Volatility Expectations and Returns

Lars Lochstoer and Tyler Muir*

October 21, 2019

Abstract

We provide evidence that agents have slow moving beliefs about stock market volatility. This is supported in survey data and is also reflected in firm level option prices. We embed these expectations into an asset pricing model and show that we jointly explain the following stylized facts (some of which are novel to this paper): when volatility increases the equity and variance risk premiums fall or stay flat at short horizons, despite higher future risk; these premiums appear to rise at longer horizons after future volatility has subsided; strategies that time volatility generate alpha; the variance risk premium forecasts stock returns more strongly than either realized variance or risk-neutral variance (VIX); changes in volatility are negatively correlated with contemporaneous returns. Slow moving expectations about volatility lead agents to initially underreact to volatility news followed by a delayed overreaction. This results in a weak, or even negative, risk-return tradeoff at shorter horizons but a stronger tradeoff at longer horizons (beyond where one can strongly forecast volatility). These dynamics are mirrored in the VIX and variance risk premium which reflect investor expectations about volatility.

1. Introduction

The link between risk and return is at the core of many asset pricing models, though there is weak empirical evidence of a risk-return tradeoff over time in the stock

*UCLA. We thank seminar participants at UCLA and Stavros Panageas, Mike Chernov, Pierre-Olivier Weill, Barney Hartman-Glaser, Andrea Eisfeldt, Valentin Haddad, Andy Atkeson, and Alan Moreira 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 providing data.

market.[1] For example, consider the canonical mean-variance representative agent model equilibrium[2]

$$E_t[r_{t+1}] - r_{f,t} = \gamma \sigma_t^2 \tag{1}$$

where γ is the representative agents risk-aversion, σ_t^2 is the conditional variance of the market and $E_t[r_{t+1}] - r_{f,t}$ is the (log) market risk premium. This equation implies a tight link between volatility and expected returns. While this model may seem like a simplistic "straw man," the strong positive link between conditional volatility and expected returns is a fairly general feature of leading structural equilibrium asset pricing models in the literature.[3]

Empirically a long literature finds that the relationship between conditional variance and risk premiums is weak at best. In particular, measures of conditional stock market variance or volatility do not strongly positively forecast returns (Glosten et al., 1993) so that risk-return ratios weaken when volatility rises, leading to profitable volatility timing strategies (Moreira and Muir, 2017). In fact, we show that current volatility relative to its average over the period of a few months if anything appears to negatively forecast future stock returns over the next month, rather than strongly positively forecasting returns as in the basic model. However, we also show that longer term lags of volatility do appear to predict returns with a positive sign. Yet, these lags of volatility forecast returns at horizons for which they no longer forecast future volatility as strongly, especially once controlling for more recent lags. Hence they no longer strongly correlate with expected volatility – the object that should be linked to expected returns in theory.[4] Measures of the variance risk premium display similar patterns. For example, increases in volatility appear to negatively predict the premium on VIX futures at short horizons of up to a month (Cheng, 2018) and we show this is true for variance swaps and other measures of the variance risk premium as well. That is, claims that provide insurance against future volatility, which are unconditionally expensive, appear "too cheap" after volatility rises. Similar to the pattern in stock returns, increases in volatility positively forecast the variance risk premium at longer horizons, peaking around six months.

[1] See, for example, Glosten, Jagannathan, and Runkle (1993) or Moreira and Muir (2017) among many others.

[2] E.g., Merton (1980)

[3] Moreira and Muir (2017) show that the basic risk-return relation is very 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 is general in a wide class of models if σ_t^2 is replaced by risk-neutral variance which we will consider empirically as well.

[4] See also Brandt and Kang (2004) for related findings using a different approach.

While the relation between conditional volatility and expected returns is weak, there are some indirect facts potentially more in favor of a risk-return tradeoff. First, realized stock returns are highly negatively related to contemporaneous innovations in volatility (French, Schwert, and Stambaugh, 1987). This strong negative relation suggests a discount rate effect, since a higher discount rate in response to higher volatility pushes current stock prices – and hence realized returns – lower. However, it is of course puzzling if this effect is a discount rate effect that returns next period are not higher. Second, the variance risk premium (VIX^2 minus realized variance) *does* strongly forecast stock returns suggesting equity risk premiums are high when the variance risk premium is high (Bollerslev, Tauchen, and Zhou, 2009).[5]

The goal of this paper is to propose a model which jointly tackles these dynamics between realized volatility, the VIX, the variance risk premium, and stock returns. We do so by making one change to an otherwise standard Epstein Zin equilibrium model with stochastic volatility: we allow the representative agent to have slow moving expectations about volatility, which we show is supported by survey evidence. In our model agents form beliefs about volatility by taking a weighted average of past volatility observations. When agents pay relatively too much attention to past volatility, they temporarily underreact to increases in volatility and then subsequently overreact. When agents see volatility increase they do react partially, driving prices down so that volatility is associated with negative returns contemporaneously. However, the initial underreaction means prices can continue to fall in the next period making ex-ante "risk premiums" appear weak, or even negative, and the subsequent overreaction keeps prices depressed for longer before eventually bouncing back, making it appear as though risk premiums are high well after the shock to volatility has largely subsided.[6] Market expectations of volatility (the VIX) mirror this, meaning the variance risk premium can initially fall before slowly rising at longer horizons. Thus the model matches the conditional dynamics of both equity and variance risk premiums following an increase in volatility, though we show this reconciles several additional pieces of evidence as well.

To see how the model delivers these dynamics, suppose that the true volatility process follows an AR(1) as in our model. This means the agents' objective best guess for next months volatility only depends on current volatility. However, suppose,

[5]See also Drechsler and Yaron (2011)

[6]We use the term risk premium as an empirical observation about the behavior of expected returns though in our model variance risk premiums only move because of agents biased expectations. Similarly, and in keeping with the literature, we often use "risk-nuetral variance" to refer to the VIX though much of the behavior in VIX in our model is driven by biased beliefs rather than standard risk-pricing.

also as in our model, that agents form a forecast for next months volatility based on a weighted average of volatility over the past several months. Then, if volatility increases, agents expectations of volatility will increase somewhat but will underreact to the news – agents will not update their forecast about next periods volatility strongly enough since they average across several lags of volatility. However, they will still demand a higher equity premium due to the higher subjective expectation of volatility, hence stock prices will fall, generating a strong negative correlation between stock returns and volatility innovations (French et al., 1987). Similarly, "risk-neutral" expectations of volatility (the VIX) will rise, though also not strongly enough relative to a rational forecast. We use the term "risk-neutral" in keeping with terminology in the literature, though in our setting the VIX also reflects expectations under the agents potentially subjective beliefs. Because the VIX does not initially rise strongly enough, VIX minus the rational expectation of realized volatility (the variance risk premium) will fall, thus volatility news will negatively forecast the variance risk premium as is true empirically (see Cheng (2018), which we extend).

When the next period arrives, suppose for simplicity there is no additional news about future volatility. With rational expectations, expected volatility declines as volatility mean reverts toward the unconditional average, and the equity premium declines as well while stock prices rise on average. Instead, in our model, the agent may again update his expectation about volatility because he averages two periods of relatively elevated volatility, and may require a relatively higher equity premium. If the initial underreaction is strong enough, this pushes equity prices down further through an additional discount rate effect. Through this channel, the ex-ante news about higher volatility in the previous period can appear to forecast negative stock returns in the next period, hence the initial innovation in volatility can negatively forecast returns in the short term. Thus, the standard risk return tradeoff appears weak or even negative (Glosten et al., 1993), and this leads to apparently profitable volatility timing strategies (Moreira and Muir, 2017, 2019) because expected volatility increases by more than the conditional expected return.[7] This also nicely reconciles the puzzling evidence that shocks to volatility do line up with contemporaneous drops in stock prices, consistent with a discount rate effect (French et al., 1987), despite the fact that a clear link between volatility and next period returns is weak and typically has the wrong sign.

Going forward, however, stock returns will be higher on average in the periods after the initial underreaction plays out. In fact, objective expected returns will remain high for longer than they would under rational expectations because the

[7]See also Brandt and Kang (2004) who find expected returns fall in response to an increase in variance.

volatility shock effectively lasts longer in the agents minds. For example, even after (true) expected volatility has returned to normal the agent may believe volatility is high because they average over a period in which past volatility had increased. Thus, the news about volatility can negatively forecast stock returns in the near term but positively forecast them in the longer term, despite the fact that news about volatility strongly forecasts future volatility in the near term, but only weakly forecasts future volatility in the longer term. The variance risk premium will mirror this behavior with a drop in the near term followed by an increase in the longer term, reflecting agents slow moving expectations. Further, and importantly, the measured variance risk premium will forecast stock returns appropriately at virtually all horizons leading to a tight link between equity and variance risk premiums as we see in the data (Bollerslev et al., 2009). However, in the model this occurs because of biased beliefs rather than because of rational risk premiums, and the model can account for the otherwise puzzling fact that while the variance risk premium is a strong forecaster of returns, neither the VIX or realized variance are individually strong forecasters of returns.

After developing the model and this intuition, we calibrate the model parameters to see if it can quantitatively account the facts outlined above and show that we can do so reasonably well. Importantly, our model nests the fully rational case but we show that slow moving expectations about volatility are important to match the data. We also compare our calibrated model to Bollerslev et al. (2009) who show that the variance risk premium can predict returns in a long-run risk model with stochastic volatility of volatility. A main point of difference is in their model while the variance risk premium does predict returns, both variance or the VIX on their own predict returns even more strongly. We show that we can account for the fact that the variance risk premium does strongly forecast returns, though either the VIX or realized variance alone do not. Further, our model helps account for the profitability of volatility timing strategies as in Moreira and Muir (2017) which typical structural asset pricing models in the literature don't match.

Next, we use survey data on volatility and uncertainty about stock returns from two sources (the Graham and Harvey CFO survey and the Shiller survey) and document that the surveys exhibit slow moving expectations as in our model. In particular, 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 recent volatility. This provides direct support for the mechanism of slow moving expectations in our paper.

We also document results at the firm level, where again 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 also Poteshman (2001) for related work). This is true even when we include time fixed effects that control for aggregate movements in firm level volatility which makes a risk based explanation difficult since this test focuses on idiosyncratic movements in firm level volatility. The firm level analysis provides further support for our story of underreaction and also provides robustness to our main empirical results which rely on aggregate market data and hence a relatively smaller sample.

Finally, we consider several potential objections to both the facts and our modeling choices. First, there is empirical evidence that volatility has two components: a higher frequency component and a lower frequency, more persistent component that our initial model ignores. It is possible that incorporating this component can match some of the longer horizon empirical results: that expected returns and variance risk premiums do rise in the long run after variance shocks. However, this feature alone still fails to account for the short term behavior, and suggests that risk premiums should be highest in the near term, counter to the data. Thus, just incorporating a long run component in variance but ignoring slow moving expectations will not easily match the patterns in the data. Second, we consider issues with extreme realizations of variance (which is positively skewed) in our main empirical facts. Third, in our baseline model volatility comes from the volatility of cash flows. This assumption is for simplicity but not crucial for our story. Fourth, we consider other explanations for our results including models with rational inattention and heterogeneous agents. While these models may indeed explain some features of the data, we explain why they do not easily generate the joint behavior of the facts we study. Finally, we study additional evidence of our channel including Nagel, Reck, Hoopes, Langetieg, Slemrod, and Stuart (2017) who show how investors respond to volatility changes empirically. In their data, more sophisticated and more experienced investors respond more quickly to changes in volatility. This makes sense if we expect these investors to have a smaller degree of bias in forming expectations of volatility.

1.1 Related Literature

Our model is similar to other models of extrapolation from past data including Barberis, Greenwood, Jin, and Shleifer (2015), Collin-Dufresne, Johannes, and Lochstoer (2016), and Nagel and Xu (2019), though our focus is on volatility rather than returns or cash flows.[8] Our paper also fits into a broader literature on under and

[8]The agents in our model also differ in spirit from other forms of extrapolation such as Nagel and Xu (2019) where agents rely too much on recent data because their memory "fades." In

overreaction, for example, Daniel, Hirshleifer, and Subrahmanyam (1998) and Barberis, Shleifer, and Vishny (1998). These papers are able to generate underreaction and delayed overreaction through potentially different underlying behavioral biases. At least qualitatively, this underreaction and delayed overreaction is precisely what is needed to match the data. We focus on extrapolation in particular to generate our results without taking a strong stand on where this extrapolation comes from at a deeper level. We find extrapolation appealing both because it appears to exist in many contexts (e.g., Barberis et al. (2015)) and also because it is analytically very tractable in our setting. Further, there is extensive empirical evidence that agents do appear to extrapolate from past experiences when forming expectations (Malmendier and Nagel, 2011; Greenwood and Shleifer, 2014; Glaeser and Nathanson, 2017). Importantly, the survey data we study suggests extrapolation of volatility, hence it suggests expectation formation similar to what we use in the model.

It may appear surprising that extrapolation in our context leads to *under*reaction as extrapolation is most often associated with *over*reaction. For example, in models where agents extrapolate from past returns or cash flows agents typically overreact to news (e.g., see Barberis et al. (2015)). However, this depends on the persistence of the true process and the noise about the conditional mean from observing past realizations. For returns or cash flow growth, whose conditional means move very slowly and are more difficult to measure using past data due to a low signal to noise ratio, it makes sense that extrapolating from recent data will lead to overreaction. Our setting is different both because volatility is easier to observe and also because it moves very quickly. If one takes a moving average of past volatility that includes many months, then one would not put *enough* weight on recent volatility and put too much weight on past volatility. This leads to *under*reaction.

In our model there is naturally underreaction in volatility expectations in the short term followed by delayed overreaction of expected volatility in the longer term. This matches the spirit of Giglio and Kelly (2017) who show focus on overreaction of long term volatility expectations. In particular, they argue that longer term expectations of volatility are too volatile relative to those at short horizons, a form of overreaction in long term expectations. However, we assume a simple structure for agents beliefs such that the dynamics for volatility under Q are easily captured by a simple AR structure – thus we can't speak directly to their result. Further, we have little to say about unconditional variance risk premiums (Dew-Becker, Giglio, Le, and Rodriguez, 2017). We could obtain facts about unconditional variance risk premiums

our context agents focus too little on recent data and too much on past data. See also Coibion and Gorodnichenko (2015) and Mankiw and Reis (2002) for models of sticky expectations more generally, and Jiang, Krishnamurthy, and Lustig (2018) for extrapolation in exchange rates.

in extensions if agents are biased on average, but in our baseline model agents beliefs are not biased on average, only conditionally. We focus on this case because we focus on matching facts related to the conditional, rather than unconditional, behavior of risk premiums.

The fact that at the aggregate level changes in volatility negatively forecast stock returns while lags positively forecast mirror the results at the firm level documented by Rachwalski and Wen (2016).[9] In fact Rachwalski and Wen (2016) suggest a story similar to ours for their findings, though there are many important differences in our work. We focus on the aggregate market since this is where the standard risk-return relation in equilibrium asset pricing models should apply and we target the additional facts on the variance risk premium as well. We also study and estimate a quantitative model. Similarly, Poteshman (2001) documents underreaction and subsequent overreaction in options markets and attributes this to cognitive biases but does not link this to equity risk premiums and does not view this through the lens of a quantitative equilibrium model. We confirm and extend these firm level results

2. Stylized Empirical Facts

We begin by focusing on the main stylized empirical facts that we will target in our model. We study US data from 1990 to 2018 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 variance for the month) and realized variance (computed as the sum of squared daily log returns within a month). Stock return data use the (log) return on the S&P500 index over the risk-free rate taken from Ken French. We also take the log price dividend ratio from CRSP based on value weighted returns.[10] In addition, we study variance swap returns and VIX futures returns from Dew-Becker et al. (2017) and Cheng (2018), respectively, though these have shorter samples (variance swap returns are 1996-2017 and VIX futures returns are 2004-2017). We define the variance risk premium (VRP) as the squared VIX minus realized variance, though to supplement this we also use the actual return series as well. When using returns (e.g., variance swap or VIX futures) we take the negative of the returns, so the

[9] "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." Rachwalski and Wen (2016)

[10] 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.

implication is the return for selling variance or being short the VIX. This means the unconditional premium for both returns is positive as the exposure to volatility is negative.

Figure 1 plots impulse responses of realized variance (RV), the log price dividend ratio on the market, excess stock returns, and implied variance (VIX squared) to a shock to RV using a VAR(1) with RV ordered first, so all variables can respond contemporaneously to a realized variance shock. We compute the path of the variance risk premium by computing the impulse response to the VIX minus the impulse response of expected RV over the next period as implied by the VAR (e.g., $VRP_k = E_t[VIX^2_{t+k}] - E_t[RV_{t+k+1}]$). We see RV follows roughly an AR(1), spiking after the shock and mean reverting back after about 6 months (consistent with persistence of about 0.72 monthly). Stock returns and stock prices fall contemporaneously with an increase in RV at time 0, consistent with a discount rate effect of volatility shocks (French et al., 1987). However, expected returns are negative in period 1 as well, meaning stock prices continue to fall next period on average. Expected returns gradually increase, eventually going positive several months out, but this is after the volatility shock has largely subsided from the perspective of the left panel. Thus the equity premium rises later, after volatility has largely mean-reverted, which is quite different from a standard risk-return tradeoff view. The variance risk premium shows very similar patterns, with the predicted premium being negative in month 1 before slowly rising and becoming positive beyond month 3. Most notably, the premiums have a hump-shaped pattern: they appear low initially but continue to rise as future volatility falls. The shaded regions indicate 95% confidence intervals based on bootstrapping the residuals in the VAR.

The results are broadly consistent with the main empirical findings in the literature. First, the risk-return tradeoff overall is fairly weak, and often estimated to be negative, for stock returns on lagged realized variance (Glosten et al., 1993; Brandt and Kang, 2004; Moreira and Muir, 2017, 2019). We note that these papers come to this conclusion over a variety of sample periods. Relative to the literature we also show that the risk-return tradeoff appears to be flat or negative at first but increases with horizon with a hump shaped response. This result is also mirrored in Brandt and Kang (2004) who study the period from 1946-1998 and find a hump shaped response with the expected return initially falling and then rising after future volatility subsides. The time 0 response for stock returns is strongly negative as in French et al. (1987). Our variance risk premium results are consistent with this type of pattern as well. A variance shock initially predicts a decline in the variance risk premium (VIX squared minus realized variance) but then gradually predicts an increase in the variance risk premium. This result is consistent with Cheng (2018)

who studies a claim on implied volatility (VIX) instead of realized variance. Finally, the variance risk premium and equity risk premium are tightly linked as in Bollerslev et al. (2009).

In the appendix, we study alternative specifications for this VAR, including using logs of VIX and RV and using weighted least squares based on lagged stock market volatility. Both help downweight the importance of high volatility observations which is important. We see largely the same patterns as in the main VAR, though these results do suggest periods of high variance are important. In particular, the VRP response at time 1 is somewhat muted, though the VRP continues to display the hump shaped pattern as before, as does the equity risk premium. Taken together, this still supports a "slow moving" response of the premiums to realized variance, despite the fact that future realized variance still mean reverts fairly quickly. 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.

To further explore these facts and understand what drives the results, we run predictive regressions in Table 1. In particular, we use current realized variance, the VIX, and the average of realized variance over the prior six months to forecast equity risk premiums, variance risk premiums, and future realized variance over the next month. The average variance over the past six months helps summarize the information in longer lags of variance and we will interpret it as a potential proxy for slow moving expectations of agents. Further, the direct regressions allow us to use actual returns on variance claims rather than taking the implied premium from the VIX.

Current realized variance negatively forecasts returns next month (column 1) and negatively forecasts variance risk premiums (columns 5-6), but positively forecasts future variance (column 7). In contrast, the average of past variance typically positively forecasts risk premiums but doesn't strongly forecast future variance (column 7). This indicates that risk premiums are high when variance in the past has been high, but not when objective conditional variance is high. Importantly, we also use actual returns on variance claims rather than the implied premium inherent in taking squared VIX minus expected realized variance as used in the VAR. Columns 5 and 6 use variance swap returns Dew-Becker et al. (2017) and VIX futures returns Cheng (2018) and show they are predictable by current and past variance with the same patterns. The negative sign indicates that it is cheap to insure against future volatility when volatility increases, similar to Cheng (2018). Column 7 predicts future variance using current variance and the average of past variance, and confirms

that the bulk of the information in expected variance comes from the first lag, consistent with variance roughly following an AR(1). Column 8 repeats this regression using the VIX on the left hand side instead of future realized variance. In contrast to column 7, in column 8 both current variance and average past variance are positively associated with the VIX and combine to explain most of the variation in the VIX with an R^2 above 80% Past variance has a large and significant coefficient of about 0.3. If we think of the VIX as capturing market expectations about variance, this suggests these expectations depend much more on longer lags of past variance than the objective forecast from column 7. This provides suggestive evidence of slow moving market expectations of variance. Columns 7 and 8 together explain the patterns in columns 5-6: because implied variance (the VIX) depends strongly on past variance, but the objective forecast of future variance depends mostly on current variance, then $VIX^2 - RV$ (the variance risk premium) will negatively load on current variance and positively load on past variance. This is exactly what we find in the regressions in 5-6.

In column 2 of Table 1 we replicate the results from Bollerslev et al. (2009) that the variance risk premium is a robust predictor of stock returns, and this result is even stronger in our sample which adds the more recent ten years of data compared to the sample used in Bollerslev et al. (2009). Thus, there is a strong link between implied variance risk premiums and equity risk premiums. Importantly, however, the variance risk premium is a strong predictor of returns, while the VIX or realized variance individually are not as show in columns 3 and 4. VIX squared has little to no forecasting power for returns, and if anything has the wrong sign with a negative coefficient. Realized variance appears to predict returns somewhat, but also with a negative sign. This is 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 contributed to the equity risk premium. In light of our results in column 8 that VIX depends strongly on past variance, this suggests an alternative explanation for the variance risk premium forecasting returns based on agents paying too much attention to past variance in forming their expectations.

In the appendix, we repeat these regressions using volatility in place of variance, which will have fewer extreme realizations and using various subsamples as robustness. The variance risk premium return results hold, with recent volatility negatively predicting risk premiums and past variance typically positively predicting, though this is somewhat weaker in some subsamples. Volatility also predicts future volatility with a fairly similar pattern as before for variances. However, the statistical significance in some of the return forecasting regressions is weaker (in particular the

result that current volatility negatively forecasts returns in a univariate regression), highlighting that much of the results in our main table come from high variance realizations. While this suggests caution in terms of reliably predicting stock returns with changes in variance, our main point is in the benchmark models the coefficient on variance would typically be strongly positive rather than negative, which is at odds with what we see empirically. Hence the null in these models is a strongly positive coefficient on current variance for the risk premium. The most robust facts are that the variance risk premium forecasts stock returns, that the risk premiums displays a hump-shaped pattern with a relative decrease at first and a rise further out, and the current variance if anything is associated with lower rather than higher risk premiums.

Standard asset pricing models typically struggle with these facts because they suggest that an increase in risk (volatility) will be associated with heightened risk premiums at all horizons, and this relationship will be strongest in the near term and will decay with horizon when 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.

Table 4, 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 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.

We also find supportive evidence in survey data: surveys that captures investors perception of volatility or uncertainty are slow moving and load significantly more on past realizations of volatility compared to optimal forecasts. This is similar to the regression in Table 1 column 9 that market expectations of variance depend strongly on past variance. We return to the survey evidence in a later section.

We return to these facts after presenting the model, and we discuss additional empirical robustness in the Appendix, including results using international data. In the Appendix, we show that the negative relation between future returns and current realized variance is robust going back to the 1950s but less so if one includes

the Great Depression (see Figure 9). However, most importantly, the point estimate is still negative, meaning even including the Great Depression there is no strong evidence of a positive risk-return tradeoff (e.g., variance does not reliably predict higher future returns) though the evidence of a statistically significant negative coefficient on the next month return does change. Hence, we interpret the negative one month prediction with some caution in light of the longer sample. In our parameter calibration, we focus on the more recent US data from 1990 when we have the VIX and variance risk premium data.

3. The Model

In this section, we develop an asset pricing model similar to that in Bollerslev, Tauchen and Zhou (2009) 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:

$$\begin{aligned} \Delta d_t &= \mu + \sigma_t \varepsilon_t, & (2) \\ \sigma_t^2 &= \bar{v} + \rho\left(\sigma_{t-1}^2 - \bar{v}\right) + \omega\eta_t, & (3) \end{aligned}$$

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 (3) implies that variance can go negative. For ease of exposition we follow, e.g., Bansal and Yaron (2004) and Bollerslev, Tauchen, and Zhou (2009), and proceed as if σ_t^2 is always non-negative. In the Appendix, 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. Bollerslev, Tauchen, and Zhou (2009) additionally let the variance of variance follow a square root process, thereby generating a time-varying variance risk premium in a rational model. We show in Section 3.4 that their model cannot account for the empirical patterns we discuss in this paper.

We assume a representative stockholder with consumption equal to aggregate dividends whose marginal utility prices all claims in the economy. The agents' expectations of the conditional variance of dividend growth are given by:

$$\begin{aligned} E_{t-1}^S\left[\sigma_t^2\right] &= \bar{v} + \lambda x_{t-1}, & (4) \\ x_t &= \phi x_{t-1} + (1-\phi)\left(\sigma_t^2 - \bar{v}\right) \\ &= (1-\phi)\sum_{j=0}^{\infty}\phi^j\left(\sigma_{t-j}^2 - \bar{v}\right). & (5) \end{aligned}$$

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. We assume $0 < \phi, \lambda < 1$.

Under agents' beliefs, the shock to variance is:

$$\begin{aligned} \omega\eta_t^S &\equiv \sigma_t^2 - \bar{v} - \lambda x_{t-1} \\ &= v\left(\sigma_{t-1}^2 - \bar{v}\right) - \lambda x_{t-1} + \omega\eta_t, \end{aligned} \tag{6}$$

where $v\left(\sigma_{t-1}^2 - \bar{v}\right) - \lambda x_{t-1} = E_{t-1}^P\left[\sigma_t^2\right] - E_{t-1}^S\left[\sigma_t^2\right] = E_{t-1}^P\left[\omega\eta_t^S\right]$ 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:

$$x_t = \left(\phi + (1-\phi)\lambda\right) x_{t-1} + (1-\phi)\,\omega\eta_t^S. \tag{7}$$

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 2 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:

$$M_t = \beta^\theta e^{-\frac{\theta}{\psi}\Delta d_t + (\theta-1)r_t}, \tag{8}$$

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 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:

$$pd_t = c - Ax_t, \tag{9}$$

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:

$$E^S_{t-1}\left[r_t - r_{f,t}\right] = (\gamma - \frac{1}{2})E^S_{t-1}\left[\sigma^2_t\right] + \delta_r, \tag{10}$$

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:

$$Var^S_{t-1}\left(r_t\right) = \Theta + E^S_{t-1}\left[\sigma^2_t\right], \tag{11}$$

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

The *objective* risk premium, however, is:

$$E^P_{t-1}\left[r_t - r_{f,t}\right] = E^S_{t-1}\left[r_t - r_{f,t}\right] - \kappa\left(1-\phi\right)A\left(E^P_{t-1}\left[\sigma^2_t\right] - E^S_{t-1}\left[\sigma^2_t\right]\right), \tag{12}$$

where the P superscript on the expectation denotes that it is taken using the true, objective variance dynamics. To see where Equation (12) comes from, recall that the shock to agents beliefs about variance is predictable (see Equation (6)). The mistake is persistent, which magnifies its effect on prices as given by the term $-\kappa A\left(1-\phi\right)$.[11]

If $\phi > 0$, agents make mistakes in their conditional variance expectations. These mistakes are reflected in current discount rates and therefore prices. Consider a

[11]This expression is found by using the Campbell-Shiller return approximation and noting that $E^P_{t-1}\left(-\kappa pd_t\right) - E^S_{t-1}\left(-\kappa pd_t\right) = -\kappa A\left(E^P_{t-1}\left[x_t\right] - E^S_{t-1}\left[x_t\right]\right) = -\kappa A\left(1-\phi\right)\left(E^P_{t-1}\left[\sigma^2_t\right] - E^S_{t-1}\left[\sigma^2_t\right]\right)$.

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, $E^P_{t-1}[\sigma^2_t] > E^S_{t-1}[\sigma^2_t]$. 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^2_t on average is higher than they had expected. This leads to a predictable decline in the price-dividend ratio under the objective measure. The upper right panel of Figure 3 shows the impulse-response of the log price-dividend ratio to a volatility shock. As expected, 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. Subsequently, given the too persistent variance expectations, agents eventually overreact to the volatility shock, which leads to $E^P_{t+j-1}\left[\sigma^2_{t+j}\right] < E^S_{t+j-1}\left[\sigma^2_{t+j}\right]$ for some $j > 0$. In this case, the second term in Equation (12) becomes positive and the conditional risk premium overshoots. This is shown in the lower right Panel of Figure 3 .

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:

$$r_t - E^P_{t-1}[r_t] = -\Theta^{1/2}\eta_t + \sigma_t\varepsilon_t, \tag{13}$$

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.

3.2 Variance risk premium dynamics

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

$$RV_t \equiv \Theta + \sigma^2_t, \tag{14}$$

where RV_t stands for realized variance at time t. 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:

$$0 = E^S_{t-1}[M_t(RV_t - IV_{t-1})]. \tag{15}$$

Thus:

$$IV_{t-1} = E^S_{t-1}[R_{f,t}M_tRV_t]. \tag{16}$$

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:

$$VRP_{t-1} = IV_{t-1} - E^P_{t-1}\left[RV_t\right]. \tag{17}$$

If realized variance is high (low) in bad times, this risk premium is positive (negative).

Our 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:

$$\left(r_t - E_{t-1}\left[r_t\right]\right)^2 = \Theta\eta_t^2 + 2\sigma_t\Theta^{0.5}\eta_t\varepsilon_t + \sigma_t^2\varepsilon_t^2. \tag{18}$$

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.[12] Setting $\eta_t^2 = \varepsilon_t^2 = 1$ and $\eta_t\varepsilon_t = 0$ in Equation (18) gives the realized variance in Equation (14).[13]

The equilibrium implied variance is:

$$IV_{t-1} = E^S_{t-1}\left[RV_t\right] + \delta_{IV}, \tag{19}$$

where $E^S_{t-1}\left[RV_t\right] = E^S_{t-1}\left[\Theta + \sigma_t^2\right] = \Theta + \bar{v} + \lambda x_{t-1}$ and $\delta_{IV} = \left(\frac{1}{2}\gamma^2 - \frac{1/\psi-\gamma}{1-1/\psi}\kappa\left(1-\phi\right)A\right)\omega^2$. 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:

$$\begin{aligned} VRP_{t-1} &= IV_{t-1} - E^P_{t-1}\left[RV_t\right] \\ &= \delta_{IV} + E^S_{t-1}\left[RV_t\right] - E^P_{t-1}\left[RV_t\right] \\ &= \delta_{IV} + E^S_{t-1}\left[\sigma_t^2\right] - E^P_{t-1}\left[\sigma_t^2\right]. \end{aligned} \tag{20}$$

[12] That is, if $W_t^{(1)}$ and $W_t^{(2)}$ are standard Brownian motions with uncorrelated innovations, $\int_t^{t+1}\left(dW_t^{(j)}\right)^2 = \int_t^{t+1} dt = 1$ for $j = \{1,2\}$ and $\int_t^{t+1} dW_t^{(1)}dW_t^{(2)} = 0$.

[13] 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.

Thus, the dynamics of the variance risk premium share a component of the dynamics of the equity risk premium (Equation (12)), 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 2018. 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), which is within the range of values used by Bollerslev, Tauchen, and Zhou (2009). 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.[14] 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.

[14]This 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.

3.4 Comparison with the data

Figure 3 shows impulse responses in the model and the data from a VAR with RV, log excess market returns, and realized VRP. The RV shock is ordered first, and we consider the impulse-responses to a one standard deviation shock to RV. The corresponding impulse-responses from the data are plotted with a black solid line, with $\pm$ two standard error bands in grey. 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.

Both calibrations of the model are consistent with the autocorrelation pattern of RV, as shown in the top left plot. A positive shock increases variance on impact and decays monotonically and relatively quickly. However, the impact of an RV shock on the conditional market and variance risk premiums (the bottom plots) is very different across the two models. While in both the rational and extrapolative models the market price decreases contemporaneously with a positive shock to variance (top left plot), the response in the rational model is to immediately increase the conditional risk premium due to the usual risk-return trade-off (Equation (12)), at odds with the empirical facts. In the extrapolative 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 2. The equity premium subsequently overshoots due to the slow-moving expectations of the extrapolative agents, consistent with the pattern in the data. The same is true for the variance risk premium, although in this case the pattern is stronger as its dynamics are only affected by the mistake in expectations (see Equation (20)). The rational version of the model has no effect on the variance risk premium from an RV shock, again at odds with the data.

Panel B of Table 2 shows unconditional moments from the model and their counterparts in the data. In addition to the moments we match, we note that stock returns and shocks to realized variance are negatively correlated as in the data. Thus, we are able to account for the discount rate effect documented in French, Stambaugh, and Schwert (1987) even though a shock to variance in fact *decreases* the equity premium in the short-run as shown in the bottom right panel of Figure 3. The long-run response, however, is positive, and it is this long-run discount rate response that dominates in the price response (see top right plot in Figure 3).

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. 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 predicts future excess market returns as in the data. To highlight how the extrapolative model of this paper differs in terms of asset price dynamics, we in Figure 4 show regression coefficients from forecasting regressions at horizons from 1 to 12 months from both models, as well as the data.[15] The solid red line shows the coefficients from the extrapolative model, while the dashed black lines show the coefficients from the BTZ model.

The top left plot shows the regression coefficients from univariate regressions of monthly log excess market returns on the k-month lagged variance risk premium ($IV_{t-k} - RV_{t-k}$; we follow BTZ here in the definition of VRP as a predictor). As in the data, the variance risk premium positively predicts excess market returns strongly with a coefficient of about 5 in both models. In the extrapolative model, the prediction power is short-lived and goes to zero after about 4 months, while in the BTZ model the persistence is somewhat higher. Overall both models do a good job matching the relationship between market returns and the VRP.

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)). The top right plot of Figure 4 shows monthly excess return forecasting regressions using RV at different lags. Empirically, there is a marginally significant *negative* response for the one-month, which turns positive as the lag length increases. This is in contrast to the strong *positive* risk premium response in BTZ. In fact, while the coefficients in the data range between -1 and 1, the BTZ model has a 1-month regression coefficient of 10, due to a strong risk-return trade-off in this model, which monotonically decreases with lag length following the AR(1) response of RV itself. The extrapolative model has a weak risk-return trade-off, much closer to the data. There is a slight positive initial response (regression coefficient around 1), and the pattern is hump-shaped as the lag length increases as in the data. The bottom left plot of Figure 4 shows the return forecasting regression coefficients with lags of VIX^2 (IV in the models) on the right hand side of the regressions. In the data, the point estimate is close to zero at all horizons, whereas in the BTZ model there is again a strong counter-factual positive response (as for RV case in this model, the coefficient is about 10 at the 1-month horizon). The extrapolative model is largely within the two standard error bands of the data, with a 1-month coefficient of 2.5. In sum, in the extrapolative model the VRP is a stronger

[15] To 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 counterfacturally high persistence of RV_t.

return predictor than either RV or the VIX^2, as in the data, whereas the opposite is the case in the BTZ model. These strong risk-return tradeoff implications are not specific to the BTZ model, but are a shared feature of many of the leading structural asset pricing models. For example, Moreira and Muir (2017) show a strong risk-return relation is an implication of models with habits, long run risk, rare disasters, and financial intermediation. Thus, our model helps reconcile the empirically weak risk-return tradeoff while keeping the predictive power of the variance risk premium.

The bottom right plot of Figure 4 shows the response of the realized variance risk premium on lagged values of RV. In the data the initial response is negative, turning positive with a hump-shaped response as the lag length of RV increases. The extrapolative model matches both the initial negative response and the following positive hump-shape, though it overshoots this shape somewhat in the 3- to 8-month lag range. In contrast, the RV coefficients in the BTZ model are counter-factually effectively zero at all lag lengths.[16] Thus our model is able to account for the facts in Cheng (2018) that the variance risk premium appears to initially fall when measures of risk rise before eventually increasing at longer horizons.

3.5 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:

$$\alpha = -\frac{c}{E\left[RV_t\right]} Cov\left(E_t\left[RV_{t+1}\right], \frac{\mu_t}{E_t\left[RV_{t+1}\right]}\right), \tag{21}$$

where $E_t\left[r_{t+1} - r_{f,t}\right] = \mu_t$ 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 extrapolative model, 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:

$$\frac{\mu_t}{E_t\left[rv_{t+1}\right]} \approx \frac{\bar{\mu}}{\bar{v}+\Theta} + \frac{1}{\bar{v}+\Theta}\left(\mu_t - \bar{\mu}\right) - \frac{\bar{\mu}}{\left(\bar{v}+\Theta\right)^2}\left(E_t\left[rv_{t+1}\right] - \bar{v} - \Theta\right), \tag{22}$$

[16]The variance risk premium *is* time-varying in the BTZ model, but the state-variable that governs this time-varition (the variance of variance) is locally uncorrelated with realized variance, which is why the regression coefficients on lagged variance are effectively zero.

and report the alpha for the volatility-managed market portfolio in Table 2 as:[17]

$$\alpha \approx -0.6 \times \frac{1}{\bar{v}+\Theta} Cov\left(E_t\left[rv_{t+1}\right], \mu_t - \frac{\bar{\mu}}{\bar{v}+\Theta} E_t\left[rv_{t+1}\right]\right), \tag{23}$$

which is equal to 2.4% annualized — in the same order of magnitude as the 4.9% Moreira and Muir (2017) document.

3.6 Sensitivity to Parameter Values

How important is the extrapolative parameter ϕ in our model for matching the data? Table 3 revisits our main stylized facts (e.g., from Table 1 and the literature) in the first column and compares model values as we vary the degree of extrapolation 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 $\lambda = 0.9$ in the remaining columns while increasing ϕ to 0.4, 0.6, 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 (six month moving 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. This is natural since current variance contains all information about expected future variance. 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 strongly on ϕ, and quantitatively is about the same for all values of ϕ we examine. The reason for this is that there are two effects which quantitatively cancel each other out in our calibrations: the first is that a higher ϕ implies a lower

[17] Moreira 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.

reaction to volatility news through slow moving expectations. On the other hand, a higher ϕ leads agents expectations to be more persistent that 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 extrapolation in our model may appear large, persistence in variance actually implies only modest mistakes. 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 0.6 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 $\phi = 0.6$ (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 extrapolation matches the moments on balance better than the rational benchmark.

4. Additional Evidence: Survey Data and Firm Level Analysis

4.1 Survey Data on Volatility

Survey data on investors expectations is especially useful because it allows us to evaluate the main mechanism in our model using direct data on expectations. In addition, Giglio, Maggiori, Stroebel, and Utkus (2019) show that survey data on investor beliefs about risk translates directly into actions in terms of portfolio allocations. Specifically, they find that investors substantially reduce their portfolio allocation to stocks when they think stocks are riskier in terms of greater probability of a significant decline in the stock market.

We bring survey data related to volatility from two sources. The first is the Graham and Harvey survey of CFOs which is quarterly from 2001. 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 does capture how spread out agents view the return distribution, and under the view of a normal distribution would perfectly capture agents expectations about volatility. Our second source of survey evidence 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. Again, we view this as correlated with agents perception of risk and volatility though it is still an imperfect measure. We use the monthly Shiller sample which begins in July of 2001.

We proceed in two ways to assess whether survey data display slow moving expectations about volatility. First, we fit survey expectations and actual realized variance over the period investors are asked to forecast as an exponential weighted average of past variance exactly as in our specification in the model:

$$y_{t \to t+k} = a + b \sum_{i=1}^{J} \phi^{i-1} \sigma^2_{t-i} + \varepsilon_t$$

where $y_{t\to t+k}$ is the actual future realized variance from time t to $t+k$, and then $y_{t\to t+k}$ is replaced by survey expectations of variance instead. When using realized variance on the left hand side, we compute forward looking realized variance over the horizon which corresponds to the survey expectations about volatility (e.g., k is 1 year for CFO survey and 6 months for the Shiller survey). We take J, the number of lags of realized variance, to be 12 periods (longer lags produce similar results). We estimate ϕ in both cases, where a higher ϕ from survey data indicates more reliance on past variance compared to the optimal forecast. This specification has the benefit that it maps exactly to our model setup for beliefs.

Table 5 gives our estimates. We find $\phi_{survey} > \phi_{RV}$ meaning survey expectations depend much more on past volatility than optimal forecasts indicate. We find ϕ_{RV} is economically and statistically fairly small, consistent with realized variance being fairly well approximated by an AR(1) as discussed earlier, while we find ϕ_{survey} to be large, between 0.75 (Shiller) and 0.77 (CFO). The estimates from these two totally separate surveys thus deliver similar degrees of extrapolation from past variance. The substantial amount of extrapolation we see in the survey data is a bit larger than what we estimate from our model, meaning the expectations bias from pure survey data appears stronger than the bias we estimate from only financial market data. The value in our model is thus conservative relative to the survey data. This provides independent evidence consistent with our model.

To further assess the degree of investors slow moving expectations, we run a vector autoregression (VAR) with future realized variance as well as the reported expectations from the survey. We order future variance first, followed by the survey expectation and plot the impulse response to a variance shock. Results are given in Figure 5. 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, consistent with expectations continuing to rise after the initial shock. 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.

4.2 Firm Level Analysis

We revisit our stylized aggregate facts at the firm level (stock level). 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 1, we forecast equity risk premiums, variance risk premiums, and future realized variance over the next month using the change in realized variance from month t to $t-6$. We use the change over six months, rather than the 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 90th 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 Poteshman (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.

4.3 Evidence on Actual Trading Behavior

Nagel et al. (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. 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.

5. Extensions and Alternative Models

5.1 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 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, Nagel 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.

Heterogenous 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 state space 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. 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).

5.2 Model Extensions

We extend the model to incorporate richer, and more realistic, volatility dynamics. In particular we assume that the volatility of volatility is time varying along the lines of Bollerslev et al. (2009). This helps us match variation in the volatility of volatility. An appendix discusses this extension.

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, which is clearly counterfactual. In particular, empirically the dividend yield is much more persistent than the VIX, though the two are positively correlated. This highlights that the dividend yield could also be driven by forces outside our model. In particular, an extension of our model with time-varying expected dividend growth would generate additional movements in the dividend yield and, if they were highly persistent, could generate the difference in persistence. 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.

6. Conclusion

We show that underreaction followed by delayed overreaction 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 extrapolate from past volatility and we estimate the degree to which they do so 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 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 much more strongly than either variance or implied variance, as in the data. Survey evidence directly supports the idea that agents have slow moving expectations about volatility, as does evidence at the firm level.

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.

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

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.

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.

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.

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

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

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.

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

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., 1980, On estimating the expected return on the market: An exploratory investigation, *Journal of Financial Economics* 8, 323 – 361.

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, Daniel Reck, Jeffrey Hoopes, Patrick Langetieg, 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.

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

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.

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

7. Tables / Figures

Table 1: Stylized Facts. We run predictive regressions of future excess stock returns (market returns over the risk free rate), future variance risk premiums (measured using variance swap returns r_{var}, and VIX futures returns r_{VIX}), and future realized variance on various measures of past variance, average of past variance over 6 months ($\overline{\sigma^2}_{t-1,t-6}$), and implied volatility from the VIX. In our notation σ_t^2 represents the realized variance of daily market returns in month t. The returns on variance swaps 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 1990-2018, the variance swap and VIX futures data are 1996-2017 and 2004-2017, respectively. Standard errors in parentheses use Newey West correction with 12 lags.

	Excess Stock Returns				Variance Risk Premium		Variance	
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
	$r^e_{M,t+1}$	$r^e_{M,t+1}$	$r^e_{M,t+1}$	$r^e_{M,t+1}$	$r_{var,t+1}$	$r_{VIX,t+1}$	σ^2_{t+1}	VIX^2_t
σ_t^2	-1.91			-1.39	-0.74	-0.37	0.71	0.51
	(0.42)			(0.49)	(0.24)	(0.10)	(0.08)	(0.04)
$\overline{\sigma^2}_{t-1,t-6}$	1.46				0.60	0.24	0.02	0.31
	(0.64)				(0.36)	(0.12)	(0.06)	(0.04)
$VIX_t^2 - \sigma_t^2$		4.58						
		(0.67)						
VIX_t^2			-0.17					
			(1.16)					
N	335	341	341	341	264	166	335	335
Adj. R^2	3.3%	7.0%	0.0%	2.0%	0.8%	5.0%	51.1%	80.9%

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

Panel A: Parameters			
Parameter	Description	Value	Targeted Moment(s)
γ	Risk Aversion	2.7	Equity Premium
ψ	Elasticity of Intertemporal Substitution	2.2	Literature / VRP
$\bar{v}$	Unconditional Variance (Monthly)	0.26%	Data
ρ	Persistence of Variance	0.72	Data
ω	Volatility of Variance Shocks (Monthly)	0.23%	Data
ϕ	Degree of Extrapolation	0.6	VRP Response
λ	Scale of Expectations	0.9	Volatility of VIX
Panel B: Moments			
Moment	Description	Model	Data
$E[r_m] - r_f$	Equity Premium (Annual)	7.1%	7.2%
$\sqrt{E[RV_t]}$	Square Root Avg. Variance (Annual)	18%	18%
$\rho(RV_t, RV_{t-1})$	Persistence of Variance (Monthly)	0.72	0.72
$\sigma(RV_t)$	Volatility of Variance (Monthly)	0.45%	0.46%
$\sigma(VIX_t^2)$	Volatility of VIX^2 (Monthly)	0.33%	0.33%
$\rho(VIX_t^2, RV_t)$	Correlation RV and VIX^2 (Monthly)	0.89	0.86
$\rho(VIX_t^2, VIX_{t-1}^2)$	Persistence of VIX^2 (Monthly)	0.92	0.81
$\alpha\left(\frac{c}{RV_{t-1}} r_{m,t}, r_{m,t}\right)$	Volatility-Managed Alpha (Moreira Muir)	2.4%	4.9%
$\rho(r_{m,t}, \widetilde{RV_t})$	Correlation of Returns and Vol Shocks	-0.24	-0.38

Table 3: Stylized Facts in Model and Sensitivity to ϕ. We compare our main facts from Table 1 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 1990-2018 as corresponding to the results in Table 1. We show how our results change as we increase the extrapolation 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 $\lambda = 0.9$.

	Data	Model			
		$\phi = 0, \lambda = \rho$	$\phi = 0.4$	$\phi = 0.6$	$\phi = 0.8$
	(1)	(2)	(3)	(4)	(5)
Risk Return Tradeoff					
σ_t^2	-1.91	2.71	1.22	0.05	-1.39
	(0.42)				
$\overline{\sigma^2}_{t-1,t-6}$	1.46	0	1.63	2.72	3.31
	(0.79)				
R^2	3.1%	5.3%	3.3%	3.3%	3.6%
Volatility-Managed Alpha					
α	4.86	-0.09	0.80	2.20	6.71
	(1.56)				
Correlation: Realized Returns and Vol Shocks					
	-0.38	-0.24	-0.24	-0.24	-0.23
Forecasting Returns with Variance Risk Premium					
VRP_t	4.58	0	7.29	4.94	3.96
	(0.67)				
R^2	7.0%	0%	2.1%	2.1%	2.8%
Expected Variance Risk Premium ($VRP_t = VIX_t^2 - E_t[\sigma_{t+1}^2]$)					
σ_t^2	-0.18	0	-0.07	-0.25	-0.47
	(0.03)				
$\overline{\sigma^2}_{t-1,t-6}$	0.22	0	0.27	0.45	0.52
	(0.03)				
Correlation: VIX^2 and Realized Variance					
	0.86	1	0.96	0.89	0.73

Table 4: Stock level analysis. We repeat our results at the stock level. We run three forecasting regressions $y_{i,t+1} = a_i + b\Delta_6\sigma_t^2 + \varepsilon_{i,t+1}$ where $\Delta_6\sigma_{i,t}$ is the 6 month change in realized variance at the stock level for firm i (the realized variance estimates on the right hand size 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, $r_{i,t+1} - r_{i,t}^f$ labeled ERP), future variance ($\sigma_{i,t+1}^2$), and the variance risk premium (difference between implied variance from option metrics and future realized variance, $VRP_t = IV_{i,t}^2 - \sigma_{i,t+1}^2$ where IV is implied volatility). Data are monthly but realized variance uses daily data with 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
$\Delta_6\sigma_t^2$	-0.129	0.253^{***}	-0.104^{***}	-0.040	0.188^{***}	-0.070^{***}
	(0.137)	(0.067)	(0.036)	(0.075)	(0.046)	(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 5: Survey Expectations. We fit the actual volatility process and the survey expectations to an exponential weighted average on past realized variance. That is, we fit: $\sigma^2_{t,t+k} = a + b\Sigma_{i=1}^{J}\phi^{i-1}\sigma^2_{t-i} + \varepsilon_t$ 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 or one year). We then repeat this replacing σ^2_t 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 (CFO) which is available quarterly and corresponds to a one year forecast horizon and the Shiller survey which is available monthly and corresponds to a six month forecast horizon. Standard errors are below in parentheses.

Dependence on Past Variance (ϕ)		
Source	Survey	Future Variance
CFO	0.77^{***}	0.05
	(0.10)	(0.31)
Shiller	0.73^{***}	0.03
	(0.04)	(0.29)

Figure 1: Impulse Response to Variance Shock, US data. VAR of realized variance, market excess returns (denoted ERP for equity risk premium), the variance risk premium (VRP), and the log price dividend ratio (labeled Prices). VRP is implied variance (VIX^2) minus next period realized variance. Responses are for a one-standard deviation shock to realized variance at time 0. Vertical dashed lines at time 1 highlight the predicted, rather than realized, ERP and VRP. X-axis is in months. Price and equity returns are in percent (monthly), while RV and the VRP are in terms of standard deviations of variance shocks. Shaded regions indicate 95% confidence intervals constructed using bootstrap. Panel B weights the observations by the inverse of lagged realized volatility (weighted least squares). Panel C uses logs of RV and the VIX in place of levels. See text for more detail.

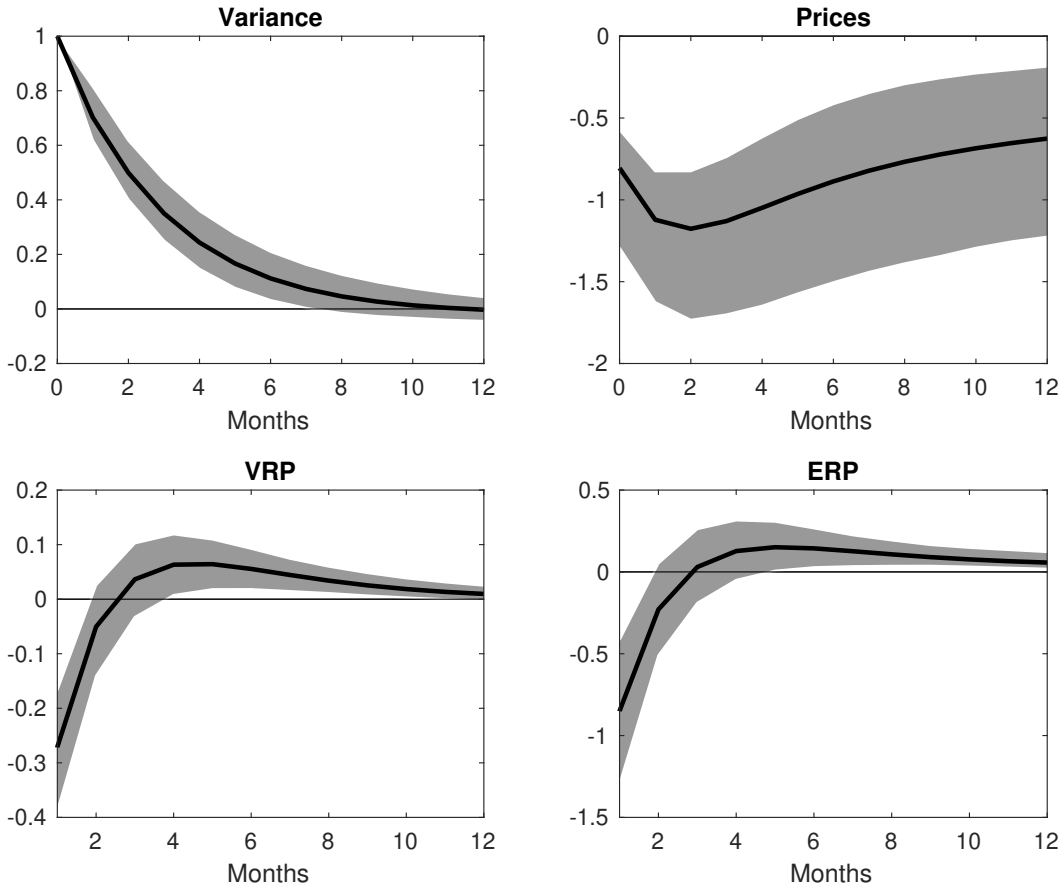


Figure 2: Dynamics of variance expectations in the model. We plot the behavior of agents expectations of volatility in our main calibration (blue line) and the true path of expected volatility (dashed red line) in response to a one standard deviation increase in variance in our model. The dot dashed black line provides an alternative calibration when we set the scale of expectations lower. Because agents extrapolate from 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 x-axis is in months.

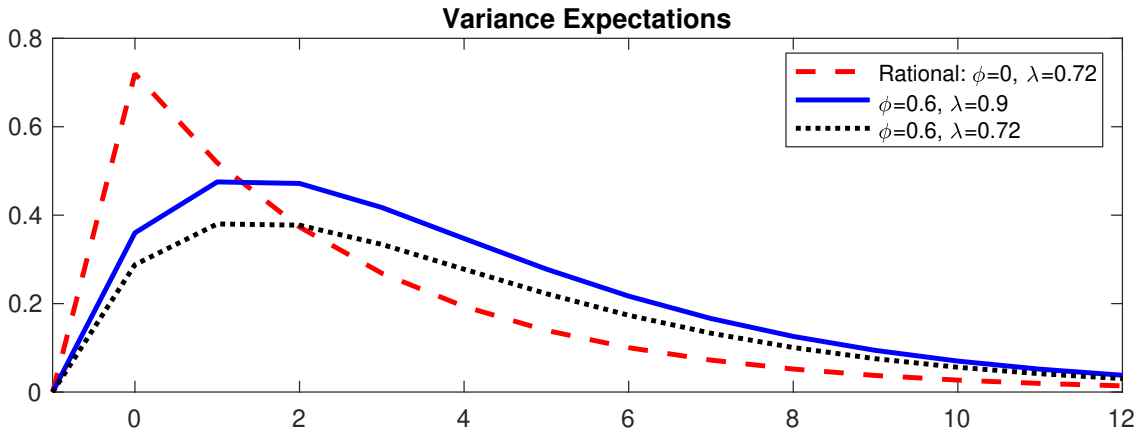


Figure 3: Impulse responses: data vs model. We plot the behavior of expected stock returns and variance risk premiums in the data vs the model at various horizons for a one standard deviation shock to variance. The black line shows the impulse response from the data using a VAR of realized variance, excess stock returns, and the variance risk premium. The blue dashed line repeats this using simulated data from the estimated model. The red dot dashed line repeats this exercise in the simulated model data but imposes no extrapolation bias (rational model). Price and equity returns are given in units of percent per month, while RV and the VRP are given in terms of standard deviations of the variance shocks. The x-axis is in months.

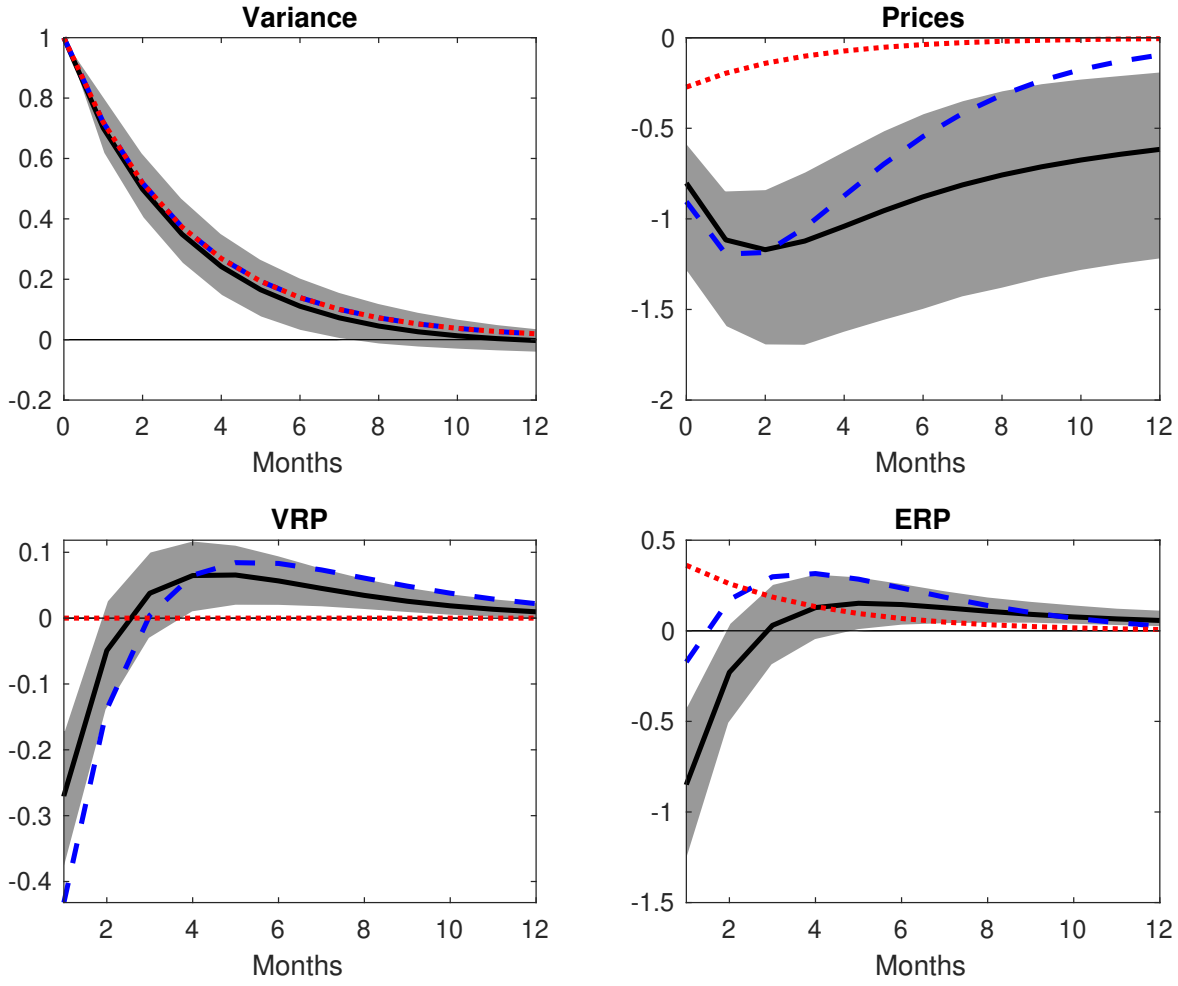


Figure 4: Regressions: data vs our calibrated model and the model of Bollerslev et al. (2009). We run forecasting regressions and plot coefficients by horizon in the data (blue line), our calibrated model (red line) and the Bollerslev et al. (2009) model (black dot dashed line). Regressions are of the form $y_t = a_k + b_k x_{t-k} + e_t$ and we plot coefficients b_k from $k = 1, .., 12$ months. Dashed blue lines indicate the 95% confidence interval in the data based on Newey-West standard errors. Returns are (log) excess returns on the market, VRP is the variance risk premium (VIX^2-RV) and RV is realized variance. Data are monthly.

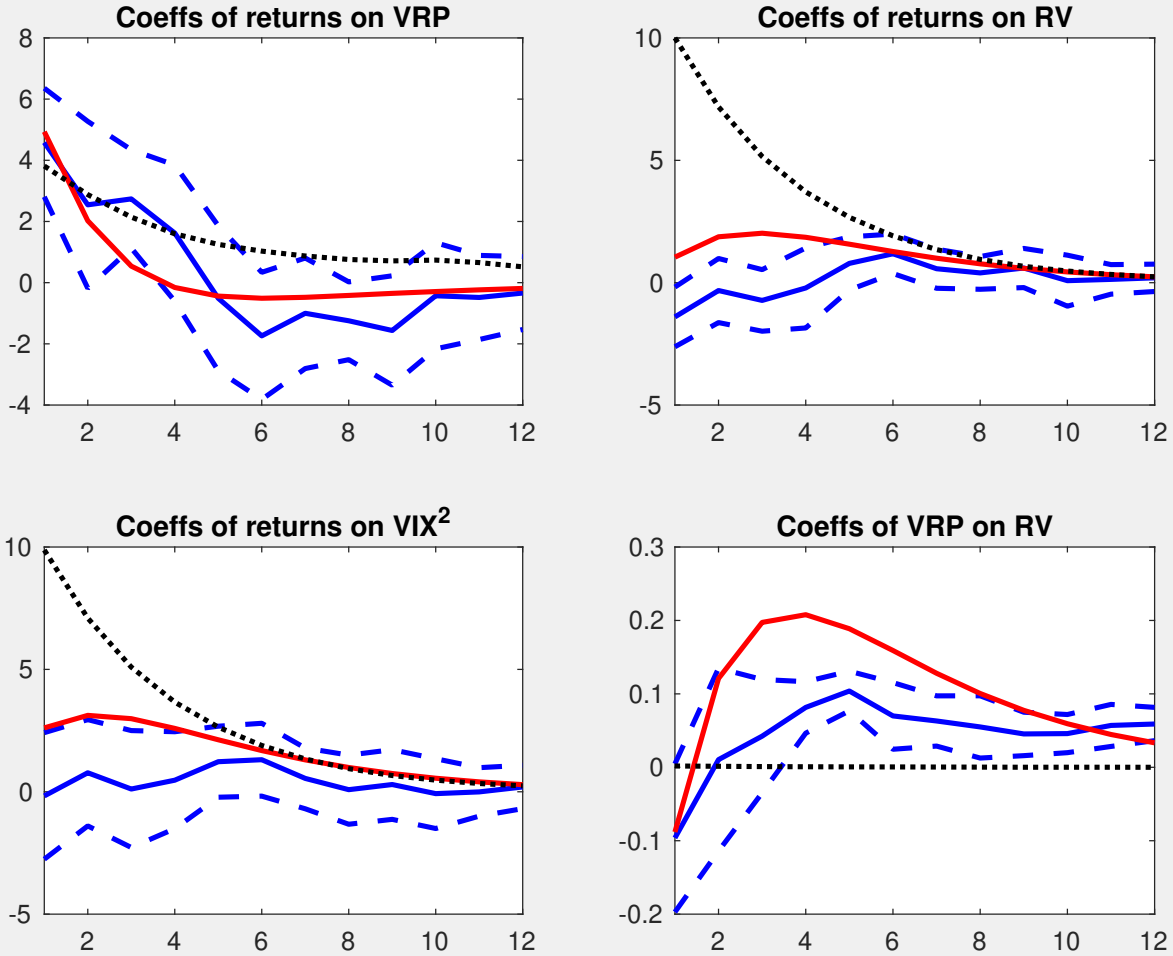
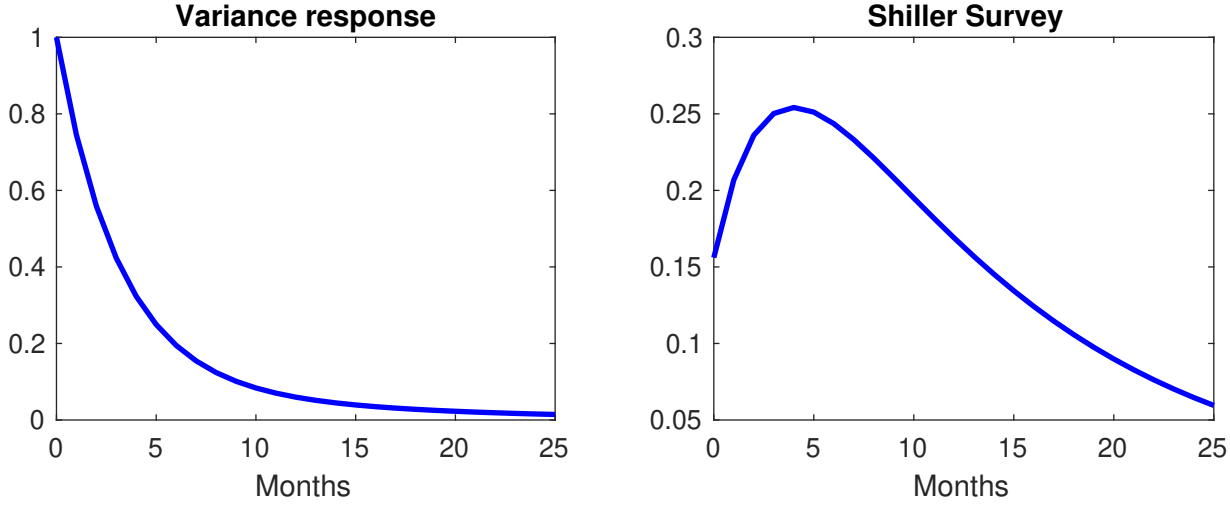
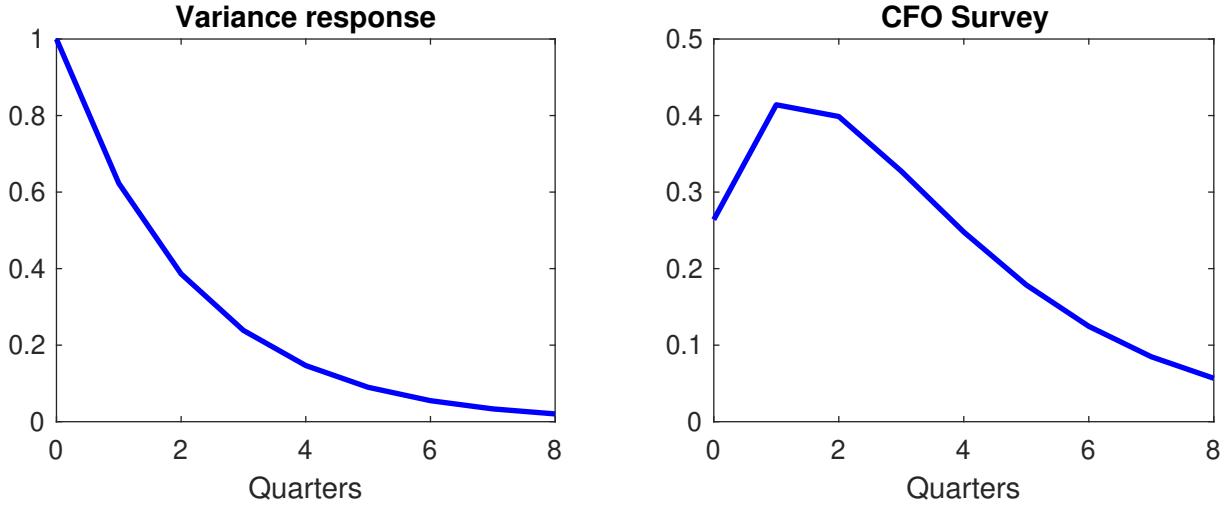


Figure 5: Survey Expectations, VAR. We run a VAR using realized volatility and survey expectations of volatility (in this order) and plot the impulse response to a volatility shock. Expected volatility rises strongly after the shock and then mean reverts fairly quickly. Survey expectations rise slowly, underreact initially and then remain elevated far longer (subsequent overreaction).

Panel A: Shiller Data



Panel B: CFO (Graham Harvey) Data



8. Appendix

Appendix contains additional derivations, tables, and figures.

9. Baseline Model solution

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

9.1 The Variance Process

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

$$\Delta d_t = \mu + \sigma_t \varepsilon_t, \tag{24}$$

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

$$\begin{aligned} \sigma_t^2 &= \bar{v} + \lambda x_{t-1} + \omega \eta_t^S, & (25) \\ x_t &= \phi x_{t-1} + (1-\phi)\left(\sigma_t^2 - \bar{v}\right) \\ &= \left(\phi + (1-\phi)\lambda\right) x_{t-1} + (1-\phi)\,\omega \eta_t^S, & (26) \end{aligned}$$

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:

$$\begin{aligned} 1 &= E_t^S\left[M_{t+1} R_{t+1}\right] \\ &= \beta^\theta E_t^S\left[e^{-\frac{\theta}{\psi}\Delta d_{t+1} + \theta r_{t+1}}\right] \\ &= \beta^\theta E_t^S\left[e^{(1-\gamma)\Delta d_{t+1} + \theta\kappa_0 + \theta\kappa pd_{t+1} - \theta pd_t}\right], & (27) \end{aligned}$$

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:

$$\begin{aligned} 1 &= \beta^\theta E_t^S\left[e^{(1-\gamma)(\mu + \sigma_{t+1}\varepsilon_{t+1}) + \theta\kappa_0 + \theta\kappa(c - Ax_{t+1}) - \theta(c - Ax_t)}\right] \\ &= \beta^\theta E_t^S\left[e^{(1-\gamma)\mu + \frac{1}{2}(1-\gamma)^2\sigma_{t+1}^2 + \theta\kappa_0 + \theta\kappa\left(c - A\left(\phi x_t + (1-\phi)\left(\sigma_{t+1}^2 - \bar{v}\right)\right)\right) - \theta(c - Ax_t)}\right]. & (28) \end{aligned}$$

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:

$$E_t^S \left[e^{(1-\gamma)\mu + \frac{1}{2}(1-\gamma)^2 \sigma_{t+1}^2 + \theta\kappa_0 + \theta\kappa\left(c - A\left(\phi x_t + (1-\phi)\left(\sigma_{t+1}^2 - \bar{v}\right)\right)\right) - \theta(c - Ax_t)} \right] =$$

$$const \times E_t^S \left[e^{(1-\gamma)^2 \frac{1}{2}\lambda x_t - \theta\kappa A(\phi x_t + (1-\phi)\lambda x_t) + \theta A x_t} \right]. \tag{29}$$

And so we have:

$$(1-\gamma)^2 \frac{1}{2}\lambda - \theta\kappa A\left(\phi + (1-\phi)\lambda\right) + \theta A = 0, \tag{30}$$

which gives:

$$A = -\frac{1}{2}\frac{\lambda(1-\gamma)(1-1/\psi)}{1-\kappa\left(\phi + (1-\phi)\lambda\right)}. \tag{31}$$

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

The conditional variance of log returns is then:

$$\begin{aligned} Var_{t-1}^S(r_t) &= Var_{t-1}^S(\kappa p d_t + \Delta d_t) \\ &= \Theta + E_{t-1}^S\left[\sigma_t^2\right], \end{aligned} \tag{32}$$

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

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

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

Then:

$$0 = \theta\ln\beta + (1-\gamma)\mu + \frac{1}{2}(1-\gamma)^2\bar{v} + \theta\kappa_0 - \theta c(1-\kappa) + \frac{1}{2}\left(\frac{1}{2}(1-\gamma)^2 - \theta\kappa A(1-\phi)\right)^2\omega^2. \tag{34}$$

And so

$$c = \frac{\ln\beta + \left(1-\psi^{-1}\right)\mu + \frac{1}{2}\left(1-\psi^{-1}\right)(1-\gamma)\bar{v} + \kappa_0 + \theta^{-1}\frac{1}{2}\left(\frac{1}{2}(1-\gamma)^2 - \theta\kappa A(1-\phi)\right)^2\omega^2}{1-\kappa}. \tag{35}$$

The risk-free rate is given by:

$$\begin{aligned}
e^{-r_{f,t}} &= E_t^S\left[M_{t+1}\right] \\
&= \beta^\theta E_t^S\left[e^{-\gamma\Delta d_{t+1}+(\theta-1)(\kappa_0+\kappa p d_{t+1}-pd_t)}\right] \\
&= \beta^\theta E_t^S\left[e^{-\gamma\Delta d_{t+1}+(\theta-1)(\kappa_0+\kappa c-A\kappa x_{t+1}-c+Ax_t)}\right] \\
&= \beta^\theta E_t^S\left[e^{-\gamma\mu+\frac{1}{2}\gamma^2\left(\bar{v}+\lambda x_t+\omega\eta_{t+1}^S\right)+(\theta-1)\left(\kappa_0+\kappa c-A\kappa\left((\phi+(1-\phi)\lambda)x_t+(1-\phi)\omega\eta_t^S\right)-c+Ax_t\right)}\right] \\
&= \beta^\theta e^{-\gamma\mu+\frac{1}{2}\gamma^2\bar{v}+(\theta-1)(\kappa_0+c(\kappa-1))+\left(\frac{1}{2}\gamma^2\lambda+(\theta-1)A(1-\kappa(\phi+(1-\phi)\lambda))\right)x_t} \\
&\quad\times e^{\frac{1}{2}\left(\frac{1}{2}\gamma^2-(\theta-1)\kappa A(1-\phi)\right)^2\omega^2}. \qquad (36)
\end{aligned}$$

Thus, plugging in for c we have:

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

The conditional expected log return is:

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

Plugging in for c we have:

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

The conditional log risk premium is then:

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

Next, note that:

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

So then

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

We then have:

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

This can be written

$$E_t^S\left[r_{t+1}-r_{f,t}\right]=\left(\gamma-\frac{1}{2}\right)E_t^S\left[\sigma_{t+1}^2\right]+\delta_r. \tag{37}$$

where

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

The objective risk-premium is:

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

We have that:

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

Thus:

$$E_t^P\left[r_{t+1}-r_{f,t}\right]=E_t^S\left[r_{t+1}-r_{f,t}\right]-\kappa A\left(1-\phi\right)\left(E_t^P\left[\sigma_{t+1}^2\right]-E_t^S\left[\sigma_{t+1}^2\right]\right), \tag{38}$$

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

Shocks to realized returns are then:

$$\begin{aligned} r_{t+1}-E_t^P\left[r_{t+1}\right] &= \Delta d_{t+1}-E_t^P\left[\Delta d_{t+1}\right]+\kappa\left(pd_{t+1}-E_t^P\left[pd_{t+1}\right]\right) \\ &= \sigma_{t+1}\varepsilon_{t+1}+\kappa A\left(-x_{t+1}+E_t^P\left[x_{t+1}\right]\right) \\ &= \sigma_{t+1}\varepsilon_{t+1}+\kappa A\left(1-\phi\right)\left(-\left(\sigma_{t+1}^2-\bar{v}\right)+E_t^P\left[\left(\sigma_{t+1}^2-\bar{v}\right)\right]\right) \\ &= \sigma_{t+1}\varepsilon_{t+1}-\kappa A\left(1-\phi\right)\omega\eta_{t+1}. \end{aligned} \tag{39}$$

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

$$\begin{aligned} IV_{t-1} &= E_{t-1}^S\left[\frac{M_t}{E_{t-1}^S\left[M_t\right]}\left(\Theta+\bar{v}+\lambda x_{t-1}+\omega\eta_t^S\right)\right] \\ &= \Theta+\delta_{IV}+E_{t-1}^S\left[\sigma_t^2\right], \end{aligned} \tag{40}$$

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:

$$E_{t-1}^S\left[\frac{M_t}{E_{t-1}^S\left[M_t\right]}\omega\eta_t^S\right]=$$

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

$$E_t^S\left[\frac{e^{\left(\frac{1}{2}\gamma^2-(\theta-1)\kappa A(1-\phi)\right)\omega\eta_{t+1}^S}}{E_t^S\left[e^{\left(\frac{1}{2}\gamma^2-(\theta-1)\kappa A(1-\phi)\right)\omega\eta_{t+1}^S}\right]}\omega\eta_{t+1}^S\right]=$$

$$E_t^S\left[\frac{e^{m\omega\eta_{t+1}^S}}{E_t^S\left[e^{m\omega\eta_{t+1}^S}\right]}\omega\eta_{t+1}^S\right]. \tag{41}$$

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

$$E_t^S\left[\frac{e^{m\omega\eta_{t+1}^S}}{E_t^S\left[e^{m\omega\eta_{t+1}^S}\right]}\omega\eta_{t+1}^S\right]=\omega^2 m. \tag{42}$$

To summarize:

$$IV_{t-1} = \Theta+\delta_{IV}+E_{t-1}^S\left[\sigma_t^2\right], \tag{43}$$

$$\delta_{IV} = \left(\frac{1}{2}\gamma^2-(\theta-1)\kappa A(1-\phi)\right)\omega^2. \tag{44}$$

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

10. 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.

10.1 The Gamma Distribution

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

$$X \sim Gamma(k,s) \tag{45}$$

$$f(x) = \frac{x^{k-1}e^{-\frac{x}{s}}}{s^k\Gamma(k)}, \tag{46}$$

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:

$$\begin{aligned} E[X] &= ks, && (47)\\ Var(X) &= ks^2 && (48)\\ E\left[e^{tX}\right] &= (1-st)^{-k} \text{ for } t < s^{-1}. && (49) \end{aligned}$$

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

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

$$\begin{aligned} E\left[e^{tx}x\right] &= \int_0^\infty e^{tx} x \frac{x^{k-1}e^{-\frac{x}{s}}}{s^k\Gamma(k)}dx \\ &= \int_0^\infty \frac{x^k e^{tx-\frac{x}{s}}}{s^k\Gamma(k)}dx \\ &= \int_0^\infty \frac{x^k e^{x\left(t-\frac{1}{s}\right)}}{s^k\Gamma(k)}dx. && (50) \end{aligned}$$

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:

$$\begin{aligned} E\left[e^{tx}x\right] &= \frac{\tilde{s}^{\tilde{k}}\Gamma\left(\tilde{k}\right)}{s^k\Gamma(k)} \int_0^\infty \frac{x^{\tilde{k}-1}e^{-\frac{x}{\tilde{s}}}}{\tilde{s}^{\tilde{k}}\Gamma\left(\tilde{k}\right)}dx \\ &= \frac{\tilde{s}^{\tilde{k}}\Gamma\left(\tilde{k}\right)}{s^k\Gamma(k)}. && (51) \end{aligned}$$

10.2 The Variance Process

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

$$\Delta d_t = \mu + \sigma_t \varepsilon_t, \tag{52}$$

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

$$\begin{aligned} \sigma_t^2 &= \bar{v} + \lambda x_{t-1} + \omega\tilde{\eta}_t^S, && (53)\\ x_t &= \phi x_{t-1} + (1-\phi)\left(\sigma_t^2 - \bar{v}\right) \\ &= \left(\phi + (1-\phi)\lambda\right)x_{t-1} + (1-\phi)\omega\tilde{\eta}_t^S, && (54) \end{aligned}$$

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

$$\frac{(1-\phi)\,\omega s^{-1}}{1-(\phi+(1-\phi)\,\lambda)} \leq \bar{v} \tag{55}$$

has to be satisfied. In our calibration, we have

$$\begin{aligned} \omega &= 0.23\%, \\ \lambda &= 0.9, \\ \phi &= 0.6 \qquad (56) \\ \bar{v} &= 0.26\%. \qquad (57) \end{aligned}$$

Thus, we need:

$$s \geq \frac{0.4 \times 0.23\%}{0.26\% \times (1-(0.6+0.4\times 0.9))} = 8.85. \tag{58}$$

Note that:

$$\begin{aligned} E_{t-1}^S[\sigma_t \varepsilon_t] &= 0, \qquad (59) \\ Var_{t-1}^S(\sigma_t \varepsilon_t) &= E_{t-1}^S\left[\sigma_t^2\right]. \qquad (60) \end{aligned}$$

10.3 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:

$$\begin{aligned} 1 &= \beta^\theta E_t^S\left[e^{(1-\gamma)(\mu+\sigma_{t+1}\varepsilon_{t+1})+\theta\kappa_0+\theta\kappa(c-Ax_{t+1})-\theta(c-Ax_t)}\right] \\ &= \beta^\theta E_t^S\left[e^{(1-\gamma)\mu+\frac{1}{2}(1-\gamma)^2\sigma_{t+1}^2+\theta\kappa_0+\theta\kappa\left(c-A\left(\phi x_t+(1-\phi)\left(\sigma_{t+1}^2-\bar{v}\right)\right)\right)-\theta(c-Ax_t)}\right]. \end{aligned} \tag{61}$$

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:

$$E_t^S\left[e^{(1-\gamma)\mu+\frac{1}{2}(1-\gamma)^2\sigma_{t+1}^2+\theta\kappa_0+\theta\kappa\left(c-A\left(\phi x_t+(1-\phi)\left(\sigma_{t+1}^2-\bar{v}\right)\right)\right)-\theta(c-Ax_t)}\right] =$$

$$const \times E_t^S \left[e^{(1-\gamma)^2 \frac{1}{2}\lambda x_t - \theta\kappa A(\phi x_t + (1-\phi)\lambda x_t) + \theta A x_t} \right]. \tag{62}$$

And so we have:

$$(1-\gamma)^2 \frac{1}{2}\lambda - \theta\kappa A\left(\phi + (1-\phi)\lambda\right) + \theta A = 0, \tag{63}$$

which gives:

$$A = -\frac{1}{2}\frac{\lambda(1-\gamma)(1-1/\psi)}{1-\kappa\left(\phi + (1-\phi)\lambda\right)}, \tag{64}$$

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:

$$\begin{aligned} Var_{t-1}^S(r_t) &= Var_{t-1}(\kappa pd_t + \Delta d_t) \\ &= \Theta + E_{t-1}^S\left[\sigma_t^2\right], \end{aligned} \tag{65}$$

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

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

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:

$$1 = \beta^\theta e^{(1-\gamma)\mu + \frac{1}{2}(1-\gamma)^2\bar{v} + \theta\kappa_0 - \theta c(1-\kappa) + z_1}, \tag{67}$$

where

$$z_1 = -\frac{1}{s^2}\ln\left(1 - s\left(\frac{1}{2}(1-\gamma)^2 - \theta\kappa A(1-\phi)\right)\omega\right), \tag{68}$$

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

$$\left(\frac{1}{2}(1-\gamma)^2 - \theta\kappa A(1-\phi)\right)\omega < s^{-1}, \tag{69}$$

to have existence of the MGF of the Gamma.

Then:

$$0 = \theta \ln \beta + (1-\gamma)\mu + \frac{1}{2}(1-\gamma)^2 \bar{v} + \theta\kappa_0 - \theta c(1-\kappa) + z_1. \tag{70}$$

And so

$$c = \frac{\theta \ln \beta + (1-\gamma)\mu + \frac{1}{2}(1-\gamma)^2 \bar{v} + \theta\kappa_0 + z_1}{\theta(1-\kappa)}. \tag{71}$$

The risk-free rate is given by:

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

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:

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

Plugging in the expression for c, we have

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

The conditional expected log return is:

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

Plugging in the expression for c, we have:

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

The conditional log risk premium is then:

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

Next, note that:

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

So then

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

and then

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

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

The objective risk-premium is:

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

We have that:

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

Thus:

$$E_{t}^{P}\left[r_{t+1}-r_{f,t}\right]=E_{t}^{S}\left[r_{t+1}-r_{f,t}\right]-\kappa A\left(1-\phi\right)\left(E_{t}^{P}\left[\sigma_{t+1}^{2}\right]-E_{t}^{S}\left[\sigma_{t+1}^{2}\right]\right), \tag{72}$$

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

Shocks to realized returns are then:

$$\begin{aligned}r_{t+1}-E_{t}^{P}\left[r_{t+1}\right] &= \Delta d_{t+1}-E_{t}^{P}\left[\Delta d_{t+1}\right]+\kappa\left(pd_{t+1}-E_{t}^{P}\left[pd_{t+1}\right]\right)\\ &= \sigma_{t+1}\varepsilon_{t+1}+\kappa A\left(-x_{t+1}+E_{t}^{P}\left[x_{t+1}\right]\right)\\ &= \sigma_{t+1}\varepsilon_{t+1}+\kappa A\left(1-\phi\right)\left(-\left(\sigma_{t+1}^{2}-\bar{v}\right)+E_{t}^{P}\left[\left(\sigma_{t+1}^{2}-\bar{v}\right)\right]\right)\\ &= \sigma_{t+1}\varepsilon_{t+1}-\kappa A\left(1-\phi\right)\omega\eta_{t+1}.\end{aligned} \tag{73}$$

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

$$\begin{aligned} IV_{t-1} &= E_{t-1}^S\left[\frac{M_t}{E_{t-1}^S[M_t]}\left(\Theta+\bar{v}+\lambda x_{t-1}+\omega\eta_t^S\right)\right] \\ &= \Theta+\delta+E_{t-1}^S\left[\sigma_t^2\right], \end{aligned} \tag{74}$$

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:

$$E_{t-1}^S\left[\frac{M_t}{E_{t-1}^S[M_t]}\omega\eta_t^S\right] =$$

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

$$E_t^S\left[\frac{e^{\left(\frac{1}{2}\gamma^2-(\theta-1)\kappa A(1-\phi)\right)\omega\eta_{t+1}^S}}{E_t^S\left[e^{\left(\frac{1}{2}\gamma^2-(\theta-1)\kappa A(1-\phi)\right)\omega\eta_{t+1}^S}\right]}\omega\eta_{t+1}^S\right] =$$

$$E_t^S\left[\frac{e^{m\omega\eta_{t+1}^S}}{E_t^S\left[e^{m\omega\eta_{t+1}^S}\right]}\omega\eta_{t+1}^S\right]. \tag{75}$$

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 $E\left[e^{tx}x\right]$ is a constant if x is Gamma and $t < s^{-1}$ (see Equation (51)):

$$E_t^S\left[\frac{e^{m\omega\eta_{t+1}^S}}{E_t^S\left[e^{m\omega\eta_{t+1}^S}\right]}\omega\eta_{t+1}^S\right] = -\omega s^{-1}+E_t^S\left[\frac{e^{m\omega\tilde{\eta}_{t+1}^S}}{E_t^S\left[e^{m\omega\tilde{\eta}_{t+1}^S}\right]}\omega\tilde{\eta}_t^S\right], \tag{76}$$

where

$$E_t^S\left[\frac{e^{m\omega\tilde{\eta}_{t+1}^S}}{E_t^S\left[e^{m\omega\tilde{\eta}_{t+1}^S}\right]}\omega\tilde{\eta}_t^S\right] = \tilde{s}\left(\frac{\tilde{s}}{\omega s}\right)^{s^{-2}}\frac{\Gamma\left(s^{-2}+1\right)}{\Gamma\left(s^{-2}\right)}, \tag{77}$$

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

$$IV_{t-1} = \Theta+\delta+E_{t-1}^S\left[\sigma_t^2\right], \tag{78}$$

$$\delta = \tilde{s}\left(\frac{\tilde{s}}{\omega s}\right)^{s^{-2}}\frac{\Gamma\left(s^{-2}+1\right)}{\Gamma\left(s^{-2}\right)}-\omega s^{-1}, \tag{79}$$

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

10.4 Data

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

Table 6: Data Sources.

Country	Index	Volatility	Source	History
USA	SP500	VIX	WRDS	From 1/2/1990
France	CAC 40	VCAC	Bloomberg	Fron 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

10.5 Appendix Tables and Figures

Table 7: Stylized Facts: Robustness. We repeat the analysis from Table 1 in the main text but split the sample into pre and post 2010. This shows robustness to the results post financial crisis and roughly splits the sample for the variance returns (variance swap and VIX futures).

Panel A: Pre 2010 Sample

	Excess Stock Returns				Variance Risk Premium		Variance	
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
	$r^e_{M,t+1}$	$r^e_{M,t+1}$	$r^e_{M,t+1}$	$r^e_{M,t+1}$	$r_{var,t+1}$	$r_{VIX,t+1}$	σ^2_{t+1}	VIX^2_t
σ^2_t	-1.78			-1.22	-0.59	-0.33	0.74	0.50
	(0.52)			(0.57)	(0.22)	(0.07)	(0.08)	(0.04)
$\overline{\sigma^2}_{t-1,t-6}$	1.42				0.57	0.12	-0.00	0.30
	(0.72)				(0.25)	(0.08)	(0.06)	(0.04)
$VIX^2_t - \sigma^2_t$		5.08						
		(1.43)						
VIX^2_t			-0.35					
			(1.19)					
N	233	239	239	239	169	71	233	233
Adj. R^2	2.6%	5.2%	0.0%	1.6%	0.5%	8.8%	54.4%	83.7%

Panel B: Post 2010

	Excess Stock Returns				Variance Risk Premium		Variance	
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
	$r^e_{M,t+1}$	$r^e_{M,t+1}$	$r^e_{M,t+1}$	$r^e_{M,t+1}$	$r_{var,t+1}$	$r_{VIX,t+1}$	σ^2_{t+1}	VIX^2_t
σ^2_t	-3.19			-2.60	-1.29	-0.67	0.40	0.51
	(0.74)			(0.66)	(0.79)	(0.25)	(0.03)	(0.06)
$\overline{\sigma^2}_{t-1,t-6}$	5.69				2.97	1.38	0.03	0.44
	(1.98)				(1.74)	(0.33)	(0.07)	(0.12)
$VIX^2_t - \sigma^2_t$		8.52						
		(0.84)						
VIX^2_t			2.26					
			(1.06)					
N	95	101	101	101	93	93	96	96
Adj. R^2	7.6%	17.0%	0.9%	2.3%	1.1%	5.9%	14.4%	58.9%

Table 8: Stylized Facts: Robustness to using volatility. We repeat our analysis using volatility (standard deviation) in place of variance. We run predictive regressions of future excess stock returns (market returns over the risk free rate), future variance risk premiums (variance swap returns r_{var}, and VIX futures returns r_{VIX}), and future realized variance on various measures of past volatility, average of past volatility over 6 months ($\overline{\sigma^2}_{t-1,t-6}$), and implied volatility from the VIX. In our notation σ_t represents the realized standard deviation of daily market returns in month t. The returns on variance swaps 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 1990-2018, the variance swap and VIX futures data are 1996-2017 and 2004-2017, respectively. Standard errors in parentheses use Newey West correction with 12 lags.

Panel A: Volatility

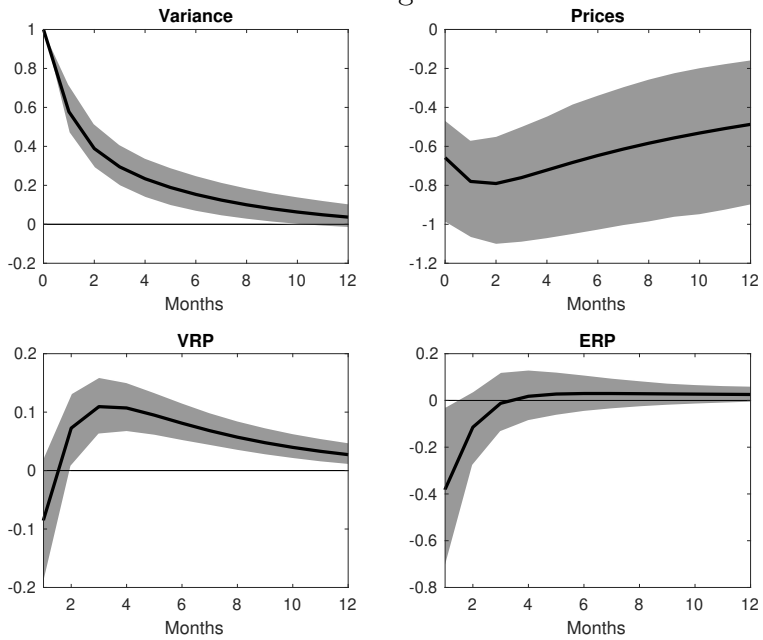
	Excess Stock Returns				Variance Risk Premium		Variance	
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
	$r^e_{M,t+1}$	$r^e_{M,t+1}$	$r^e_{M,t+1}$	$r^e_{M,t+1}$	$r_{var,t+1}$	$r_{VIX,t+1}$	σ_{t+1}	VIX_t
σ_t	-0.35			-0.20	-0.16	-0.09	0.68	0.56
	(0.16)			(0.13)	(0.07)	(0.03)	(0.09)	(0.04)
$\overline{\sigma}_{t-1,t-6}$	0.27				0.07	0.06	0.12	0.30
	(0.14)				(0.07)	(0.02)	(0.07)	(0.04)
$VIX_t - \sigma_t$		0.81						
		(0.21)						
VIX_t			-0.02					
			(0.16)					
N	335	341	341	341	264	166	335	335
Adj. R^2	2.3%	5.5%	0.0%	1.2%	0.8%	5.5%	55.6%	67.7%

Table 9: Change in volatility, equity risk premium, and variance risk premium. We run three forecasting 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, $r_{i,t+1} - r^f_{i,t}$ 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)	(2)	(3)	(4)	(5)	(6)
	ERP	Vol	VRP	ERP	Vol	VRP
$\Delta_6\sigma_t$	-0.15*	0.27***	-0.06***	-0.25***	0.32***	-0.09***
	(0.09)	(0.08)	(0.02)	(0.08)	(0.05)	(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)	(2)	(3)	(4)	(5)	(6)
	ERP	Vol	VRP	ERP	Vol	VRP
$\Delta_6\sigma_t^2$	-0.88***	0.23**	0.04	-1.66***	0.36***	-0.13***
	(0.34)	(0.10)	(0.10)	(0.39)	(0.04)	(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 6: Impulse Response to Variance Shock: Robustness. VAR of realized variance, market excess returns (denoted ERP for equity risk premium), the variance risk premium (VRP), and the log price dividend ratio (labeled Prices). VRP is implied variance (VIX^2) minus next period realized variance. Responses are for a one-standard deviation shock to realized variance at time 0. Vertical dashed lines at time 1 highlight the predicted, rather than realized, ERP and VRP. X-axis is in months. Price and equity returns are in percent (monthly), while RV and the VRP are in terms of standard deviations of variance shocks. Shaded regions indicate 95% confidence intervals constructed using bootstrap. Panel A weights the observations by the inverse of lagged realized volatility (weighted least squares). Panel B uses logs of RV and the VIX in place of levels. See text for more detail.

Panel A: Weighted VAR



Panel B: Logs of Realized Variance and VIX

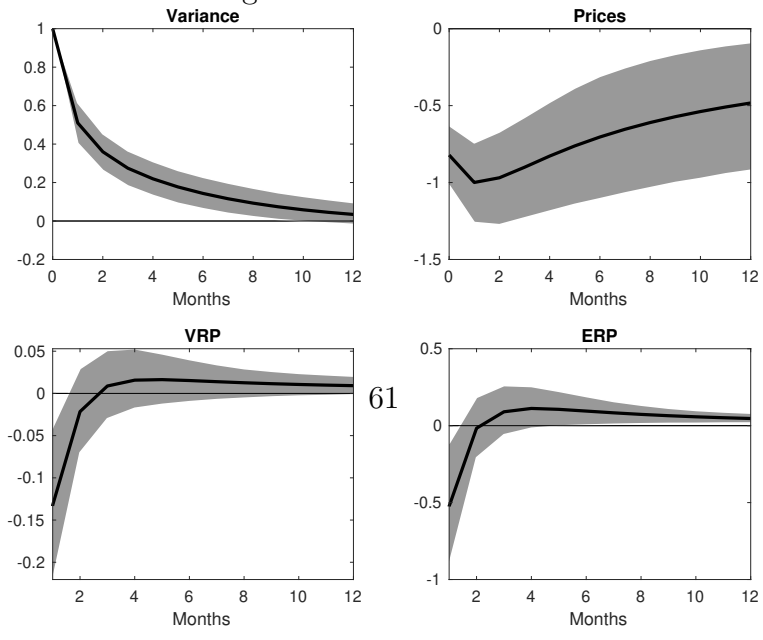


Figure 7: Stylized facts for US data. We run regressions of returns, variance risk premiums, and realized variance on lags of realized variance and plot coefficients by horizon. Variance risk premiums are measured either as squared VIX minus realized variance, using the negative for variance swap returns (e.g., selling variance), or using the negative of VIX futures (shorting the VIX). We also plot stock returns on the lagged variance risk premium (squared VIX minus realized variance). The x-axis is in months.

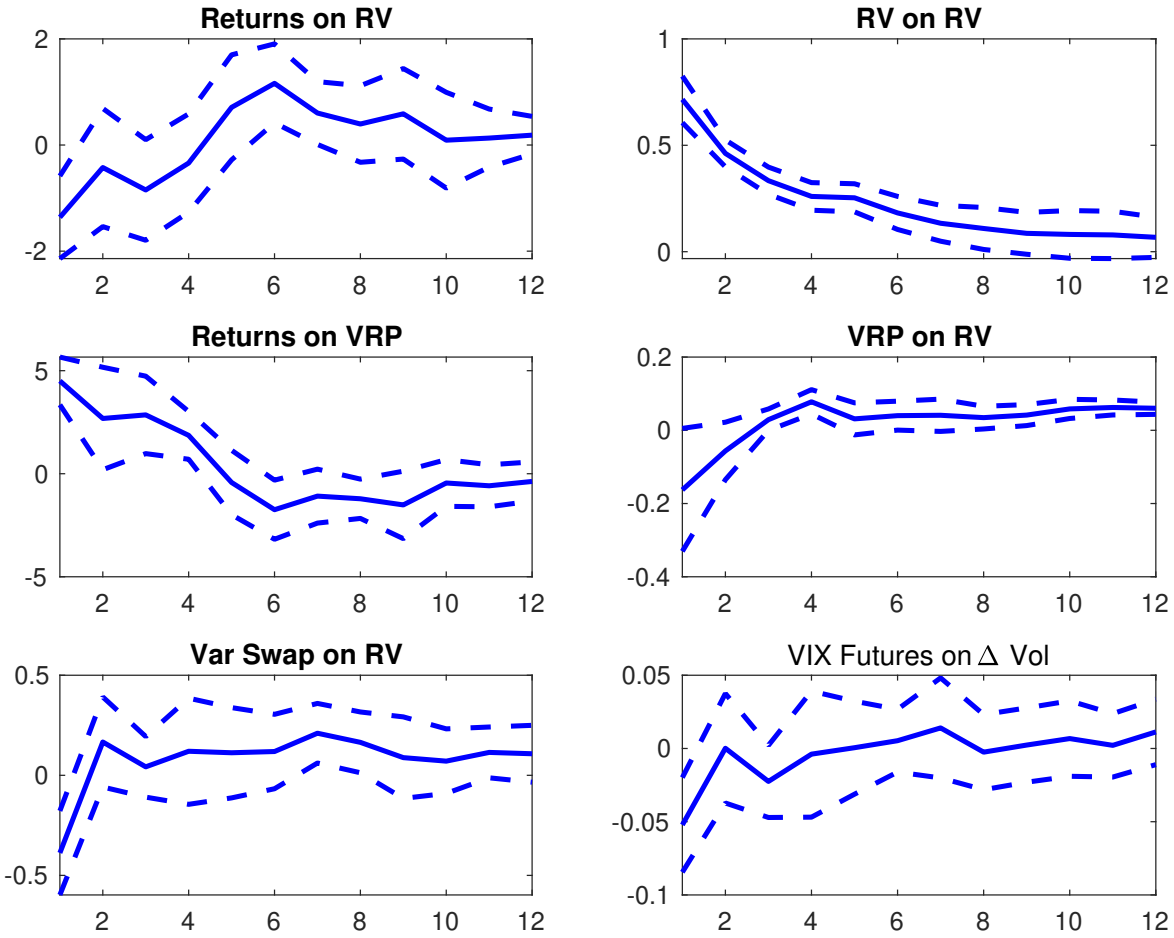


Figure 8: Stylized facts for US data: Post 2010. We replicate our main stylized facts using only US data from 2010 onwards, thus excluding the financial crisis. We run regressions of returns, variance risk premiums, and realized variance on lags of realized variance and plot coefficients by horizon. Variance risk premiums are measured either as squared VIX minus realized variance, using the negative for variance swap returns (e.g., selling variance), or using the negative of VIX futures (shorting the VIX). We also plot stock returns on the lagged variance risk premium (squared VIX minus realized variance). The x-axis is in months.

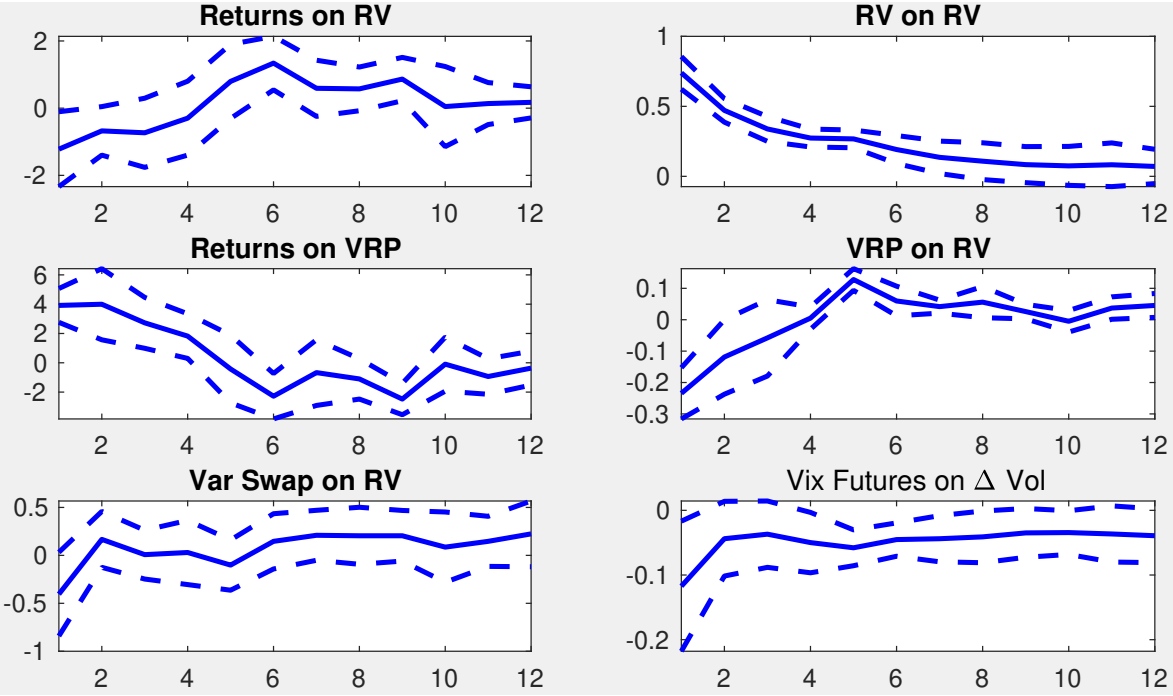


Figure 9: Longer Sample of US Data. We plot regression coefficients of returns on lags of realized variance and of realized variance on lags of realized variance. Our top sample includes all US data from 1926 while the bottom sample includes only post War US data (since 1950). Note the negative coefficient of returns on realized variance is weaker using the longer sample of returns, though even in this sample there is no strong evidence of a positive risk-return tradeoff.

