^{\mathbb{P}_M} N(\eta - \frac{1}{2}\sigma^2 T, \sqrt{T}\sigma) \), and hence \( R_{0,T} \sim^{\mathbb{Q}_{T,M}} \).
\[ N\left(-\frac{1}{2}\sigma^2 T, \sqrt{T}\sigma\right). \]

Then, choosing \( J = 2 \) we have

\[
\mathcal{L}^{(2)}_{\mathcal{M}_t}(R_{0,T}; \gamma) = 1 + \frac{T\gamma}{4} \left(2\sigma - \gamma - \frac{4\eta}{\sigma}\right) + \left(\frac{T\gamma}{4}\right)^2 \left(\frac{4\eta^2}{\sigma^2} - 4\eta + \sigma^2\right) + \left(\frac{\gamma}{\sigma} - \frac{\gamma^2\eta T}{2\sigma^2} + \frac{\gamma^2 T}{4}\right) \cdot R_{0,T} + \frac{\gamma^2}{4\sigma^2} \cdot R_{0,T}^2.
\]

The buy-and-hold excess return at time \( t \) from this instrument is then

\[
\left(\frac{\gamma}{\sigma} - \frac{\gamma^2\eta T}{2\sigma^2} + \frac{\gamma^2 T}{4}\right) \cdot (R_{0,T} - \mathbb{E}^{\mathcal{Q}_t}_0 [R_{0,T}]) + \left(\frac{\gamma^2}{4\sigma^2}\right) \cdot (R_{0,T}^2 - \mathbb{E}^{\mathcal{Q}_t}_0 [R_{0,T}^2])
\]

where the swap payoffs \( R_{0,T} - \mathbb{E}^{\mathcal{Q}_t}_0 [R_{0,T}] \) and \( R_{0,T}^2 - \mathbb{E}^{\mathcal{Q}_t}_0 [R_{0,T}^2] \) can be replicated through the KNS trading strategies from Section 3.2. A power-utility investor is always long variance, but the exposure to first return moments can change sign.

4. Assessing the Value of VIX and Valuation-Ratios for Predictability

In this section I apply the concept of generalized risk premia to measure the economic value of predictability to S&P 500 data. To this end I introduce three competing models. The Heston (1993) model for the joint time-series of log S&P 500 forwards and the VIX index, the homoscedastic Bilateral Gamma model from Küchler and Tappe (2008a) for log S&P 500 forward returns, and the same Bilateral Gamma model, but centered around a conditional expected value based on the dividend-to-price ratio suggested in Campbell and Thompson (2008). The former model accommodates stochastic volatility, while the

\[ \text{This implies } \gamma = \frac{\eta}{\sigma} \text{ and } C = \exp\left(\frac{T\eta(y - \gamma^2)}{2\sigma^2}\right). \]
latter two do not. For all models under consideration, the distribution of log index returns can be parameterized to exhibit flexible skewness and sizable excess kurtosis. Importantly, the tails of these distributions are heavy enough to render expansion (11) meaningful. The Heston model has 6 parameters, the Bilateral Gamma models have 4 parameters.\textsuperscript{11}

4.1. Data and Statistical Properties. The S&P 500 and the VIX index are taken from Bloomberg. Options data are from OptionMetrics. The data set includes closing bid and ask quotes for each option contract along with the corresponding strike price, Black-Scholes implied volatility, the zero-yield curve, and dividend yield. From the data I filter out all entries with non-standard settlements and with implied volatility less than 0.001 or higher than 9. The options mature every third Friday each month, and I use this maturity in a monthly time grid. Joint option and S&P 500 data is available from January 1996 up to January 2012. I use the sample period from January 1996 to August 2002 as a burn-in period and the remainder for the out-of-sample study.

With the yield curve and dividend yield information from OptionMetrics I construct a time series of forwards on the S&P 500 spot index. To estimate the conditional $Q_T$ moments $G$ for the trading strategy from Section 3.2, I first fit a cubic spline through the available implied volatilities along the strike dimension and subsequently integrate numerically equations (32) and (33). As time evolves, the maturity of these contracts decreases until the next exercise date (third Friday each month) is reached, where I switch to next month’s maturity.

Figure 3 shows the time evolution of monthly excess returns on the first four moments of log S&P 500 forward returns. Panel (a) of Table 1 displays the correlations between

\textsuperscript{11}The Heston model is comprised of parameters controlling the 1) unconditional growth rate of the forward price, 2) (instantaneously) predictable correlation with stochastic volatility, 3) instantaneous (unpredictable) correlation between the two driving Brownian motions, 4) mean-reversion, and 5) unconditional mean of stochastic variance, and 6) instantaneous volatility of stochastic variance. The Bilateral Gamma model has 4 parameters with no obvious interpretation.
them. To appreciate the non-zero correlations in Table 1, it is important to keep in mind that the monomials $R^i$ appearing in (13) are not orthogonal\footnote{Suppose there was a distribution $M$ for which the canonical basis $\langle 1, x, x^2, \ldots \rangle$ represented an orthogonal system in $L^2_M$. Then for $n \in \mathbb{N}$ we would have $\|x^n\|^2_{L^2_M} > 0$, but $\langle x^{2n}, x^0 \rangle_{L^2_M} = 0$, a contradiction.}, while the corresponding polynomials $H$ from (12) are (Appendix B.1). Hence, we should ex-ante expect non-zero correlations for the excess returns on the moment swaps. Risk premia on first and second log return moments are highly negatively correlated at -65%. Compensation for second-moment and third-moment risk is even more highly correlated at -75%. This confirms the finding in Kozhan et al. (2013) that variance and skew risk premia are manifestations of the same risk factor.\footnote{Their definition of the skew and variance risk premium is slightly different than the one used in this paper, explaining the difference in the correlation coefficient.} Third-moment and fourth-moment compensation is almost

Figure 3. Excess Returns from Return Moment Swaps: The figure shows excess returns from trading moment swaps on monthly log forward returns on the S&P 500 index. The excess returns, the difference between realized and implied return moments are defined in (22). Realized return moments are measured with daily data.
(A) Correlations Between Excess Returns on Moment Contracts

<table>
<thead>
<tr>
<th></th>
<th>$RP^{(1)}$</th>
<th>$RP^{(2)}$</th>
<th>$RP^{(3)}$</th>
</tr>
</thead>
<tbody>
<tr>
<td>$RP^{(2)}$</td>
<td>-64.90</td>
<td></td>
<td></td>
</tr>
<tr>
<td>$RP^{(3)}$</td>
<td>29.83</td>
<td>-75.45</td>
<td></td>
</tr>
<tr>
<td>$RP^{(4)}$</td>
<td>-7.60</td>
<td>52.09</td>
<td>-94.78</td>
</tr>
</tbody>
</table>

(b) Bilateral Gamma Excess Returns

<table>
<thead>
<tr>
<th></th>
<th>$l_1$</th>
<th>$l_2$</th>
<th>$l_3$</th>
<th>$l_4$</th>
</tr>
</thead>
<tbody>
<tr>
<td>Mean</td>
<td>25.71</td>
<td>48.66</td>
<td>-371.66</td>
<td>397.28</td>
</tr>
<tr>
<td>Correlations</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$l_2$</td>
<td>-82.55</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$l_3$</td>
<td>95.40</td>
<td>-93.27</td>
<td></td>
<td></td>
</tr>
<tr>
<td>$l_4$</td>
<td>-97.56</td>
<td>89.22</td>
<td>-99.44</td>
<td></td>
</tr>
</tbody>
</table>

Table 1. **Risk Premia and Conditional Pricing Kernel Factor Loadings:**
Table (a) shows correlations between excess returns on the moment swaps defined in (22). Panel (b) shows the relative contributions of excess returns on moment swaps to excess returns on the likelihood ratio swap $l_i := a_i(\theta, Q_T)RP_i/(RP_{0,T}^{L})$, where $M$ denotes the Bilateral Gamma model from Küchler and Tappe (2008a). The parameters $\theta$ are taken from an expanding out-of-sample estimation on the log returns of S&P 500 forwards.

mechanically connected with a correlation of -95%. This suggests that the dependence between compensation for moments becomes stronger with the order of the moments. The degree of persistence in the excess return time series is moderate, and a unit root test suggests stationarity.

4.2. **The Economic Value of Stochastic Volatility and Valuation Ratios.** Starting from the initial burn-in sample span January 1996 to August 2002 I estimate all three models with maximum likelihood. For this purpose the likelihood function of the Bilateral Gamma density is given in Küchler and Tappe (2008b), while for the Heston model I use fourth-order likelihood expansions from Filipović et al. (2013). To check robustness against econometric methodology I also use the method of moment estimator described in
Küchler and Tappe (2008a) for the Bilateral Gamma and the likelihood expansions from Aït-Sahalia (2008) for the Heston model with no qualitative effect on the subsequent results. For the Bilateral Gamma model centered around the valuation-based return predictions I use the data provided on Amit Goyal’s website which is available until the end of 2011 and follow exactly the construction from Campbell and Thompson (2008, eq. (4)) for the S&P 500 return forecast using the dividend-price ratio, modified only to accommodate a prediction of a forward return. For the remainder of the paper I denote by Heston the Heston (1993) model, by Bilateral the Küchler and Tappe (2008a) model, and by BilateralCT the same model centered around the valuation-based return prediction.

Using the point estimates I then compute the conditional n-power moments $E^{\mathcal{M}}_{t_0} [R^n_{t_0,t_1}]$ for $\mathcal{M} \in \{\text{Bilateral, BilateralCT, Heston}\}$, and together with the $E^{Q_T}_{t_0} [R^n_{t_0,t_1}]$ moments estimated from options data I obtain likelihood ratio approximations $L^{(4)}_{\mathcal{M}}(R_{t_0,t_1})$, where $t_0$ is 15 August 2002 and $t_1$ is 19 September 2002. Collecting terms and using the conditional $Q_T$ moments estimated from option prices again, I then compute the fixed leg $E^{Q_T}_{t_0} \left[ L^{(4)}_{\mathcal{M}}(R_{t_0,t_1}) \right]$ of the likelihood ratio swap written on model $\mathcal{M}$. The differences between the realized counterparts developed in Section 3.2 and the fixed legs give the excess returns on the likelihood ratio swap approximation. I then move on to estimate the models again including the data point at time $t_1$, make forecasts for time $t_2$ and so forth. The predictions used in this exercise are therefore entirely out-of-sample.

On statistical grounds the Heston (1993) model is dominated by the Bilateral Gamma models on the basis of bias and RMSE of the out-of-sample forecast errors. Table 2 indicates this uniformly across the first four moments. The differences are sizable, and stay qualitatively the same when changing estimation methodology to method of moments.
or using different likelihood approximations. The Bilateral Gamma model with valuation-based return predictions performs slightly better in first moments than the standard Bilateral Gamma model which reflects exactly the sample moments.

The big differences according to standard statistical criteria suggest that the investor trusting the Bilateral Gamma model should be able to make more money than an investor using Heston’s model when using the predictions of the model as signals in a trading strategy. From Leitch and Tanner (1991) we know that this is not necessarily the case. To study this in detail for the S&P 500 index and options market, Figure 4 shows the outcome of a trading experiment. The investor enters into a short position in an n-moment swap if \( E^{P,M}_{t_i} \left[ R_{t_i,t_{i+n}} \right] < E^{Q,T}_{t_i} \left[ R_{t_i,t_{i+n}} \right] \) for \( M \in \{ \text{Bilateral}, \text{BilateralCT}, \text{Heston} \} \), otherwise into a long position. A fourth strategy, which I term Simple, goes short if average excess

<table>
<thead>
<tr>
<th></th>
<th>Bias</th>
<th>RMSE</th>
</tr>
</thead>
<tbody>
<tr>
<td>Heston</td>
<td>-7.08e-03</td>
<td>6.02e-02</td>
</tr>
<tr>
<td></td>
<td>-2.14e-03</td>
<td>1.11e-02</td>
</tr>
<tr>
<td></td>
<td>-6.27e-04</td>
<td>3.40e-03</td>
</tr>
<tr>
<td></td>
<td>-1.94e-04</td>
<td>1.48e-03</td>
</tr>
<tr>
<td>Bilateral</td>
<td>-1.75e-03</td>
<td>5.56e-02</td>
</tr>
<tr>
<td></td>
<td>6.40-04</td>
<td>8.07e-03</td>
</tr>
<tr>
<td></td>
<td>-1.39e-04</td>
<td>1.84e-03</td>
</tr>
<tr>
<td></td>
<td>3.85e-05</td>
<td>4.23e-04</td>
</tr>
<tr>
<td>BilateralCT</td>
<td>-6.80e-04</td>
<td>5.55e-02</td>
</tr>
<tr>
<td></td>
<td>6.40-04</td>
<td>8.07e-03</td>
</tr>
<tr>
<td></td>
<td>-1.39e-04</td>
<td>1.84e-03</td>
</tr>
<tr>
<td></td>
<td>3.85e-05</td>
<td>4.23e-04</td>
</tr>
</tbody>
</table>

Table 2. Bias and RMSE of Forecast Errors: The table shows bias and RMSE of the out-of-sample forecast errors obtained from maximum likelihood estimations of the Bilateral Gamma model from Kückler and Tappe (2008a, Bilateral), the Bilateral Gamma model centered around earnings/price ratio forecast from Campbell and Thompson (2008, BilateralCT), and the Heston (1993, Heston) model. The forecast errors are obtained from the difference \( E^{P,M}_{t_i} \left[ R_{t_i,T} \right] - R_{t_i,T} \) for \( M \in \{ \text{Bilateral}, \text{BilateralCT}, \text{Heston} \} \) and \( n = 1, \ldots, 4 \). The estimates obtained from historical averages agree with the estimates from the Bilateral Gamma distribution and are therefore omitted. The estimations are redone every month, starting from an initial window of monthly data ranging from January 1996 to August 2002, and expanding up to January 2012.
returns up to time $t_i$ have been positive, and short otherwise. In the present data set none of the simple average-based strategies switches sign.
The results from the trading strategies do not agree with the Bias and RMSE statistics from Table 2. Panel 4a shows that the Bilateral Gamma investor looses on a first moment strategy to the Heston investor despite her superior bias and RMSE results.\textsuperscript{14} This is an example of the non-monotonic mapping between common statistical criteria of predictability and economic value. The Heston investor gets a good prediction of returns at the “right times” which more than compensates a loss on average against the Bilateral investor. The importance of stochastic volatility in the conditioning information for predicting equity returns is also found important in Johannes et al. (2013). Another interesting result concerns the results of the quadratic strategy using the signals of the models. Panel 4b shows that the Heston investor looses to the Bilateral Gamma investor despite her stochastic volatility specification. Even more surprisingly, both, Heston and Bilateral Gamma investors loose to the investor basing her decisions on simple averages of past squared excess returns. No big differences can be seen for moments 3 and 4 in panels 4c and 4d.

Given the confirmation of the Leitch and Tanner (1991) result of the possible divergence between statistical and economic criteria also for the S&P 500, the question remains how to weight the conflicting statistical and economic evidence collected so far. How severe is a failure in predicting skewness compared to a failure to predict variance and how does this translate into excess returns from trading strategies? The likelihood ratio swap from (12) helps addressing these questions along two dimensions. First, the likelihood ratio trading rule is based on both the $P$ density of the model as well as the $Q_{T}$ density. The portfolio decision imposed by the likelihood ratio swap thereby takes into account both predictability as well as profitability. Second, the likelihood is benchmarked against the

\textsuperscript{14}Here, the valuation-based Bilateral Gamma model delivers exactly the same wealth trajectory as the sample-moment based one and is therefore omitted from the picture.
Figure 5. **Excess Returns from Likelihood Ratio Swap:** The figure shows excess returns from $\mathcal{M} \in \{\text{Bilateral}, \text{BilateralCT}, \text{Heston}\}$ from the likelihood ratio swap trading strategy from (23) in 100% terms. The time series are computed from September 2002 until January 2012 on the basis of out-of-sample predictions from the Bilateral Gamma model from Küchler and Tappe (2008a) with and without valuation information and the Heston (1993) model.

Hansen-Jagannathan bound through the Sharpe ratio, and thus not require the specification of a separate loss function. If a specific utility function was desired, then the extension outlined in Section 3.3 could be used, however.

Figure 5 shows excess returns from the Bilateral Gamma (with and without valuation information) and the Heston likelihood ratio swaps. There are regimes in which the time series co-move almost one-to-one, and others, in particular crises, in which the excess returns differ dramatically. It can be seen that during these crisis times the Heston investor can generate large negative excess returns, indicating a significant advantage over the homoscedastic models. But these large negative spikes also drive up volatility,
so that the absolute value of the sample Sharpe ratio tends to be lower for the Heston investor.

The time series of excess returns are serially uncorrelated, allowing an independence bootstrap. Figure 6 shows the sampling distribution of the model-implied Sharpe ratios of the likelihood-ratio strategy. The result suggests that according to the Sharpe ratio criterion the homoscedastic models outperform Heston’s stochastic volatility model in most states of the world. The point estimates of the strategies are monthly Sharpe ratios of -0.33, -0.32, and -0.17 for the Bilateral Gamma, the Bilateral Gamma with valuation information, and Heston’s model, respectively. The result does not question the importance of stochastic volatility in the S&P 500 market, which is found to be important in Buraschi and Jackwerth (2001), but the inferior return-risk ratio implied by Heston’s model. The large spike on 17 March 2011 in Figure 5 raises concerns about the robustness of the result with respect to outliers in small samples. I therefore bootstrap the difference
between the Sharpe ratio implied by the Bilateral Gamma model and Heston’s model with, and also without the spike. In Panel 6b it can be seen that the distribution of the difference is robust to this large negative return in the Heston strategy.

4.3. **Dissecting the Variance of the Pricing Kernel.** The series representation (12) yields a decomposition of the conditional variance of the pricing kernel into the prices of risk for moments. Specifically, from (13) we can investigate the relative contribution of the price of moment risk \( j \) defined as

\[
\frac{a_j(\theta, Q_T)\mathbb{E}^Q_T [R^j_{0,T}]}{\mathbb{E}^Q_T [\mathcal{L}^{(j)}(R_{0,T})]}
\]

(28)

to the fixed leg of the likelihood ratio swap. This quantity depends on the model \( \mathcal{M} \) through the functional form of the orthonormal \( H \), and the parameter \( \theta \) in the coefficients. In this section I use the most successful model from the previous model selection exercise, the Bilateral Gamma model, for this purpose. The price of moment 0 risk is interpreted as an investment in the riskless bond market, since it does not depend on \( R \).

The relative contributions of moments \( j = 0, \ldots, 4 \) exhibit a pronounced factor structure. A principal component analysis reveals that a single factor can explain 95% of the variation of the variance of the pricing kernel. The contribution of the price of moment 0 to the variance of the pricing kernel is negative throughout the troubled economy around 2002 and the later crisis starting from 2008, indicating flight-to-quality of the Bilateral Gamma investor.

Figure 7 shows how the relative contributions of the moments 0 to 4 to the fixed leg of the likelihood ratio swap change over time. Exposure to first moments of S&P 500 log returns is surprisingly small. The trajectories of the exposures to moments 2 to 4 are the most revealing. The Bilateral Gamma investor is generally long kurtosis, short skewness, but her variance exposure switches sign. The three time series are highly correlated and
net out to selling insurance in bad times and buying insurance in good times. This strengthens the role of S&P 500 variance risk as an indicator for investor fear. Panel (b) in Table 1 shows that similar properties also hold for excess returns. Compensation for skew risk is a near-perfect substitute for compensation for kurtosis risk.

5. Conclusions

This paper is based on a benchmark trading strategy that yields the maximal Sharpe ratio possible in an arbitrage free economy. The strategy has an interpretation as a likelihood ratio swap, trading implied pricing kernel variance for realized. Analogous to a moment-generating function having a series representation in terms of moments of all
orders, the risk premium associated with the likelihood ratio swap is the sum of risk premia on all moments of the asset of interest. Accordingly, the equity premium and the variance premium are shown to be of first-order importance.

For a given time series model, I show how to trade the model-implied likelihood ratio swap using dynamically rebalanced portfolios of options and the underlying asset. This trading strategy can be thought of as an insurance contract against model misspecification, because it quantifies economically relevant model failures by relating them to observed prices. Comparing the Sharpe ratios of the realized profits from model-implied likelihood ratio swaps from two competing models yields a predictability criterion with a market-implied loss function.

In an out-of-sample study using S&P 500 data from 1996 to January 2012 extant statistical and economic value-based criteria give conflicting results on the Bilateral Gamma model from Küchler and Tappe (2008a) for S&P 500 log forward returns, the same model centered around valuation-based predictions suggested in Campbell and Thompson (2008), and the Heston (1993) model for the joint time series of S&P forwards and the VIX implied volatility index. The methodology developed in this paper consolidates the predictive and price information into one trading strategy and gives preference to the homoscedastic Bilateral Gamma model.

Using the technology developed in this paper, the conditional second moment of the pricing kernel can be dissected into the price of variance, skew, and kurtosis risk. This approach, in connection with the Bilateral Gamma model, reveals a strong inverse relation between a bond investment and implied variance and kurtosis, hinting at flight to quality.

The framework introduced in this paper can be used to trade pricing kernels induced by economic as well as statistical models. As such it lends itself for investigating, for instance, under which circumstances a long-run-risk investor looses money compared to
a power-utility investor. This notion could be used to extend the tests performed in Ferson et al. (2013) to a combined economic-statistical framework. Perhaps even more compelling are the tradeable implications of differences in belief model such as the one from (Dumas et al., 2009).

Finally, it is not clear why an agent should rely on standard inference methods to estimate models for investment purposes. This paper shows that the predictive density jointly with the forward-neutral density are important when it comes to implementing trading strategies. The size of, and the dependence between risk premia determines the performance of a trading strategy. The out-of-sample behavior of excess returns generated from models estimated on economic and statistical criteria jointly, an idea originating from Brandt and Chapman (2002) and Aït-Sahalia and Brandt (2001), therefore suggests to be a fruitful future research topic.

References


A.1. Proof of Result 2.1.

Part 1. The case of no risk premia is covered by $\mathcal{L}(R_0, T) = 1$ for all $R_0, T$ ($\mathbb{P}$ and $\mathbb{Q}_T$ are the same). In the non-constant case assume by contradiction that $\mathcal{L}(R_0, T) \leq 1$ for all $R_0, T$. By continuity it has at least one interval strictly smaller than one. We would hence have $\langle \mathcal{L}, 1 \rangle_{L^2_p} < 1$ and $\mathcal{L}$ can not be a likelihood ratio. The case $\mathcal{L}(R_0, T) \geq 1$ for all $R_0, T$ can be treated analogously. This means that $\mathcal{L}(R_0, T)$ must be, both, strictly greater and strictly smaller than one. The claim then follows from the fact that $0 \leq \mathbb{E}^0_{\mathbb{P}} [\mathcal{L}^2] - 1 = \mathbb{E}^0_{\mathbb{P}} [\mathcal{L}(R_0, T)] - 1$.

Part 2.

$$\mathbb{E}^0_{\mathbb{P}} [\mathcal{L}(R_0, T)] - \mathbb{E}^{Q_T}_0 [\mathcal{L}(R_0, T)] = \mathbb{E}^0_{\mathbb{P}} [\mathbb{E}^{Q_T} [\mathcal{L} | R_0, T]] - \mathbb{E}^{Q_T}_0 [\mathbb{E}^Q [\mathcal{L} | R_0, T]]$$

$$= \mathbb{E}^{Q_T}_0 \left[ \frac{d\mathbb{P}}{d\mathbb{Q}_T} \cdot \mathbb{E}^{Q_T} [\mathcal{L} | R_0, T] \right] - \mathbb{E}^{Q_T}_0 [\mathcal{L}]$$

$$= \mathbb{E}^{Q_T}_0 \left[ \mathbb{E}^{Q_T} \left[ \frac{d\mathbb{P}}{d\mathbb{Q}_T} \cdot \frac{d\mathbb{Q}_T}{d\mathbb{P}} | R_0, T \right] \right] - \mathbb{E}^{Q_T}_0 [\mathcal{L}]$$

$$= 1 - \mathbb{E}^{Q_T}_0 [\mathcal{L}] .$$

Boundedness follows from $\mathbb{E}^{Q_T}_0 [\mathcal{L}] = \mathbb{E}^0_{\mathbb{P}} [\mathcal{L}^2] = ||\mathcal{L}||_{L^2_p}^2 \in L^2_p$ by assumption. Negativity from the inequality $0 \leq \mathbb{E}^{Q_T}_0 [\mathcal{L}(R_0, T)] - 1$ (from the end of Part 1.).

Part 3. We consider the conditional version of the Hansen Jagannathan bound. Introducing the price of a bond at time $t$ paying one unit currency at time $T$, $p_{0,T}$, the conditional expected gross return on the likelihood ratio contract over the riskfree investment is

$$\mathbb{E}^0_{\mathbb{P}} \left[ \frac{\mathcal{L}(R_0, T)}{p_{0,T} \mathbb{E}^0_{\mathbb{P}} [\mathcal{L}^2]} - \frac{1}{p_{0,T}} \right] .$$
The standard deviation of the gross return on the likelihood ratio contract is
\[
\frac{1}{p_{0,T} \mathbb{E}_0^P [\mathcal{L}^2]} \sqrt{\mathbb{E}_0^P [\mathcal{L}^2] - 1}.
\]

Computing the ratio of the two we obtain
\[
\frac{\mathbb{E}_0^P \left[ \frac{\mathcal{L}}{p_{0,T} \mathbb{E}_0^P [\mathcal{L}^2]} - \frac{1}{p_{0,T}} \right]}{p_{0,T} \mathbb{E}_0^P [\mathcal{L}^2]} = \frac{1 - \mathbb{E}_0^P [\mathcal{L}^2]}{\sqrt{\mathbb{E}_0^P [\mathcal{L}^2] - 1}} = -\sqrt{\mathbb{E}_0^P [\mathcal{L}^2] - 1}.
\]

This is minus the inverse Sharpe ratio of the pricing kernel.

A.2. Proof of Result 3.1. Under the assumptions made we have \( \mathcal{L} \in L^2_P \) and exponential moments exist. Hence, \( \mathcal{L} \) has a series expansion analogous to (11) (see Filipović et al., 2013, Lemma 3.1). This means that to first order the likelihood ratio is of the form \( a_0 + a_1 R_{0,T} \) where the coefficients \( a_0 \) and \( a_1 \) depend on \( Q_T \) and \( \mathbb{P} \). \( VIX^2 \) is defined as \( -2 \mathbb{E}_0^{Q_T} [R_{0,T}] \) and the expected realized variance under \( \mathbb{P} \) is \( 2 \mathbb{E}_0^P \left[ e^{R_{0,T}} - 1 - R_{0,T} \right] \).

Therefore
\[
\mathbb{E}_0^P \left[ \mathcal{L}^{(1)}(R_{0,T}) \right] - \mathbb{E}_0^{Q_T} \left[ \mathcal{L}^{(1)}(R_{0,T}) \right] \propto \left( \mathbb{E}_0^P [R_{0,T}] - \mathbb{E}_0^{Q_T} [R_{0,T}] \right)
\]
\[
\propto \mathbb{E}_0^P \left[ e^{R_{0,T}} - 1 - R_{0,T} \right] - \mathbb{E}_0^P [R_{0,T}] + \mathbb{E}_0^{Q_T} [R_{0,T}] - \mathbb{E}_0^P \left[ e^{R_{0,T}} - 1 \right]
\]
\[
= \mathbb{E}_0^P \left[ e^{R_{0,T}} - 1 - R_{0,T} \right] + \mathbb{E}_0^{Q_T} [R_{0,T}] - \mathbb{E}_0^P \left[ e^{R_{0,T}} - 1 \right] + \mathbb{E}_0^{Q_T} \left[ e^{R_{0,T}} - 1 \right].
\]

The equity premium here is represented as the payoff to a simple first moment swap. With forwards under the forward measure, the expected simple return is zero, \( \mathbb{E}_t^F \left[ e^{R_{0,T}} - 1 \right] = 0 \), different from the expected return under the risk-neutral measure which would be the risk free bond. The equity premium is therefore just the expected simple forward return under the physical measure.

\(^{15}\)cf. Neuberger (2012); Bondarenko (2010)
Appendix B. Expansion of Likelihood Ratio

By existence of exponential moments of the $Q_T$ and the $P_M$ and technical conditions on the tails of $P_M$ there exists an orthonormal basis of $L^2_{P_M}$. To compute the basis we can employ the Gram-Schmidt process reviewed below.

### B.1. $P_M$ Orthonormal Polynomials.

The orthonormal polynomials $H$ from (12) can be computed with

**Algorithm B.1** (Gram-Schmidt Process).

\[
H_0(x \mid \theta) = 1,
\]

\[
\tilde{H}_i(x \mid \theta) = x^i - \sum_{j=0}^{i-1} \int_{\mathbb{R}} \xi^j H_j(\xi \mid \theta) dP_M(\xi \mid \theta) \cdot H_j(x \mid \theta),
\]

\[
H_i(x \mid \theta) = \frac{\tilde{H}_i(x \mid \theta)}{\sqrt{\int_{\mathbb{R}} \tilde{H}_i^2(\xi \mid \theta) dP_M(\xi \mid \theta)}}
\]

Denoting the moments of $P_M$ generically by $\mu$

### B.2. Expansion Coefficients.

The coefficients in the expansion (12) are obtained through the formula

\[
c_i(\theta, Q_T) = \frac{\mathbb{E}^{Q_T}_t \left[ \tilde{H}_i(R_{t,T} \mid \theta) \right]}{\sqrt{\int_{\mathbb{R}} \tilde{H}_i^2(\xi \mid \theta) dP_M(\xi \mid \theta)}},
\]

where for twice-differentiable $f$ we have from Carr and Madan (2001)

\[
\mathbb{E}^{Q_T}_t [f(F_{T,t})] = f(F_{t,T}) + \frac{1}{P_{t,T}} \left( \int_{0}^{F_{t,T}} f''(K) P_{t,T}(K) dK + \int_{F_{t,T}}^{\infty} f''(K) C_{t,T}(K) dK \right).
\]

(31)
Appendix C. Implied Moments of the Return Distribution

For $G_{t,T} = \mathbb{E}_{t}^{Q_{t}} [g(R_{t,T})]$ we have from Carr and Madan (2001)

\[
G_{t,T} = \frac{1}{p_{t,T}} \int_{0}^{F_{t,T}} \frac{g'' \left( \log \frac{K}{F_{t,T}} \right) - g' \left( \log \frac{K}{F_{t,T}} \right)}{K^2} P_{t,T}(K) dK
+ \frac{1}{p_{t,T}} \int_{F_{t,T}}^{\infty} \frac{g'' \left( \log \frac{K}{F_{t,T}} \right) - g' \left( \log \frac{K}{F_{t,T}} \right)}{K^2} C_{t,T}(K) dK. \tag{32}
\]

For $G_{\tau,t,T} = \mathbb{E}_{\tau}^{Q_{\tau}} [g(R_{t,T})]$ the formula is

\[
G_{\tau,t,T} = g \left( \log \frac{F_{\tau,T}}{F_{t,T}} \right) + \frac{1}{p_{\tau,T}} \int_{0}^{F_{\tau,T}} \frac{g'' \left( \log \frac{K}{F_{\tau,T}} \right) - g' \left( \log \frac{K}{F_{\tau,T}} \right)}{K^2} P_{\tau,T}(K) dK
+ \frac{1}{p_{\tau,T}} \int_{F_{\tau,T}}^{\infty} \frac{g'' \left( \log \frac{K}{F_{\tau,T}} \right) - g' \left( \log \frac{K}{F_{\tau,T}} \right)}{K^2} C_{\tau,T}(K) dK, \tag{33}
\]

where $C_{t,T}(K)$ and $P_{t,T}(K)$ denote European Calls and Puts written on the spot underlying at time $t$ with maturity $T$ and strike price $K$. 