

NBER WORKING PAPER SERIES

VOLATILITY EXPECTATIONS AND RETURNS

Lars A. Lochstoer
Tyler Muir

Working Paper 28102
<http://www.nber.org/papers/w28102>

NATIONAL BUREAU OF ECONOMIC RESEARCH
1050 Massachusetts Avenue
Cambridge, MA 02138
November 2020

We thank seminar participants at NBER Behavioral Finance, Virtual Finance Seminar, UCLA, UT Austin, Tilburg, Toronto, Insead, Norwegian School of Economics, Maastricht, Nova, and Amsterdam, and Andy Atkeson, Nick Barberis, John Campbell, Ing-Haw Cheng, Mike Chernov, Itamar Dreschler (discussant), Andrea Eisfeldt, Stefano Giglio (discussant), Valentin Haddad, Barney Hartman-Glaser, Hanno Lustig, Alan Moreira, Stefan Nagel, Stavros Panageas, Andrei Shleifer, Kelly Shue, and two anonymous referees for comments. We thank James O'Neill and Alvaro Boitier for excellent research assistance. We thank Ing-Haw Cheng, Stefano Giglio, Ian Dew-Becker, and Travis Johnson for making data available. The views expressed herein are those of the authors and do not necessarily reflect the views of the National Bureau of Economic Research.

NBER working papers are circulated for discussion and comment purposes. They have not been peer-reviewed or been subject to the review by the NBER Board of Directors that accompanies official NBER publications.

© 2020 by Lars A. Lochstoer and Tyler Muir. All rights reserved. Short sections of text, not to exceed two paragraphs, may be quoted without explicit permission provided that full credit, including © notice, is given to the source.

Volatility Expectations and Returns
Lars A. Lochstoer and Tyler Muir
NBER Working Paper No. 28102
November 2020
JEL No. G0,G12,G4

ABSTRACT

We provide evidence that agents have slow-moving beliefs about stock market volatility that lead to initial underreaction to volatility shocks followed by delayed overreaction. These dynamics are mirrored in the VIX and variance risk premiums which reflect investor expectations about volatility and are also supported in surveys and in firm-level option prices. We embed these expectations into an asset pricing model and find that the model can account for a number of stylized facts about market returns and return volatility which are difficult to reconcile, including a weak, or even negative, risk-return tradeoff.

Lars A. Lochstoer
University of California at Los Angeles
lars.lochstoer@anderson.ucla.edu

Tyler Muir
University of California at Los Angeles
Anderson School of Management
110 Westwood Plaza
Los Angeles, CA 90024
and NBER
tyler.muir@anderson.ucla.edu

1. Introduction

Agents perceptions of risk play a critical role in asset pricing models. However, a long literature finds that the empirical risk-return tradeoff is weak at best (Glosten, Jagannathan, and Runkle, 1993), despite this tradeoff being strong in leading asset pricing models (Moreira and Muir, 2017).¹ This paper proposes a model where the representative agent may have biased, slow moving expectations about volatility and we show that these expectations can help explain a weak relation between risk and return. We discipline the expectations about volatility in the model in three ways: we microfound beliefs based on sticky and extrapolative expectations (shown to be present in many other contexts), we use survey data to directly assess agents expectations of volatility, and we study prices and returns of volatility dependent claims (VIX futures, variance swaps, and straddles) which help assess if mistakes in beliefs are present in financial market prices.

Expectations about volatility appear to initially underreact to news about volatility followed by a delayed overreaction. This pattern matches direct empirical facts about volatility dependent claims and the variance risk premium. Cheng (2018) finds that prices of VIX futures do not respond strongly enough to changes in volatility, so that increases in volatility negatively predict the premium on a short-position in VIX futures. We extend these results to variance swaps and straddle returns over a longer sample. Claims that provide insurance against future volatility, which are unconditionally expensive, initially appear “too cheap” after volatility rises but appear expensive later on. This is hard to square with a rational risk premium because these claims are typically riskier after volatility rises (Cheng, 2018), suggesting the risk premium should go up rather than down.

We then embed these beliefs into an otherwise standard Epstein-Zin (1989) equilibrium model with stochastic volatility and argue we can explain many empirical facts

¹Moreira and Muir (2017) show that the basic risk-return relation is strong in calibrations of leading equilibrium asset pricing models (including habits, long run risk, time-varying disasters, and intermediary based models). Martin (2016) argues this relationship holds in a wide class of models if σ_t^2 is replaced by risk-neutral variance which we will consider empirically as well.

about stock market volatility, the VIX, the variance risk premium, and stock returns. First, the model can generate a weak – and potentially even negative – conditional risk-return tradeoff. When volatility rises, agents only partially react by requiring a higher expected return on stocks which pushes current stock prices down. This matches the negative correlation between realized returns and volatility innovations (French, Schwert, and Stambaugh, 1987) and is consistent with a discount rate effect from volatility shocks. Given the initial underreaction, however, agents on average again update positively about volatility next period even without additional news. This effect can push prices next period down further, and make it appear as though the initial increase in volatility is associated with a future decline in the observed equity risk premium in the short term. Through this channel, the model can simultaneously match that realized stock returns are strongly negatively correlated with contemporaneous innovations in volatility (French et al., 1987) while also generating a weak risk-return tradeoff (Glosten et al., 1993). This also leads to volatility timing strategies which generate positive alpha (Moreira and Muir, 2017). The price decline is hump-shaped and prices take longer to recover than in the rational benchmark, implying that objective equity risk premiums are high well after the volatility shock subsides. We provide evidence consistent with this view, as in Brandt and Kang (2004).

The variance risk premium – defined as VIX^2 minus an objective forecast of variance – will also feature underreaction and delayed overreaction directly through the beliefs channel which shows up in market implied volatility (VIX). Because the mistake in beliefs shows up both in volatility claims and equity claims in the same direction, the observed variance risk premium will positively forecast stock returns (Bollerslev, Tauchen, and Zhou, 2009; Drechsler and Yaron, 2011). However, compared to rational models that link equity and variance risk premiums (Bollerslev et al., 2009), our model can account for the otherwise puzzling evidence that while the variance risk premium positively forecasts stock returns, neither the VIX nor realized variance are individually strong forecasters of returns.

We microfound the beliefs in our model through sticky expectations and extrapolation

tion, both of which have been found in many other contexts (i.e., Coibion and Gorodnichenko (2015), Mankiw and Reis (2002), Bordalo, Gennaioli, Ma, and Shleifer (2018) and Landier, Ma, and Thesmar (2019)).² These forces combine to generate initial underreaction and delayed overreaction to volatility news. Intuitively, sticky expectations lead to underreaction in beliefs particularly at short horizons, while overextrapolation leads to overreaction. While these forces are prevalent in other work and lead to a convenient and tractable belief process, we do not take a strong stand on these microfoundations – other behavioral factors that lead to initial underreaction and delayed overreaction might also generate similar dynamics (Daniel, Hirshleifer, and Subrahmanyam, 1998; Barberis, Shleifer, and Vishny, 1998; Hong and Stein, 1999).³

We use survey data on volatility and uncertainty about stock returns from the Graham and Harvey CFO survey (as well as the Shiller survey) and document that the surveys exhibit slow moving expectations as in our model. In particular, respondents are asked about the 90th and 10th percentile of stock returns over the next year, which we use to infer beliefs about volatility. We regress survey expectations about volatility on past volatility realizations and show that expectations look like a weighted average of past volatility realizations as our model predicts, whereas optimal forecasts mainly load only on current volatility. Thus, the surveys expectations appear sticky or slow-moving. The evidence of slow updating toward volatility is consistent with the dynamics of CFO learning found in Boutros, Ben-David, Graham, Harvey, and Payne (2019). We also use longer horizon survey evidence based on a 10 year horizon which allows us to assess agents views on the persistence of volatility. Similar to Landier et al. (2019), agents use a persistence parameter that is too large, consistent with overextrapolation. This can generate overreaction in long run claims on volatility in

²See Payzan-LeNestour, Pradier, and Putniňš (2018) for related work on expectations of volatility based on past volatility. Bouchaud, Krueger, Landier, and Thesmar (2019) and Jiang, Krishnamurthy, and Lustig (2018) apply sticky expectations to the profitability anomaly and exchange rates, respectively.

³Our model is related to other models of extrapolation from past data or experience effects including Barberis, Greenwood, Jin, and Shleifer (2015), Collin-Dufresne, Johannes, and Lochstoer (2016), Nagel and Xu (2019), Malmendier and Nagel (2011), Greenwood and Shleifer (2014), and Glaeser and Nathanson (2017).

the spirit of Giglio and Kelly (2017) and Stein (1989).

The sharpest place to identify such beliefs in financial market prices is in volatility claims and option markets. While survey evidence is useful, a concern is that biases in surveys may not show up in prices (i.e., if respondents don't actively trade or other rational agents trade sufficiently to eliminate respondents' impact on prices). We follow and extend (Cheng, 2018) and show that the mistakes in the model show up in prices and in predictable variation in the returns of variance swaps, VIX futures, and straddles. We also document results at the firm level, where we show implied volatility from firm level options does not react strongly enough to recent changes in volatility, leading to underreaction and a lower variance risk premium following increases in volatility (see Potoshman (2001) for related work). This is true even when we include time fixed effects that control for aggregate movements in firm volatility which makes a risk based explanation even more difficult. The firm level analysis provides further support for our story of initial underreaction and also provides robustness to our main empirical results which rely on aggregate market data and hence a relatively smaller sample.⁴

We use the survey data and evidence from option markets to calibrate our model. The calibrated model does fairly well at quantitatively matching the main stylized facts in the literature on the relation between volatility, stock returns, and the variance risk premium (which we extend to a more recent sample). Most importantly, we show that slow-moving expectations about volatility are key to matching these dynamics – the nested rational version of the model fails to account for the evidence. We come to similar conclusions for the rational model of Bollerslev et al. (2009). A natural concern is that biases in beliefs will lead to excessive trading profits for a rational investor in the model. The mistakes in the calibrated model are modest: agents beliefs about volatility in the model have a correlation of about 0.9 with an objective forecast. In an extension, we show that Sharpe ratio gains for a rational investor who trades on

⁴Rachwalski and Wen (2016) also provide firm level evidence on the risk-return tradeoff: “Stocks with increases in idiosyncratic risk tend to earn low subsequent returns for a few months. However, high idiosyncratic risk stocks eventually earn persistently high returns.”

these mistakes are modest compared to other anomalies in the literature. We also explicitly consider learning in our objective volatility forecasts where we construct estimates of variance using only data available to the agents at the time.

Our model is parsimonious, stylized, and tractable and our main point is this simple change in volatility expectations can help match key features of the data. However, we also outline several shortcomings of the model and discuss extensions that add complication but better fit certain features of the data. Finally, we present additional model implications on the term structure of volatility dependent returns and discuss alternative models.

Section 2 provides a microfoundation for the expectations in this paper and brings in survey data about volatility. Section 3 presents the model, while Section 4 compares the model to the stylized facts in the literature on stock returns, volatility, the VIX, and the variance risk premium. Section 5 provides additional evidence, discusses shortcomings of the model, and considers alternative explanations.

2. Belief Microfoundations and Survey Evidence

2.1 A simple microfounded model of variance belief formation

There is a burgeoning literature estimating investor beliefs from both surveys and experiments. For instance, Coibion and Gorodnichenko (2015) find that average survey expectations display dynamics consistent with the sticky information model of Mankiw and Reis (2003). In this model, average expectations are sticky since only a fraction of agents update their information set at each point in time. At the same time, evidence on individual agents expectations (e.g., Bordalo, Gennaioli, Ma, and Sheifer (2020), Landier, Ma, and Thesmar (2019)) find that agents tend to overextrapolate in the sense that they believe shocks have a more persistent effect than they truly do. These two biases thus appear pervasive in survey data.

We provide a simple model of investor expectations of stock market variance that

allows for both stickiness and overextrapolation. In particular, as in Mankiw and Reis (2003), let a fraction $1 - \tau$ of agents update beliefs at each point in time. Denote their time $t - j$ expectation of realized variance (rv) at time $t + 1$ as $E_{j,t-j}^S[rv_{t+1}]$, where the S superscript stands for subjective expectations. The average forecast, denoted $E_t^S[rv_{t+1}]$, is then:

$$E_t^S[rv_{t+1}] = (1 - \tau) \sum_{j=0}^{\infty} \tau^j E_{j,t-j}^S[rv_{t+1}]. \quad (1)$$

Further, we let agents believe realized variance follows an AR(1) process with autocorrelation coefficient $\tilde{\rho}$. If $\tilde{\rho}$ is too high relative to the true process for expected variance, agents overextrapolate. That is, consistent with the findings in Landier, Ma, and Thesmar (2019), we allow individual agents to believe that the true process is more persistent than it really is. This gives a simple model that relates expectations to lagged realized variance:

$$E_t^S[rv_{t+1}] = \bar{v} + (1 - \tau)\tilde{\rho} \sum_{j=0}^{\infty} \tau^j \tilde{\rho}^j (rv_{t-j} - \bar{v}), \quad (2)$$

where \bar{v} is the unconditional mean. This form of belief formation is also used in Brooks, Katz, and Lustig (2018). A higher value of τ implies more information stickiness. If the underlying true process is persistent, stickiness can lead to initial underreaction in aggregate beliefs as these beliefs load too heavily on past lags of realized variance. If the underlying true process is not sufficiently persistent (e.g., if it is i.i.d.), the beliefs process only generates overreaction. A higher value of $\tilde{\rho}$ leads to overreaction and too persistent beliefs.

This is particularly clear if we consider agents' longer-run expectations about variance, as the expected variance k periods from today is:

$$E_t^S[rv_{t+k}] = \bar{v} + \tilde{\rho}^{k-1}(E_t^S[rv_{t+1}] - \bar{v}). \quad (3)$$

This model provides a useful way to understand our empirical estimates and gives a microfoundation for the belief dynamics we use in our main model and that we uncover

in survey beliefs. Stickiness leads to initial underreaction, while the overextrapolation bias leads to subsequent overreaction. We note, however, that belief patterns which generate initial underreaction and delayed overreaction can be generated assuming other investor biases, as has been done in previous literature (Daniel et al., 1998; Barberis et al., 1998; Hong and Stein, 1999). We leave the exercise of distinguishing these potentially different microfoundations for future work.

2.2 Survey Data

We bring survey data related to volatility which allows us to evaluate the main mechanism in our paper using direct data on expectations. Our main source is the Graham and Harvey survey of CFOs which is quarterly from 2001 to 2018. The survey asks respondents for a mean forecast for the stock market over the next year as well as 10th and 90th percentiles. We construct the 90th minus 10th percentile as a measure of volatility or uncertainty and square this number to get a measure of expected variance. While this measure has limitations, it captures how spread out agents view the return distribution, and under the view of a normal distribution would perfectly capture agents expectations about variance.

2.2.1 Stickiness of variance expectations

We fit survey expectations and actual realized variance over the period investors are asked to forecast as an exponential weighted average of past variance as in the micro-founded model above. In particular, we project the expected 12-month variance from the survey onto lagged monthly realized variance using the same functional form as that in our model (Equation (2)):

$$Survey_t^{(1yr)} = a + b \sum_{j=0}^J \phi^j rv_{t-j} + \epsilon_{t,t+12}^s. \quad (4)$$

Since the variance-measure from the survey is only proportional to variance, we focus on the estimated value of ϕ in this regression. Relative to the model above, $\phi = \tau\tilde{\rho}$. We emphasize, however, that a non-zero ϕ for the survey alone does not necessarily imply non-rational expectations.⁵

Note that we allow survey expectations to embody signals other than lagged realized variance. In particular, we do not require that the error term is i.i.d. and we compute standard errors with block bootstrap with a 6 quarter block length. We set $J = 11$ so that we use one year of lagged realized variance given our finite sample. Our results are not sensitive to small variation in the number of lags. The survey is quarterly, $t = 3, 6, 9, \dots, T$, but we use the monthly frequency for rv (sum of squared daily returns to the S&P500 in a given month) as our model and later data analysis is at the monthly frequency.

Table 1 reports that the estimated value of ϕ is 0.87 with a standard error of 0.11. Thus, survey respondents effectively take into account many lags of realized variance when forming expectations about future variance. As a benchmark, we also report the results from projecting realized variance over the next 12 months onto lagged rv using the same functional form on the right hand side as in Equation (4). If the survey expectations are rational, these two projections will yield the same coefficients.⁶ In this benchmark case, ϕ is estimated to be -0.16 . The difference between this estimate and the estimate using the survey expectations is statistically significant at the 1%-level as reported in the table. In other words, the full-information rational expectations process for realized variance loads much more heavily on current variance and less on past variance over this sample, while agents' expectations appear sticky in the sense that they assign higher than optimal weights to additional lags.

Next, we evaluate the extent to which survey expectations reflect the information

⁵The lag structure is that of an ARMA(1,1) in realized variance, which is a parsimonious way to capture short- and long-run dynamics in expected variance. Thus, the model also allows for a commonly used process for the objective dynamics of realized variance. See Appendix D.1 and D.1.1 for further details about the survey, as well as derivations showing how the lag structure in Equation (4) relates to an ARMA(1,1) and the above model of beliefs.

⁶To see this, note that $E[E[rv_{t,t+12}|y^t, rv^t]|rv_t, \dots, rv_{t-11}] = E[rv_{t,t+12}|rv_t, \dots, rv_{t-11}]$ where y^t is the history of other signals and rv^t is the history of rv .

in the VIX, and vice versa. In Panel B of Table 1, we regress the squared VIX on the contemporaneous survey expectation of variance, as well as the most recent monthly realized variance. Both variables come in positive and significant at the 1%-level with an R^2 of 74%. If we instead put next year's realized variance on the left-hand-side (rightmost column of Panel B), the survey comes in with an insignificant negative sign, while the lagged variance comes in positive and strongly significant. Thus, a rational forecast does not load on the survey at all after controlling for the current level of realized variance, while the squared VIX is in fact strongly related to the survey. The survey expectations should show up in the VIX if they capture the expectations of market participants, and the results in Panel B of Table 1 indicates that they indeed do.

As a more direct way to explore this, we also use the VIX^2 in place of the survey as a measure of market expected variance and again estimate our ϕ parameter. We estimate a ϕ of 0.42 (std err of 0.12) using VIX data from January 1990 to April 2020. Thus, the dependence on past variance that leads agents expectations to respond too slowly (underreaction) is present not just in surveys, but in actual market prices of expected variance.⁷ However, the lower value for ϕ suggests that this bias is much lower in actual financial market prices.

An additional source of survey evidence on volatility is from Robert Shiller who asks investors the probability of a stock market crash over the next 6 months such as that seen in 1987. This survey is monthly from July of 2001. While this measure is not as direct as a measure of variance expectations, it does gauge agents beliefs about risk in the stock market in general. We use this survey as a robustness check and relegate the findings to a Appendix Table 16. The Shiller survey produces an estimate of ϕ of around 0.77, close to the value found in the CFO survey. This provides further support for stickiness in variance expectations (e.g., $\tau > 0$). The evidence of slow updating

⁷Other evidence also suggests agents do take action based on their reported beliefs about risk, which is important for survey beliefs to affect prices. Giglio, Maggiori, Stroebel, and Utkus (2019) show that survey data on investor beliefs about risk translates directly into actions in terms of portfolio allocations, while Ben-David, Graham, and Harvey (2013) find that CFO expectations about volatility translate into investment decisions.

toward volatility is consistent with the findings of Boutros et al. (2019).

To further assess the degree of investors slow moving expectations, we run a first order vector autoregression (VAR) with future realized variance as well as the reported expectations from the survey. Both variables are normalized to have unit standard deviation. We order future variance first, followed by the survey expectation and plot the impulse response to a one standard deviation variance shock. Results are given in Figure 3. As before, future variance increases substantially after this shock then subsequently declines as it mean reverts. The survey expectations, however, show a hump shaped response, with expectations continuing to rise after the initial shock. This is directly consistent with finding a large value of ϕ in the survey beliefs. The expectations initially do not rise as much (underreaction) but then subsequently remain elevated long after expected variance declines (subsequent overreaction), consistent with the dependence of survey expectation on longer lags of past variance. The pattern from both surveys is similar, and this prediction is exactly what we expect from our model of slow moving expectations of volatility. Thus, two independent surveys provide consistent evidence in favor of the mechanism we propose.

2.2.2 Persistence of variance expectations

Next, we evaluate the persistence of agents' beliefs as implied by the CFO survey. Importantly, respondents are asked the same questions about the return distribution at the 10 year horizon. Equation (3) shows how long-run forecasts relate to the short-run forecast under the illustrative model of aggregate expectations given earlier. The 10-year and 1-year survey forecasts are for *cumulative* 10- and 1-year variance expressed in annual terms. Thus, we have:

$$\begin{aligned} E_t^S[rv_{t+1} + \dots + rv_{t+12}] &= 12\bar{v} + \sum_{k=0}^{11} \tilde{\rho}^k (E_t^S[rv_{t+1}] - \bar{v}), \\ \frac{1}{10} E_t^S[rv_{t+1} + \dots + rv_{t+120}] &= 12\bar{v} + \sum_{k=0}^{119} \tilde{\rho}^k (E_t^S[rv_{t+1}] - \bar{v}), \end{aligned}$$

which implies that

$$Survey_t^{(10yr)} = \tilde{a} + \tilde{b} \times Survey_t^{(1yr)} + \eta_t,$$

where $\tilde{b} = \frac{\sum_{k=0}^{119} \tilde{\rho}^k}{10 \sum_{k=0}^{11} \tilde{\rho}^k} = \frac{1}{10} \frac{1-\tilde{\rho}^{120}}{1-\tilde{\rho}^{12}}$ and where the error term allows this relation to not be exact.

Panel C of Table 1 shows that $\tilde{b} = 0.25$ with a standard error of 0.04 in the sample available. Under the belief model above, the implied monthly autocorrelation, $\tilde{\rho}$, is 0.96 with a 95% confidence interval from 0.92 to 0.97. This is much more persistent than the autocorrelation of realized variance, which is only 0.72 in this sample, which implies that agents overextrapolate when forecasting future variance. This number is close to the experimental evidence in Landier, Ma, and Thesmar (2019) who provide agents reported belief about persistence as a function of the true persistence of the process. These results echo the evidence in Stein (1989) and Giglio and Kelly (2017) who show overreaction of long term volatility expectations from options data and variance swap prices, respectively. In particular, they argue that longer term expectations of volatility are too volatile relative to those at short horizons, a form of relative overreaction in long term expectations.

In summary, the survey-based variance expectations indicate that agents' expectations are sticky in the sense that they use too many lags of variance to form expectations of future variance $\tau > 0$, and that agents perceive the persistence of variance to be substantially higher than it truly is $\tilde{\rho} > \rho$. A reasonable concern is that survey forecasts do not reflect expectations of agents in the market and that these beliefs therefore are not important for asset prices. However, we have shown that these patterns show up in the VIX. Later, we strengthen this evidence further by studying the implications of biased beliefs on return predictability for claims on stock market variance. In particular, this finding directly links to the work of Cheng (2018) who shows that the price of volatility claims don't respond strongly enough to increases in volatility. These "mistakes," or forecast errors, are better assessed in terms of return

predictability, i.e., if agents make mistakes in their expectations about volatility this should feed into the prices of volatility dependent claims and generate return predictability, as Cheng (2018) shows. We reevaluate this evidence in light of our model in Section 4.

3. The Model

In this section, we develop an asset pricing model similar to that in Bansal and Yaron (2004) except that we allow the representative investor to have biased beliefs regarding the dynamics of stock return volatility. This simple modification enables the model to account for the empirical evidence discussed earlier.

Let the objective process for aggregate log dividend growth be given by:

$$\Delta d_t = \mu + \sigma_t \varepsilon_t, \quad (5)$$

$$\sigma_t^2 = \bar{v} + \rho (\sigma_{t-1}^2 - \bar{v}) + \omega \eta_t, \quad (6)$$

where σ_t^2 is the realized variance of dividend growth innovations, observed at time t , and ε_t and η_t are uncorrelated i.i.d. standard Normal shocks. Variance is persistent with $0 < \rho < 1$. Equation (6) implies that variance can go negative. For ease of exposition we follow Bansal and Yaron (2004) and proceed as if σ_t^2 is always non-negative. In Appendix B, we show that this simplification is unimportant for our conclusions by solving a model with Gamma distributed variance shocks, where variance is guaranteed to always be positive.

We assume a representative stockholder with consumption equal to aggregate dividends whose marginal utility prices all claims in the economy but whose beliefs potentially depart from rationality. The agents' expectations of the conditional variance

of dividend growth are given by:

$$E_{t-1}^S [\sigma_t^2] = \bar{v} + \lambda x_{t-1}, \quad (7)$$

$$\begin{aligned} x_t &= \phi x_{t-1} + (1 - \phi) (\sigma_t^2 - \bar{v}) \\ &= (1 - \phi) \sum_{j=0}^{\infty} \phi^j (\sigma_{t-j}^2 - \bar{v}). \end{aligned} \quad (8)$$

The S superscript on the expectations operator highlights that the expectation is taken under the agent's *subjective* beliefs. If $\phi = 0$ and $\lambda = \rho$, the agent has rational expectations about the volatility dynamics, while if $\phi > 0$ the agent has slow-moving volatility expectations, allowing an exponentially weighted average of past variance to affect the current expectation, as opposed to only the current value as the physical volatility dynamics prescribe. The scale of agents' expectations is set by λ , and we return to the interpretation of this parameter shortly. We assume $0 < \phi, \lambda < 1$. Section 2.1 provides direct microfoundations for these beliefs using a model of sticky expectations and overextrapolation. In this case, $\lambda = (1 - \tau)\tilde{\rho}$ and $\phi = \tau\tilde{\rho}$, where $\tau > 0$ governs the stickiness of expectations and $\tilde{\rho} > \rho$ implies that investors overextrapolate. These two features have been found in many other settings analyzing survey evidence (e.g., Coibion and Gorodnichenko (2015), Bordalo, Gennaioli, Ma, and Sheifer (2020), Landier, Ma, and Thesmar (2019)).

Under agents' beliefs, the shock to variance is:

$$\begin{aligned} \omega\eta_t^S &\equiv \sigma_t^2 - \bar{v} - \lambda x_{t-1} \\ &= \rho (\sigma_{t-1}^2 - \bar{v}) - \lambda x_{t-1} + \omega\eta_t, \end{aligned} \quad (9)$$

where $\rho (\sigma_{t-1}^2 - \bar{v}) - \lambda x_{t-1} = E_{t-1}^P [\sigma_t^2] - E_{t-1}^S [\sigma_t^2] = E_{t-1}^P [\omega\eta_t^S]$ is the mistake agents make when forecasting variance. Here a P superscript on the expectations operator means the expectation is taken under the objective measure. We can thus write the

dynamics of x_t under agents' beliefs as:

$$x_t = (\phi + (1 - \phi) \lambda) x_{t-1} + (1 - \phi) \omega \eta_t^S. \quad (10)$$

Note that investors' variance expectations are sticky relative to the true variance dynamics if $\phi > 0$ and $\lambda \geq \rho$, as the persistence of x_t then is higher than the true persistence of σ_t^2 (that is, $\phi + (1 - \phi) \lambda > \rho$). Also note that the shock itself is moderated by a factor of $1 - \phi$. Figure 1 shows the impulse-response from a positive variance shock (η_0) for objective and subjective expected variance. The parameter values are calibrated to the data as described below. The true AR(1) dynamics of variance are reflected in the monotonically decaying response in the rational case (dashed red line). The solid blue line give the impulse-response of agents' expected variance as reflected in the dynamics of x_t . Agents' initially underreact, as ϕ is greater than zero in this case, but the higher persistence of x_t leads to subsequent overreaction.

Following Bollerslev, Tauchen, and Zhou (2009), the agent has Epstein-Zin utility (Epstein and Zin, 1989) where β , γ , and ψ are the time-discounting, risk aversion, and intertemporal substitution parameters, respectively. The stochastic discount factor is therefore:

$$M_t = \beta^\theta e^{-\frac{\theta}{\psi} \Delta d_t + (\theta-1)r_t}, \quad (11)$$

where $\theta = \frac{1-\gamma}{1-1/\psi}$ and r_t is the log return to the aggregate dividend claim. We use the standard log-linearization techniques of Campbell and Shiller (1988) and Bansal and Yaron (2004) to derive equilibrium asset prices (see Appendix A for details). In particular, we assume aggregate log returns are $r_t = \kappa_0 + \kappa pd_t - pd_{t-1} + \Delta d_t$, where pd is the aggregate log price-dividend ratio and κ is a constant close to but less than one that arises from the log-linearization. We then obtain:

$$pd_t = c - Ax_t, \quad (12)$$

where $A = -\frac{1}{2} \frac{\lambda(1-\gamma)(1-1/\psi)}{1-\kappa(\phi+(1-\phi)\lambda)}$. Notice that if $\gamma, \psi > 1$ we have that $A > 0$. This is the

standard preference parameter configuration for asset pricing models with Epstein-Zin preferences. It implies that the price-dividend ratio is low when agents perceive variance to be high, as in the data.

3.1 Equity risk premium dynamics

Let r_t and $r_{f,t}$ denote the aggregate log return and risk-free rate in period t , respectively. The subjective conditional risk premium of log returns in this economy is:

$$E_{t-1}^S [r_t - r_{f,t}] = (\gamma - \frac{1}{2}) E_{t-1}^S [\sigma_t^2] + \delta_r, \quad (13)$$

where δ_r is a constant given in the Appendix that captures the price effect of discount rate shocks due to the variance shocks (η_t). The first term reflects the standard risk-return trade-off that is linear in the conditional variance of dividend growth, where the $-1/2$ part arises as this is the log return risk premium.

The conditional variance of log returns is determined both by the conditional variance of dividend growth and the impact of the variance shock on the price-dividend ratio:

$$Var_{t-1}^S (r_t) = \Theta + E_{t-1}^S [\sigma_t^2], \quad (14)$$

where $\Theta = (\kappa A(1 - \phi)\omega)^2$.

The *objective* risk premium, however, is:

$$E_{t-1}^P [r_t - r_{f,t}] = E_{t-1}^S [r_t - r_{f,t}] + \kappa (1 - \phi) A (E_{t-1}^S [\sigma_t^2] - E_{t-1}^P [\sigma_t^2]), \quad (15)$$

where the P superscript on the expectation denotes that it is taken using the true, objective variance dynamics. Note that the risk premium loads *negatively* on true conditional variance as $\kappa (1 - \phi) A > 0$ for our calibrated parameters — a major departure from earlier literature. To see where Equation (15) comes from, recall that the shock to agents beliefs about variance is predictable (see Equation (9)). The mistake

is persistent, which magnifies its effect on prices as given by the term $\kappa A (1 - \phi)$.⁸

The mistakes in agents' conditional variance expectations are reflected in current discount rates and therefore prices. Consider a positive shock to variance ($\eta_{t-1} > 0$). With $\phi > 0$ investors' expectations are sticky, meaning investors do not update their beliefs sufficiently and initially underreact to the variance shock. Thus, $E_{t-1}^P [\sigma_t^2] > E_{t-1}^S [\sigma_t^2]$. Since $A > 0$ in the relevant calibrations, this means a *positive* shock to variance can, if the mistake is sufficiently large, *decrease* next period's objective risk premium. The reason is that investors will on average perceive a positive shock to discount rates next period as the realized value of σ_t^2 on average is higher than they had expected. This leads to a predictable decline in the price-dividend ratio under the objective measure. Recall that the price dividend ratio is given by $c - Ax_t$. Hence, the price-dividend ratio falls at the impulse, but note that it keeps falling in the following period due to the increase in discount rates when agents learn variance is higher than expected and $E_{t-1}^S [\sigma_t^2]$ rises. Subsequently, given the too persistent variance expectations, agents eventually overreact to the volatility shock, which leads to $E_{t+j-1}^P [\sigma_{t+j}^2] < E_{t+j-1}^S [\sigma_{t+j}^2]$ for some $j > 0$. In this case, the second term in Equation (15) becomes positive and the conditional risk premium overshoots.

Upon impact, a positive shock to variance decreases prices as the long-run impact on discount rates is positive when A is positive. This is consistent with the negative contemporaneous correlation of realized variance and returns in the data (e.g., French, Schwert, and Stambaugh, 1987). In particular, shocks to returns are:

$$r_t - E_{t-1}^P [r_t] = -\Theta^{1/2} \eta_t + \sigma_t \varepsilon_t, \quad (16)$$

where $\Theta^{1/2} = \kappa A (1 - \phi) \omega$ encodes the present value impact of the shock to variance (η_t) due to its effect on the discount rates agents require for holding the risky asset.

⁸This expression is found by using the Campbell-Shiller return approximation and noting that $E_{t-1}^P (-\kappa pd_t) - E_{t-1}^S (-\kappa pd_t) = -\kappa A (E_{t-1}^P [x_t] - E_{t-1}^S [x_t]) = -\kappa A (1 - \phi) (E_{t-1}^P [\sigma_t^2] - E_{t-1}^S [\sigma_t^2])$.

3.2 Variance risk premium dynamics

In addition to the equity claim, we also price a variance claim with payoff:

$$RV_t \equiv \Theta + \sigma_t^2, \quad (17)$$

where RV_t stands for realized variance at time t .⁹ We define the time $t-1$ implied variance (IV_{t-1}) as the swap rate that gives a one-period variance swap a present value of zero:

$$0 = E_{t-1}^S [M_t (RV_t - IV_{t-1})]. \quad (18)$$

Thus:

$$IV_{t-1} = E_{t-1}^S [R_{f,t} M_t RV_t]. \quad (19)$$

As is standard in the literature, we denote the (objective) expected payoff of a position in the variance swap where you are paying the realized variance and receiving the implied variance as the variance risk premium:

$$VRP_{t-1} = IV_{t-1} - E_{t-1}^P [RV_t]. \quad (20)$$

⁹Our model-definition of realized variance is motivated by industry practice for variance swap payoffs, where monthly realized variance is the sum of squared daily log returns within the month. In the model, squared monthly log returns are:

$$(r_t - E_{t-1} [r_t])^2 = \Theta \eta_t^2 + 2\sigma_t \Theta^{0.5} \eta_t \varepsilon_t + \sigma_t^2 \varepsilon_t^2.$$

To approximate the use of higher frequency data to estimate realized variance within our model, we assume that the second moments of realized shocks equal their continuous-time limit. Setting $\eta_t^2 = \varepsilon_t^2 = 1$ and $\eta_t \varepsilon_t = 0$ in the above gives the realized variance in Equation (17). In benchmark equilibrium models, typically calibrated at the monthly frequency (e.g., Bollerslev, Tauchen, and Zhou (2009), Drechsler and Yaron (2011)), there is no clear counterpart to this multi-frequency approach where IV and RV are monthly, but where RV is estimated using daily data. In the models cited above, the definition of the IV_t is the risk-neutral expectation of the market return variance in month $t+2$. For example, IV at the end of January is the risk-neutral expectation at the end of January of market return variance in March. We define RV in a manner that avoids this one-month offset that is at odds with the data definitions. This brings the model closer to the moments from the data we use for calibration of the model parameters. While it is convenient to align the model definitions more closely to the timings used in the data, we note that our model results would also go through with alternate definitions of the variance risk premium used in earlier literature.

The dynamics of this risk premium depends directly on investors' variance expectations per Equation (19).

The equilibrium implied variance is:

$$IV_{t-1} = E_{t-1}^S [RV_t] + \delta_{IV}, \quad (21)$$

where $E_{t-1}^S [RV_t] = E_{t-1}^S [\Theta + \sigma_t^2] = \Theta + \bar{v} + \lambda x_{t-1}$ and $\delta_{IV} = \left(\frac{1}{2} \gamma^2 - \frac{1/\psi - \gamma}{1-1/\psi} \kappa (1-\phi) A \right) \omega^2$. The second term is an unconditional risk premium required by the agents due to the variance claim's exposure to shocks to variance. The conditional variance risk premium is then:

$$\begin{aligned} VRP_{t-1} &= IV_{t-1} - E_{t-1}^P [RV_t] \\ &= \delta_{IV} + E_{t-1}^S [RV_t] - E_{t-1}^P [RV_t] \\ &= \delta_{IV} + E_{t-1}^S [\sigma_t^2] - E_{t-1}^P [\sigma_t^2]. \end{aligned} \quad (22)$$

Thus, the dynamics of the variance risk premium share a component of the dynamics of the equity risk premium (Equation (15)), namely the mistakes agents' make in their variance expectation. Thus, agents will initially underreact to the variance shock, but subsequently overreact due to their sticky expectations, which leads to time-variation in the variance risk premium similar to that in the data. In fact, the lagged variance risk premium forecasts equity returns, as it does in the data and as it does in Bollerslev, Tauchen, and Zhou (2009). However, in the their model this is due to time-varying variance of variance, which we abstract from in this baseline version of our model.

Next, we calibrate the parameters of the model to assess if it can quantitatively account for the empirical observations discussed earlier.

3.3 Model calibration

We calibrate the model to moments that are at the heart of the issues we seek to address with the model. The data is monthly and from 1990 through April 2020. We

use the VIX_t^2 as the proxy for IV_t , where VIX_t is the option-implied risk-neutral volatility of stock returns over the next month. RV_t is calculated as the sum of daily squared log excess market returns in month t .

Panel A of Table 2 gives the parameters of the baseline model. We match the mean, autocorrelation, and variance of RV_t in the data with the parameters governing the objective variance dynamics in the model (\bar{v} , ρ , and ω). We set the risk aversion parameter γ by matching the equity premium and we take the elasticity of substitution ψ to be 2.2 as estimated in Bansal, Kiku, and Yaron (2016). Finally, we set ϕ to match the response of the variance risk premium (VRP) to a shock to RV in the model to that in the data, and we set λ to match the variance of IV_t in the model to the variance of the VIX_t^2 in the data. We set $\kappa = 0.97^{1/12}$, consistent with values used in earlier the literature and the average level of the price-dividend ratio in the data. Our moments of interest do not require us to estimate the time-discounting parameter, β , or the mean of dividend growth, μ .¹⁰ Table 2 gives the parameter values as well as the moments from the data used in the calibration. Note that the chosen value of ϕ is conservative in the sense that it is lower than that estimated using the survey data. We choose a ϕ of 0.5 which is in between the estimate from the survey data (0.87) and the estimate we get if we use the VIX (0.4) in place of surveys as a measure of market variance. Our choice of λ implies a value of the persistence of volatility dynamics under agents beliefs of $\phi + (1 - \phi)\lambda = 0.9$ which is again close to but more conservative than the survey estimates based on long term and shorter term volatility expectations (0.96).

Figure 1 shows the impulse-response of a one-standard deviation shock to RV under the rational beliefs (red, dashed line) and the calibrated subjective beliefs about next period's variance (solid, blue line). We show an alternative calibration in the black line that gives intuition for a lower value of ϕ . The subjective beliefs display a hump-shaped "slow-moving" response due to the beliefs loading on too many lags of variance and not enough on current variance. The difference between the rational

¹⁰This is why we set the log-linearization parameter κ exogenously to a standard value in the literature. In our monthly calibration, κ is very close to 1 and there is little sensitivity to reasonable variation in this parameter to the moments we target.

and the subjective beliefs – initial underreaction followed by longer-term overreaction – is what gives rise to deviations from the standard models in terms of the observed risk-return trade-off. The figure also shows how λ scales the impulse-response (see dotted black line versus solid blue line). The bottom panel plots the price dividend ratio. Notably, in our model calibration prices fall initially by more than in the rational case – this is because of the overextrapolation leading to too high of persistence for expected variance. However, consistent with underreaction, prices can continue to fall after the initial shock which breaks the standard risk-return tradeoff. In this way, we match that realized returns fall substantially when variance increases (consistent with a discount rate effect), yet next period returns are not high on average.

4. Model Comparison to Stylized Facts

4.1 Stylized Facts

We outline the main stylized empirical facts from the literature that we will target in our model, which we extend using more recent data. Since these facts are mainly already documented in the empirical literature, we relegate discussion of robustness to the appendix.

Data and Sources. We study US data from December 1991 to April 2020 for which we have stock market excess returns, the VIX (taken as the VIX on the last day of the month, thus representing forward looking volatility for the month), realized variance (computed as the sum of squared daily log returns within a month), and a measure of expected variance. Stock return data use the return on the S&P500 index over the risk-free rate taken from Ken French. In addition, we study variance swap returns, VIX futures returns, and straddle returns to capture claims on future volatility or variance from several sources. Our main data source for variance swap returns is Dew-Becker, Giglio, Le, and Rodriguez (2017) which provides variance swap returns based on dealer quotes from 1996-2017. VIX futures returns are from Cheng

(2018) from 2004-2017 and we supplement this with returns on the VXX ETF from 2017-2020 (as Cheng (2018) notes the VXX ETF tracks VIX futures returns and has a correlation near 1 with one month VIX futures in the overlapping sample). We further measure the variance risk premium (VRP) as the squared VIX minus realized variance as in Bollerslev et al. (2009). When using returns (e.g., variance swap or VIX futures) we take the negative of the returns, so the implication is the return for selling variance or being short the VIX. This means the unconditional premium for these returns is positive as the exposure to volatility is negative. Finally, we add data from Johnson (2017) which contains daily straddle returns and synthetic variance swap returns from underlying options which we cumulate to monthly returns from 1996-2019. This data has the advantage of avoiding over the counter prices or quotes that may have illiquidity concerns.

Expected Variance. To get a measure of objective expected variance at time $t - 1$, which we denote $\hat{\sigma}_{t-1}^2$, we use high frequency intraday data on the S&P500 and follow the HAR model used in Bekaert and Hoerova (2014). This uses squared five-minute returns at the daily, weekly, and monthly horizon to forecast next month's realized variance. Importantly, our forecast is done in real time so that at time $t - 1$ the forecast uses only information up to $t - 1$. Our intraday data are available in 1990 but we use a two year burn in period to construct the expected variance forecast. Further details are contained in Appendix D. Following our model, we proxy for subjective expected variance as a weighted average of past expected variance over the past six months, denoted $\frac{1}{\sum_{j=1}^6 \phi^j} \sum_{j=1}^6 \phi^j \hat{\sigma}_{t-j}^2$. The term in front of the sum ensures weights sum to one. We use $\phi = 0.5$ as in the model calibration, and only 6 lags since weights ϕ^j are close to negligible beyond this.

Empirical Results. Table 3 Panel A shows forecasting regressions for volatility claims on current expected variance and the exponential weighted average of past expected variance. In the first five columns, for each of the (short) volatility dependent returns, we see a negative coefficient on expected variance and a positive coefficient on the weighted average term, with the magnitude of the coefficients being roughly

similar. That is, the regression in all cases emphasizes the difference between the weighted average and objective expected variance as the model predicts. This says that when agents beliefs about variance are high relative to an objective forecast of variance, volatility claims are “expensive” such that being short volatility or variance is profitable over the next month. The monthly R^2 implies a reasonably high degree of predictability of all the volatility claims and the coefficients are significant in each of the columns. These results are closely related to those in Cheng (2018) who shows that increases in expected volatility negatively predict the VIX futures return. Following Cheng (2018), in the first column for VIX futures returns we use expected volatility in place of expected variance, and the weighted average of expected volatility. Using variance instead leads to similar results but allows less of a direct comparison to Cheng (2018). While the magnitudes of coefficients in columns (5)-(7) are easily interpretable in our model, (1)-(4) are less so. For straddle returns, the coefficient on the difference between expected variance and a weighted average of expected variance (the “mistake” in beliefs), is about -30. The standard deviation of the difference is 0.16% per month, so a 1 standard deviation increase in the difference translates to nearly a 5% change in the straddle risk premium. This is large given the unconditional monthly average return of 9.5% for straddles. Similarly, for variance swaps, a 1 standard deviation increase in the difference translates to about a 9% change vs the unconditional premium of 25%.

Column (6) shows that only expected variance predicts future realized variance, while the additional lags of expected variance do not. The coefficient is not statistically different from one and the R^2 is high at 47% confirming our real time expected variance forecast does predict future variance well (when we run this as a univariate regression, without the weighted average, we obtain a coefficient of 1.10 and standard error of 0.20). Column (7) shows that market implied variance (the squared *VIX*), however, loads substantially on the weighted average of past expected variance in addition to current expected variance as our model predicts. The differential pattern in columns 6 and 7 helps explain the patterns in the first five columns. The R^2 on the *VIX* is very high, indicating that we capture the majority of *VIX* variation from

these two components. While we have interpreted all these results through biases in beliefs as in our model, an alternative interpretation is through rational risk premiums. However, as noted in Cheng (2018), this interpretation is not natural because volatility claims are typically much riskier when expected volatility is high relative to the past. In Appendix Table 15 we show these return regressions at a weekly frequency to assess where the monthly return predictability is most concentrated. Consistent with Cheng (2018), the predictability is stronger in the first two weeks relative to that in weeks three and four.¹¹ This is consistent with our model because underreaction in our model is most pronounced in the near term.

Panel B shows the risk-return tradeoff regressions of the market excess return on expected variance and the VIX. Column (1) shows that expected variance has no predictive power for future returns, so that the risk return tradeoff is weak, echoing a much longer literature on this finding (Glosten et al., 1993; Whitelaw, 1994; Lettau and Ludvigson, 2003).¹² We note that these papers come to this conclusion over a variety of sample periods. This result also holds when using the VIX to forecast market returns in column (2). However, as column (4) shows, the variance risk premium ($VIX_{t-1}^2 - \hat{\sigma}_{t-1}^2$) predicts returns with a positive sign, while neither the VIX or RV predict returns by themselves. This confirms results from Bollerslev et al. (2009) in our extended sample. However, we note that the first two columns are puzzling from the perspective of the model in Bollerslev et al. (2009) in which the VIX alone is a strong predictor of returns, both because it embeds the variance risk premium and because it reflects expected future variance, and both of these strongly contribute to the equity risk premium in that model.

These facts are in line with findings from the empirical literature and this table should be mainly viewed as extending them in a more recent sample. As such, we leave extensive robustness checks to the appendix. We show subsample analysis to the financial crisis, we use volatility in place of variance to reduce dependence on high

¹¹See Cheng (2018) Figure 4.

¹²See also Brandt and Kang (2004); Moreira and Muir (2017, 2019); Eraker (2020) as a non exhaustive list of this literature.

variance observations, and we run weighted least squares to downweight high volatility periods which might overly influence our results. An important takeaway is that high variance periods are important, especially for the results in Panel B predicting equity market returns, while results predicting the variance claim returns (first 4 columns of Panel A) appear more robust. The latter evidence is true in our model as well – variance return predictability identifies our main mechanism more sharply while the equity return is exposed to additional sources of shocks. These results are echoed in Johnson (2019) who argues the evidence for the variance risk premium predicting stock returns is weaker than previously recognized. However, Bollerslev, Marrone, Xu, and Zhou (2014) study the ability of the variance risk premium to predict returns across eight different countries and argue it is a robust feature of the data. We provide international evidence along these lines in Appendix Table 18. Zhou (2018), a review article, finds that the variance risk premium helps predict returns across many asset classes including stocks, credit, currencies, and bonds and contains many additional references that find the variance risk premium helps predict equity returns (Drechsler and Yaron, 2011; Bollerslev, Todorov, and Xu, 2015).

Table 6, which we return to later, repeats this analysis at the firm level for US data which gives us significantly more observations compared to the aggregate results and hence provides further robustness to our main results. We see strikingly similar results for the variance risk premium with increases in firm-level variance negatively predicting the firm-level variance risk premium. This is true even when including time fixed effects suggesting that the facts we document hold even when removing aggregate movements in volatility.

4.2 Stylized Facts in the Model

We show these facts in our calibrated model in Table 4. To emphasize intuition and the importance of the bias to match the data, we show results as we vary the parameter ϕ . We consider the fully rational case in the model as a benchmark ($\phi = 0, \lambda = 0.72$) in column 2 and then use our calibration of $\lambda = 0.8$ in the remaining columns while

increasing ϕ to 0.3, 0.5, and 0.8.

We show the relation between risk and return (regression of future market return on current and past variance), volatility-managed alphas, the correlation between realized returns and variance shocks, the forecasting regressions of stock returns using the variance risk premium, the relation of the conditional variance risk premium with current and past variance, and the correlation of the model implied variance (VIX^2) and realized variance.

We first note that the dependence of future returns on current variance declines as we increase ϕ , while the dependence on past variance (ϕ weighted average) increases as we increase ϕ . The rational case in our model implies only current variance should predict returns, with zero weight on the past average, consistent with the basic risk return tradeoff intuition. With high enough ϕ , current variance can have zero or even negative relation to next period returns, while the average of past variance comes in positively for larger values of ϕ . These results are mirrored in the next row which documents the volatility managed alpha, with empirical numbers taken from Moreira and Muir (2017). The alpha is positive in the data, reflecting a weak risk-return tradeoff. As we increase ϕ and the risk-return tradeoff weakens, we increase the volatility timing alpha as well.

The contemporaneous correlation between realized returns and shocks to variance doesn't depend too strongly on ϕ , and quantitatively decreases slightly as we increase ϕ . The reason for this is that there are two effects which go in opposite directions in our model: the first is that a higher ϕ implies a lower reaction to volatility news through slow moving expectations. On the other hand, a higher ϕ leads agents expectations to be more persistent than the true volatility process. This second effect results in an effectively larger discount rate response to volatility shocks as they last longer in agents expectations, and tends to move prices more when volatility changes, while the first effect dampens the response to volatility news. Thus, our model keeps the negative correlation between returns and variance shocks even when underreaction to volatility is large.

Next, we show implications for the variance risk premium. First, the the variance risk premium forecasts stock returns strongly in the data, and the model can account for this once $\phi > 0$. The variance risk premium itself (here measured as VIX^2 minus a forecast of realized variance based on current and past variance) is negatively related to current variance and positively related to past variance. This is just the result from Table 1 that, relative to future variance, VIX loads more on past variance and less on current variance. The model can generate this pattern with $\phi > 0$ (which is required for the variance risk premium to have time-variation). As ϕ increases, so that expectations are slow moving, current variance forecasts this premium more negatively and past variance forecasts the premium more positively. This is simply because the mistake in expectations is larger when we increase ϕ .

Finally, empirically there is a strong correlation between implied variance (VIX^2) and realized variance (0.86). If VIX is influenced market by beliefs about variance this suggests that such beliefs are highly correlated with an objective forecast. In the model, this correlation weakens as we increase ϕ as it implies investors make larger mistakes. However, notably this correlation remains fairly high even for large values of ϕ . This may at first seem surprising, since it implies the subjective forecast of variance is strongly correlated with the objective measure in the model, meaning mistakes are actually fairly small, even when we increase ϕ . But note that volatility is persistent, and agents beliefs still put most weight on recent variance. Because volatility is fairly persistent, putting weight on lagged variance results in only a modest mistake, and these weights decay fairly quickly for longer lags (which have weight ϕ^k). This is an important point since it highlights that while the degree of bias in our model may appear large, persistence in variance actually implies only modest mistakes and, as we show shortly, modest profits from trading. Only in the case where ϕ is highest at 0.8 is this correlation in the model lower than what we see empirically.

Having discussed this intuition, we note that ϕ of around 0.5 does fairly well jointly accounting for the facts in the data in terms of the risk-return tradeoff, volatility managed alpha, variance risk premium dependence on past variance, and correlation

between VIX and realized variance. However, these results also suggest some tension in the model in terms of jointly matching all facts quantitatively. In particular, the risk-return tradeoff is even weaker in the data than the model with $\phi = 0.5$ (this is also reflected in the volatility managed alpha), and a large ϕ is needed to match this moment. On the other hand, the variance risk premium results favor a more modest value of ϕ for the magnitudes of the variance risk premium on past variance to not be too large. Most important, however, the model with biased beliefs matches the moments on balance better than the rational benchmark.

4.3 Volatility managed portfolios

Moreira and Muir (2017) document that volatility-managed factor portfolios yield positive alpha in standard Gibbons, Ross, Shanken (1987) type return regressions. For the market factor they consider a strategy that each period has a portfolio weight in the market that is inversely proportional to RV . They show that the alpha of such a strategy relative to the buy-and-hold market factor can be approximated by:

$$\alpha = -\frac{c}{E[RV_t]} \text{Cov} \left(E_t[RV_{t+1}], \frac{\mu_t}{E_t[RV_{t+1}]} \right), \quad (23)$$

where $E_t[r_{t+1} - r_{f,t}] = \mu_t$ and where c is a constant that scales the timing portfolio to have the same return variance as the market. Since there is no strong risk-return trade-off in the model with biased beliefs, the covariance above is negative, which gives rise to a positive alpha as in the data. Our simple variance process allows negative values for variance, therefore to calculate this covariance we use the approximation:

$$\frac{\mu_t}{E_t[rv_{t+1}]} \approx \frac{\bar{\mu}}{\bar{v} + \Theta} + \frac{1}{\bar{v} + \Theta} (\mu_t - \bar{\mu}) - \frac{\bar{\mu}}{(\bar{v} + \Theta)^2} (E_t[rv_{t+1}] - \bar{v} - \Theta), \quad (24)$$

and report the alpha for the volatility-managed market portfolio in Table 2 as:¹³

$$\alpha \approx -0.6 \times \frac{1}{\bar{v} + \Theta} \text{Cov} \left(E_t [rv_{t+1}] , \mu_t - \frac{\bar{\mu}}{\bar{v} + \Theta} E_t [rv_{t+1}] \right), \quad (25)$$

which is equal to 1.4% annualized — in the same order of magnitude as the 4.9% Moreira and Muir (2017) document. As emphasized in the last section, to fully match this alpha would require a larger value of ϕ .

4.4 Comparison to Bollerslev Tauchen Zhou

Standard asset pricing models typically struggle with the facts outlined above because they suggest that an increase in risk (volatility) will be associated with heightened risk premiums at all horizons, and in particular this relationship will be strongest in the near term but decays with horizon as volatility is mean-reverting. For example, Moreira and Muir (2017) show the risk return tradeoff in leading models is strong, including models with habit formation (Campbell and Cochrane, 1999), long run risk (Bansal and Yaron, 2004; Drechsler and Yaron, 2011), rare disasters (Barro, 2006; Wachter, 2013) and intermediary models (He and Krishnamurthy, 2013). Further, expected returns will typically rise most on impact and will gradually fade through time as volatility fades. We are not aware of leading equilibrium asset pricing models which produce a temporary decline in risk premiums followed by a delayed increase.

Bollerslev, Tauchen, and Zhou (2009; BTZ hereafter) provide a rational benchmark model of the dynamics of the variance risk premium and the conditional equity premium and therefore overlaps with some of the stylized facts we seek to explain. In this model, the representative agent has rational expectations and the volatility of volatility follows a mean-reverting process. The time-variation in the amount of variance risk gives rise to time-varying expected returns to variance swaps and the market risk premium. In particular, the lagged variance risk premium in their model

¹³Moreira and Muir find that $\frac{c}{E[RV_t]} \approx 0.6$ in the data, and we simply use this value to compute the volatility timed portfolio alpha implied by our model.

predicts future excess market returns as in the data. To highlight how the subjective beliefs model of this paper differs in terms of asset price dynamics, Table 5 compares our model directly to BTZ based on the stylized facts we target.¹⁴

Both models match that the variance risk premium positively predicts stock returns, which is the main fact that BTZ is calibrated to match. Where the models strongly differ is on the risk-return tradeoff. First, the BTZ model implies a coefficient of 10 for the risk-return tradeoff, while in the data this is much weaker. Because of this, their model can't match the return predictability regressions on past variance or generate a positive volatility managed alpha. Next, a salient fact in the data is that while the difference $VIX_t^2 - RV_t$ is a strong predictor of market returns, neither lagged RV nor the VIX (or the VIX^2) are strong return predictors on their own — a fact that BTZ documents in their Table 3 (page 4482 in BTZ (2009)). In the BTZ model, including both VIX_t^2 and RV_t to forecast stock returns, both coefficients are strongly positive, while in the data the coefficients have opposite signs.

4.5 Impulse Responses: Data vs Model

To succinctly summarize the empirical patterns in a way that is easy to relate to our model, we estimate a first order vector autoregression (VAR) with expected variance $\hat{\sigma}^2$ (formed as before), the variance risk premium $VIX^2 - \hat{\sigma}^2$, excess market returns, and the log market price-dividend ratio as the variables in the state vector.¹⁵ Expected variance is ordered first, so all variables can respond contemporaneously to a shock to $\hat{\sigma}^2$. Figure 2 plots impulse responses of the equity premium and variance risk premium to a shock to $\hat{\sigma}^2$.

The variance risk premium goes negative after the shock before slowly rising and

¹⁴To give the BTZ model a better chance at matching these patterns, we recalibrate the objective variance process in their model to match that in the data. Their calibration implies a counterfactually high persistence of RV_t .

¹⁵The log price dividend ratio is from CRSP based on value weighted returns. We construct the price dividend ratio as the sum of dividends over the past year divided by the current price. We find similar results using other price measures for example the cyclically adjusted price to earnings ratio (CAPE) from Robert Shiller's website.

becoming positive beyond month 3 (Cheng, 2018). Similarly, the equity risk premium, if anything, initially falls but rises further out. While the initial response of the equity risk premium is not statistically different from zero here (consistent with a weak or negative risk-return tradeoff), note that the upper bound is small which we will show relative to leading rational models. Most notably, the premiums have a hump-shaped pattern: they appear low initially but rise as future volatility falls. This is in contrast to the standard benchmark model, with the equity premium being affine in expected variance. In this setting the risk premium response should peak immediately and roughly mirror the response of future variance from period 1 onwards, with a spike upwards followed by a decline as future variance mean reverts. The shaded regions indicate 95% confidence intervals based on bootstrapping the residuals in the VAR. We consider alternative specifications of the VAR in the appendix that rely less on high variance periods and argue that the main patterns documented here are robust.

Figure 2 also shows impulse responses in the model vs the data. The impulse-responses from the calibrated model are given in the blue dashed lines, while the impulse-responses from the model assuming $\phi = 0$ and $\lambda = \rho$ (the rational case) are given in the red dash-dotted lines for comparison.

The impact of a shock to expected variance on the conditional market and variance risk premiums is very different across the two models. In particular, the response in the rational model is to immediately increase the conditional risk premium due to the usual risk-return trade-off (Equation (15)), at odds with the empirical facts. In our model, however, the response of the conditional equity premium as measured in the VAR is, as in the data, initially negative. This is due to the mistake investors are making in their variance forecast as shown in Figure 1. The equity premium subsequently overshoots due to the slow-moving expectations of the agents, consistent with the pattern in the data. The same is true for the variance risk premium, although in this case the pattern is slightly stronger than the data as its dynamics are only affected by the mistake in expectations (see Equation (22)). The rational version of the model has no effect on the variance risk premium from a shock to expected variance, again at odds with the

data. While one could change this, for example in a model where the volatility of volatility shocks varied over time, it would be difficult to generate both the negative initial response and hump shaped dynamics. In particular, because the volatility of volatility shocks tend to be positively correlated with the level of volatility, this would tend to push the initial response of the variance risk premium up. In other words, volatility claims tend to be riskier after a shock to expected variance (Cheng, 2018), and so a rational model with this feature would have a harder time matching the data.

5. Additional Evidence, and Alternative Explanations

5.1 Firm Level Analysis

We revisit our stylized aggregate facts at the firm level (stock level) in Table 6. We take implied volatility from OptionMetrics at the stock level from 1996-2017 and use daily and monthly return data from CRSP for the stocks in the merged OptionMetrics sample (6,489 unique stocks over the sample). Implied volatility is measured on the last day of the month and measures option implied volatility over the subsequent month (30 days) for at the money options. Realized variance is computed using the daily returns within a given month. Our measure of the variance risk premium is then $IV_{i,t}^2 - RV_{t+1}$ which is the implied variance over the next month minus the actual realized variance over the next month. We use daily log stock returns from CRSP and computed the sum of squared log returns over the following month's trading days as our measure of realized variance.

Similar to our results in Table 3, we forecast equity risk premiums, variance risk premiums, and future realized variance over the next month but we use the change in realized variance from month t to $t - 6$. We use the change over six months, rather than the weighted average of all realizations over six months, for several reasons. Most importantly, this helps account for quarterly earnings announcements at the firm level

which are a big driver of firm level volatility and result in quarter fixed effects at the firm level with realized volatility being high during months with earnings announcements. By differencing the six month lag we both account for unconditional firm level effects and effects of quarterly earnings announcements on firm level volatility. We winsorize lags of realized variance at the 95th percentile, though importantly we do not winsorize future realized variance so that the left hand side in this case is still the realized variance risk premium. In unreported results we find similar result without winsorization, but the main advantage is we find much stronger predictive power for future variance with winsorization due to substantially more noise in firm level realized volatility estimates compared to the aggregate. We also find qualitatively similar results in several other specifications, including using log of realized variance or using volatility in place of variance, though these results are omitted for space.

The results show that increases in volatility over 6 months negatively forecast variance risk premiums, but positively forecast future variance. The coefficients for predicting future variance and future variance risk premiums are highly statistically significant with or without time fixed effects (standard errors are double clustered by time and firm). The results with time fixed effects are especially important because these remove any aggregate movements in firm level variance or variance risk premiums. By removing aggregate effects, we are more likely capturing purely idiosyncratic movements in realized variance that helps push against a risk-based story for our results. These results are also similar in spirit to Potoshman (2001) who argues for underreaction in option prices in an earlier sample.

The firm level analysis achieves two things. First, it provides robustness to our aggregate results which rely on fewer observations. Second, it provides more insight into whether the variance risk premium results we document are likely driven by true economic risk premiums (compensation for risk) or whether they are instead more likely driven by biased expectations and underreaction to changes in volatility. As stressed earlier, the aggregate results are not consistent with standard risk based models since higher risk (more variance) should, if anything, imply a higher rather

than lower risk premium. Nevertheless, it is always possible to construct a model in which investor preferences move in such a way to match the aggregate evidence. The firm level evidence is more powerful since we think of firm level variance as largely idiosyncratic, especially in our second specification where we include time fixed effects in the regression to remove any common components of firm level variance. Hence, we would likely expect a much smaller effect at the firm level from a risk premium story due to variance shocks being more idiosyncratic at the firm level. Instead, we recover a coefficient of around -0.1 for the firm level VRP, which is in line with the magnitudes we observe in the aggregate results.

Further in this dimension, at the firm level we see a weakly negative but not significant coefficient for the equity risk premium. This is exactly what we expect in the model if agents do not price idiosyncratic firm level risk. In our aggregate results, investors should require more compensation for the increase in variance and this mechanism combined with biased beliefs results in the negative coefficient on the equity risk premium. Absent this channel, we would only expect the results to hold for the variance risk premium. Taken together, the firm level results support our main hypothesis that agents initially underreact to changes in variance and that this is reflected in implied volatilities.

5.2 Evidence on Actual Trading Behavior

Hoopes, Langetieg, Nagel, Reck, Slemrod, and Stuart (2017) show evidence that investors do react to changes in volatility with more sophisticated investors and older investors responding more strongly. Specifically, they show that higher income and older investors sell more aggressively following increases in volatility. This is reasonable in our model if one takes higher income investors to be more sophisticated and less prone to the expectations bias in our paper. Similarly, it is possible that investors learn more about the volatility process with time (as the evidence on investor experience suggests they would) and hence exhibit less of a bias as they are older. Giglio et al. (2019) also find that agents views about risk are directly informative for their

portfolio decisions, with higher expectations of stock market risk being associated with a lower allocation to stocks. A shortcoming of our model is that it features a representative investor and so does not speak directly to this evidence (as there is no trade in equilibrium), though modest extensions of the model which allow for differences in the amount of bias would naturally be consistent with the evidence on trading behavior in Hoopes et al. (2017).

5.3 Term Structure Evidence

The model has implications for the conditional term structure of variance risk premiums. We showed that the model produces underreaction to volatility for 1 month variance claims. A natural question is how this extends to longer horizon variance claims.

Empirically, we have return data for variance swaps, VIX futures, and straddles for multiple maturities. We show these implications in Table 7 where we run the same predictive regressions as before using all maturities and normalize the coefficients by dividing by 100. For VIX futures and straddles, we focus only on the sample where we have non-missing observations at all maturities using data provided by Johnson (2017). We do the analogous procedure in the model where we regress the one month return of a variance swap with maturity k on expected variance and a weighted average of expected variance. Overall, current expected variance continues to negatively predict returns for all maturities while past variance positively predicts. The slow-moving nature of the extrapolative expectations means agents believe variance is highly persistent, which in turn implies that mistakes in conditional variance expectations also matter for long-horizon claims. This generates excess volatility in the long end of the variance term structure, matching the spirit of the empirical findings in Giglio and Kelly (2018).

5.4 Credit Returns

Table 8 again runs predictive regressions but uses returns that depend on credit risk. These returns are particularly interesting for our story because credit risk is sensitive to changes in volatility which affect default risk (Merton, 1974). Credit returns are constructed from three sources. The first uses CDS excess returns from He, Kelly, and Manela (2017) who form 20 CDS portfolios based on credit risk. We equally weight across these 20 portfolios to form a single CDS excess return. The second column uses the Barclays total return index for high yield and investment grade corporate bonds. We compute the excess return as the difference between the return of these two series. The third is the (negative) change in the BaaAaa spread from Moody's. This computes the change in this yield spread as a proxy for the credit return as used in López-Salido, Stein, and Zakrajšek (2017).

All three credit return series point in the same direction and support our main results. Current volatility negatively predicts the returns and the moving average of past variance positively predicts. Similar to our earlier regressions, the absolute magnitudes of the coefficients are similar, meaning the change in volatility relative to the moving average explains the predictability. Both coefficients are highly significant in all three cases.

5.5 Model Shortcomings and Extensions

Our model is intentionally simplified to focus on one particular channel – beliefs about volatility – in influencing the prices of financial market claims. Here we outline limitations of the baseline model and discuss useful extensions.

Our model has implications for the price dividend ratio that are clearly rejected in the data. Most importantly the model – if taken literally – says the dividend yield is perfectly correlated with the VIX in levels, which is clearly counterfactual. In particular, empirically the dividend yield is much more persistent than the VIX, though the two are correlated (in particular, VIX and the dividend yield both tend to

go up in bad times). This highlights that the dividend yield is likely also influenced by forces outside our model. An extension of our model with time-varying expected dividend growth would generate additional movements in the dividend yield and, if these growth rates were highly persistent, could generate the difference in persistence and also lower the correlation between the dividend yield and the VIX in levels. Still, in the data a much more robust fact is that changes in prices are highly negatively correlated with changes in the VIX (prices go down when VIX goes up) as in our model, and this result can still hold even in an extension with slow time-varying dividend growth. This highlights why we choose not to use dividend yield related moments when targeting the parameters in our calibration even though our baseline model has implications for these moments.

In the main model we put the stochastic volatility on the cash flow process. However, this is not particularly important and our paper doesn't have much to say whether this is discount rate or cash flow volatility. In our model discount rate volatility would still be priced and would still imply the highest premium at shorter horizons. However, a lower price of discount rate shocks could lower the risk-return tradeoff further. We make the assumption of stochastic cash flow volatility for convenience.

In our model, RV follows an AR(1) process. Thus, the optimal predictor of variance in the model is simply lagged RV, while the VIX has no marginal predictive power. In the sample we consider (1990-2020), this is in fact a close approximation – the increase in R^2 in a forecasting regression for RV is negligible when adding the VIX as a predictive variable in addition to lagged RV. In other samples, however, the VIX has stronger marginal predictive power. Chernov (2007) shows that this is indeed the case in the 1986-2001 sample, but he also points out that the VIX alone cannot be the optimal predictor of future variance since the variance risk premium is time-varying in the data.¹⁶ This latter fact is consistent with our model, where the VIX reflects the expectational errors of the agents in addition to the objective variance forecast. In

¹⁶Related, Chernov (2007) finds that the coefficient on the VIX in variance forecasting regressions is significantly below 1, consistent with substantial variation in the variance risk premium. This is also the case in our sample.

terms of the former, it is straightforward to extend our model to make the VIX have marginal predictive power for future variance. In particular, if we assume investors each period observe a noisy signal about next period’s RV that is uncorrelated with current RV (i.e., a time t signal correlated with η_{t+1}), the VIX will reflect this added information as the signal affects agents’ beliefs. Importantly, such an extension will not affect the projection of the VIX or agents’ beliefs about future variance onto lagged RVs, which is the focus of the bias we consider in our model. We provide this extension in Appendix C.

Finally, both the individual survey evidence from Giglio et al. (2019) and the evidence on actual trading behavior in Hoopes et al. (2017) suggest that heterogeneity in beliefs about risk is important, though our model features a representative agent. A limitation of our analysis is we can’t speak directly to this evidence.

A natural extension of our model is to introduce a set of agents (risk averse arbitrageurs) that have rational expectations. In such a model the wealth share of the arbitrageurs becomes a new state-variable. Generally speaking, the degree of mispricing will increase with impediments to arbitrage and decrease with the arbitrageur wealth share. There are several features of the data that are consistent with this. For instance, we document a stronger pattern when current variance is high. Since arbitrageurs on average are short volatility, these are times when they recently have suffered losses on their arbitrage positions and therefore hold less wealth, scaling back their positions. This is consistent with data on hedge fund positions as shown by Cheng (2018). Further, our results are strong after 2010, which coincides with a period of tighter bank regulation after the Financial Crisis. Also, our results that a higher ϕ is needed to match the moments in the stock market than those in the variance market is consistent as there are many other shocks to stock prices than variance shocks, making the arbitrage riskier in this market. While an extension to such a heterogeneous agent model would be quite interesting, it is beyond the scope of this paper.

One implication and limitation of our representative agent model is that expectations of returns and variance should be strongly positively related. However, we

find empirically that survey expectations of returns and variance from the Graham and Harvey CFO survey are essentially uncorrelated: the adjusted R-square of survey expected returns on our measure of expected variance is zero (results untabulated but available on request). Giglio et al. (2019) show at the *individual* level expectations returns and expectations of a crash in the stock market are strongly *negatively* correlated. These results are natural if agents have heterogeneous beliefs, which our representative agent model doesn't capture. For example, as Giglio et al. (2019) mention, if an individual investor believes there is a higher chance of a market crash than other agents, they will think future risk is high and also that expected returns are low. The reason is this agent is small relative to the market, so their beliefs barely affect market prices if at all. The agent will thus view stock prices as too high and future returns as too low since they don't reflect the risk of a crash and these agents decrease their portfolio allocation in their data as expected. This would not be the case if this were the only investor in the market, as prices would then need to fall substantially to keep the agent from selling.

However, as we aggregate up the views of agents, the correlation between expectations of returns and expectations of risk will increase. If a large subset of investors believes the risk of a crash is high, then this view will be reflected in aggregate market prices and thus raise their view of expected returns. Our CFO survey results are consistent with this view: the correlation between aggregated CFO views on variance and expected returns in our survey evidence is much higher than the individual level results of Giglio et al. (2019), though still not nearly as high as in our model with a single investor. There are typically about 300 responses to the CFO survey (Graham and Harvey, 2008) which is far from the aggregate expectations of the entire market. This view is also related to Greenwood and Shleifer (2014) who acknowledge not all investors can have low expectations of returns or be return extrapolators – there must be other investors on the other side which are absent in these surveys.

We view this heterogeneity in beliefs as important, and feel that extensions of our model which account for these facts at the individual level are fruitful areas to

study, as are extensions where some agents extrapolate risk and some have rational beliefs. However, accommodating this in our current framework would add a substantial amount of complexity, and we therefore leave this study to future research.

5.6 Rational Arbitrageurs and Trading Profits

The biased beliefs of the representative agent in our model leads to errors in expectations about conditional stock market variance. In this section, we analyze the gains a rational agent would achieve through optimal timing of the variance claim, assuming this agent trades only an infinitesimally small amount and thus does not affect prices. We focus on the variance claim because this is where trading on the mistake is the most direct and has the highest sharpe ratio compared to the equity claim.

From Equation 22, we have that variation in the variance risk premium is due to differences between the agent's subjective belief about conditional stock market variance and the true conditional stock market variance. Consider a myopic mean-variance optimizing rational agent that is timing the variance claim based on the current expectational errors of the representative agent. The rational agent's position in the variance claim is then:

$$\omega_t = \frac{VRP_t}{\gamma\sigma_\eta^2}, \quad (26)$$

where γ is a risk aversion coefficient and ω_t is the number of units short in the variance claim. That is, when the expected return to shorting variance is high (low) the short variance claim position is scaled up (down). To focus on the effects of timing, we consider the Sharpe ratio of a strategy that is on average variance neutral. That is, we compute the Sharpe ratio for the strategy with returns $(\omega_t - E[\omega_t])(IV_t - RV_{t+1})$ both in the model and in the data, noting that the risk aversion parameter γ does not need to be specified for the Sharpe ratio calculation.

To estimate the VRP in the data, we run the regression:

$$IV_{t-1} - RV_t = \alpha + \beta_1 \hat{\sigma}_{t-1}^2 + \beta_2 IV_{t-1} + \eta_t, \quad (27)$$

where IV is the squared VIX and $\hat{\sigma}^2$ is the HAR-estimate of the true conditional market variance. The estimated VRP is then $\widehat{VRP}_t = \hat{\alpha} + \hat{\beta}_2 \hat{\sigma}_t^2 + \hat{\beta}_2 IV_t$. This is the optimal regression to run within the model for estimating the conditional variance risk premium. Table 9 shows the results from this regression both in the data and in the model. The first column shows that the $\hat{\beta}_1$ and $\hat{\beta}_2$ coefficients in the data are -0.85 and 0.87 , while the R^2 is 19% . We note that the regression coefficients are not statistically different from -1 and 1 , respectively. The annualized sample Sharpe ratio from the above timing strategy is 0.41 .

In column 2, we run the same regression on simulated model data. The regression coefficients in the model are indeed -1 and 1 , while the R^2 is 12% , slightly lower than the regression in the data. The timing Sharpe ratio in the model is, however, quite a bit larger than that in the data at 1.17 .

This suggests that the mistakes in variance expectations within our model are too large. However, the high timing Sharpe ratio is not a robust feature of our model. In particular, in the baseline model we for analytical convenience assume conditional variance is Normally distributed. Realized variance in the data, however, has sample skewness and kurtosis of 6.89 and 69 , respectively. To illustrate the effect of a more realistic data process for variance, without altering the expectation formation mechanism, we simulate variance as:

$$\sigma_t^2 = \bar{v} + \rho (\sigma_{t-1}^2 - \bar{v}) + \omega \sqrt{\sigma_{t-1}^2} \tilde{\eta}_t, \quad (28)$$

where $\tilde{\eta}_t = \eta_t^{(1)} + \eta_t^{(2)} J_t$ is a mixture of Normals. In particular, $\eta_t^{(1)} \sim N(0, 1)$, $\eta_t^{(2)} \sim N(0, \sigma_J^2)$, and J_t is a random indicator variable that each period equals 1 with probability p and 0 otherwise. Thus, we have both time-varying volatility and 'jumps' in the variance process that along with the square root process allow us to achieve high skewness and kurtosis. Eraker, Johannes, and Polson (2003) shows that such dynamics are important to account for the dynamic behavior of conditional market return variance. In continuous-time this process never goes negative, and in our discrete-time

simulations we simply set any negative variances to zero, though we note that such observations are rare and small in magnitude. We calibrate this process to match the first four moments of realized variance by setting $\omega = 0.0738$, $\sigma_J = 5$, $p = 1/30$, using the same persistence parameter as before, $\rho = 0.71$. Under the assumption that variance risk is not priced and with the same process for x_t as before, the impulse-response of the variance risk premium to a variance shock is the same as in the baseline model at all horizons. In other words, the main features of our model remain the same in this simulation.

The third column of Table 9 shows that this more realistic specification of variance dynamics again gives regression coefficients of -1 and 1 as in the baseline case. The R^2 is 9% and the Sharpe ratio from timing is now much smaller at 0.30, slightly lower than that in the data. We note that an annualized Sharpe ratio around 0.5 is similar to that of other 'anomalies' such as momentum or carry.

5.7 Alternative Explanations

Moreira and Muir (2017) show that leading equilibrium asset pricing models (e.g., habits models, intermediary models, long run risk, and rare disasters) typically imply a strong risk return tradeoff and hence won't match the facts that volatility is a weak predictor of returns.

What other models could explain our results? While some models can indeed match some of our stylized facts, we are not aware of models that can quantitatively jointly match them. This is especially true regarding the firm level analysis which relies solely on idiosyncratic movement in firm level variance, and our survey expectation data which suggests slow moving volatility expectations. We briefly discuss models with rational inattention and heterogeneity in terms of which facts they can explain.

Models featuring infrequent rebalancing and/or rational inattention (Abel, Eberly, and Panageas, 2013) at first appear promising but won't easily match the facts that we document. Essentially, even if a small fraction of traders is attentive at any given time, they will still price in changes to volatility. Similarly, even if agents know they

will not rebalance again soon they will still ensure a risk-return tradeoff at the horizon at which they expect rebalance. This will result in a risk-return tradeoff that resembles the standard case. Further, Hoopes et al. (2017) show evidence that investors do react to changes in volatility with more sophisticated investors (e.g., those in highest income brackets) responding most quickly. That is, it does not appear agents are not aware and do not act on changes in volatility. Finally, we are unaware of these models being able to easily match the variance risk premium dynamics, and particularly the firm level facts or the survey expectation data.

Heterogeneous agents models can potentially explain the weak risk-return relation, and in these models this risk-return relation can even go negative depending on the wealth distribution (e.g., Gârleanu and Panageas (2015), Longstaff and Wang (2012)). These models feature a conditional risk-return tradeoff that is typically positive for most parts of the stationary distribution but can turn negative in the worst states. For the unconditional risk-return tradeoff to be weak, calibrations of the models would typically also require that the correlation between contemporaneous returns and volatility would be weak, which is not the case. Further, in typical calibrations that aim to match other moments (e.g., the equity premium) the risk-return tradeoff is positive. It is not obvious these models would be able to explain the mismatch in frequencies that we observe, e.g., with risk premiums initially declining but then rising further out after volatility increases. Further, it is less clear that these models can match the variance risk premium results, the firm level results we document (which rely on firm level idiosyncratic variance rather than aggregate variance), and the slow moving expectations from our survey data. In these models the relation between volatility and expected returns is only weak or negative in bad times though we don't find such a conditional relation in the data (for example, the strong negative correlation of returns and realized variance innovations is robust in good and bad market conditions).

Since high current volatility reflects high uncertainty about asset values, it is natural to consider learning about parameters and/or latent states governing economic dynamics as a possible explanation for the empirical facts. Consider as an example a

model where there are two latent high volatility regimes – one regime is short-lived, the other regime is long-lived. Let’s say current volatility is high, so investors know the economy is in one of these high volatility regimes, but they do not know which one. In a stationary learning environment with a long sample, investors prediction for future volatility will on average be right given available information so there should not be return predictability in returns other than that coming from higher expected risk. However, the learning may be nonstationary and/or the sample may be relatively small when it comes to these types of events. Thus, ex post predictability that was not ex ante actionable may appear in forecasting regressions. For instance, at the onset of the financial crises there was anecdotally substantial fear of a long-lived crisis as volatility spiked to extreme levels. In fact, volatility instead mean-reverted quite quickly as investors learned that the financial meltdown-scenario was averted. Note that investors in this case would have overestimated future variance, as the high volatility regime ex post turned out to be of the short-lived variety. Thus, investors would appear to have overreacted to the initial high volatility shock, which would lead to high variance forecasting high returns. This is the opposite of what we find. Thus, it is not immediately clear that a learning channel will explain the empirical results. That said, the concern of small-sample issues is a valid one, which is one of the reasons we also examine the cross-section of firm-level options. The consistent results in a pure cross-sectional analysis suggests that our results are not due to small-sample concerns.

6. Conclusion

We show that underreaction followed by delayed overreaction to volatility news can match many empirical facts surrounding volatility and risk premiums that are puzzling from leading equilibrium asset pricing models. We achieve this feature by assuming agents expectations of volatility are slow moving and extrapolative and we show this is directly consistent with expectations in survey data. In particular, our model matches the weak overall risk-return tradeoff and matches the dynamic responses of both the

equity premium and variance risk premium following shocks to variance.

We are able to account for the fact that shocks to volatility are indeed associated with negative contemporaneous realized returns through a discount rate channel though still the relation between volatility and next period returns are weak. Finally, in our model the variance risk premium predicts returns more strongly than either variance or implied variance, as in the data. Survey evidence directly supports slow moving expectations about volatility, as does evidence using firm level option prices.

References

Abel, Andrew B, Janice C Eberly, and Stavros Panageas, 2013, Optimal inattention to the stock market with information costs and transactions costs, *Econometrica* 81, 1455–1481.

Bansal, Ravi, and Amir Yaron, 2004, Risks for the long run: A potential resolution of asset pricing puzzles, *Journal of Finance* 59, 1481–1509.

Barberis, Nicholas, Robin Greenwood, Lawrence Jin, and Andrei Shleifer, 2015, X-capm: An extrapolative capital asset pricing model, *Journal of Financial Economics* 115, 1 – 24.

Barberis, Nicholas, Andrei Shleifer, and Robert Vishny, 1998, A model of investor sentiment, *Journal of Financial Economics* 49, 307 – 343.

Barro, Robert J., 2006, Rare disasters and asset markets in the twentieth century, *The Quarterly Journal of Economics* 823–866.

Bekaert, Geert, and Marie Hoerova, 2014, The vix, the variance premium and stock market volatility, *Journal of econometrics* 183, 181–192.

Ben-David, Itzhak, John R Graham, and Campbell R Harvey, 2013, Managerial mis-calibration, *The Quarterly Journal of Economics* 128, 1547–1584.

Bollerslev, Tim, James Marrone, Lai Xu, and Hao Zhou, 2014, Stock return predictability and variance risk premia: statistical inference and international evidence, *Journal of Financial and Quantitative Analysis* 633–661.

Bollerslev, Tim, George Tauchen, and Hao Zhou, 2009, Expected Stock Returns and Variance Risk Premia, *The Review of Financial Studies* 22, 4463–4492.

Bollerslev, Tim, Viktor Todorov, and Lai Xu, 2015, Tail risk premia and return predictability, *Journal of Financial Economics* 118, 113–134.

Bordalo, Pedro, Nicola Gennaioli, Yueran Ma, and Andrei Shleifer, 2018, Over-reaction in macroeconomic expectations, Working paper, National Bureau of Economic Research.

Bouchaud, Jean-philippe, Philipp Krueger, Augustin Landier, and David Thesmar, 2019, Sticky expectations and the profitability anomaly, *The Journal of Finance* 74, 639–674.

Boutros, Michael, Itzhak Ben-David, John R Graham, Campbell R Harvey, and John W Payne, 2019, The persistence of miscalibration, *Available at SSRN 3462107* .

Brandt, Michael W., and Qiang Kang, 2004, On the relationship between the conditional mean and volatility of stock returns: A latent var approach, *Journal of Financial Economics* 72, 217 – 257.

Brooks, Jordan, Michael Katz, and Hanno Lustig, 2018, Post-fomc announcement drift in us bond markets, Working paper, National Bureau of Economic Research.

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

Cheng, Ing-Haw, 2018, The VIX Premium, *The Review of Financial Studies* 32, 180–227.

Coibion, Olivier, and Yuriy Gorodnichenko, 2015, Information rigidity and the expectations formation process: A simple framework and new facts, *American Economic Review* 105, 2644–78.

Collin-Dufresne, Pierre, Michael Johannes, and Lars A. Lochstoer, 2016, Asset Pricing When ‘This Time Is Different’, *The Review of Financial Studies* 30, 505–535.

Daniel, Kent, David Hirshleifer, and Avanidhar Subrahmanyam, 1998, Investor psychology and security market under- and overreactions, *The Journal of Finance* 53, 1839–1885.

Dew-Becker, Ian, Stefano Giglio, Anh Le, and Marius Rodriguez, 2017, The price of variance risk, *Journal of Financial Economics* 123, 225 – 250.

Drechsler, Itamar, and Amir Yaron, 2011, What’s vol got to do with it, *Review of Financial Studies* 24, 1–45.

Eraker, Bjorn, 2020, Predictability puzzles, *Available at SSRN 3625709* .

Eraker, Bjørn, Michael Johannes, and Nicholas Polson, 2003, The impact of jumps in volatility and returns, *The Journal of Finance* 58, 1269–1300.

French, Kenneth R., G. William Schwert, and Robert F. Stambaugh, 1987, Expected stock returns and volatility, *Journal of Financial Economics* 19, 3–29.

Gârleanu, Nicolae, and Stavros Panageas, 2015, Young, old, conservative, and bold: The implications of heterogeneity and finite lives for asset pricing, *Journal of Political Economy* 123, 670–685.

Giglio, Stefano, and Bryan Kelly, 2017, Excess Volatility: Beyond Discount Rates*, *The Quarterly Journal of Economics* 133, 71–127.

Giglio, Stefano, Matteo Maggiori, Johannes Stroebel, and Stephen Utkus, 2019, Five facts about beliefs and portfolios, Working paper, National Bureau of Economic Research.

Glaeser, Edward L., and Charles G. Nathanson, 2017, An extrapolative model of house price dynamics, *Journal of Financial Economics* 126, 147 – 170.

Glosten, Lawrence R., Ravi Jagannathan, and David E. Runkle, 1993, On the relation between the expected value and the volatility of the nominal excess return on stocks, *Journal of Finance* 48, 1779–1801.

Graham, John R., and Campbell R Harvey, 2008, The equity risk premium in 2008: evidence from the global cfo outlook survey, *Available at SSRN 1162809* .

Greenwood, Robin, and Andrei Shleifer, 2014, Expectations of Returns and Expected Returns, *The Review of Financial Studies* 27, 714–746.

He, Zhiguo, Bryan Kelly, and Asaf Manela, 2017, Intermediary asset pricing: New evidence from many asset classes, *Journal of Financial Economics* 126, 1–35.

He, Zhiguo, and Arvind Krishnamurthy, 2013, Intermediary asset pricing, *The American Economic Review* 103, 732–770.

Hong, Harrison, and Jeremy C Stein, 1999, A unified theory of underreaction, momentum trading, and overreaction in asset markets, *The Journal of finance* 54, 2143–2184.

Hoopes, Jeffrey, Patrick Langetieg, Stefan Nagel, Daniel Reck, Joel Slemrod, and Bryan Stuart, 2017, Who sold during the crash of 2008-9? evidence from tax return data on daily sales of stock, Working paper, National Bureau of Economic Research.

Jiang, Zhengyang, Arvind Krishnamurthy, and Hanno Lustig, 2018, Foreign safe asset demand and the dollar exchange rate, Working Paper 24439, National Bureau of Economic Research.

Johnson, Travis L, 2017, Risk premia and the vix term structure, *Journal of Financial and Quantitative Analysis* 52, 2461–2490.

Johnson, Travis L, 2019, A fresh look at return predictability using a more efficient estimator, *The Review of Asset Pricing Studies* 9, 1–46.

Landier, Augustin, Yueran Ma, and David Thesmar, 2019, Biases in expectations: Experimental evidence, Working paper, working paper, May.

Lettau, M, and S Ludvigson, 2003, Measuring and modelling variation in the risk–return tradeoff, prepared for the handbook of financial econometrics, edited by y, *Ait-Sahalia and Lars-Peter Hansen* .

Longstaff, Francis A., and Jiang Wang, 2012, Asset Pricing and the Credit Market, *The Review of Financial Studies* 25, 3169–3215.

López-Salido, David, Jeremy C Stein, and Egon Zakrajšek, 2017, Credit-market sentiment and the business cycle, *The Quarterly Journal of Economics* 132, 1373–1426.

Malmendier, Ulrike, and Stefan Nagel, 2011, Depression Babies: Do Macroeconomic Experiences Affect Risk Taking?*, *The Quarterly Journal of Economics* 126, 373–416.

Mankiw, N. Gregory, and Ricardo Reis, 2002, Sticky Information versus Sticky Prices: A Proposal to Replace the New Keynesian Phillips Curve*, *The Quarterly Journal of Economics* 117, 1295–1328.

Martin, Ian, 2016, What is the expected return on the market?, *Quarterly Journal of Economics* 132(1), 367–433.

Merton, Robert C, 1974, On the pricing of corporate debt: The risk structure of interest rates, *The Journal of finance* 29, 449–470.

Moreira, Alan, and Tyler Muir, 2017, Volatility-managed portfolios, *The Journal of Finance* 72, 1611–1644.

Moreira, Alan, and Tyler Muir, 2019, Should long-term investors time volatility?, *Journal of Financial Economics* 131, 507 – 527.

Nagel, Stefan, and Zhengyang Xu, 2019, Asset pricing with fading memory, Working paper, University of Chicago.

Payzan-LeNestour, Elise, Lionnel Pradier, and Tālis J Putniņš, 2018, The 'waterfall illusion' in the financial markets: Risk perception is distorted after prior exposure to extreme risk, *UNSW Business School Research Paper* .

Poteshman, Allen M., 2001, Underreaction, overreaction, and increasing misreaction to information in the options market, *The Journal of Finance* 56, 851–876.

Rachwalski, Mark, and Quan Wen, 2016, Idiosyncratic Risk Innovations and the Idiosyncratic Risk-Return Relation, *The Review of Asset Pricing Studies* 6, 303–328.

Stein, Jeremy, 1989, Overreactions in the options market, *The Journal of Finance* 44, 1011–1023.

Wachter, Jessica A., 2013, Can time-varying risk of rare disasters explain aggregate stock market volatility?, *Journal of Finance* 68, 987–1035.

Whitelaw, Robert F., 1994, Time variations and covariations in the expectation and volatility of stock market returns, *Journal of Finance* 49, 515–541.

Zhou, Hao, 2018, Variance risk premia, asset predictability puzzles, and macroeconomic uncertainty, *Annual Review of Financial Economics* 10, 481–497.

7. Tables / Figures

Table 1: Survey Expectations. We fit the actual variance process and the survey expectations to an exponential weighted average on past realized variance. That is, we fit: $y_{t,t+k} = a + b\sum_{i=1}^J \phi^{i-1} \sigma_{t-i}^2 + \varepsilon_t$ and report the estimated ϕ where we choose J to be 12 periods, and k as the horizon at which investors forecast variance in the survey (one year). We then repeat this replacing σ_t^2 on the left hand side with the expectation of variance from the survey over the same horizon. A higher ϕ from the expectations data signifies that expectations rely more on variance farther in the past compared to the optimal forecast for volatility. We use the Graham and Harvey CFO survey which is available quarterly and corresponds to a one year forecast horizon. Standard errors are below in parentheses.

Panel A: Dependence on Past Variance (ϕ)			
Source	Survey	Future Variance	
CFO	0.87*** (0.11)	-0.16 (0.38)	
N	69	69	
R^2	0.74	0.36	
$t(\phi_{\text{survey}} - \phi_{fv})$		(2.58)	
Panel B: Surveys, VIX, and Future Variance			
	VIX_t^2	Future Variance	$VIX_t^2 - RV_{t+1}$
CFO_t	1.62*** (0.30)	-0.78 (1.04)	2.10*** (0.77)
σ_t^2	0.60*** (0.09)	0.67*** (0.13)	
R^2	0.74	0.36	0.05
N	69	69	68
Panel C: Regression of 10 year Expectations on 1 year Expectations			
	Volatility	Variance	
1 Year	0.38*** (0.06)	0.25*** (0.04)	
R^2	0.47	0.44	
N	69	69	
$\tilde{\rho}$	0.98	0.96	
CI $\tilde{\rho}$	[0.96,0.98]	[0.92,0.97]	

Table 2: Calibration. We describe the calibration of parameter values.

Panel A: Parameters			
Parameter	Description	Value	Targeted Moment(s)
γ	Risk Aversion	3	Equity Premium
ψ	Elasticity of Intertemporal Substitution	2.2	Literature / VRP
\bar{v}	Unconditional Variance (Monthly)	0.25%	Data
ρ	Persistence of Variance	0.71	Data
ω	Volatility of Variance Shocks (Monthly)	0.31%	Data
ϕ	Expectation Stickiness	0.5	VIX / Surveys
λ	Scale of Expectations	0.8	Vol of VIX / Surveys
Panel B: Moments			
Moment	Description	Model	Data
$E[r_m] - r_f$	Equity Premium (Annual)	7.9%	7.7%
$\sqrt{E[RV_t]}$	Square Root Avg. Variance (Annual)	18%	18%
$\rho(RV_t, RV_{t-1})$	Persistence of Variance (Monthly)	0.71	0.71
$\sigma(RV_t)$	Volatility of Variance (Monthly)	0.44%	0.44%
$\sigma(VIX_t^2)$	Volatility of VIX^2 (Monthly)	0.33%	0.35%
$\rho(VIX_t^2, RV_t)$	Correlation RV and VIX^2 (Monthly)	0.88	0.85
$\rho(VIX_t^2, VIX_{t-1}^2)$	Persistence of VIX^2 (Monthly)	0.91	0.84
$\alpha \left(\frac{c}{RV_{t-1}} r_{m,t}, r_{m,t} \right)$	Volatility-Managed Alpha (Moreira Muir)	1.4%	4.9%
$\rho(r_{m,t}, \widetilde{RV_t})$	Correlation of Returns and Vol Shocks	-0.20	-0.38

Table 3: Stylized Facts. Panel A runs predictive regressions of variance dependent returns, the VIX, and future realized variance on expected variance ($\hat{\sigma}_{t-1}^2$) and a weighted average of expected variance over the past six months $(1 - \phi)\sum_{k=1}^6 \phi^{k-1} \hat{\sigma}_{t-k}^2$. $\hat{\sigma}_{t-1}^2$ represents expected variance at time $t - 1$ while σ_t^2 is the realized variance of daily market returns in month t . Variance dependent returns are short positions in VIX Futures, straddles, and variance swaps. These returns represent the premium for insuring against future increases in VIX, variance, or volatility (so that the variance risk premium is positive on average). Panel B runs excess stock returns (market returns over the risk free rate) on expected variance, the average of past expected variance, and the implied variance from the VIX. Data are monthly from 1992-2020, the variance swap, VIX futures, and straddle return data are 1996-2019, 2004-2020, 1996-2020, respectively. Standard errors in parentheses use Newey West correction with 12 lags.

Panel A: Variance Returns

	(1) Vix Fut	(2) Straddle	(3) Var Swap	(4) Var Swap (TJ)	(5) $VIX_{t-1}^2 - \sigma_t^2$	(6) σ_t^2	(7) VIX_{t-1}^2
$\hat{\sigma}_{t-1}^2$	-5.04***** (1.78)	-29.18** (11.49)	-40.62*** (14.16)	-55.90*** (14.52)	-1.26*** (0.49)	1.53*** (0.54)	0.30** (0.12)
$\sum_{j=1}^6 \phi^j \hat{\sigma}_{t-j}^2$	5.11** (2.13)	36.57** (14.24)	42.90** (20.19)	58.96*** (16.86)	1.23*** (0.46)	-0.58 (0.46)	0.64*** (0.07)
N	194	292	264	282	334	334	335
R^2	0.0523	0.0240	0.00682	0.00496	0.172	0.469	0.790

Panel B: Stock Market Returns

	(1) Market	(2) Market	(3) Market	(4) Market
$\hat{\sigma}_{t-1}^2$	-0.59 (0.80)		-2.21 (1.38)	-3.44*** (0.87)
VIX_{t-1}^2		0.41 (1.16)		3.69*** (1.25)
$\sum_{j=1}^6 \phi^j \hat{\sigma}_{t-j}^2$			2.17 (2.11)	
N	340	363	335	340
R^2	2.04e-05	-0.00170	0.00105	0.0211

Table 4: Stylized Facts in Model and Sensitivity to ϕ . We compare our main facts from Table 3 in the data (first column) vs in our model (remaining columns). The risk-return tradeoff regresses one month ahead market excess returns on current variance and an average of variance over the past 6 months. The volatility-managed alpha is taken from Moreira and Muir (2017) based on their volatility timing strategy (see text for details). All other data used are monthly from 1990-2020 as corresponding to the results in Table 3. We show how our results change as we increase the parameter ϕ across the columns. The second column is the rational model case, $\phi = 0, \lambda = 0.72$ while in other columns we use our calibrated value of $\lambda = 0.8$.

	Data		Model		
	(1)	(2)	$\phi = 0.3$	$\phi = 0.5$	$\phi = 0.8$
Risk Return Tradeoff					
σ_{t-1}^2	-2.21 (1.38)	2.71	-0.23	-1.52	-1.05
$\Sigma_{j=1}^6 \phi^j \hat{\sigma}_{t-j}^2$	2.17 (1.42)	0	3.02	3.74	1.78
R^2	0.1% (1.56)	5.3%	3.5%	3.4%	1.4%
Volatility-Managed Alpha					
α	4.86 (1.56)	-0.09	0.30	1.40	5.33
Correlation: Realized Returns and Vol Shocks					
	-0.38	-0.24	-0.20	-0.16	-0.15
Forecasting Returns with Variance Risk Premium					
VRP_{t-1}	3.47 (0.92)	0	7.40	2.58	1.42
R^2	2.3% (0.46)	0%	1.9%	1%	0.5%
Expected Variance Risk Premium ($VRP_{t-1} = VIX_{t-1}^2 - E_{t-1}[\sigma_t^2]$)					
σ_{t-1}^2	-1.26 (0.49)	0	-0.25	-0.70	-0.92
$\Sigma_{j=1}^6 \phi^j \hat{\sigma}_{t-j}^2$	1.23 (0.46)	0	0.39	0.80	0.58
Correlation: VIX^2 and Realized Variance					
	0.86	1	0.98	0.92	0.73

Table 5: BTZ Comparison. We compare our model to the model of (Bollerslev et al., 2009), given in column 2. The text discusses the calibration of BTZ. Remaining columns show how our model results change as we increase the parameter ϕ .

	Data		Model		
	(1)	BTZ (2)	$\phi = 0.3$ (3)	$\phi = 0.5$ (4)	$\phi = 0.8$ (5)
Risk Return Tradeoff					
$\hat{\sigma}_{t-1}^2$	-2.21 (1.38)	10	-0.23	-1.52	-1.05
$\Sigma_{j=1}^6 \phi^j \hat{\sigma}_{t-j}^2$	2.17 (1.42)	0	3.02	3.74	1.78
R^2	0.1% 7.9%	3.5%	3.4%	1.4%	
Forecasting Returns with Variance Risk Premium					
VRP_{t-1}	3.47 (0.92)	3.94	7.40	2.58	1.42
R^2	2.3% 1.2%	1.9%	1%	0.5%	
Forecasting Returns with $E[RV]$ and VIX					
$\hat{\sigma}_{t-1}^2$	-3.44 (0.87)	6.06	-2.28	-2.21	-2.02
VIX_{t-1}^2	3.69 (1.25)	3.94	4.78	4.71	4.52
R^2	2.1% 8%	4%	3.6%	3.4%	3%
Expected Variance Risk Premium ($VRP_{t-1} = VIX_{t-1}^2 - E_{t-1}[\sigma_t^2]$)					
$\hat{\sigma}_{t-1}^2$	-1.26 (0.49)	0	-0.25	-0.70	-0.92
$\Sigma_{j=1}^6 \phi^j \hat{\sigma}_{t-j}^2$	1.23 (0.46)	0	0.39	0.80	0.58
Correlation: VIX^2 and Realized Variance					
	0.86	0.99	0.96	0.89	0.73

Table 6: Stock level analysis. We repeat our results at the stock level. We run three forecasting regressions $y_{i,t} = a_i + b\Delta_6\sigma_{t-1}^2 + \varepsilon_{i,t}$ where $\Delta_6\sigma_{i,t-1}^2$ is the (lagged) 6 month change in realized variance at the stock level for firm i (the realized variance estimates on the right hand side are winsorized at the 95% level, see text for discussion). As dependent variables, y , we use the equity risk premium (stock return over the risk free rate, $r_{i,t} - r_t^f$ labeled ERP), future variance ($\sigma_{i,t}^2$), and the variance risk premium (difference between implied variance from option metrics and future realized variance, $VRP = IV_{i,t-1}^2 - \sigma_{i,t}^2$ where IV is option implied volatility). Data are monthly but realized variance uses daily data within the month. The last three columns repeat the regression using time fixed effects. In our panel regressions standard errors are double clustered by stock and time.

	ERP	Vol	VRP	ERP	Vol	VRP
$\Delta_6\sigma_{t-1}^2$	−0.129 (0.137)	0.253*** (0.067)	−0.104*** (0.036)	−0.040 (0.075)	0.188*** (0.046)	−0.070*** (0.022)
N	536,726	536,726	536,726	536,726	536,726	536,726
Adj R ²	0.001	0.010	0.002	0.159	0.060	0.024
Time FE	N	N	N	Y	Y	Y

Table 7: Term Structure. We regress excess returns to variance claims at different maturities on lagged variance and a weighted average of past variance. Data and returns are monthly, standard errors in parentheses are Newey-West with 12 lags. In the data coefficients are normalized for comparison to the model. See text for more detail.

Panel A: Model Coefficients by Maturity					
	(1)	(2)	(3)	(6)	(12)
σ_{t-1}^2	-0.34	-0.20	-0.15	-0.14	-0.12
$\Sigma_{j=1}^6 \phi^j \hat{\sigma}_{t-j}^2$	0.54	0.31	0.24	0.17	0.14
Adj. R^2	15.9%	15.9%	15.9%	15.9%	15.9%
Panel B: Data Variance Swaps					
	(1)	(2)	(3)	(6)	(12)
σ_{t-1}^2	-0.90*** (0.24)	-0.78*** (0.19)	-0.66*** (0.15)	-0.45*** (0.09)	-0.22*** (0.06)
$\Sigma_{j=1}^6 \phi^j \hat{\sigma}_{t-j}^2$	1.02** (0.47)	0.85** (0.36)	0.69*** (0.26)	0.40** (0.17)	0.15 (0.12)
N	264	264	264	264	264
Adj. R^2	0.7%	1.0%	1.4%	1.3%	0.0%
Panel C: Data Straddle Returns					
	(1)	(2)	(3)	(6)	(12)
σ_{t-1}^2	-0.40*** (0.12)	-0.41*** (0.07)	-0.27*** (0.07)	-0.17*** (0.05)	-0.16*** (0.03)
$\Sigma_{j=1}^6 \phi^j \hat{\sigma}_{t-j}^2$	0.58*** (0.21)	0.47*** (0.10)	0.32*** (0.09)	0.22*** (0.06)	0.15*** (0.04)
N	264	264	264	264	264
Adj. R^2	1.1%	2.9%	1.7%	1.4%	1.7%
Panel D: Data VIX Futures					
	(1)		(3)	(6)	
σ_{t-1}^2	-0.35*** (0.08)		-0.33*** (0.06)	-0.20*** (0.04)	
$\Sigma_{j=1}^6 \phi^j \hat{\sigma}_{t-j}^2$	0.26** (0.13)		0.03 (0.09)	-0.17* (0.09)	
N	166		166	166	
Adj. R^2	2.2%		3.6%	5.1%	

Table 8: Additional Returns. We run predictive regressions of future excess returns on past expected variance, and a weighted average of expected variance over the past six months $(1 - \phi)\sum_{k=1}^6 \phi^{k-1} \hat{\sigma}_{t-k}^2$. We use credit returns as credit is a volatility sensitive asset class. We use CDS returns from He et al. (2017), the difference in Barclays high yield and investment grade returns (HY-IG), and the (negative) change in the Baa Aaa spread from Moody's. Data are monthly from 1990-2018, the CDS data and High Yield total return data are 2004-2012 and 1995-2015, respectively. Standard errors in parentheses use Newey West correction with 12 lags.

	Credit Returns		
	(1)	(2)	(3)
	CDS	HY-IG	BaaAaa
σ_{t-1}^2	-0.54*** (0.05)	-3.08*** (0.62)	-0.17*** (0.02)
$\sum_{j=1}^6 \phi^j \hat{\sigma}_{t-j}^2$	0.66*** (0.10)	2.68*** (1.14)	0.16*** (0.04)
<i>N</i>	143	240	313
Adj. R^2	19.2%	15.9%	25.9%

Table 9: Timing Sharpe Ratios. We run predictive regressions of the realized variance risk premium ($VIX_{t-1}^2 - RV_t$) on expected variance and the squared VIX at $t - 1$. We then take the predicted value from the regression and compute the Sharpe ratio of the optimal timing portfolio. The first column is the data and the second two columns are the main model and our extended model with jumps and time-varying volatility of volatility. Standard errors in parentheses use Newey West correction with 12 lags.

	(1)	(2)	(3)
	Data	Model	Model Extended
$\hat{\sigma}_{t-1}^2$	-0.85*** (0.20)	-1	-1
VIX_{t-1}^2	0.87*** (0.18)	1	1
<i>N</i>	338		
R^2	0.19	0.124	0.09
Sharpe	0.41	1.17	0.30

Figure 1: Dynamics of variance expectations in the model. We plot the behavior of agents expectations of volatility in our model (blue line) and the true path of expected volatility (dashed red line) in response to a one standard deviation increase in variance in our model. The dot dashed black line provides an alternative calibration when we set the scale of expectations lower. Because agents effectively take a weighted average of past volatility they initially underreact and then subsequently overreact. The variance risk premium then reflects the difference between agents expectations of volatility minus the rational forecast of volatility, hence it goes negative initially then becomes positive. The bottom panel shows the behavior of stock prices (the price dividend ratio) which also responds slowly in the model, reflecting agents slow moving beliefs. This slow response makes it appear as if equity risk premiums don't initially rise (and potentially even fall) after an increase in volatility but then rise later after the volatility shock has largely subsided. The x-axis is in months. The black dashed line shows an alternative model calibration.

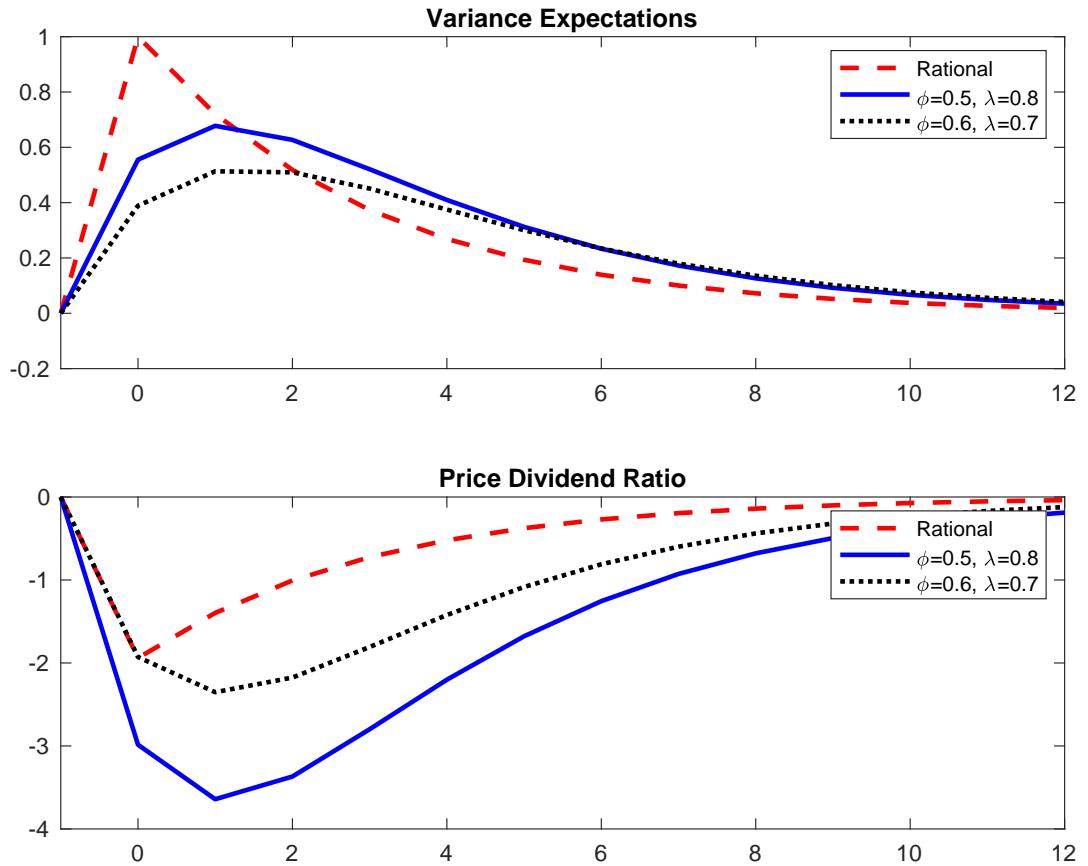


Figure 2: Impulse responses: data vs model. We plot the behavior of expected stock returns and variance risk premiums in the data vs the model at various horizons for a one standard deviation shock to variance. The black line shows the impulse response in the data from a VAR(1) of expected variance, market excess returns (denoted ERP for equity risk premium), the variance risk premium (VRP), and the log price dividend ratio. VRP is implied variance (VIX^2) minus expected variance. Responses are for a one-standard deviation shock to expected variance at time 0 and gray shaded region represents 95% confidence interval. The blue dashed line repeats this using simulated data from the calibrated model. The red dot dashed line repeats this exercise in the simulated model data but imposes no bias in beliefs (rational model). Equity returns are given in units of percent per month. The x-axis is in months.

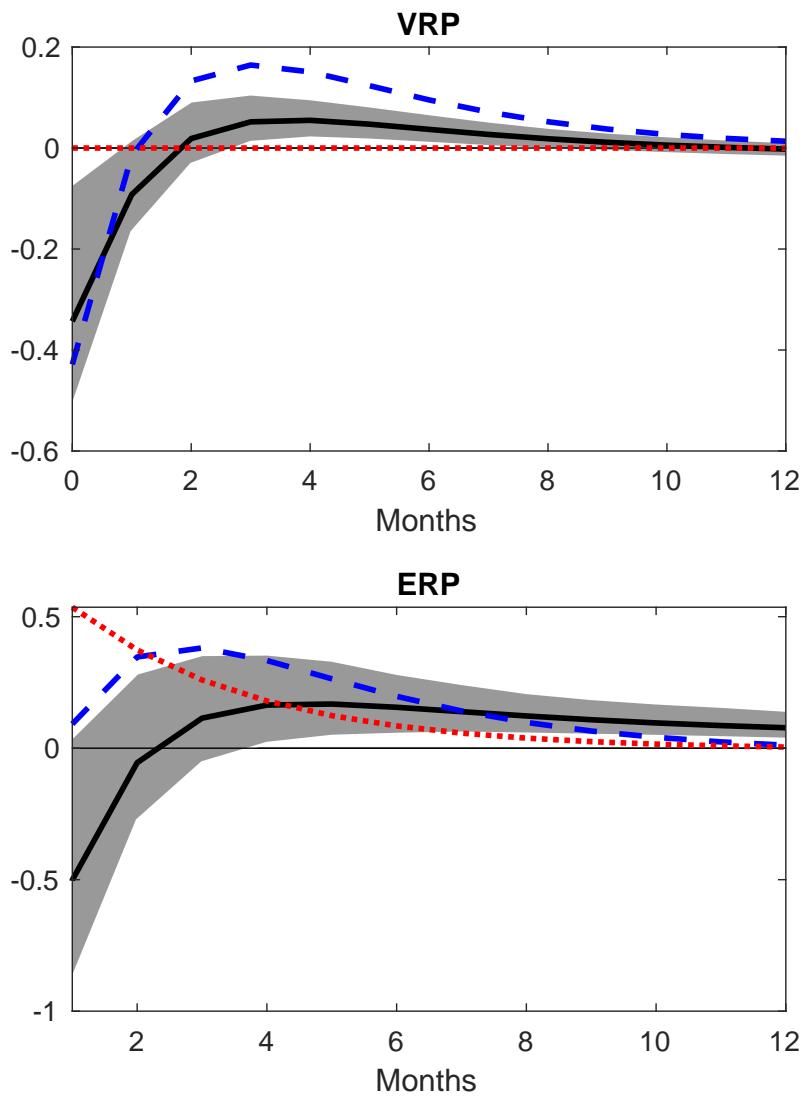
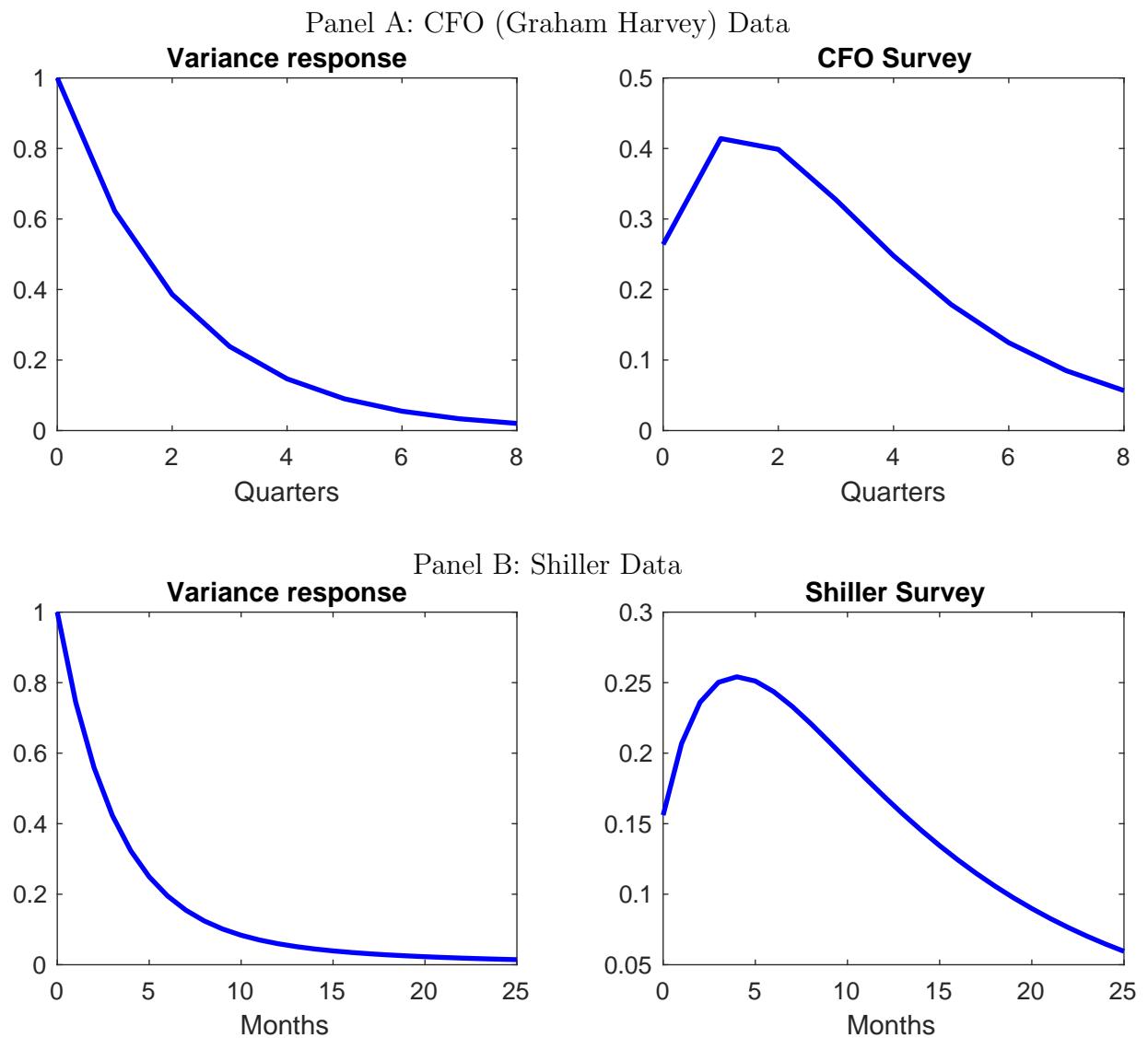


Figure 3: Survey Expectations, VAR. We run a VAR(1) using realized variance and survey expectations of variance and plot the impulse response to a 1 standard deviation shock to realized variance. Expected variance rises strongly after the shock and then mean reverts fairly quickly. Survey expectations rise slowly, underreact initially and then remain elevated for longer. Both variables are standardized to have unit variance.



8. Internet Appendix: Not Intended for Publication

The appendix contains additional derivations, tables, and figures. In particular,

- A. Baseline model derivation
- B. Derivation of model with Gamma distributed shocks
- C. VIX derivation with additional variance signal
- D. Expected variance construction and additional data sources
- E. Robustness: Additional tables and figures
 - (a) Sample splits (Table 13)
 - (b) Volatility instead of variance in regressions (Table 11)
 - (c) Weighted Least Squares (Table 12)
 - (d) International data (Table 18)
 - (e) Long-sample regressions at different horizons (Table 17)

A. Baseline Model solution

In this section, we provide more detailed solutions for the baseline model in the paper, where variance shocks are Normally distributed.

A.1 The Variance Process

From the main text, agents beliefs about the dividend process are as follows:

$$\Delta d_t = \mu + \sigma_t \varepsilon_t, \quad (29)$$

where ε_t is i.i.d. standard Normal and

$$\sigma_t^2 = \bar{v} + \lambda x_{t-1} + \omega \eta_t^S, \quad (30)$$

$$\begin{aligned} x_t &= \phi x_{t-1} + (1 - \phi) (\sigma_t^2 - \bar{v}) \\ &= (\phi + (1 - \phi) \lambda) x_{t-1} + (1 - \phi) \omega \eta_t^S, \end{aligned} \quad (31)$$

where η_t^S is an i.i.d. standard Normal shock uncorrelated with ε . Both variance σ_t^2 and ε_t are observed at time t .

We assume an exchange economy where the agent has Epstein-Zin preferences, and aggregate log dividend growth is denoted Δd and the agent's consumption equal aggregate dividends. The first order condition is then:

$$\begin{aligned} 1 &= E_t^S [M_{t+1} R_{t+1}] \\ &= \beta^\theta E_t^S \left[e^{-\frac{\theta}{\psi} \Delta d_{t+1} + \theta r_{t+1}} \right] \\ &= \beta^\theta E_t^S \left[e^{(1-\gamma) \Delta d_{t+1} + \theta \kappa_0 + \theta \kappa p d_{t+1} - \theta p d_t} \right], \end{aligned} \quad (32)$$

where r is the log return on the dividend claim and where pd is the log price-dividend ratio. Also, $\theta = \frac{1-\gamma}{1-1/\psi}$, where γ and ψ are the risk aversion and intertemporal elasticity of substitution parameters, respectively.

We proceed with the conjecture $pd_t = c - Ax_t$. Then:

$$\begin{aligned} 1 &= \beta^\theta E_t^S \left[e^{(1-\gamma)(\mu + \sigma_{t+1} \varepsilon_{t+1}) + \theta \kappa_0 + \theta \kappa (c - Ax_{t+1}) - \theta (c - Ax_t)} \right] \\ &= \beta^\theta E_t^S \left[e^{(1-\gamma)\mu + \frac{1}{2}(1-\gamma)^2 \sigma_{t+1}^2 + \theta \kappa_0 + \theta \kappa (c - A(\phi x_t + (1-\phi)(\sigma_{t+1}^2 - \bar{v}))) - \theta (c - Ax_t)} \right]. \end{aligned} \quad (33)$$

Now, ignoring any terms that don't multiply x and using $\sigma_{t+1}^2 = \bar{v} + \lambda x_t + \omega \eta_{t+1}^S$, we have that:

$$\begin{aligned} E_t^S \left[e^{(1-\gamma)\mu + \frac{1}{2}(1-\gamma)^2 \sigma_{t+1}^2 + \theta \kappa_0 + \theta \kappa (c - A(\phi x_t + (1-\phi)(\sigma_{t+1}^2 - \bar{v}))) - \theta (c - Ax_t)} \right] &= \\ const \times E_t^S \left[e^{(1-\gamma)^2 \frac{1}{2} \lambda x_t - \theta \kappa A(\phi x_t + (1-\phi) \lambda x_t) + \theta A x_t} \right]. \end{aligned} \quad (34)$$

And so we have:

$$(1-\gamma)^2 \frac{1}{2} \lambda - \theta \kappa A (\phi + (1-\phi) \lambda) + \theta A = 0, \quad (35)$$

which gives:

$$A = -\frac{1}{2} \frac{\lambda(1-\gamma)(1-1/\psi)}{1-\kappa(\phi+(1-\phi)\lambda)}. \quad (36)$$

Thus, with $\gamma, \psi > 1$, we have that $A > 0$.

The conditional variance of log returns is then:

$$\begin{aligned} Var_{t-1}^S(r_t) &= Var_{t-1}^S(\kappa p d_t + \Delta d_t) \\ &= \Theta + E_{t-1}^S [\sigma_t^2], \end{aligned} \quad (37)$$

where $\Theta = (\kappa A (1-\phi) \omega)^2$. To get the equity risk premium, we need to solve for the risk-free rate which in turn requires solving for c . Going back to the first-order equation for the risky asset:

$$\begin{aligned} 1 &= \beta^\theta E_t^S \left[e^{(1-\gamma)\mu + \frac{1}{2}(1-\gamma)^2 \sigma_{t+1}^2 + \theta \kappa_0 + \theta \kappa (c - A(\phi x_t + (1-\phi)(\sigma_{t+1}^2 - \bar{v}))) - \theta (c - Ax_t)} \right] \\ &= \beta^\theta E_t^S \left[e^{(1-\gamma)\mu + \frac{1}{2}(1-\gamma)^2 (\bar{v} + \omega \eta_{t+1}^S) + \theta \kappa_0 + \theta \kappa (c - A(1-\phi) \omega \eta_{t+1}^S) - \theta c} \right] \\ &= \beta^\theta e^{(1-\gamma)\mu + \frac{1}{2}(1-\gamma)^2 \bar{v} + \theta \kappa_0 - \theta c(1-\kappa) + \frac{1}{2} (\frac{1}{2}(1-\gamma)^2 - \theta \kappa A(1-\phi))^2 \omega^2}, \end{aligned} \quad (38)$$

where the second equality uses the fact from above that terms in the exponential that multiplies x_t add to zero.

Then:

$$0 = \theta \ln \beta + (1 - \gamma) \mu + \frac{1}{2} (1 - \gamma)^2 \bar{v} + \theta \kappa_0 - \theta c (1 - \kappa) + \frac{1}{2} \left(\frac{1}{2} (1 - \gamma)^2 - \theta \kappa A (1 - \phi) \right)^2 \omega^2. \quad (39)$$

And so

$$c = \frac{\ln \beta + (1 - \psi^{-1}) \mu + \frac{1}{2} (1 - \psi^{-1}) (1 - \gamma) \bar{v} + \kappa_0 + \theta^{-1} \frac{1}{2} \left(\frac{1}{2} (1 - \gamma)^2 - \theta \kappa A (1 - \phi) \right)^2 \omega^2}{1 - \kappa}. \quad (40)$$

The risk-free rate is given by:

$$\begin{aligned} e^{-r_{f,t}} &= E_t^S [M_{t+1}] \\ &= \beta^\theta E_t^S [e^{-\gamma \Delta d_{t+1} + (\theta-1)(\kappa_0 + \kappa p d_{t+1} - p d_t)}] \\ &= \beta^\theta E_t^S [e^{-\gamma \Delta d_{t+1} + (\theta-1)(\kappa_0 + \kappa c - A \kappa x_{t+1} - c + A x_t)}] \\ &= \beta^\theta E_t^S \left[e^{-\gamma \mu + \frac{1}{2} \gamma^2 (\bar{v} + \lambda x_t + \omega \eta_{t+1}^S) + (\theta-1)(\kappa_0 + \kappa c - A \kappa ((\phi + (1 - \phi) \lambda) x_t + (1 - \phi) \omega \eta_t^S) - c + A x_t)} \right] \\ &= \beta^\theta e^{-\gamma \mu + \frac{1}{2} \gamma^2 \bar{v} + (\theta-1)(\kappa_0 + c(\kappa-1)) + \left(\frac{1}{2} \gamma^2 \lambda + (\theta-1) A (1 - \kappa (\phi + (1 - \phi) \lambda)) \right) x_t} \\ &\quad \times e^{\frac{1}{2} \left(\frac{1}{2} \gamma^2 - (\theta-1) \kappa A (1 - \phi) \right)^2 \omega^2}. \end{aligned} \quad (41)$$

Thus, plugging in for c we have:

$$\begin{aligned} r_{f,t} &= -\ln \beta + \psi^{-1} \mu - \frac{1}{2} (1 - \psi^{-1}) (1 - \gamma) \bar{v} - \frac{1}{2\theta} \left(\frac{1}{2} (1 - \gamma)^2 - \theta \kappa A (1 - \phi) \right)^2 \omega^2 \dots \\ &\quad + \frac{1}{2} (1 - 2\gamma) \bar{v} + \frac{1}{2} \left(\frac{1}{2} (1 - \gamma)^2 - \theta \kappa A (1 - \phi) \right)^2 \omega^2 \dots \\ &\quad - \left(\frac{1}{2} \gamma^2 \lambda + (\theta - 1) A (1 - \kappa (\phi + (1 - \phi) \lambda)) \right) x_t \\ &\quad - \frac{1}{2} \left(\frac{1}{2} \gamma^2 - (\theta - 1) \kappa A (1 - \phi) \right)^2 \omega^2 \end{aligned}$$

The conditional expected log return is:

$$\begin{aligned} E_t^S [\kappa_0 + \kappa p d_{t+1} - p d_t + \Delta d_{t+1}] &= \kappa_0 + \kappa c - c + A x_t + \mu + E_t^S [-\kappa A x_{t+1}] \\ &= \kappa_0 + c(\kappa - 1) + \mu + A(1 - \kappa (\phi + (1 - \phi) \lambda)) x_t. \end{aligned}$$

Plugging in for c we have:

$$\begin{aligned} E_t^S [r_{t+1}] &= \kappa_0 + c(\kappa - 1) + \mu + A(1 - \kappa (\phi + (1 - \phi) \lambda)) x_t \\ &= -\ln \beta + \psi^{-1} \mu - \frac{1}{2} (1 - \psi^{-1}) (1 - \gamma) \bar{v} \dots \\ &\quad - \frac{1}{2\theta} \left(\frac{1}{2} (1 - \gamma)^2 - \theta \kappa A (1 - \phi) \right)^2 \omega^2 \dots \\ &\quad + A(1 - \kappa (\phi + (1 - \phi) \lambda)) x_t \end{aligned}$$

The conditional log risk premium is then:

$$\begin{aligned}
E_t^S [r_{t+1} - r_{f,t}] &= \left(\frac{1}{2} \gamma^2 \lambda + \theta A (1 - \kappa (\phi + (1 - \phi) \lambda)) \right) x_t \dots \\
&\quad - \frac{1}{2} (1 - 2\gamma) \bar{v} - \frac{1}{2} \left(\frac{1}{2} (1 - \gamma)^2 - \theta \kappa A (1 - \phi) \right)^2 \omega^2 \dots \\
&\quad + \frac{1}{2} \left(\frac{1}{2} \gamma^2 - (\theta - 1) \kappa A (1 - \phi) \right)^2 \omega^2
\end{aligned}$$

Next, note that:

$$\begin{aligned}
\theta A (1 - \kappa (\phi + (1 - \phi) \lambda)) &= -\frac{1}{2} \theta \frac{\lambda (1 - \gamma) (1 - 1/\psi)}{1 - \kappa (\phi + (1 - \phi) \lambda)} (1 - \kappa (\phi + (1 - \phi) \lambda)) \\
&= -\frac{1}{2} \lambda \theta (1 - \gamma) (1 - 1/\psi) \\
&= \lambda \left(-\frac{1}{2} + \gamma - \frac{1}{2} \gamma^2 \right)
\end{aligned}$$

So then

$$\frac{1}{2} \gamma^2 \lambda + \theta A (1 - \kappa (\phi + (1 - \phi) \lambda)) = \lambda \left(\gamma - \frac{1}{2} \right).$$

We then have:

$$\begin{aligned}
E_t^S [r_{t+1} - r_{f,t}] &= \left(\gamma - \frac{1}{2} \right) \bar{v} + \lambda \left(\gamma - \frac{1}{2} \right) x_t \dots \\
&\quad - \frac{1}{2} \left(\frac{1}{2} (1 - \gamma)^2 - \theta \kappa A (1 - \phi) \right)^2 \omega^2 \dots \\
&\quad + \frac{1}{2} \left(\frac{1}{2} \gamma^2 - (\theta - 1) \kappa A (1 - \phi) \right)^2 \omega^2.
\end{aligned}$$

This can be written

$$E_t^S [r_{t+1} - r_{f,t}] = \left(\gamma - \frac{1}{2} \right) E_t^S [\sigma_{t+1}^2] + \delta_r. \quad (42)$$

where

$$\begin{aligned}
\delta_r &= -\frac{1}{2} \left(\frac{1}{2} (1 - \gamma)^2 - \theta \kappa A (1 - \phi) \right)^2 \omega^2 \dots \\
&\quad + \frac{1}{2} \left(\frac{1}{2} \gamma^2 - (\theta - 1) \kappa A (1 - \phi) \right)^2 \omega^2.
\end{aligned}$$

The objective risk-premium is:

$$\begin{aligned}
E_t^P [r_{t+1} - r_{f,t}] &= E_t^S [r_{t+1} - r_{f,t}] + \kappa (E_{t+1}^P [pd_{t+1}] - E_{t+1}^S [pd_{t+1}]) \\
&= E_t^S [r_{t+1} - r_{f,t}] - \kappa A (E_{t+1}^P [x_{t+1}] - E_{t+1}^S [x_{t+1}]) .
\end{aligned}$$

We have that:

$$\begin{aligned} E_t^S[x_{t+1}] &= (\phi + (1 - \phi)\lambda)x_t \\ E_t^P[x_{t+1}] &= (\phi + (1 - \phi)\lambda)x_t + (1 - \phi)E_t^P[\omega\eta_{t+1}^S] \\ &= E_t^S[x_{t+1}] + (1 - \phi)(E_t^P[\sigma_{t+1}^2] - E_t^S[\sigma_{t+1}^2]). \end{aligned}$$

Thus:

$$E_t^P[r_{t+1} - r_{f,t}] = E_t^S[r_{t+1} - r_{f,t}] - \kappa A(1 - \phi)(E_t^P[\sigma_{t+1}^2] - E_t^S[\sigma_{t+1}^2]), \quad (43)$$

which is the same equation we get in the case with Normal variance shocks.

Shocks to realized returns are then:

$$\begin{aligned} r_{t+1} - E_t^P[r_{t+1}] &= \Delta d_{t+1} - E_t^P[\Delta d_{t+1}] + \kappa(p d_{t+1} - E_t^P[p d_{t+1}]) \\ &= \sigma_{t+1}\varepsilon_{t+1} + \kappa A(-x_{t+1} + E_t^P[x_{t+1}]) \\ &= \sigma_{t+1}\varepsilon_{t+1} + \kappa A(1 - \phi)(-(\sigma_{t+1}^2 - \bar{v}) + E_t^P[(\sigma_{t+1}^2 - \bar{v})]) \\ &= \sigma_{t+1}\varepsilon_{t+1} - \kappa A(1 - \phi)\omega\eta_{t+1}. \end{aligned} \quad (44)$$

Next, turning the the variance risk premium (VRP), note that the error in variance expectation will feed through in the VRP. In particular:

$$\begin{aligned} IV_{t-1} &= E_{t-1}^S \left[\frac{M_t}{E_{t-1}^S[M_t]} (\Theta + \bar{v} + \lambda x_{t-1} + \omega\eta_t^S) \right] \\ &= \Theta + \delta_{IV} + E_{t-1}^S[\sigma_t^2], \end{aligned} \quad (45)$$

where $\delta_{IV} = E_{t-1}^S \left[\frac{M_t}{E_{t-1}^S[M_t]} \omega\eta_t^S \right]$. To see that this is indeed a constant, note that:

$$\begin{aligned} E_{t-1}^S \left[\frac{M_t}{E_{t-1}^S[M_t]} \omega\eta_t^S \right] &= \\ E_t^S \left[\frac{\beta^\theta e^{-\gamma\mu + \frac{1}{2}\gamma^2(\bar{v} + \lambda x_t + \omega\eta_{t+1}^S) + (\theta-1)(\kappa_0 + \kappa c - \kappa A(\phi x_t + (1-\phi)(\lambda x_t + \omega\eta_{t+1}^S)) - c + A x_t)}}{E_t^S \left[\beta^\theta e^{-\gamma\mu + \frac{1}{2}\gamma^2(\bar{v} + \lambda x_t + \omega\eta_{t+1}^S) + (\theta-1)(\kappa_0 + \kappa c - \kappa A(\phi x_t + (1-\phi)(\lambda x_t + \omega\eta_{t+1}^S)) - c + A x_t} \right]} \omega\eta_{t+1}^S \right] &= \\ E_t^S \left[\frac{e^{(\frac{1}{2}\gamma^2 - (\theta-1)\kappa A(1-\phi))\omega\eta_{t+1}^S}}{E_t^S \left[e^{(\frac{1}{2}\gamma^2 - (\theta-1)\kappa A(1-\phi))\omega\eta_{t+1}^S} \right]} \omega\eta_{t+1}^S \right] &= \\ E_t^S \left[\frac{e^{m\omega\eta_{t+1}^S}}{E_t^S \left[e^{m\omega\eta_{t+1}^S} \right]} \omega\eta_{t+1}^S \right]. \end{aligned} \quad (46)$$

where $m = \frac{1}{2}\gamma^2 - (\theta - 1)\kappa A(1 - \phi)$. From Stein's Lemma we have that:

$$E_t^S \left[\frac{e^{m\omega\eta_{t+1}^S}}{E_t^S \left[e^{m\omega\eta_{t+1}^S} \right]} \omega\eta_{t+1}^S \right] = \omega^2 m. \quad (47)$$

To summarize:

$$IV_{t-1} = \Theta + \delta_{IV} + E_{t-1}^S [\sigma_t^2], \quad (48)$$

$$\delta_{IV} = \left(\frac{1}{2} \gamma^2 - (\theta - 1) \kappa A (1 - \phi) \right) \omega^2. \quad (49)$$

The results given in the main text follow from the derivations shown here.

B. Model with Gamma-Distributed Variance Shocks

In this section, we show that a model where the variance process has Gamma distributed shocks – which allow us to guarantee positive variance – lead to the same expressions for risk premium dynamics as in the model in the main text. The only thing that changes somewhat are the unconditional levels of risk premiums (the intercepts in the expressions), but these are not our main focus. Before we get into the model, it is useful to establish some general properties of the Gamma distribution.

B.1 The Gamma Distribution

If $X > 0$ is a Gamma distributed random variable, we have that:

$$X \sim \text{Gamma}(k, s) \quad (50)$$

$$f(x) = \frac{x^{k-1} e^{-\frac{x}{s}}}{s^k \Gamma(k)}, \quad (51)$$

where k is the shape and s is the scale parameter, respectively, $\Gamma(k)$ is the Gamma function, $f(x)$ is the probability density function, and $k, s > 0$. Then:

$$E[X] = ks, \quad (52)$$

$$Var(X) = ks^2 \quad (53)$$

$$E[e^{tX}] = (1 - st)^{-k} \text{ for } t < s^{-1}. \quad (54)$$

Imposing $Var(X) = 1$ implies that $k = s^{-2}$.

Note that the analogue of Stein's Lemma for a Gamma distributed variable is:

$$\begin{aligned} E[e^{tx} x] &= \int_0^\infty e^{tx} x \frac{x^{k-1} e^{-\frac{x}{s}}}{s^k \Gamma(k)} dx \\ &= \int_0^\infty \frac{x^k e^{tx - \frac{x}{s}}}{s^k \Gamma(k)} dx \\ &= \int_0^\infty \frac{x^k e^{x(t - \frac{1}{s})}}{s^k \Gamma(k)} dx. \end{aligned} \quad (55)$$

Next, define $\tilde{k} \equiv k + 1$ and $\tilde{s} \equiv -\frac{1}{t-s^{-1}}$. Recall that we always need $t < s^{-1}$ so $t - s^{-1} < 0$ and thus $\tilde{s} > 0$, which is required. Clearly, $\tilde{k} > 0$ given that $k > 0$. We can then write:

$$\begin{aligned} E[e^{tx}x] &= \frac{\tilde{s}^{\tilde{k}}\Gamma(\tilde{k})}{s^k\Gamma(k)} \int_0^\infty \frac{x^{\tilde{k}-1}e^{-\frac{x}{\tilde{s}}}}{\tilde{s}^{\tilde{k}}\Gamma(\tilde{k})} dx \\ &= \frac{\tilde{s}^{\tilde{k}}\Gamma(\tilde{k})}{s^k\Gamma(k)}. \end{aligned} \quad (56)$$

B.2 The Variance Process

As in the main text, agents beliefs about the dividend process are as follows:

$$\Delta d_t = \mu + \sigma_t \varepsilon_t, \quad (57)$$

where ε_t is i.i.d. standard Normal and

$$\sigma_t^2 = \bar{v} + \lambda x_{t-1} + \omega \tilde{\eta}_t^S, \quad (58)$$

$$\begin{aligned} x_t &= \phi x_{t-1} + (1 - \phi) (\sigma_t^2 - \bar{v}) \\ &= (\phi + (1 - \phi) \lambda) x_{t-1} + (1 - \phi) \omega \tilde{\eta}_t^S, \end{aligned} \quad (59)$$

where $\tilde{\eta}_t^S = \eta_t^S - s^{-1}$ where η_t^S is a Gamma distributed variable scale and shape parameters s and k , respectively, that is independent of ε_t . We normalize the shock to have unit variance and so $k = s^{-2}$, which in turns implies that its mean is mean s^{-1} and so $E_{t-1}[\tilde{\eta}_t] = 0$. Both variance σ_t^2 and ε_t are observed at time t .

In order for variance to always be non-negative, the restriction

$$\frac{(1 - \phi) \omega s^{-1}}{1 - (\phi + (1 - \phi) \lambda)} \leq \bar{v} \quad (60)$$

has to be satisfied. In our calibration, we have

$$\omega = 0.23\%, \quad (61)$$

$$\lambda = 0.9, \quad (62)$$

$$\phi = 0.6 \quad (63)$$

$$\bar{v} = 0.26\%. \quad (64)$$

Thus, we need:

$$s \geq \frac{0.4 \times 0.23\%}{0.26\% \times (1 - (0.6 + 0.4 \times 0.9))} = 8.85. \quad (65)$$

Note that:

$$E_{t-1}^S[\sigma_t \varepsilon_t] = 0, \quad (66)$$

$$Var_{t-1}^S(\sigma_t \varepsilon_t) = E_{t-1}^S[\sigma_t^2]. \quad (67)$$

B.3 Solving the Gamma-Model

As in the main text, we assume an exchange economy where the agent has Epstein-Zin preferences, and aggregate log dividend growth is denoted Δd and the agent's consumption equal aggregate dividends. We proceed as before with the conjecture $pd_t = c - Ax_t$ and consider the first-order condition for the risky asset:

$$\begin{aligned} 1 &= \beta^\theta E_t^S \left[e^{(1-\gamma)(\mu+\sigma_{t+1}\varepsilon_{t+1})+\theta\kappa_0+\theta\kappa(c-Ax_{t+1})-\theta(c-Ax_t)} \right] \\ &= \beta^\theta E_t^S \left[e^{(1-\gamma)\mu+\frac{1}{2}(1-\gamma)^2\sigma_{t+1}^2+\theta\kappa_0+\theta\kappa(c-A(\phi x_t+(1-\phi)(\sigma_{t+1}^2-\bar{v})))-\theta(c-Ax_t)} \right]. \end{aligned} \quad (66)$$

Now, ignoring any terms that don't multiply x and using $\sigma_{t+1}^2 = \bar{v} + \lambda x_t + \omega \eta_{t+1}^S$, we have that:

$$\begin{aligned} E_t^S \left[e^{(1-\gamma)\mu+\frac{1}{2}(1-\gamma)^2\sigma_{t+1}^2+\theta\kappa_0+\theta\kappa(c-A(\phi x_t+(1-\phi)(\sigma_{t+1}^2-\bar{v})))-\theta(c-Ax_t)} \right] &= \\ const \times E_t^S \left[e^{(1-\gamma)^2\frac{1}{2}\lambda x_t-\theta\kappa A(\phi x_t+(1-\phi)\lambda x_t)+\theta A x_t} \right]. \end{aligned} \quad (67)$$

And so we have:

$$(1-\gamma)^2 \frac{1}{2} \lambda - \theta \kappa A (\phi + (1-\phi) \lambda) + \theta A = 0, \quad (68)$$

which gives:

$$A = -\frac{1}{2} \frac{\lambda(1-\gamma)(1-1/\psi)}{1-\kappa(\phi+(1-\phi)\lambda)}, \quad (69)$$

which is the same as for the case of Normally distributed variance shocks. Thus, with $\gamma, \psi > 1$, we have that $A > 0$.

The conditional variance of log returns is:

$$\begin{aligned} Var_{t-1}^S(r_t) &= Var_{t-1}(\kappa pd_t + \Delta d_t) \\ &= \Theta + E_{t-1}^S [\sigma_t^2], \end{aligned} \quad (70)$$

where $\Theta = (A(1-\phi)\omega\kappa)^2$. To get the equity risk premium, we need to solve for the risk-free rate which in turn requires solving for c . Going back to the first-order equation for the risky asset:

$$\begin{aligned} 1 &= \beta^\theta E_t^S \left[e^{(1-\gamma)\mu+\frac{1}{2}(1-\gamma)^2\sigma_{t+1}^2+\theta\kappa_0+\theta\kappa(c-A(\phi x_t+(1-\phi)(\sigma_{t+1}^2-\bar{v})))-\theta(c-Ax_t)} \right] \\ &= \beta^\theta E_t^S \left[e^{(1-\gamma)\mu+\frac{1}{2}(1-\gamma)^2(\bar{v}+\omega\eta_{t+1}^S)+\theta\kappa_0+\theta\kappa(c-A(1-\phi)\omega\eta_{t+1}^S)-\theta c} \right] \\ &= \beta^\theta e^{(1-\gamma)\mu+\frac{1}{2}(1-\gamma)^2\bar{v}+\theta\kappa_0-\theta c(1-\kappa)} E_t^S \left[e^{[\frac{1}{2}(1-\gamma)^2-\theta\kappa A(1-\phi)]\omega\eta_{t+1}^S} \right], \end{aligned} \quad (71)$$

where the second equality uses the fact that terms in the exponential that multiplies x_t add to zero.

From the moment generating function (MGF) of the Gamma distributed shock we then have that:

$$1 = \beta^\theta e^{(1-\gamma)\mu + \frac{1}{2}(1-\gamma)^2\bar{v} + \theta\kappa_0 - \theta c(1-\kappa) + z_1}, \quad (72)$$

where

$$z_1 = -\frac{1}{s^2} \ln \left(1 - s \left(\frac{1}{2} (1-\gamma)^2 - \theta\kappa A (1-\phi) \right) \omega \right), \quad (73)$$

and where s is the scale parameter for the Gamma distribution as given above. This imposes the parameter restriction

$$\left(\frac{1}{2} (1-\gamma)^2 - \theta\kappa A (1-\phi) \right) \omega < s^{-1}, \quad (74)$$

to have existence of the MGF of the Gamma.

Then:

$$0 = \theta \ln \beta + (1-\gamma) \mu + \frac{1}{2} (1-\gamma)^2 \bar{v} + \theta\kappa_0 - \theta c (1-\kappa) + z_1. \quad (75)$$

And so

$$c = \frac{\theta \ln \beta + (1-\gamma) \mu + \frac{1}{2} (1-\gamma)^2 \bar{v} + \theta\kappa_0 + z_1}{\theta (1-\kappa)}. \quad (76)$$

The risk-free rate is given by:

$$\begin{aligned} e^{-r_{f,t}} &= E_t^S [M_{t+1}] \\ &= \beta^\theta E_t^S [e^{-\gamma \Delta d_{t+1} + (\theta-1)(\kappa_0 + \kappa p d_{t+1} - p d_t)}] \\ &= \beta^\theta e^{-\gamma \mu + \frac{1}{2} \gamma^2 \bar{v} + (\theta-1)(\kappa_0 + c(\kappa-1)) + \left(\frac{1}{2} \gamma^2 \lambda + (\theta-1) A (1-\kappa(\phi+(1-\phi)\lambda)) \right) x_t} \\ &\quad \dots E_t^S \left[e^{\left(\frac{1}{2} \gamma^2 - (\theta-1) \kappa A (1-\phi) \right) \omega \eta_{t+1}^S} \right]. \end{aligned}$$

Again using the MGF of the Gamma distribution, we have:

$$e^{-r_{f,t}} = \beta^\theta e^{-\gamma \mu + \frac{1}{2} \gamma^2 \bar{v} + (\theta-1)(\kappa_0 + c(\kappa-1)) + \left(\frac{1}{2} \gamma^2 \lambda + (\theta-1) A (1-\kappa(\phi+(1-\phi)\lambda)) \right) x_t + z_2}$$

where

$$z_2 = -\frac{1}{s^2} \ln \left(1 - s \left(\frac{1}{2} \gamma^2 - (\theta-1) \kappa A (1-\phi) \right) \omega \right).$$

This yields the second parameter restriction:

$$\left(\frac{1}{2} \gamma^2 - (\theta-1) \kappa A (1-\phi) \right) \omega < s^{-1}.$$

We then have:

$$\begin{aligned} r_{f,t} &= -\theta \ln \beta + \gamma \mu - \frac{1}{2} \gamma^2 \bar{v} - (\theta-1) (\kappa_0 + c(\kappa-1)) \dots \\ &\quad - \left(\frac{1}{2} \gamma^2 \lambda + (\theta-1) A (1-\kappa(\phi+(1-\phi)\lambda)) \right) x_t - z_2. \end{aligned}$$

Plugging in the expression for c , we have

$$\begin{aligned}
r_{f,t} &= -\theta \ln \beta + \gamma \mu - \frac{1}{2} \gamma^2 \bar{v} - (\theta - 1) \kappa_0 + \dots \\
&\quad (\theta - 1) \left(\ln \beta + (1 - \gamma) \mu \theta^{-1} + \frac{1}{2} (1 - \gamma)^2 \bar{v} \theta^{-1} + \kappa_0 + \theta^{-1} z_1 \right) \dots \\
&\quad - \left(\frac{1}{2} \gamma^2 \lambda + (\theta - 1) A (1 - \kappa (\phi + (1 - \phi) \lambda)) \right) x_t - z_2 \\
&= \kappa_0 + \mu + \frac{1}{2} \bar{v} - \gamma \bar{v} + z_1 \dots \\
&\quad - \left(\ln \beta + (1 - \gamma) \mu \theta^{-1} + \frac{1}{2} (1 - \gamma)^2 \bar{v} \theta^{-1} + \kappa_0 + \theta^{-1} z_1 \right) \\
&\quad - \left(\frac{1}{2} \gamma^2 \lambda + (\theta - 1) A (1 - \kappa (\phi + (1 - \phi) \lambda)) \right) x_t - z_2
\end{aligned}$$

The conditional expected log return is:

$$\begin{aligned}
E_t^S [\kappa_0 + \kappa p d_{t+1} - p d_t + \Delta d_{t+1}] &= \kappa_0 + \kappa c - c + A x_t + \mu + E_t^S [-\kappa A x_{t+1}] \\
&= \kappa_0 + c (\kappa - 1) + \mu + A [1 - \kappa (\phi + (1 - \phi) \lambda)] x_t.
\end{aligned}$$

Plugging in the expression for c , we have:

$$\begin{aligned}
E_t^S [r_{t+1}] &= \kappa_0 + \frac{\theta \ln \beta + (1 - \gamma) \mu + \frac{1}{2} (1 - \gamma)^2 \bar{v} + \theta \kappa_0 + z_1}{\theta (1 - \kappa)} (\kappa - 1) \dots \\
&\quad + \mu + A [1 - \kappa (\phi + (1 - \phi) \lambda)] x_t \\
&= \kappa_0 - \left(\ln \beta + \theta^{-1} (1 - \gamma) \mu + \theta^{-1} \frac{1}{2} (1 - \gamma)^2 \bar{v} + \kappa_0 + \theta^{-1} z_1 \right) \dots \\
&\quad + \mu + A [1 - \kappa (\phi + (1 - \phi) \lambda)] x_t.
\end{aligned}$$

The conditional log risk premium is then:

$$\begin{aligned}
E_t^S [r_{t+1} - r_{f,t}] &= \kappa_0 - \left(\ln \beta + \theta^{-1} (1 - \gamma) \mu + \theta^{-1} \frac{1}{2} (1 - \gamma)^2 \bar{v} + \kappa_0 + \theta^{-1} z_1 \right) \dots \\
&\quad + \mu + A [1 - \kappa (\phi + (1 - \phi) \lambda)] x_t \\
&\quad - \left(- \left(\ln \beta + (1 - \gamma) \mu \theta^{-1} + \frac{1}{2} (1 - \gamma)^2 \bar{v} \theta^{-1} + \kappa_0 + \theta^{-1} z_1 \right) \dots \right. \\
&\quad \left. - \left(\frac{1}{2} \gamma^2 \lambda + (\theta - 1) A (1 - \kappa (\phi + (1 - \phi) \lambda)) \right) x_t - z_2 \right) \\
&= \gamma \bar{v} - \frac{1}{2} \bar{v} - z_1 + z_2 + \left(\frac{1}{2} \gamma^2 \lambda + \theta A (1 - \kappa (\phi + (1 - \phi) \lambda)) \right) x_t
\end{aligned}$$

Next, note that:

$$\begin{aligned}
\theta A (1 - \kappa (\phi + (1 - \phi) \lambda)) &= -\frac{1}{2} \theta \frac{\lambda (1 - \gamma) (1 - 1/\psi)}{1 - \kappa (\phi + (1 - \phi) \lambda)} (1 - \kappa (\phi + (1 - \phi) \lambda)) \\
&= -\frac{1}{2} \lambda \theta (1 - \gamma) (1 - 1/\psi) \\
&= -\frac{1}{2} \lambda (1 - \gamma)^2 \\
&= \lambda \left(-\frac{1}{2} + \gamma - \frac{1}{2} \gamma^2 \right)
\end{aligned}$$

So then

$$\frac{1}{2} \gamma^2 \lambda + \theta A (1 - \kappa (\phi + (1 - \phi) \lambda)) = \lambda \left(\gamma - \frac{1}{2} \right)$$

and then

$$E_t^S [r_{t+1} - r_{f,t}] = z_2 - z_1 + \left(\gamma - \frac{1}{2} \right) E_t^S [\sigma_{t+1}^2],$$

which is the same as that we get in the Normal variance shock case, up to an intercept.

The objective risk-premium is:

$$\begin{aligned}
E_t^P [r_{t+1} - r_{f,t}] &= E_t^S [r_{t+1} - r_{f,t}] + \kappa (E_{t+1}^P [pd_{t+1}] - E_{t+1}^S [pd_{t+1}]) \\
&= E_t^S [r_{t+1} - r_{f,t}] - \kappa A (E_{t+1}^P [x_{t+1}] - E_{t+1}^S [x_{t+1}]) .
\end{aligned}$$

We have that:

$$\begin{aligned}
E_t^S [x_{t+1}] &= (\phi + (1 - \phi) \lambda) x_t \\
E_t^P [x_{t+1}] &= (\phi + (1 - \phi) \lambda) x_t + (1 - \phi) E_t^P [\omega \eta_{t+1}^S] \\
&= E_t^S [x_{t+1}] + (1 - \phi) (E_t^P [\sigma_{t+1}^2] - E_t^S [\sigma_{t+1}^2]) .
\end{aligned}$$

Thus:

$$E_t^P [r_{t+1} - r_{f,t}] = E_t^S [r_{t+1} - r_{f,t}] - \kappa A (1 - \phi) (E_t^P [\sigma_{t+1}^2] - E_t^S [\sigma_{t+1}^2]), \quad (77)$$

which again is the same equation we get in the case with Normal variance shocks.

Shocks to realized returns are then:

$$\begin{aligned}
r_{t+1} - E_t^P [r_{t+1}] &= \Delta d_{t+1} - E_t^P [\Delta d_{t+1}] + \kappa (pd_{t+1} - E_t^P [pd_{t+1}]) \\
&= \sigma_{t+1} \varepsilon_{t+1} + \kappa A (-x_{t+1} + E_t^P [x_{t+1}]) \\
&= \sigma_{t+1} \varepsilon_{t+1} + \kappa A (1 - \phi) (-(\sigma_{t+1}^2 - \bar{v}) + E_t^P [(\sigma_{t+1}^2 - \bar{v})]) \\
&= \sigma_{t+1} \varepsilon_{t+1} - \kappa A (1 - \phi) \omega \eta_{t+1} .
\end{aligned} \quad (78)$$

Next, turning the the variance risk premium (VRP), note that the error in variance expectation will feed through in the VRP. In particular:

$$\begin{aligned}
IV_{t-1} &= E_{t-1}^S \left[\frac{M_t}{E_{t-1}^S [M_t]} (\Theta + \bar{v} + \lambda x_{t-1} + \omega \eta_t^S) \right] \\
&= \Theta + \delta + E_{t-1}^S [\sigma_t^2] ,
\end{aligned} \quad (79)$$

where $\delta = E_{t-1}^S \left[\frac{M_t}{E_{t-1}^S[M_t]} \omega \eta_t^S \right]$. To see that this is indeed a constant, note that:

$$\begin{aligned}
E_{t-1}^S \left[\frac{M_t}{E_{t-1}^S[M_t]} \omega \eta_t^S \right] &= \\
E_t^S \left[\frac{\beta^\theta e^{-\gamma\mu + \frac{1}{2}\gamma^2(\bar{v} + \lambda x_t + \omega \eta_{t+1}^S) + (\theta-1)(\kappa_0 + \kappa c - \kappa A(\phi x_t + (1-\phi)(\lambda x_t + \omega \eta_{t+1}^S)) - c + Ax_t)}}{E_t^S \left[\beta^\theta e^{-\gamma\mu + \frac{1}{2}\gamma^2(\bar{v} + \lambda x_t + \omega \eta_{t+1}^S) + (\theta-1)(\kappa_0 + \kappa c - \kappa A(\phi x_t + (1-\phi)(\lambda x_t + \omega \eta_{t+1}^S)) - c + Ax_t} \right]} \omega \eta_{t+1}^S \right] &= \\
E_t^S \left[\frac{e^{\left(\frac{1}{2}\gamma^2 - (\theta-1)\kappa A(1-\phi)\right)\omega \eta_{t+1}^S}}{E_t^S \left[e^{\left(\frac{1}{2}\gamma^2 - (\theta-1)\kappa A(1-\phi)\right)\omega \eta_{t+1}^S} \right]} \omega \eta_{t+1}^S \right] &= \\
E_t^S \left[\frac{e^{m\omega \eta_{t+1}^S}}{E_t^S \left[e^{m\omega \eta_{t+1}^S} \right]} \omega \eta_{t+1}^S \right]. & \tag{80}
\end{aligned}$$

where $m = \frac{1}{2}\gamma^2 - (\theta-1)\kappa A(1-\phi)$. Recall that agents believe $\tilde{\eta}_t^S = \eta_t^S + s^{-1}$ is Gamma distributed with mean s^{-1} and variance 1 (i.e., $k = s^{-2}$). Note that in this case, $\omega \tilde{\eta}_t^S$ is Gamma with shape parameter $k = s^{-2}$ and scale parameter ωs . Also, recall that $E[e^{tx}x]$ is a constant if x is Gamma and $t < s^{-1}$ (see Equation (56)):

$$E_t^S \left[\frac{e^{m\omega \eta_{t+1}^S}}{E_t^S \left[e^{m\omega \eta_{t+1}^S} \right]} \omega \eta_{t+1}^S \right] = -\omega s^{-1} + E_t^S \left[\frac{e^{m\omega \tilde{\eta}_t^S}}{E_t^S \left[e^{m\omega \tilde{\eta}_t^S} \right]} \omega \tilde{\eta}_t^S \right], \tag{81}$$

where

$$E_t^S \left[\frac{e^{m\omega \tilde{\eta}_t^S}}{E_t^S \left[e^{m\omega \tilde{\eta}_t^S} \right]} \omega \tilde{\eta}_t^S \right] = \tilde{s} \left(\frac{\tilde{s}}{\omega s} \right)^{s^{-2}} \frac{\Gamma(s^{-2} + 1)}{\Gamma(s^{-2})}, \tag{82}$$

where $\tilde{s} = -(m - (\omega s)^{-1})^{-1}$. To summarize:

$$IV_{t-1} = \Theta + \delta + E_{t-1}^S [\sigma_t^2], \tag{83}$$

$$\delta = \tilde{s} \left(\frac{\tilde{s}}{\omega s} \right)^{s^{-2}} \frac{\Gamma(s^{-2} + 1)}{\Gamma(s^{-2})} - \omega s^{-1}, \tag{84}$$

which is the same expression as that we get in the case of Normal variance shocks, except for the intercept term.

C. Additional variance signal

In our main model, lagged realized variance is the best forecaster of future realized variance. At the firm-level this is clearly unrealistic as, for instance, earnings announcements are pre-scheduled and a known source of return variance. Even at the

aggregate level investors may have access to additional sources of information than just lagged variance that inform their forecasts of future variance.

In this section we show that it is straightforward to relax this assumption without altering the main mechanism we highlight in this paper. In particular, let the true variance dynamics be the same as in the main paper (Equations (5) and (6)). Assume investors form beliefs based on lagged variance as in the main paper:

$$\begin{aligned} E_{t-1}^S [\sigma_t^2 | \sigma_{t-1}^2, \sigma_{t-2}^2, \dots] &= \bar{v} + \lambda x_{t-1,t} \\ x_t &= \phi x_{t-1} + (1 - \phi) (\sigma_t^2 - \bar{v}) . \end{aligned} \quad (85)$$

In addition, however, let agents at time $t - 1$ also observe a signal that is informative about the time t shock to variance, $\omega \eta_t$:

$$s_{t-1} = \omega \eta_t + k \eta_{t-1}^\perp,$$

where η_{t-1}^\perp is a standard Normal shock uncorrelated with all other shocks. Thus, $\bar{v} + \lambda x_{t-1}$ is a signal about σ_t^2 that has variance ω^2 , while s_{t-1} is a signal about $\omega \eta_t$ that has variance k^2 . Agents' combine these sources of information using Bayes rule:

$$E_{t-1}^S [\sigma_t^2 | s_{t-1}, \sigma_{t-1}^2, \sigma_{t-2}^2, \dots] = \bar{v} + \lambda x_{t-1} + \frac{k^{-2}}{\omega^{-2} + k^{-2}} s_{t-1}. \quad (86)$$

Note that lagged values of the signal s are not useful as lagged σ_{t-j}^2 are observed.

If agents are risk-neutral, we have that:

$$VIX_{t-1}^2 = \bar{v} + \lambda x_{t-1} + \frac{k^{-2}}{\omega^{-2} + k^{-2}} s_{t-1}. \quad (87)$$

Thus, the VIX_{t-1}^2 contains information about σ_t^2 not contained in the history of realized variance, namely $\omega \eta_t$. Since s_{t-1} is orthogonal to the history of realized variance, a projection of VIX^2 onto lags of realized variance is the same as in the baseline model.

D. Expected variance measure and additional data sources

There is a large literature on estimating the conditional stock market variance. Early approaches include the ARCH and GARCH models of Engle (1982) and Bollerslev (1986). More recently, attention has focused on the use of multifrequency data in variance forecasting (see, e.g., Corsi (2009), Chen and Ghysels (2012)).

Bekaert and Hoerova (2014) compare a wide array of variance forecasting models to find the best forecaster of monthly realized market variance, including multifrequency models, models that consider asymmetric response of variance to stock returns, and models that separately include jump and diffusion components in realized variance.

In the sample from 1990, which is the sample where the VIX is available that we also focus on in our analysis, the leading model is the Heterogeneous Autoregressive (HAR) model of Corsi (2009), which uses different frequency lags of realized stock market variance, extended to include the squared VIX as a predictor variable, which they denote CV8. Since the VIX is a key endogenous variable in our model, we use the HAR model without the VIX as our baseline model for estimating expected market variance. However, in this appendix we show that our results are not driven by this specific model, but are essentially unchanged when using the CV8 model as well as simply using lags of monthly realized variance.

The baseline version of the HAR model considered in Bekaert and Hoerova (2014) is:

$$RV_{t+1}^{(22)} = \alpha + \beta_1 RV_t^{(1)} + \beta_2 RV_t^{(5)} + \beta_3 RV_t^{(22)} + \varepsilon_{t+1}, \quad (88)$$

where t measures time in months and $RV_t^{(k)}$ is the sum of the last k days' realized variance. Daily realized variance is measured using 5-minute S&P500 futures returns. In words, the conditional monthly stock market variance for next month is assumed to be affine in the realized variance the last day, the last five days, and the last twenty-two days of the current month. The CV8 model adds the end of month squared VIX to this projection. The time t forecast we use in our baseline analysis (see Table 14) is then:

$$\sigma_t^2 = \hat{\alpha}_t + \hat{\beta}_{1t} RV_t^{(1)} + \hat{\beta}_{2t} RV_t^{(5)} + \hat{\beta}_{3t} RV_t^{(22)}, \quad (89)$$

where the t subscripts on the parameter estimates highlights that we re-estimate the model each month, using only data up until time t to form the time t variance forecast. In Panels A and B of Table 14 we show the same regressions as in Table 3 in the main paper using the CV8 measure estimated in the same way using an expanding sample. To allow reasonable estimation of the initial coefficients, the first variance forecasts are for January 1992. The data for the traded variance claims starts in 1996, which implies a six-year initial estimation period.

D.1 Survey Data

Our main survey measure comes from the Duke-CFO Survey (Graham and Harvey).

Each quarter respondents are asked the following questions.

Over the next year, I expect the annual S&P 500 return will be:

- There is a 1-in-10 chance the actual return will be less than %.
- There is a 1-in-10 chance the actual return will be greater than %.

We take the difference of these two numbers as the 80% confidence interval for CFOs which we average across respondents. Importantly, this is not disagreement but the difference between the mean of the 90th and 10th percentiles. We square this to convert to a measure proportional to variance.

D.1.1 Relation to GARCH (1,1)

Note that if we write:

$$rv_{t+1} = a + b \sum_{j=0}^{\infty} \phi^j rv_{t-j} + \varepsilon_{t+1}$$

we have:

$$\begin{aligned} rv_{t+1} &= a + brv_t + \phi b \sum_{j=0}^{\infty} \phi^j rv_{t-1-j} + \varepsilon_{t+1} \\ &= a + brv_t + \phi (rv_t - a - \varepsilon_t) + \varepsilon_{t+1} \\ &= \tilde{a} + (\phi + b) rv_t - \phi \varepsilon_t + \varepsilon_{t+1} \end{aligned}$$

Thus, the survey expectations and those that we use in our model can be thought of as a GARCH(1,1) process. This suggests another interpretation for how agents form expectations, as using a GARCH process is a reasonable way to forecast volatility. However, our results suggest that while this is a plausible motivation, it would still mean agents use too large a number for ϕ relative to a rational or objective forecast as we show in the survey evidence.

D.2 Additional Data Sources

The table below details our international data sources including starting time periods for each series.

Table 10: Data Sources.

Country	Index	Volatility	Source	History
USA	SP500	VIX	WRDS	From 1/2/1990
France	CAC 40	VCAC	Bloomberg	From 1/3/2000
Canada	sptsx60	VIXC	Montreal Exchange	From 10/2010
UK	FTSE 100	VFTSE	Bloomberg	From 1/4/2000
Germany	DAX	DAX New Volatility (V1XI)	Bloomberg	From 1/2/1992
Japan	Nikkei 225	VXJ	Bloomberg	From 1/5/1998
South Korea	KOSPI	VKOSPI	Bloomberg	From 1/2/2003
Netherlands	AEX	VAEX	Bloomberg	From 1/2000
Switzerland	SMI	V3X	Bloomberg	From 6/28/1999

E. Appendix Tables and Figures: Robustness Checks

In Panels A and B of Table 11 we show the same regressions volatility instead of variances, where volatility is formed as the square root of variance.

Panel A and B of Table 12 show a weighted least squares version of Table 3. In particular, we first run the OLS regressions in Table 3, next we for each regression run

a GARCH(1,1) on the residuals, and then we re-run the regression using the inverse of the conditional variance estimate as weights. Thus, these regressions down weight period with high estimated conditional variance of the estimated residuals from the OLS regressions. The results are qualitatively the same across the board and, though most of the coefficients remain significant, we note that the significance levels are typically lower. Since the standard errors in Table 3 are heteroskedasticity consistent, these results indicate that high variance periods (and periods with high variance of variance, which itself is correlated with variance) indeed are periods where the bias is stronger. That said, the results are not driven exclusively by, say, the financial crisis.

Panels A and B of Table 13 show the results from same regressions again, splitting the sample into before and after 2010. We note that the results are in fact stronger in the period after the financial crisis than before.

Table 11: Stylized Facts: Robustness to using volatility. We repeat our analysis using volatility (standard deviation) in place of variance. Panel A runs predictive regressions of variance dependent returns, the VIX, and future realized variance on expected volatility ($\hat{\sigma}_{t-1}$), a weighted average of expected volatility over the past six months $(1-\phi)\sum_{k=1}^6 \phi^{k-1} \hat{\sigma}_{t-k}$, and implied volatility VIX . In our notation σ_t represents the realized standard deviation of daily market returns in month t . The returns on variance swaps, straddles, and VIX futures have a negative sign, thus representing the premium for insuring against future increases in VIX or variance (so that the variance risk premium is positive on average). Data are monthly from 1990-2020, the variance swap, VIX futures, and straddle return data are 1996-2017, 2004-2020, and 1996-2020, respectively. Standard errors in parentheses use Newey West correction with 12 lags.

Panel A: Variance Returns

	(1) Vix Fut	(2) Straddle	(3) Var Swap	(4) Var Swap (TJ)	(5) $VIX_{t-1} - \sigma_t$	(6) σ_t	(7) VIX_{t-1}
$\hat{\sigma}_{t-1}$	-5.04*** (1.78)	-7.88*** (2.07)	-10.29*** (3.88)	-14.78*** (4.39)	-0.90*** (0.33)	1.39*** (0.33)	0.51*** (0.11)
$\Sigma_{j=1}^6 \phi^j \hat{\sigma}_{t-j}$	5.11** (2.13)	9.66*** (2.67)	10.94** (4.64)	15.20*** (3.83)	0.92*** (0.31)	-0.39 (0.31)	0.52*** (0.09)
N	194	292	264	282	334	334	335
R^2	0.0523	0.0438	0.0104	0.0102	0.101	0.560	0.805

Panel B: Stock Market Returns

	(1) Market	(2) Market	(3) Market	(4) Market
$\hat{\sigma}_{t-1}$	-0.04 (0.18)		-0.52 (0.35)	-0.41 (0.29)
VIX_{t-1}		0.06 (0.15)		0.39* (0.22)
$\Sigma_{j=1}^6 \phi^j \hat{\sigma}_{t-j}$			0.60 (0.40)	
N	340	363	335	340
R^2	-0.00252	-0.00193	0.00196	0.00345

Table 12: Stylized Facts: Robustness using Weighted Least Squares. Panel A runs predictive regressions of variance dependent returns, the VIX, and future realized variance on expected variance ($\hat{\sigma}_{t-1}^2$) and a weighted average of expected variance over the past six months $(1-\phi)\sum_{k=1}^6 \phi^{k-1} \hat{\sigma}_{t-k}^2$. $\hat{\sigma}_{t-1}^2$ represents expected variance at time $t-1$ while σ_t^2 is the realized variance of daily market returns in month t . Variance dependent returns are short positions in VIX Futures, straddles, and variance swaps. These returns represent the premium for insuring against future increases in VIX, variance, or volatility (so that the variance risk premium is positive on average). Panel B runs excess stock returns (market returns over the risk free rate) on expected variance, the average of past expected variance, and the implied variance from the VIX. Data are monthly from 1992-2020, the variance swap, VIX futures, and straddle return data are 1996-2019, 2004-2020, 1996-2020, respectively. We estimate a GARCH(1,1) for each set of first pass residuals and then in the second stage regressions use the inverse conditional variance as the weight. Standard errors in parentheses use Newey West correction with 12 lags.

Panel A: Variance Returns

	(1)	(2)	(3)	(4)	(5)	(6)	(7)
	Vix Fut	Straddle	Var Swap	Var Swap (TJ)	$VIX_{t-1}^2 - \sigma_t^2$	σ_t^2	VIX_{t-1}^2
$\hat{\sigma}_{t-1}^2$	-3.58*	-24.05**	-26.13*	-22.14**	-0.52	1.55**	0.40***
	(2.14)	(11.22)	(14.79)	(10.40)	(0.51)	(0.63)	(0.11)
$\Sigma_{j=1}^6 \phi^j \hat{\sigma}_{t-j}^2$	3.74*	31.12**	31.52	30.58*	0.65	-0.38	0.72***
	(2.19)	(13.64)	(19.73)	(15.70)	(0.46)	(0.50)	(0.09)
N	194	292	264	282	334	334	335

Panel B: Stock Market Returns

	(1)	(2)	(3)	(4)
	Market	Market	Market	Market
$\hat{\sigma}_{t-1}^2$		-1.12	-3.21	-2.62*
		(0.87)	(2.25)	(1.40)
VIX_{t-1}^2		-0.22	1.61	
		(0.98)	(1.46)	
$\Sigma_{j=1}^6 \phi^j \hat{\sigma}_{t-j}^2$			2.78	
			(3.10)	
N	340	363	335	340

Table 13: Stylized Facts: Robustness to Sample Splits. We repeat the analysis from Table 3 in the main text but split the sample into pre and post 2010 (Panels A and C). This shows robustness to the post financial crisis and roughly splits the sample for the variance returns (variance swap and VIX futures). Panel B additionally shows the Results dropping the COVID crisis by omitting 2020.

Panel A: Post 2010

	(1) Vix Fut	(2) Straddle	(3) Var Swap	(4) Var Swap (TJ)	(5) Market	(6) Market	(7) Market	(8) Market
$\hat{\sigma}_{t-1}^2$	-9.03*** (4.20)	-99.93*** (24.16)	-126.42 (87.15)	-79.28 (49.02)	0.93 (1.16)		-14.49** (6.25)	-11.28*** (4.04)
$\Sigma_{j=1}^6 \phi^j \hat{\sigma}_{t-j}^2$	12.60*** (4.83)	137.01*** (32.71)	181.23* (107.73)	137.67** (55.99)			23.81** (9.85)	
VIX_{t-1}^2					2.58*** (0.90)			11.66*** (3.68)
N	123	123	95	113	123	123	123	123
R^2	0.0786	0.0702	0.0103	0.00410	-0.00492	0.0253	0.0926	0.0951

Panel B: Post 2010 (Excluding COVID)

	(1) Vix Fut	(2) Straddle	(3) Var Swap	(4) Var Swap (TJ)	(5) Market	(6) Market	(7) Market	(8) Market
$\hat{\sigma}_{t-1}^2$	-3.57 (3.36)	-61.49** (24.81)	-126.42 (87.15)	-79.28 (49.02)	1.37 (1.75)		-8.28 (5.27)	-7.34** (3.41)
$\Sigma_{j=1}^6 \phi^j \hat{\sigma}_{t-j}^2$	7.35*** (3.15)	103.73*** (29.41)	181.23* (107.73)	137.67** (55.99)			13.78** (5.40)	
VIX_{t-1}^2					3.26*** (1.12)			7.96*** (1.77)
N	120	120	95	113	120	120	120	120
R^2	0.0305	0.0284	0.0103	0.00410	-0.00517	0.0234	0.0240	0.0437

Panel C: Pre 2010

	(1) Vix Fut	(2) Straddle	(3) Var Swap	(4) Var Swap (TJ)	(5) Market	(6) Market	(7) Market	(8) Market
$\hat{\sigma}_{t-1}^2$	-3.99*** (1.05)	-21.73*** (5.79)	-36.54*** (9.66)	-56.20*** (14.35)	-0.83 (0.71)		-1.52 (1.27)	-2.70*** (0.69)
$\Sigma_{j=1}^6 \phi^j \hat{\sigma}_{t-j}^2$	3.15*** (1.05)	28.42*** (7.80)	39.27*** (13.02)	59.87*** (15.06)			0.89 (1.52)	
VIX_{t-1}^2					-0.15 (1.19)			2.59** (1.13)
N	71	169	169	169	217	240	212	217
R^2	0.101	0.0157	0.00497	0.00157	0.00297	-0.00404	-0.000777	0.0117

Table 14: Stylized Facts: Robustness to Alternative Measure of Expected Variance. We proxy for expected variance using the “CV8” model of Bekaert and Hoerova (2014). Panel A runs predictive regressions of variance dependent returns, the VIX, and future realized variance on expected variance ($\hat{\sigma}_{t-1}^2$) and a weighted average of expected variance over the past six months $(1 - \phi)\sum_{k=1}^6 \phi^{k-1} \hat{\sigma}_{t-k}^2$. $\hat{\sigma}_{t-1}^2$ represents expected variance at time $t - 1$ while σ_t^2 is the realized variance of daily market returns in month t . Variance dependent returns are short positions in VIX Futures, straddles, and variance swaps. These returns represent the premium for insuring against future increases in VIX, variance, or volatility (so that the variance risk premium is positive on average). Panel B runs excess stock returns (market returns over the risk free rate) on expected variance, the average of past expected variance, and the implied variance from the VIX. Data are monthly from 1992-2020, the variance swap, VIX futures, and straddle return data are 1996-2019, 2004-2020, 1996-2020, respectively. Standard errors in parentheses use Newey West correction with 12 lags.

Panel A: Variance Returns

	(1) Vix Fut	(2) Straddle	(3) Var Swap	(4) Var Swap (TJ)	(5) $VIX_{t-1}^2 - \sigma_t^2$	(6) σ_t^2	(7) VIX_{t-1}^2
$\hat{\sigma}_{t-1}^2$	-4.83*** (1.89)	-28.51** (11.97)	-37.8*** (13.65)	-48.87*** (12.91)	-1.25** (0.52)	1.58*** (0.59)	0.36** (0.15)
$\Sigma_{j=1}^6 \phi^j \hat{\sigma}_{t-j}^2$	5.04** (2.10)	36.59** (14.64)	41.79** (20.16)	55.14*** (16.71)	1.26** (0.49)	-0.63 (0.51)	0.62*** (0.08)
Observations	194	292	264	282	334	334	335
R2	0.0466	0.0219	0.00401	0.00145	0.155	0.457	0.828

Panel B: Stock Market Returns

	(1) Mkt	(2) Mkt	(3) Mkt	(4) Mkt
$\hat{\sigma}_{t-1}^2$	-0.49 (0.79)		-2.50* (1.38)	-3.72*** (1.12)
VIX_{t-1}^2		0.41 (1.16)		3.97** (1.61)
$\Sigma_{j=1}^6 \phi^j \hat{\sigma}_{t-j}^2$			2.70 (2.26)	
Observations	340	363	335	340
R2	-0.00102	-0.00170	0.00186	0.0193

Table 15: Higher Frequency Weekly Data. We run the same predictive regression of variance dependent returns (straddles, VIX futures, and variance swaps) using weekly return data rather than monthly. We regress future returns at the 1 to 4 week horizon, with column numbers representing the lead in weekly returns (non-cumulative). The right hand side variables are $\hat{\sigma}_{t-1}^2$ and a weighted average of expected variance over the past six months (26 weeks) $(1 - \phi)\sum_{k=1}^{26} \phi^{k-1} \hat{\sigma}_{t-k}^2$. Our value of $\phi = 0.5^{1/4}$ to be consistent with higher frequency data compared to our monthly regressions. Data are weekly from 1992-2019, the variance swap and straddle returns are from 2004-2019 and VIX futures data are from 1996-2019. Standard errors in parentheses use Newey West correction with a 6 month lag.

Panel A: Straddle Returns by Horizon

	(1)	(2)	(3)	(4)
$\hat{\sigma}_{t-1}^2$	-6.50*** (1.35)	-6.32*** (1.14)	-3.60*** (1.24)	-2.24* (1.23)
$\sum_{j=1}^{26} \phi^j \hat{\sigma}_{t-j}^2$	7.15*** (1.69)	7.44*** (1.62)	5.08*** (1.50)	3.93*** (1.52)
N	1,222	1,222	1,222	1,222
R^2	0.0127	0.0125	0.0043	0.0022

Panel B: VIX Futures Returns by Horizon

	(1)	(2)	(3)	(4)
$\hat{\sigma}_{t-1}^2$	-3.47*** (1.23)	-4.12*** (1.29)	-2.57** (1.04)	-2.25** (1.01)
$\sum_{j=1}^{26} \phi^j \hat{\sigma}_{t-j}^2$	2.86** (1.27)	3.79** (1.50)	2.48** (1.11)	2.34* (1.20)
N	716	716	716	716
R^2	0.0130	0.0176	0.0050	0.0031

Panel C: Variance Swap Returns by Horizon

	(1)	(2)	(3)	(4)
$\hat{\sigma}_{t-1}^2$	-7.57** (3.49)	-12.06*** (3.22)	-5.72* (3.19)	-3.12 (3.12)
$\sum_{j=1}^{26} \phi^j \hat{\sigma}_{t-j}^2$	7.64** (3.56)	12.88*** (4.09)	7.18** (3.36)	4.89 (3.46)
N	1,222	1,222	1,222	1,222
R^2	0.0018	0.0072	0.0006	-0.0006

Table 16: Shiller Survey. We fit the actual volatility process and the survey expectations to an exponential weighted average on past realized variance. That is, we fit: $y_{t,t+k} = a + b\sum_{i=1}^J \phi^{i-1} \sigma_{t-i}^2 + \varepsilon_t$ and report the estimated ϕ where we choose J to be 12 periods, and k as the horizon at which investors forecast variance in the survey (six months). We then repeat this replacing σ_t^2 on the left hand side with the expectation of variance from the survey over the same horizon. A higher ϕ from the expectations data signifies that expectations rely more on variance farther in the past compared to the optimal forecast for volatility. The Shiller survey is available monthly and corresponds to a six month forecast horizon. Standard errors are below in parentheses.

Shiller Survey: Dependence on Past Variance (ϕ)			
	Survey	Future Variance	
ϕ	0.73*** (0.04)	0.03 (0.29)	
N	193	193	

Table 17: Long Sample. This Table regresses excess stock returns (market returns over the risk free rate) on lagged realized variance and the lagged weighted average of past expected variance $(1 - \phi)\sum_{k=1}^6 \phi^{k-1} \hat{\sigma}_{t-k}^2$. Data are monthly the first two columns (Post War) are post 1945 US data while the last two columns are post 1926. Standard errors in parentheses use Newey West correction with 12 lags.

	Post War Data		Great Depression	
	(1)	(2)	(3)	(4)
$\hat{\sigma}_{t-1}^2$	-0.98*** (0.43)	-1.34*** (0.46)	-0.18 (0.73)	-0.33 (0.69)
$\sum_{j=1}^6 \phi^j \hat{\sigma}_{t-j}^2$		1.50*** (0.67)		0.52 (0.54)
N	878	878	1,111	1,111
R^2	0.006	0.013	0.00	0.00

Table 18: Change in volatility, equity risk premium, and variance risk premium. We run three predictive regressions $y_{i,t+1} = a_i + b\Delta_6\sigma_t + \varepsilon_{i,t+1}$ where $\Delta_6\sigma_{i,t}$ is the 6 month change in volatility of the stock market index for country i . As dependent variables, y , we use the equity risk premium (future index return over the risk free rate, $r_{i,t+1} - r_{i,t}^f$ labeled ERP), future volatility ($\sigma_{i,t+1}$), and the volatility risk premium (difference between volatility index and future realized volatility, $VIX_{i,t} - \sigma_{i,t+1}$ labeled VRP). Data are monthly. The first columns use all countries, the last use only US data. In our panel regressions standard errors are clustered by time.

Panel A: Volatility						
	All Countries			US Only		
	(1) ERP	(2) Vol	(3) VRP	(4) ERP	(5) Vol	(6) VRP
$\Delta_6\sigma_t$	-0.15* (0.09)	0.27*** (0.08)	-0.06*** (0.02)	-0.25*** (0.08)	0.32*** (0.05)	-0.09*** (0.02)
N	1,786	1,786	1,786	340	340	340
R-squared	0.01	0.15	0.05	0.03	0.13	0.04
Country	All	All	All	USA	USA	USA
Panel B: Variance						
	All Countries			US Only		
	(1) ERP	(2) Vol	(3) VRP	(4) ERP	(5) Vol	(6) VRP
$\Delta_6\sigma_t^2$	-0.88*** (0.34)	0.23** (0.10)	0.04 (0.10)	-1.66*** (0.39)	0.36*** (0.04)	-0.13*** (0.03)
N	1,786	1,786	1,786	340	340	340
R-squared	0.02	0.10	0.01	0.05	0.19	0.05
Country	All	All	All	USA	USA	USA

Figure 4: Impulse Response to Variance Shock: Robustness. We run a first order VAR of expected variance, market excess returns (denoted ERP for equity risk premium), the variance risk premium (VRP), and the log price dividend ratio, following Figure 2 in the main text, and plot the response to an expected variance shock. Shaded regions indicate 95% confidence intervals constructed using bootstrap. Panel A uses volatility in place of variance. Panel B weights the observations by the inverse of lagged volatility (weighted least squares). See text for more detail.

