

The price of variance risk

Ian Dew-Becker, Stefano Giglio, Anh Le, and Marius Rodriguez*

December 21, 2014

Abstract

The average investor in the variance swap market is indifferent to news about future variance at horizons ranging from 1 month to 14 years. It is only purely transitory and unexpected realized variance that is priced. These results present a challenge to most structural models of the variance risk premium, such as the intertemporal CAPM, recent models with Epstein–Zin preferences and long-run risks, and models where institutional investors have value-at-risk constraints. The results also have strong implications for macro models where volatility affects investment decisions, suggesting that investors are not willing to pay to hedge shocks in expected economic uncertainty.

*Northwestern Kellogg, Chicago Booth, UNC Kenan–Flagler, and the Federal Reserve Board. The views in this paper are solely the those of the authors and do not reflect the views of the Board of Governors of the Federal Reserve System. We appreciate helpful comments from and conversations with Dante Amen-gual, Jean-Francois Bacmann, Brandon Bates, Nick Bloom, John Campbell, and Cam Harvey, and seminar participants at the Booth Econometrics Workshop, University of New South Wales, University of Sydney, University of Technology, Sydney, University of Adelaide, Northwestern Kellogg, and the 2014 Macro Finance Society Meeting. We are particularly grateful to Gerald Lim and Markit Group for providing variance swap data.

1 Introduction

The recent explosion of research on the effects of volatility in macroeconomics and finance shows that economists care about macroeconomic volatility. Investors, on the other hand, do not. We show in this paper that it is costless on average to hedge news about future volatility in aggregate stock returns; in other words investors are not willing to pay for insurance against volatility news. In recent macroeconomic models, uncertainty about the future, or expectations of high future volatility, can induce large fluctuations in the economy. But if increases in economic uncertainty can drive the economy into a recession, as in, e.g., Bloom (2009) and Gourio (2012, 2013), we would expect that investors would want to hedge those shocks. The fact that volatility shocks are unpriced thus presents a challenge to the recent macro literature on the effects of volatility shocks.

As a concrete example, consider the legislative battles over the borrowing limit of the United States in the summers of 2010 and 2011. Those periods were associated with increases in both financial measures of uncertainty, e.g. the VIX, and also the measure of policy uncertainty from Bloom, Baker, and Davis (2014)¹. Between June and July, 2011, the 1-month variance swap rate – a measure of investor expectations for S&P 500 volatility over the next month – rose from 16.26 to 25.96 percent (annualized). However, those shocks also had small effects on *realized* volatility in financial markets: annualized realized volatility in June and July, 2011, was 14.59 and 15.23 percent, respectively. The debt ceiling debate caused uncertainty about the future to be high, but did not correspond to high contemporaneous volatility.¹ **[what is the timing of the numbers here and in the footnote? When is the vix and rv calculated?]**

Those facts make the debt-ceiling shocks the exact type of shock that is studied in the recent literature. It is precisely changes in expectations of *future* uncertainty that can have strong macroeconomic effects, because they affect all forward-looking decisions. In this paper, we directly measure how much people are willing to pay to hedge shocks to expectations of future volatility. We find that those news shocks are unpriced: any investor

¹The table below reports realized volatility and the 1-month variance swap rate (nearly identical to the VIX) for June to October of 2011:

	June	July	Aug.	Sept.	Oct.
1-month variance swap	16.26	25.96	31.68	42.32	28.53
Realized volatility	14.59	15.23	47.18	28.80	29.80

Except for August, when both the variance swap rate and realized volatility rose in tandem, for all other months the changes in the two series are essentially unrelated.

Some of the volatility in the fall of 2011 was also due to uncertainty about the state of the European economy.

can buy insurance against volatility shocks for free, and therefore any investor could have freely hedged the increases in uncertainty during the debt ceiling debate.

[Could we come up with a way to replicate or price shocks to the BBD index? For example, what if we projected innovations in the index (e.g. relative to an AR) onto innovations in the VS market? We could then look at the pricing and the implied risk premium. The R^2 would also be interesting. This would help give us policy relevance and connect us to the macro literature more. On the other hand, maybe this should be reserved for another paper...]

We measure the price of variance risk using novel data on a wide range of volatility-linked assets both in the US and around the world, focusing primarily on variance swaps with maturities between 1 month and 14 years. Variance swaps are assets that pay to their owner the sum of daily squared stock market returns from their inception to maturity. They thus give direct exposure to future stock market volatility and are the most natural and direct hedge for the risks associated with increases in aggregate economic uncertainty.

The analysis of the pricing of variance swaps yields two simple but important results. First, news about future volatility is unpriced – exposure to volatility news does not earn a risk premium. Second, exposure to *realized variance* is strongly priced, with an annualized Sharpe ratio of -1.7 – five times larger than the Sharpe ratio on equities.² Since Bollerslev and Todorov (2011)? show that realized variance is priced due to its correlation with large negative jumps, we conclude that investors are willing to pay a large amount of money for protection from extreme negative shocks to the economy (which mechanically generate spikes in realized volatility), but they will not pay to hedge news that uncertainty or the probability of a disaster has changed.

The results present a challenge to a wide range of models. In macroeconomics, there is now a large literature following Bloom (2009) (who also studies the variance of aggregate stock returns) arguing that shocks to uncertainty can have important effects on the aggregate economy. If increases in future uncertainty have sufficiently important effects on the economy that they affect investor utility, though, we would expect them to carry a risk premium. The fact that they do not implies that volatility shocks are not a major driver of welfare.

From a finance perspective, Merton's (1973?) intertemporal capital asset pricing model says that assets that have high returns in periods with good news about future investment opportunities are viewed as hedges and thus earn low average returns. Since expected future volatility is a natural state variable for the investment opportunity set, the covariance of an asset's returns with shocks to future volatility should affect its expected return, but it does

²The Sharpe ratio is equal to the average excess return divided by the standard deviation. It thus gives a scale-free measure of the risk premium on an asset.

not.³

Consumption-based models with Epstein–Zin (1991)? preferences have similar predictions. Under Epstein–Zin preferences, marginal utility depends on lifetime utility, so that assets that covary positively with innovations to lifetime utility earn high average returns because they have low payouts in bad states of the world.⁴ If high expected volatility is bad for lifetime utility (either because volatility affects the path of consumption or because volatility reduces utility simply due to risk aversion), then volatility news should be priced.⁵

As a specific parameterized example with Epstein–Zin preferences, we study variance swap prices in Drechsler and Yaron’s (2011) calibrated long-run risk model. Drechsler and Yaron (2011) is a key benchmark because it is a quantitative model that can match a wide range of features of the dynamics of consumption growth, stock returns, and volatility. While the model represents a major innovation in being able to both generate a large variance risk premium (the risk premium of short-term realized volatility shocks) and match results about the predictability of market returns, we find that its implications for the term structure of variance swap prices and returns are strongly at odds with the data: as one would expect, it predicts that shocks to future expected volatility should be strongly priced, counter to what we observe empirically.

We obtain similar results in Wachter’s (2013)? model of time-varying disaster risk with Epstein–Zin preferences: the combination of predictability in the long-run probability of disaster and Epstein–Zin preferences results in a counterfactually high price for insurance against shocks to expected future volatility relative to current volatility. In both Wachter (2013) and Drechsler and Yaron (2011), Sharpe ratios earned by claims on future variance from 3 months to 14 years ahead are similar to those earned by claims to realized variance over the next month, whereas in the data the Sharpe ratios are all near zero (or positive) for claims to variance more than two months in the future. So both models fail to match our key stylized fact that only very short-term variance claims earn large negative Sharpe ratios.⁶

³Recently, Campbell et al. (2014)? and Bansal et al. (2013)? estimate an ICAPM model with stochastic volatility and find that shocks to expected volatility (and especially long-run volatility) are priced in the cross-section of returns of equities and other asset classes. Although the focus on their paper is not the variance swap market, Campbell et al. (2014) test their specification of the ICAPM model also on straddle returns and synthetic volatility claims, and find that the model manages to explain only part of the returns on these securities. This suggests that the model is missing some high-frequency features of the volatility market.

⁴This is true in the most common calibrations with a preference for early resolution of uncertainty. When investors prefer a late resolution of uncertainty the risk prices are reversed.

⁵Also see Branger and Völkert (2010?) and Zhou and Zhu (2012?) for discussions. Barras and Malkhozov (2014)? study the determinants of changes in the variance risk premium over time.

⁶Similar problems with matching term structures of Sharpe ratios in structural models have been studied

More positively, we show that Gabaix’s (2012)? model of rare disasters can match the stylized fact that Sharpe ratios on variance claims fall to zero rapidly with maturity. In his model, the probability of a disaster is constant, but the exposure of the stock market – its expected decline if a disaster occurs – varies over time. The realization of a disaster is inevitably a state with high realized volatility (since if returns are highly negative, squared daily returns will mechanically be high), so variance swaps provide a direct hedge against the occurrence of a disaster, meaning they earn a large negative Sharpe ratio. But since changes in the exposure of the stock market to consumption disasters, which drive expected future return variance, are uncorrelated with the *current level* of consumption, they are not priced shocks. The model is thus able to simultaneously generate a large negative premium on realized stock return variance and zero premium on news about future variance, just like in the data.⁷ That said, Gabaix’s (2012) model is not a complete quantitative description of financial markets; we simply view it as giving a set of sufficient conditions a model must satisfy to match the behavior of the variance swaps.

Our work is related to three main strands of the literature. First, there is the recent work in macroeconomics on the consequences of shocks to volatility, such as Bloom (2009)?, Bloom et al. (2014), Christiano, Motto, and Rostagno (2014)?, Fernandez-Villaverde et al. (2011)?, and Gourio (2012, 2013)??. We argue that if shocks to volatility are important to the macroeconomy, then investors should be willing to pay to hedge them. The lack of a risk premium on volatility news thus argues against theories in which aggregate volatility news is a major driver of business cycles.

Second, we build on the consumption-based asset pricing literature that has recently focused on the pricing of volatility, including Bansal and Yaron (2004), Drechsler and Yaron (2011), Wachter (2013), Campbell et al. (2014)?, and Bansal et al. (2013)?. We argue that consumption-based models with Epstein–Zin preferences are unlikely to explain the pricing of volatility claims.

Finally, there is a large extant literature studying the pricing of volatility in financial markets.⁸ Most closely related to us is a small number of recent papers with data on variance

in the context of claims to aggregate market dividends by van Binsbergen, Brandt, and Koijen (2012)?. Our results thus support and complement theirs in a novel context. Our paper also relates to a large literature that looks at derivative markets to learn about general equilibrium asset pricing models, for example Backus, Chernov and Martin (2011)? and Martin (2013, 2014)??.

⁷An alternative possibility is that the variance market is segmented from other markets, as in, e.g., Gabaix, Krishnamurthy, and Vigneron (2007)?. In that case, the pricing of risks might not be integrated between the variance market and other markets. We show, however, that our results hold not only with variance swaps, but also in VIX futures and in the options market, which is large, liquid, and integrated with equity markets, making it less likely that our results are idiosyncratic.

⁸A number of papers study the pricing of volatility in options markets, e.g. Jackwerth and Rubinstein (1996)?, Coval and Shumway (2001)?, Bakshi and Kapadia (2003)?, Broadie, Chernov and Johannes (2009)?,

swaps with maturities from two to 24 months, including Egloff, Leippold, and Wu (2010) and Aït-Sahalia, Karaman, and Mancini (2014)?, who study no-arbitrage term structure models. The pricing models we estimate are less technically sophisticated than that of Aït-Sahalia, Karaman, and Mancini (2014), but we complement and advance their work in three ways. First, we examine a vast and novel range of data sources. For S&P 500 variance swaps, our panel includes data at both shorter and longer maturities than in previous studies – from one month to 14 years. The one-month maturity is important for giving a claim to shorter-term realized variance, which is what we find is actually priced. Having data at very long horizons is important for testing models, like Epstein–Zin preferences, in which expectations at very long horizons are the main drivers of asset prices. In addition, we are the first to examine the term structure of variance swaps for major international indexes, as well as for the term structure of the VIX obtained from options on those indexes. We are thus able to confirm that our results hold across a far wider range of markets, maturities, and time periods than previously studied.

Our second contribution to the previous term structure literature is that rather than working exclusively within the context of a particular no-arbitrage pricing model for the term structure of variance claims, we derive from the data more general and model-independent facts about pricing in this market. Our pricing results can be directly compared against the pricing implications of different structural economic models, which would be more difficult if the pricing results were only derived within a specific no-arbitrage framework. Our key result, that purely transitory realized variance is priced while innovations to expectations are not, can be obtained from a simple reduced-form analysis and in data both for the United States and other countries. Nevertheless, we also confirm our results in a more formal no-arbitrage setting, whose main advantage is to yield much more precise estimates of risk prices.

Our third and most important contribution is to explore the ability of variance swaps to test structural economic models. Our theoretical analysis leads us to the conclusion that the empirical facts in the variance swap market are most consistent with a model in which variance swaps are used to hedge the realization of market crashes and in which variation in expected future stock market volatility is not priced by investors, counter to the predictions of most standard asset pricing and macroeconomic models.

The remainder of the paper is organized as follows. Section 2 describes the novel datasets

Christofferson, Jacobs, Ornathanalai, and Wang (2008)?, and Kelly, Pastor, and Veronesi (2014)?. Lu and Zhu (2010)? and Mencia and Sentana (2013)? study VIX futures markets, while Bakshi, Panayotov, and Skoulakis (2011)? show how to construct forward claims on variance with portfolios of options. In the Treasury bond market, Cieslak and Povala (2014)? find, similar to us, that short-run volatility is more strongly priced than long-run volatility. See also Amengual and Xiu (2014)? for an important recent study of jumps in volatility.

we obtain for variance swap prices. Section 3 reports unconditional means for variance swap prices and returns, which demonstrate our results in their simplest form. Section 4 analyzes the cross-sectional and time-series behavior of variance swap prices and returns more formally in standard asset pricing frameworks. In section 5, we discuss what structural general-equilibrium models can fit the data. We calibrate three leading models from the literature, comparing them to our data, showing that only one matches the key stylized facts. Section 6 concludes.

2 The data

This section discusses various ways that an investor can obtain exposure to volatility. They are all obviously closely related. We have data on each of the major markets, both in the US and internationally.

2.1 Variance swaps

We focus primarily on variance swaps. Variance swaps are contracts in which one party pays a fixed amount at maturity, which we refer to as price of the variance swap, in exchange for a payment equal to the sum of squared daily log returns of the underlying occurring until maturity. In this paper, the underlying is the S&P 500 index unless otherwise specified. The payment at expiration of a variance swap initiated at time τ and with maturity m is

$$Payoff f_{\tau}^m = \sum_{j=\tau+1}^{\tau+m} r_j^2 - VS_{\tau}^m \quad (1)$$

where time here is indicated in days, r_j is the log return on the underlying on date j , and VS_{τ}^m is the price on date τ of an m -day variance swap. We focus on variance swaps because they give pure exposure to variance, their payoffs are transparent and easy to understand, they have a relatively long time-series, and they are relatively liquid.

Our main analysis focuses on two proprietary datasets of quoted prices for S&P 500 variance swaps.⁹ Dataset 1 contains monthly variance swap prices for contracts expiring in 1, 2, 3, 6, 12, and 24 months, and includes data from December, 1995, to October, 2013. Dataset 2 contains data on variance swaps with expirations that are fixed in calendar time,

⁹Both datasets were obtained from industry sources. Dataset 2 is obtained from Markit Totem, and reports averages of quotes obtained from dealers in the variance swap market. Since the prices we observe are a composite of quotes from many different dealers (on average 11), the quality of this dataset is very high, and comparable to that of the widely used CDS dataset from Markit.

instead of fixed maturities. Common maturities are clustered around 1, 3, and 6 months, and 1, 2, 3, 5, 10, and 14 years. Dataset 2 contains prices of contracts maturities up to five years starting in September, 2006, and up to 14 years starting in August, 2007, and runs up to February, 2014. We apply spline interpolation to each dataset to obtain the prices of variance swaps with standardized maturities covering all months between 1 and 12 months for Dataset 1 and between 1 and 120 months for Dataset 2 (though in the no-arbitrage model below we use the original price data without interpolation).¹⁰

Both variance swap datasets are novel to the literature. Variance swap data with maturities up to 24 months as in Dataset 1 has been used before (Egloff, Leippold, and Wu, 2010?, Ait-Sahalia, Karaman, and Mancini, 2014?, and Amengual and Xiu, 2014?), but the shortest maturity previous studies observed was two months. We show that the one-month variance swap is special in this market because it is the exclusive claim to next month’s realized variance, which is the *only* risk priced in this market. Observing the one-month variance swap is critical for precisely measuring the price of realized-variance risk.

This is also the first paper to observe and use variance swap data with maturity longer than two years. Since Epstein–Zin preferences imply that it is the very low-frequency components of volatility that should be priced (Branger and Volkert, ?2010; Dew-Becker and Giglio, 2014?), having claims with very long maturities is important for effectively testing the central predictions of Epstein–Zin preferences.

The variance swap market is large: the notional value of outstanding variance swaps at the end of 2013 was \$4 billion of vega.¹¹ This means that an increase in annualized realized volatility of one percentage point induces total payments of \$4 billion. This market is thus small relative to the aggregate stock market, but it is non-trivial economically. Current bid/ask spreads in the variance swap market average 1 to 3 percent, depending on maturity and trade size.

Table ?? shows the total volume (in notional vega) for all transactions between March 2013 and June 2014. In little more than a year, the variance swap market saw \$7.2 billion of notional vega traded. Only 11 percent of the volume was traded in short maturity contracts (1-3 months); the bulk of the transactions occurred for maturities between 6 months and 5 years, and the median maturity was 12 months.

To check the accuracy of the quoted prices that we obtained, we compare them to those

¹⁰For the times and maturities for which we have both datasets, the prices are effectively identical: the correlations between the two datasets are never below 0.996. We will also show below that the prices are well explained by only two principal components, suggesting that interpolation should accurately recover prices.

¹¹See the Commodity Futures Trading Commission’s (CFTC) weekly swap report. The values reported by the CFTC are consistent with data obtained from the Depository Trust & Clearing Corporation that we discuss below.

reported for actual trades by Depository Trust & Clearing Corporation (DTCC), which has collected data on all trades of variance swaps in the US since 2013.¹² Appendix Figure ?? shows the distribution of the percentage difference between our quotes and the transaction prices for different maturity baskets. Quotes and transaction prices are in most cases very close, with the median absolute percentage difference across all maturities approximately 1 percent.¹³

[What is the median absolute difference? I don't think the number above is right. What is the standard deviation of the errors across all the claims?]

In addition to the prices of S&P 500 variance swaps, we also obtained prices for variance swaps in 2013 and 2014 for the FTSE 100 (UK), Euro Stoxx 50 (Europe), and DAX (Germany) indexes. This is the first paper to examine volatility claims in international markets and we show that our main results are consistent globally.

2.2 Options

The VIX index is constructed using all available out-of-the money options and, if the underlying follows a diffusion, measures the risk-neutral expectation of variance integrated over the life of the options (Carr and Wu (2009)?). The VIX is usually reported for a 30-day maturity, but the formulas are valid at any horizon.

While the VIX has the drawback that it requires extra assumptions in order to represent a claim on volatility (i.e. no jumps) it has the advantage that it is calculated based on an extraordinarily deep market. Options are traded in numerous venues, have notional values outstanding of trillions of dollars, and have been thoroughly studied.¹⁴ Since options are exchange-traded, they involve no counterparty risk, so we can use them to whether our results for variance swaps are affected by counterparty risk.

We construct VIX-type portfolios for the S&P 500, FTSE 100, Euro Stoxx 50, DAX, and

¹²DTCC was the only swap data repository registered under the Dodd–Frank act to collect data on variance swaps in 2013. The Dodd–Frank act requires that all swaps be reported to a registered data repository.

¹³Since variance swaps are traded over the counter, it is possible that counterparty risk could influence their prices. Given that variance swaps are standardized contracts covered by the ISDA Master Agreement, their margining follows standard procedures: an initial margin is posted by both parties, and variational margin is exchanged regularly depending on the value of the position. The residual counterparty risk in these contracts depends on the possibility of jumps in the value of the contracts between exchanges of collateral, and is therefore only a material issue when returns have high skewness and kurtosis at short horizons. As we will discuss later, only short-term variance swaps have payoffs that are far from Gaussian, and are therefore exposed to counterparty risk, and we argue below that for these contracts counterparty risk would push *against* the results we observe. We conclude that if counterparty risk was indeed priced by market participants, accounting for it would in fact make our results stronger.

¹⁴Even in 1990, Vijh (1990)? noted that the CBOE was highly liquid and displayed little evidence of price impact for large trades.

CAC 40 indexes using data from Optionmetrics. We confirm our main results by showing that term structures and returns obtained from investments in options are similar to those obtained from variance swaps.¹⁵

2.3 VIX futures

Futures have been traded on the VIX since 2004. The VIX futures market is significantly smaller than the variance swap market, with current outstanding notional vega of roughly \$500 million.¹⁶ Bid/ask spreads are smaller than what we observe in the variance swap market, at roughly 0.1 percent, but as the market is smaller, we would expect price impact to be larger (and market participants claim that it is). We collected data on VIX futures prices from Bloomberg since their inception and show below that they yield nearly identical results to variance swaps.

More recently, a market has developed in exchange-traded notes and funds available to retail investors that are linked to VIX futures prices. These funds currently have an aggregate notional exposure to the VIX of roughly \$5 billion, making them comparable in size to the variance swap market.

3 The term structure of variance claims

In this section we study average prices and returns of variance swaps. The key result that emerges is that only very short-duration variance claims earn a risk premium.

3.1 Variance Swap Prices

The shortest maturity variance swap we consistently observe has a maturity of one month, so we treat a month as the fundamental period of observation. We define RV_t to be realized variance ($\sum r_j^2$) during month t . The subscript from here forward always indexes months, rather than days

Given a risk-neutral (pricing) measure Q , the price of an n -month variance swap at the

¹⁵Recently, Boguth et al. (2012a, b)?? argue that returns measured on options portfolios can be substantially biased by noise, one potential source of which is the bid/ask spread. The majority of our results pertain directly to prices of volatility claims, as opposed to their returns, meaning that the issues noted by Boguth et al. are unlikely to affect our analysis. Furthermore, when we analyze returns, the portfolios are not levered to the degree that Boguth et al. argue causes biases in results.

¹⁶According to the CBOE futures exchange market statistics. See: <http://cfe.cboe.com/Data/HistoricalData.aspx>

end of month t , VS_t^n , is

$$VS_t^n = E_t^Q \left[\sum_{j=1}^n RV_{t+j} \right] \quad (2)$$

where E_t^Q denotes the mathematical expectation under the risk-neutral measure conditional on information available at the end of month t .¹⁷

Since an n -month variance swap is a claim to the sum of realized variance over months $t + 1$ to $t + n$, it is straightforward to compute prices of zero-coupon claims on realized variance. Specifically, we define an n -month zero-coupon variance claim as an asset with a payoff equal to realized volatility in month $t + n$. The absence of arbitrage implies

$$Z_t^n \equiv E_t^Q [RV_{t+n}] \quad (3)$$

$$= VS_t^n - VS_t^{n-1} \quad (4)$$

Z_t^n represents the market's risk-neutral expectation of realized variance n months in the future. We use the natural convention that

$$Z_t^0 = RV_t \quad (5)$$

so that Z_t^0 is the variance realized *during* the current month t . A one-month zero-coupon variance claim is exactly equivalent to a one-month variance swap, $Z_t^1 = VS_t^1$.

Figure ?? plots the time series of zero-coupon variance claim prices for maturities between one month and ten years. The figure shows all series in annualized percentage volatility units, rather than variance units: $100 \times \sqrt{12 \times Z_t^n}$ instead of Z_t^n . It also plots annualized realized volatility, $100 \times \sqrt{12 \times Z_t^0}$, in each panel. The top panel plots zero-coupon variance claim prices for maturities below one year, while the bottom panel focuses on maturities longer than one year.

The term structure of variance claim prices is usually weakly upward sloping. In times of distress, though, such as during the financial crisis of 2008, the short end of the curve spikes, temporarily inverting the term structure. Volatility obviously was not going to continue at crisis levels, so markets priced variance swaps with the expectation that it would fall in the future.

¹⁷In the absence of arbitrage, there exists a probability measure Q such that the price of an asset with payoff X_{t+1} is $\frac{1}{R_{f,t+1}} E_t^Q [X_{t+1}]$, where $R_{f,t+1}$ is the risk-free interest rate. Under power utility, for example, we have $E_t^Q [X_{t+1}] = E^P \left[\frac{(C_{t+1}/C_t)^{-\rho}}{E_t^P [(C_{t+1}/C_t)^{-\rho}]} X_{t+1} \right]$, where ρ is the coefficient of relative risk aversion, C is consumption, and P is the physical probability measure. The price of a variance swap does not involve the interest rate because money only changes hands at the maturity of the contract.

Figure ?? reports the average term structure of zero-coupon variance claims for two different subperiods – 2008–2014, a relatively short sample for which we have data for longer maturities, is in the top panel, while the full sample, 1996–2013, is in the bottom panel. The term structure of zero-coupon variance claim prices is upward sloping on average, but the figure also shows that it is concave, flattening out very quickly as the maturity increases.

The average zero-coupon term structures in Figure ?? provide the first indication that the compensation for bearing risk associated with news about future volatility is small in this market. The return on holding a zero-coupon variance claim for a single month is $\frac{Z_{t+1}^{n-1} - Z_t^n}{Z_t^n}$. The average return is therefore closely related to the slope of the variance term structure. If the variance term structure is upward sloping then zero-coupon claims will have negative average returns, implying that it is costly to buy insurance against increases in future expected volatility. The fact that it is steep at short horizons and flat at long horizons is a simple way to see that it is only the claims to variance in the very near future that earn significant negative returns.

3.2 Returns on zero-coupon variance claims

We now study the monthly returns on zero-coupon variance claims. The return on an n -month claim corresponds to a strategy that buys the n -month claim and sells it one month later as an $(n - 1)$ -month claim, reinvesting then again in new n -month zero-coupon variance claims. We define the excess return of an n -period variance claim following Gorton, Hayashi, and Rouwenhorst (2013)?¹⁸

$$R_{t+1}^n = \frac{Z_{t+1}^{n-1} - Z_t^n}{Z_t^n} \quad (6)$$

Given the definition that $Z_t^0 = RV_t$, the return on a one-month claim, R_{t+1}^1 is simply the percentage return on a one-month variance swap. We focus here on the returns for maturities of one to 12 months, for which we have data since 1995. All the results extend to higher maturities in the shorter sample.

Table ?? reports descriptive statistics for our panel of monthly returns. Only the average returns for the first and the second zero-coupon claims are negative, while all the others are zero or slightly positive. Return volatilities are also much higher at short maturities, though the long end still displays significant variability – returns on the 12-month zero-coupon claim

¹⁸Note that $Z_{t+1}^{n-1} - Z_t^n$ is also an excess return on a portfolio since no money changes hands at the inception of a variance swap contract. Following Gorton, Hayashi, and Rouwenhorst (2013), we scale the return by the price of the variance claim bought. This is the natural scaling if the amount of risk scales proportionally with the price, as in Cox, Ingersoll, and Ross (1985)?. We have reproduced all of our analysis using the unscaled excess return $Z_{t+1}^{n-1} - Z_t^n$ as well and confirmed that all the results hold in that case.

have an annual standard deviation of 17 percent, which indicates that markets' expectations of 12-month volatility fluctuates significantly over time.

Finally, note that only very short-term returns have high skewness and kurtosis. A buyer of short-term variance swaps is therefore potentially exposed to counterparty risk if realized variance spikes and the counterparty defaults. This should induce her to pay less for the insurance, i.e. we should expect the average return to be *less negative*. Therefore, the presence of counterparty risk on the short end of the term structure would bias our estimate towards not finding the large negative expected returns that we instead find. On the other hand, returns on longer-maturity zero-coupon claims have much lower skewness and kurtosis, which indicates that counterparty risk is substantially less relevant for longer maturities.

Given the different volatilities of the returns at different ends of the term structure, it is perhaps more informative to examine Sharpe ratios (average excess returns scaled by standard deviations), which measure compensation earned per unit of risk. Figure ?? shows the annual Sharpe ratios of the 12 zero-coupon claims. The Sharpe ratios are negative for the one- and two-month claims (at -1.4 and -0.5, respectively), but all other Sharpe ratios are insignificantly different from zero.

The results at the short end of the curve indicate that investors are willing to pay a large premium to hedge realized volatility. What is new and surprising in this picture is the fact that agents are *not* willing to pay to hedge any innovations in *expected* volatility, even two or three months ahead. A claim to volatility at a horizon beyond one month is purely exposed to news about future volatility: its return corresponds exactly to the change in expectations about volatility at its maturity. Pure news about future expected volatility will therefore affect its return, whereas purely transitory shocks to volatility that disappear before its maturity will not affect it at all. Our results therefore show that news about future volatility commands a small to zero risk premium.¹⁹

3.3 Evidence from other markets

The results for variance swaps can also be confirmed in the options market. We exploit the well-known fact that if the S&P 500 follows a diffusion, a variance swap can be replicated by a portfolio of options with the same maturity.²⁰ The term structure of synthetic zero-

¹⁹The declining term structure of Sharpe ratios on short positions in volatility is consistent with the finding of van Binsbergen, Brandt, and Koijen (2012?) that Sharpe ratios on claims to dividends decline with maturity, and that of Duffee (2011?) that Sharpe ratios on Treasury bonds decline with maturity.

²⁰The most famous use of that result is in the construction of the VIX index, which uses 1-month options, and corresponds to the price of a 1-month variance swap. If returns are not a diffusion, the corresponding portfolio of options will not perfectly replicate the payoff of the variance swap. While the difference in prices of variance swaps and option-based synthetic contracts is economically informative, its magnitude is

coupon variance claim prices constructed from options should then align well with the term structure of actual variance swap prices. The appendix reports details of the construction of the synthetic variance swap prices.

Figure ?? shows the term structure and Sharpe ratios of zero-coupon variance claims obtained from the variance swap data compared to the synthetic claims for maturities up to 1 year. While the curves obtained using options data seem noisier, they curves deliver the same message: the volatility term structure is extremely steep at the very short end but quickly flattens out for maturities above two months, and Sharpe ratios rapidly approach zero as the maturity passes two months.²¹ Appendix Figure ?? shows that we obtain similar results with VIX futures.

Figure ?? shows that our results also extend to international markets. Figure ?? plots average term structures obtained from both variance swaps and synthetic option-based variance claims for the Euro Stoxx 50, FTSE 100, CAC 40 and DAX indexes.²² To ease the comparison across markets, in this figure we plot the term structures relative to the prices of the respective 2-month claims, so that all the curves are equal to 1 at the two-month maturity.

Both panels of the figure show that the international term structures have an average shape that closely resembles the one observed for the US (the solid line in both panels), demonstrating that our results using US variance swaps extend to the international markets.²³

4 Asset pricing

We now formally examine the pricing of risks in the variance market.

4.1 Reduced-form estimates

We begin by exhibiting our main pricing result in a simple reduced-form setting: investors pay to hedge the immediate realized volatility but not shocks to expected volatility. To test

far smaller than the differences in prices of zero-coupon volatility claims across maturities (Bollerslev and Todorov (2011)? and Ait-Sahalia et al. (2014)).

²¹Given the high liquidity of the options market, we might have expected option-based oprtfolios to be less noisy. However, the synthetic variance portfolios load heavily on options very far out of the money where liquidity is relatively low. This demonstrates another advantage of studying variance swaps instead of options.

²²We do not plot Sharpe ratios for these markets because the data is of relatively poor quality, and the series of returns are very noisy.

²³In the appendix (Figure ??) we also confirm that for the indexes for which we have both variance swap prices and synthetic prices obtained from options, the two curves align well.

that claim, we need to disentangle shocks to realized variance from shocks to expectations of future volatility. This subsection focuses just on the returns of the variance claims with maturity of 12 months or less since they require less interpolation; all the economically interesting results are clearly visible in this maturity range.

4.1.1 Extracting innovations

As usual in the term structure literature, we begin by extracting principal components from the term structure of zero-coupon variance claims. The first factor explains 97.1 percent of the variation in the term structure and the second explains an additional 2.7 percent. The loadings of the variance swaps on the factors are plotted in the top panel of Figure ??, while the time series of the factors are shown in the bottom panel of the figure. The first factor captures the level of the term structure, while the second measures the slope. As we would expect, during times of crisis, the slope turns negative. The level factor captures the longer-term trend in volatility and clearly reverts to its mean more slowly.

The two factors explain 99.9 percent of the variation in variance swap prices and thus encode essentially all the information contained in variance swap prices. So if we find that the shocks to both factors are unpriced, then that means that no forward-looking information in the term structure, whether it is driven by expectations for volatility or risk premia, is priced.

To extract shocks to variance and expectations, we estimate a first-order vector autoregression (VAR) with the two principal components and realized variance (RV). Including RV in the VAR allows us to separately identify shocks to the term structure of variance swaps and transitory shocks to realized volatility. The three estimated innovations are positively correlated: the correlation between RV and level shocks is 0.7, and that between RV and slope shocks is 0.6.

We rotate the three shocks using a Cholesky factorization where the first shock affects all three variables, the second affects only the slope and RV , and the third shock affects only RV . We will therefore refer to the third shock as the pure RV shock. The pure RV shock allows us to measure the price of risk for a shock that has only a transitory effect on realized variance and no effect on the term structure of variance swap prices, while the other two rotated shocks affect both current realized variance and also expectations of future variance. Impulse response functions are reported in Appendix Figure ??.

4.1.2 Risk prices

We estimate risk prices for the three shocks using the Fama–MacBeth (1973)? procedure on 1- to 12-month zero-coupon variance claims.²⁴ The top panel of Table ?? reports the loadings of each variance swap return on the three orthogonalized shocks. Short-maturity variance swaps are exposed to all three shocks with the expected signs. The higher maturities are mostly exposed to the level and slope shocks, with essentially no exposure to the pure *RV* shock.

The bottom panel of Table ?? reports the estimated annualized risk prices. Of the three shocks, only the pure *RV* shock has a statistically significant risk price. The risk price is also economically highly significant: it implies that an asset that was exposed only to the pure *RV* shock would earn an annualized Sharpe ratio of -2.72. Since the three shocks all have the same standard deviation, the magnitudes of the risk prices are directly comparable. Those for shocks 1 and 2 are five to eight times smaller than that for the pure *RV* shock, and thus economically far less important.

[Can we do a GRS test or something here to claim that the model fits?]

Table ?? thus shows that investors do not price shocks to the level and slope, but they accept large negative returns to hedge transitory *RV* shocks. No forward-looking information about volatility is priced.

4.1.3 Controlling for the market return

One possible explanation for why realized variance is priced is that it provides a good hedge for aggregate market shocks. To test that possibility, we add the market return as an additional factor in the estimation.²⁵ The first column of Table ?? shows that indeed the zero-coupon volatility claims are heavily exposed to the market return. But when the pure *RV* shock is included, the market return is no longer significantly priced. The R^2 of the model for the cross-section of average returns also rises from 37.7 to 99.7 percent when the pure *RV* shock is included.

[Again, it would be nice to have a GRS statistic]

²⁴The results are robust to estimating the risk prices using one- and two-step GMM.

²⁵We add the market return as a test asset to impose discipline on its risk premium. For readability and to ensure that the risk premium on the market is matched relatively closely, we increase the weight on the market return by a factor of 12 as a test asset in our cross-sectional tests. That way, the market return carries as much weight in the pricing tests as do all the variance claims combined. The market factor, though, is still the monthly market return, as are all our zero-coupon variance returns.

4.2 The predictability of volatility

Since the key result of the paper concerns the pricing of volatility shocks at different horizons, a natural question is how much news there actually is about future volatility. Perhaps the reason that we do not estimate significant risk prices for volatility news is that there simply is not much news (which would increase standard errors, though the estimates would still be unbiased). We would first note that if there is not much news, then the macro literature showing that volatility news can drive the business cycle would seem irrelevant. Second, though, we now show that there is in fact substantial predictability of future volatility.

First, a large literature has shown that realized volatility is strongly predictable at horizons of a few months using high-frequency data, and that univariate and multivariate predictability extends to longer horizons as well.²⁶ Building on that literature, we report in Table ?? R²s from predictive regressions of realized volatility at different frequencies and horizons. The first pair of columns focuses on forecasts of monthly realized variance, while the second pair repeats the exercise at the annual frequency. The R²s for monthly volatility range from 45 percent at the 1-month horizon to 20 percent at the 12-month horizon. In predicting annual volatility, R²s range between 56 and 21 percent for horizons of 1 to 10 years.

The third pair of columns in Table ?? reports, as a comparison, the results of forecasts of dividend growth.²⁷ R²s for dividend growth are never higher than 9 percent. So in the context of financial markets, there is an economically large amount of predictability of volatility. The appendix takes an extra step beyond Table ?? and shows, using Fama and Bliss (1987) and Campbell and Shiller (1991) regressions, that nearly all the variation in variance swap prices is actually due to variations in expected volatility, rather than risk premia. We thus conclude that investors' expectations of volatility in fact vary substantially over time.

4.3 A no-arbitrage model

In this section, we extend the pricing results reported above by considering a more formal estimation. We analyze a standard no-arbitrage term structure model for variance swaps.

²⁶See for example Andersen et al. (2003)?, Ait-Sahalia and Mancini (2006), Bandi, Russell and Yang (2008)?, and Brownlees, Engle and Kelly (2011)?. Recently, Campbell et al. (2014) focus explicitly on longer horizons (up to 10 years) and show evidence of predictability of realized volatility in a multivariate setting: in particular, they show that both the aggregate price-earnings ratio and the BAA-AAA default spread are useful predictors of long-run volatility.

²⁷We compare predictability of volatility to that of dividends since realized variance in each month is the stochastic payment of the variance swap contract in that month.

The model delivers implications strongly supportive of our reduced-form results. Because the no-arbitrage model uses the prices of the variance swaps, rather than just their returns, and because it uses a full no-arbitrage structure, it is able to obtain much more precise estimates of risk prices. We show that not only are the risk prices on the level and slope factors statistically insignificant, but they are also economically small.

The no-arbitrage model has three additional advantages over the reduced-form analysis: it explicitly allows for time-variation in the volatility of shocks to the economy and risk prices, the standard errors for the risk prices take into account uncertainty about the dynamics of the economy (through the VAR), and it links us more directly to the previous literature. Furthermore, because the inputs to the estimation of the no-arbitrage model are the observed variance swap prices rather than monthly returns, the results in this section do not rely on any interpolation and we can simultaneously use the full time series from 1996 to 2013 and every maturity from one month to 14 years.

4.3.1 Risk-neutral dynamics

As above, we assume that the term structure of variance swaps is governed by a bivariate state vector $(s_t^2, l_t^2)'$. Rather than state the factors as a level and slope, we now treat them as a short- and a long-term component, which will aid in the estimation process. s_t^2 is the one-month variance swap price: $s_t^2 = E_t^Q [RV_{t+1}]$. The other state variable, l_t^2 , governs the central tendency of s_t^2 .

We begin by specifying the conditional risk-neutral mean of the states,

$$E_t^Q \left[\begin{pmatrix} s_{t+1}^2 \\ l_{t+1}^2 \\ RV_{t+1} \end{pmatrix} \right] = \begin{pmatrix} \rho_s^Q & 1 - \rho_s^Q & 0 \\ 0 & \rho_l^Q & 0 \\ 1 & 0 & 0 \end{pmatrix} \begin{pmatrix} s_t^2 \\ l_t^2 \\ RV_t \end{pmatrix} + \begin{pmatrix} 0 \\ v_l^Q \\ 0 \end{pmatrix} \quad (7)$$

where v_l^Q is a constant to be estimated which captures the unconditional mean of realized variance. l_t^2 can be viewed as the risk-neutral trend of s_t^2 . The first two rows of (7) are the discrete-time counterpart to the standard continuous-time setup in the literature, e.g. Egloff, Leippold, and Wu (2010) and Ait-Sahalia, Karaman, and Mancini (2014).²⁸ We diverge from Egloff, Leippold, and Wu (2010) and Ait-Sahalia, Karaman, and Mancini (2014) in explicitly specifying a separate process for realized variance, noting that it is not spanned by the other shocks. The specification of a separate shock to RV_{t+1} allows us to ask how shocks to both

²⁸For admissibility, we require that $0 < \rho_s^Q < 1$, $\rho_l^Q > 0$, and $v_l^Q > 0$. These restrictions ensure that risk-neutral forecasts of s_t^2 and l_t^2 , hence variance swap prices at various maturities, are strictly positive.

realized variance and the term structure factors are priced.²⁹

Given the assumption that $s_t^2 = E_t^Q [RV_{t+1}]$, the price of an n -period variance swap VS_t^n is

$$VS_t^n = E_t^Q \left[\sum_{i=1}^n RV_{t+i} \right] = E_t^Q \left[\sum_{i=1}^n s_{t+i-1}^2 \right] \quad (8)$$

which can be computed by applying (7) repeatedly, and which implies that VS_t^n is affine in s_t^2 and l_t^2 for any maturity.

4.3.2 Physical dynamics and risk prices

Define $X_t \equiv (s_t^2, l_t^2, RV_t)'$. We assume that X follows a VAR(1) under the physical measure:³⁰

$$\begin{pmatrix} s_{t+1}^2 \\ l_{t+1}^2 \\ RV_{t+1} \end{pmatrix} = \begin{pmatrix} 0 \\ v_l^Q \\ 0 \end{pmatrix} + \begin{pmatrix} \rho_s & 1 - \rho_s^Q & 0 \\ 0 & \rho_l & 0 \\ \rho_{s,RV} & 0 & 0 \end{pmatrix} \begin{pmatrix} s_t^2 \\ l_t^2 \\ RV_t \end{pmatrix} + \varepsilon_{t+1} \quad (9)$$

$$\varepsilon_{t+1} \sim N \left(\begin{pmatrix} 0 \\ 0 \\ 0 \end{pmatrix}, V_t(X_{t+1}) \right) \quad (10)$$

In our main results, we follow Egloff, Leippold, and Wu (2010) and Ait-Sahalia, Karaman, and Mancini (2014) and assume that the market prices of risk are proportional to the states, so that the log SDF, m_{t+1} , is

$$m_{t+1} - E_t[m_{t+1}] = \Lambda_t' V_t(X_{t+1})^{-1/2} \varepsilon_{t+1} \quad (11)$$

$$\text{where } \Lambda_t = \begin{pmatrix} \lambda_s s_t \\ \lambda_l l_t \\ \lambda_{RV} s_t \end{pmatrix} \quad (12)$$

where the superscript $^{1/2}$ indicates a lower triangular Cholesky decomposition. The term $V_t(X_{t+1})^{-1/2}$ standardizes and orthogonalizes the shocks ε_{t+1} . Λ_t thus represents the price

²⁹From a continuous-time perspective, it is not completely obvious how to think about a "shock" to realized variance that is completely transitory. There are two standard interpretations. One is that the innovation in RV_{t+1} represents the occurrence of jumps in the S&P 500 price. Alternatively, there could be a component of the volatility of the diffusive component of the index that has shocks that last less than one month. At some point, the practical difference between a jump and an extremely short-lived change in diffusive volatility is not obvious. The key feature of the specification is simply that there are shocks to the payout of variance swaps that are orthogonal to both past and future information contained in the term structure.

³⁰Admissibility requires that v_l^Q and the feedback matrix in (9) be non-negative, which ensures that the forecasts of X_t , and hence future volatility, be strictly positive.

of exposure to a unit standard deviation shock to each component of X_{t+1} .

To maintain the affine structure of the model, we need the product $V_t(X_{t+1})^{1/2}\Lambda_t$ to be affine in X_t . The specification for Λ_t in (12) is therefore typically accompanied by a structure for the conditional variance similar to that of Cox, Ingersoll, and Ross (1985)?,

$$V_t(X_{t+1}) = \begin{pmatrix} \sigma_s^2 s_t^2 & 0 & \sigma_{s,RV} s_t^2 \\ 0 & \sigma_l^2 l_t^2 & 0 \\ \sigma_{s,RV} s_t^2 & 0 & \sigma_{RV}^2 s_t^2 \end{pmatrix} \quad (13)$$

which guarantees that $V_t(X_{t+1})^{1/2}\Lambda_t$ is affine in X_t .³¹

4.3.3 Empirical results

The estimation uses standard likelihood-based methods. The appendix describes the details. We use both Dataset 1 and Dataset 2, meaning that the number of variance swap prices used in the estimation varies over time depending on availability.

Model fit Table ?? reports the means and standard deviations of the variance swap prices observed and fitted by our model together with the corresponding root mean squared errors (RMSE). The average RMSE across maturities up to 24 months is 0.73 annualized volatility points (i.e. the units in Figure ??).³² For maturities longer than 24 months, since we do not have time series of variance swap prices with fixed maturities for the entire sample, we cannot report the sample and fitted moments for any fixed maturity. Instead, we stack all contracts with more than 24 months to maturity into one single series and compute the RMSE from the observed and fitted values of this series. The corresponding RMSE is reported in the last row of Table ?. At 0.87 percentage points, it compares favorably with the RMSE for the shorter maturities. Table ? suggests that our models with two term structure factors plus RV are capable of pricing the cross-section of variance swap prices for an extended range of maturities. Even when maturities as long as 14 years are included in estimation, the data does not seem to call for extra pricing factors.

³¹It is important to note that the specifications of Λ_t in (??) and $V_t(X_{t+1})$ in (13) introduce tight restrictions on the difference $E_t(X_{t+1}) - E_t^Q(X_{t+1})$. In the appendix, we therefore consider two alternative specifications for the variance process $V_t(X_{t+1})$ and the risk prices Λ_t that are more flexible in certain dimensions. The results, both in terms of point estimates and standard errors, are essentially identical across the various specifications that we consider, so we report results for this specification here since it is most common in the literature.

³²When we exclude the financial crisis, using a sample similar to that of Egloff, Leippold, and Wu (2010), we obtain an RMSE of 0.33 percentage points, which is nearly identical to their reported value. The increase in fitting error in the full sample is, not surprisingly, brought about by the large volatility spikes that occurred during the crisis.

Risk prices The steady-state risk prices in the model are reported in Table ?? along with their standard errors. As in the previous analysis, we find clearly that it is the purely transitory shock to realized variance that is priced (*RV*-risk). The Sharpe ratio associated with an investment exposed purely to the transitory *RV* shock – analogous to the pure *RV* shock above – is -1.70.

In the VAR analysis in the previous section, the pure *RV* shock had no immediate impact on the level and slope factors, but it could potentially indirectly affect future expected variance through the VAR feedback. In the no-arbitrage model, that effect is shut off through the specification of the dynamics. That is, the *RV* shock here is completely transitory – it has no impact on expectations of volatility on any future date. The other two shocks are forced to account for all variation in expectations. The fact that the results are consistent between the no-arbitrage model and the reduced-form analysis in the previous section helps underscore the robustness of our findings to different modeling assumptions.

The short- and long-term factors earn risk premia of only -0.11 and -0.18, respectively, neither of which is significantly different from zero. The lack of statistical significance is not due to particularly large standard errors; the standard errors for the risk prices for the s_t^2 and l_t^2 shocks are in fact substantially smaller than that for the *RV* shock. Moreover, Sharpe ratios of -0.11 and -0.18 are also economically small. For comparison, the Sharpe ratio of the aggregate stock market in the 1996–2013 period is 0.43. So the risk premia on the short- and long-term components of volatility are between 25 and 42 percent of the magnitude of the Sharpe ratio on the aggregate stock market. On the other hand, the Sharpe ratio for the *RV* shock is nearly four times larger than that for the aggregate stock market and 10 to 15 times larger than the risk prices on the other two shocks. Our no-arbitrage model thus clearly confirms the results from the previous sections.

Time-series dynamics The estimated parameters determining the dynamics of the state variables under the physical measure are (equation 9):

$$\begin{pmatrix} s_{t+1}^2 \\ l_{t+1}^2 \\ RV_{t+1} \end{pmatrix} = \begin{pmatrix} 0 \\ 0.99^{***} \\ 0 \end{pmatrix} + \begin{pmatrix} 0.82^{***} & 0.16^{**} & 0 \\ 0 & 0.98^{***} & 0 \\ 0.75^{***} & 0 & 0 \end{pmatrix} \begin{pmatrix} s_t^2 \\ l_t^2 \\ RV_t \end{pmatrix} + \varepsilon_{t+1}$$

The key parameter to focus on is the persistence of l_t^2 . The point estimate is 0.9814, with a standard error of 0.0013. At the point estimate, long-term shocks to variance have a half-life of 37 months. That level of persistence is actually *higher* than the persistence of consumption growth shocks in Bansal and Yaron’s (2004) long-run risk model, and only slightly smaller than the persistence they calibrate for volatility, 0.987. Our empirical model

thus allows us to estimate risk prices on exactly the type of long-run shocks that have been considered in calibrations. As we discuss further below, the fact that we find that the long-term shock to volatility is unpriced is strongly at odds with Epstein–Zin preferences.

5 Economic interpretation

The key message of our empirical analysis is that the average investor in the variance swap market is not willing to pay for protection against news about high future volatility. In other words, they do not hedge volatility intertemporally. That fact immediately suggests that models based on Epstein–Zin (1991)? preferences, where intertemporal hedging effects are central, will struggle to match the data. To confirm that intuition, we simulate two models with Epstein–Zin preferences. The first is the long-run risk model proposed by Drechsler and Yaron (2011)?, and the second is a discrete-time version of the model with time-varying disaster risk proposed by Wachter (2013)?. In both cases, we show that the models imply that the Sharpe ratios earned from rolling over long-term zero-coupon variance claims are nearly as negative as those earned from holding just the one-month variance swap, counter to what we observe empirically in Figure ??.

The evidence we provide that there is no intertemporal hedging runs counter to many models beyond Epstein–Zin. Merton’s ICAPM, for example, implies that shocks to expected volatility should be priced since volatility affects the investment opportunity set.³³ The variance swap market thus is not well explained by the ICAPM. Similarly, in models with value-at-risk or leverage constraints, the constraint on financial intermediaries depends on expected volatility, rather than realized volatility.³⁴ In general, then, it is forward-looking volatility that is relevant in most asset pricing models.

The lack of intertemporal hedging in the variance swap market suggests a myopic model of investors. We therefore consider a simple model in which investors have power utility. While it is well known that the power utility model fails to match many asset pricing facts when consumption follows a process with low volatility, Rietz (1988)?, Barro (2006)?, Martin (2013)?, and others show that allowing for a small probability of a large decline in consumption can render the power utility model consistent with standard asset pricing moments. Gabaix (2012)? extends the disaster model to allow for a time-varying exposure of the stock market to disasters. We find that Gabaix’s model is able to match both the qualitative and

³³This is true even if volatility does not predict returns. If volatility rises but expected returns remain constant, then the investment opportunity set has deteriorated.

³⁴For example, financial intermediaries might be limited in the total amount of risk they may take. When expected volatility is higher, their demand for risky securities will fall.

quantitative features of the variance swap market. This suggests that investors in the variance swap market are mostly worried about large negative shocks to the economy in which returns collapse and variance spikes, and are purchasing variance swaps to hedge these, and only these, shocks.

5.1 Structural models of the variance premium

5.1.1 A long-run risk model

Drechsler and Yaron (2011) extend Bansal and Yaron's (2004) long-run risk model to allow for jumps in both the consumption growth rate and volatility. DY show that the model can match the mean, volatility, skewness, and kurtosis of consumption growth and stock market returns, and generates a large variance risk premium that forecasts market returns, as in the data. DY is thus a key quantitative benchmark in the literature.

The structure of the endowment process is

$$\Delta c_t = \mu_{\Delta c} + x_{t-1} + \varepsilon_{c,t} \quad (14)$$

$$x_t = \mu_x + \rho_x x_{t-1} + \varepsilon_{x,t} + J_{x,t} \quad (15)$$

$$\bar{\sigma}_t^2 = \mu_{\bar{\sigma}} + \rho_{\bar{\sigma}} \bar{\sigma}_{t-1}^2 + \varepsilon_{\bar{\sigma},t} \quad (16)$$

$$\sigma_t^2 = \mu_{\sigma} + (1 - \rho_{\sigma}) \bar{\sigma}_{t-1}^2 + \rho_{\sigma} \bar{\sigma}_{t-1}^2 + \varepsilon_{\sigma,t} + J_{\sigma,t} \quad (17)$$

where Δc_t is log consumption growth, the shocks ε are mean-zero and normally distributed, and the shocks J are jump shocks. σ_t^2 controls both the variance of the normally distributed shocks and also the intensity of the jump shocks. There are two persistent processes, x_t and $\bar{\sigma}_t^2$, which induce potentially long-lived shocks to consumption growth and volatility. We follow DY's calibration for the endowment process exactly.

Aggregate dividends are modeled as

$$\Delta d_t = \mu_d + \phi x_{t-1} + \varepsilon_{d,t} \quad (18)$$

Dividends are exposed to the persistent but not the transitory part of consumption growth. Equity is a claim on the dividend stream, and we treat variance claims as paying the realized variance of the return on equities.

DY combine that endowment process with Epstein–Zin preferences, and we follow their calibration. Because there are many parameters to calibrate, we refer the reader to DY for the full details. However, the parameters determining the volatility dynamics are obviously critical to our analysis. Note that the structure of equations (16) and (17) is the same as

the VAR in our no-arbitrage model in equation (9). The parameters governing volatility in DY’s calibration and the corresponding values from our estimation are:

	DY	Estimates
ρ_σ	0.87	0.82
$\rho_{\bar{\sigma}}$	0.987	0.9814
$\text{stdev}(\varepsilon_{\bar{\sigma},t})$	0.10	0.05
$\text{stdev}(\varepsilon_{\sigma,t} + J_{\sigma,t})$	1.10	1.48

The two feedback coefficients, ρ_σ and $\rho_{\bar{\sigma}}$, are nearly identical to our estimated values. Their long-term component, $\bar{\sigma}^2$, has a persistence of 0.987, which compares favorably with our estimate of 0.9814. Similarly, their calibration of $\rho_\sigma = 0.87$ is comparable to our estimate of 0.82. The calibration deviates somewhat more in the standard deviations of the innovations.

Overall, though, DY’s calibration implies volatility dynamics highly similar to what we observe empirically. The close match is not surprising as DY’s model was calibrated to fit the behavior of the (one-month) VIX and realized variance. As a robustness check, though, in the appendix we also simulate DY setting the standard deviations of the innovations to match our empirical estimates and obtain implications for variance swap prices that are essentially unchanged.

Given the high quality of DY’s calibration, if the long-run risk model fails to match the term structure of variance swap prices, it is not because it has an unreasonable description of the dynamics of volatility. Rather, we would conclude that the failure is due to the specification of the preferences, namely Epstein–Zin.

5.1.2 Time-varying disaster risk

The second model we study is a discrete-time version of Wachter’s (2013) model of time-varying disaster risk. In this case, consumption growth follows the process,

$$\Delta c_t = \mu_{\Delta c} + \sigma_{\Delta c} \varepsilon_{\Delta c,t} + J_{\Delta c,t} \tag{19}$$

where $\varepsilon_{\Delta c,t}$ is a mean-zero normally distributed shock and J_t is a disaster shock. The probability of a disaster in any period is F_t , which follows the process

$$F_t = (1 - \rho_F) \mu_F + \rho_F F_{t-1} + \sigma_F \sqrt{F_{t-1}} \varepsilon_{F,t} \tag{20}$$

The CIR process ensures that the probability of a disaster is always positive in the continuous-time limit, though it can generate negative values in discrete time. We calibrate the model

similarly to Wachter (2013) and Barro (2006). Details of the calibration are reported in the appendix. The model is calibrated at the monthly frequency. In the calibration, the steady-state annual disaster probability is 1.7 percent as in Wachter (2013). σ_F is set to 0.0075 (ε_F is a standard normal), and $\rho_F = 0.87^{1/12}$, which helps generate realistically volatile stock returns and a persistence for the price/dividend ratio that matches the data. If there is no disaster in period t , $J_t = 0$. Conditional on a disaster occurring, $J_t \sim N(-0.30, 0.15^2)$, as in Barro (2006). Finally, dividends are a claim to aggregate consumption with a leverage ratio of 2.8.³⁵

Wachter (2013) combines this specification of disasters with Epstein–Zin preferences. One of her key results is that a model with time-varying disaster risk and power utility has strongly counterfactual predictions for the behavior of interest rates and other asset prices. She thus argues that time-varying disaster risk should be studied in the context of Epstein–Zin preferences. We follow her in assuming the elasticity of intertemporal substitution is 1, and we set risk aversion to 3.6.³⁶

5.1.3 Time-varying recovery rates

The final model we study is a version of Gabaix’s (2012) model of disasters with time-varying recovery rates. Because the probability of a disaster is constant, power utility and Epstein–Zin are equivalent in terms of their implications for risk premia. We use power utility in our calibration, which eliminates the intertemporal hedging motives present in the two previous models. In this model, the expected value of firms following a disaster is variable. Specifically, we model the consumption process identically to equation (19) above, but with the probability of a disaster, F_t , fixed at 1 percent per year (Gabaix’s calibration). Following Gabaix, dividend growth is

$$\Delta d_t = \mu_{\Delta d} + \lambda \varepsilon_{\Delta c,t} - L_t \times 1 \{J_{\Delta c,t} \neq 0\} \quad (21)$$

λ here represents leverage. $1 \{\cdot\}$ is the indicator function. Dividends are thus modeled as permanently declining by an amount L_t on the occurrence of a disaster. The value of L is

³⁵The occurrence of a disaster shock implies that firm value declines instantaneously. To calculate realized variance for periods in which a disaster occurs, we assume that the shock occurs over several days with maximum daily return of -5 percent. For example, a jump of 20% would occur over 4 consecutive days, with a 5% decline per day. This allows for a slightly delayed diffusion of information and also potentially realistic factors such as exchange circuitbreakers. The small shocks $\varepsilon_{\Delta c,t}$ are treated as though they occur diffusively over the month, as in Drechsler and Yaron (2011).

³⁶Given the calibration of the endowment, if risk aversion is raised any higher the model does not have a solution. The upper bound on risk aversion is a common feature of models in which the riskiness of the economy varies over time.

allowed to change over time and follows the process

$$L_t = (1 - \rho_L) \bar{L} + \rho_L L_{t-1} + \varepsilon_{L,t} \quad (22)$$

We calibrate $\bar{L} = 0.5$ and $\rho_L = 0.87^{1/12}$ as in the previous model, and $\varepsilon_{L,t} \sim N(0, 0.16)$. We set the coefficient of relative risk aversion to 7 to match the Sharpe ratio on one-month variance swaps. Other than the change in risk aversion, our calibration of the model is nearly identical to Gabaix’s (2012), which implies that we will retain the ability to explain the same ten puzzles that he examines. He did not examine the ability of his model to match the term structure of variance claims, so this paper provides a new test of the theory.

5.2 Results

We now examine the implications of the three models for the zero-coupon variance curve. Figure ?? plots population moments from the models against the values observed empirically. The top panel reports annualized Sharpe ratios for zero-coupon variance claims with maturities from 1 to 12 months. Our calibration of Gabaix’s model with time-varying recovery rates matches the data well: it generates a Sharpe ratio for the one-month claim of -1.3, while all the forward claims earn Sharpe ratios of zero, similarly to what we observe in the data.

The two models with Epstein–Zin preferences, on the other hand, both generate Sharpe ratios for claims on variance more than one month ahead that are counterfactually large, especially when compared to the Sharpe ratio of the 1-month variance swap. For both the long-run risk and the time-varying disaster model, the Sharpe ratio on the three-month variance claim is roughly three-fourths as large as that on the one-month claim, whereas the three-month claim actually earns a slightly positive return empirically.

The economic intuition for the result is straightforward. If investors are risk-averse, then periods of high volatility are periods of low utility. And under Epstein–Zin preferences, periods with low lifetime utility are periods with high marginal utility. Investors thus desire to hedge news about future volatility, and forward variance claims allow them to do so. Moreover, volatility in all future periods affects lifetime utility symmetrically (discounted by the rate of pure time preference), which is why investors in these models pay the same amount to hedge volatility at any horizons.

The expected returns on the variance claims are closely related to the average slope of the term structure. The bottom panel of Figure ?? reports the average term structure in the data and in the models. The figure shows, as we would expect, that neither model with Epstein–

Zin preferences generates a curve that is as concave as we observe in the data. Instead, the DY model generates a curve that is too steep everywhere (including on the very long end), while the time-varying disaster model generates a curve that is too flat everywhere.³⁷ On the other hand, the average term structure in the model with time-varying recovery rates qualitatively matches what we observe in the data – it is steep initially and then perfectly flat after the first month.

The comparison between the calibrated models and the data reported in Figure ?? does not take into account the statistical uncertainty due to the fact that we only observe variance swap prices in a specific sample. To directly test the models against the data, we simulate the calibrated models and verify how likely we would be to see a period in which the variance swap curve looks like it does in our data. In particular, the left-hand column of Figure ?? plots results from 10,000 215-month simulations of the models. In each simulation, we calculate the average term structure of the variance curve, and normalize the value at the third month to 1 (so that we are sampling the shape of the term structure rather than its level). We then plot the median and 95-percent sampling interval of the term structure from the simulations.

Both the long-run risk model and time-varying disaster risk model have a hard time in matching the empirical shape of the variance term structure in the simulations, and particularly producing a steep slope at the short end of the curve and a small slope for higher maturities. Gabaix’s model with time-varying recovery rates performs qualitatively better: the 1-month variance swap is priced significantly higher than average realized volatility, but the slope is zero for all the rest of the curve. The long-run risk and time-varying disaster models are statistically rejected at the short end of the curve, while the long-run risk model is rejected at the long end of the curve.

The right-hand column of figure ?? simulates variance swap prices in the models out to maturities of 10 years. The sampling intervals are wider because our sample with 10-year maturities only runs for 70 months. The story is similar to that in the right-hand column though: all three models fail quantitatively, but the time-varying recovery model is the one that best matches the qualitative features of the term structure.

DY’s long-run risk model is calibrated in such a way that it nearly exactly matches our estimated volatility dynamics and it still fails to match the basic features of the term structure of variance claims. We therefore conclude from Figures ?? and ?? that models with a major intertemporal hedging motive, such as Epstein–Zin preferences, do not match the features of the variance swap market. On the other hand, a model in which investors

³⁷The models have similar Sharpe ratios but different slopes of the term structure because the latter depends on the expected return, not the Sharpe ratio.

have power utility, and hence make investment choices myopically, is able to better match our data.

The main features of the models that affect their ability to match our data can be summarized as follows. In models with Epstein–Zin preferences, investors will pay to hedge shocks to expected future volatility, especially at long horizons. Long-term zero-coupon variance claims should thus have large negative returns because they provide such a hedge. But in the data, we observe shocks to future expected volatility and find that their price is close to zero. Models with power utility, or where the variation in expected stock market volatility is independent of consumption volatility, solve that problem since investors are myopic and shocks to future expected volatility are not priced. However, the models also need to explain the high risk price associated with the realized volatility shock. In a power utility framework, this can be achieved if states of the world with high volatility are associated with large drops in consumption, as in a disaster model.

5.3 The behavior of volatility during disasters

In order for variance swaps to be useful hedges in disasters, realized volatility must be high during large market declines. A number of large institutional asset managers sell products meant to protect against tail risk that use variance swaps, which suggests that they or their investors believe that realized volatility will be high in future market declines.³⁸

In the spirit of Barro (2006), we now explore the behavior of realized volatility during consumption disasters and financial crises using a panel data of 17 countries, covering 28 events. We obtain two results. First, volatility is indeed significantly higher during disasters. Second, the increase in volatility is not uniform during the disaster period; rather, volatility spikes for one month only during the disaster and quickly reverts. It is those short-lived but extreme spikes in volatility that make variance swaps a good product to hedge tail risk.

We collect daily market return data from Datastream for a total of 37 countries since 1973. We compute realized volatility in each month for each country. To identify disasters, we use both the years marked by Barro (2006) as consumption disasters and the years marked by Schularick and Taylor (2012)?, Reinhart and Rogoff (2009)? and Bordo et al. (2001)? as financial crises.³⁹ Given the short history of realized volatility available, our final sample contains 17 countries for which we observe realized volatility and that experienced a disaster during the available sample. Table ?? shows for each country the first year of our RV sample

³⁸In particular, see Man Group’s TailProtect product (Man Group (2014)?), Deutsche Bank’s ELVIS product (Deutsche Bank, 2010?) and the JP Morgan Macro Hedge index.

³⁹See Giglio et al. (2014)? for a more detailed description of the data sources.

and the years we identify as consumption or financial disasters.

The first three columns of Table ?? compare the monthly annualized realized volatility during disaster and non-disaster years. Column 1 shows the maximum volatility observed in any month of the year identified as a disaster averaged across all disasters for each country. Column 2 shows the average volatility during the disaster years, and column 3 shows the average volatility in all other years.

Comparing columns 2 and 3, we can see that in almost all cases realized volatility is indeed higher during disasters. For example, in the US the average annualized realized volatility is 25 percent during disasters and 15 percent otherwise. Column 1 reports the average across crises of the highest observed volatility. Within disaster years there is large variation in realized volatility: the maximum volatility is always much higher than the average volatility, even during a disaster. Disasters are associated with large spikes in realized volatility, rather than a generalized increase in volatility during the whole period.

To confirm this result, in Figure ?? we perform an event study around the peak of volatility during a disaster. For each country and for each disaster episode, we identify the month of the volatility peak during that crisis (month 0) and the three months preceding and following it. We then scale the volatility behavior by the value reached at the peak, so that the series for all events are normalized to 1 at the time of the event. We then average the rescaled series across our 28 events.

The figure shows that indeed, the movements in volatility that we observe during disasters are short-lived spikes, where volatility is high for essentially only a single month. In the single months immediately before and after the one with the highest volatility, volatility is 40 percent lower than its peak, and it is lower by half both three months before and after the worst month.

Table ?? and Figure ?? show that an asset that provides protection for news about high future volatility provides only weak protection against market crashes since volatility is not particularly high *on average* during crashes. The asset that provides the best tail protection is one that provides protection against high realized volatility, rather than high expected volatility. That said, though, it is important to note that not all periods of high volatility coincide with large declines in consumption. The 1987 stock market crash is the prime example of an episode with high volatility in financial markets that had little or no effects on real activity. So while variance swaps clearly provide a hedge against crashes, their returns are not perfectly correlated with crashes.⁴⁰

[This last paragraph is not very convincing. Does this section rally show that

⁴⁰That fact also implies that there must be assets with even larger Sharpe ratios than one-month variance swaps, and thus the annual Hansen–Jagannathan (1991?) bound must be greater than 1.5.

a claim to expected volatility wouldn't provide a hedge?]

6 Conclusion

This paper shows that it is only the transitory part of realized variance that is priced. That fact is not consistent with a broad range of structural asset pricing models. It is qualitatively consistent with a model in which investors desire to hedge rare disasters, but even that model does not match all the quantitative features of the data. Interestingly, the data is not consistent with all disaster models. The key feature that we argue models need in order to match our results is that variation in expected stock market volatility is not priced by investors, whereas the transitory component of volatility is strongly priced.

The idea that variance claims are used to hedge crashes is consistent with the fact that many large asset managers, such as Deutsche Bank, JP Morgan, and Man Group sell products meant to hedge against crashes that use variance swaps and VIX futures. These assets have the benefit of giving tail protection, essentially the form of a long put, but also being delta neutral (in an option-pricing sense). They thus require little dynamic hedging and yield powerful protection against large declines.

In the end, we conclude that the variance swap market is well described as a set of assets that investors use to hedge crashes. They do not seem to desire to hedge changes in the probability of a crash (or any other sort of volatility). That fact is at odds with models that imply that investors hedge intertemporally. An investor who does have an intertemporal hedging motive and wants protection against increases in future volatility would be well served to purchase that protection, essentially for free, from financial markets.