NBER WORKING PAPER SERIES

THE VIX, THE VARIANCE PREMIUM AND STOCK MARKET VOLATILITY

Geert Bekaert
Marie Hoerova

Working Paper 18995
http://www.nber.org/papers/w18995

NATIONAL BUREAU OF ECONOMIC RESEARCH
1050 Massachusetts Avenue
Cambridge, MA 02138
April 2013

The views expressed herein are those of the authors and do not necessarily reflect the views of the National Bureau of Economic Research. Geert Bekaert acknowledges financial support from the NetSpar foundation.

The VIX, the Variance Premium and Stock Market Volatility
Geert Bekaert and Marie Hoerova
NBER Working Paper No. 18995
April 2013
JEL No. C22,C52,E32,G12

ABSTRACT

We decompose the squared VIX index, derived from US S&P500 options prices, into the conditional variance of stock returns and the equity variance premium. The latter is increasing in risk aversion in a wide variety of economic settings. We tackle several measurement issues assessing a plethora of state-of-the-art volatility forecasting models. We then examine the predictive power of the VIX and its two components for stock market returns and economic activity. The variance premium predicts stock returns but the conditional stock market variance predicts economic activity, and is more contemporaneously correlated with financial instability than is the variance premium.

Geert Bekaert
Graduate School of Business
Columbia University
3022 Broadway, 411 Uris Hall
New York, NY 10027
and NBER
gb241@columbia.edu

Marie Hoerova
European Central Bank
Kaiserstrasse 29 D-60311
Frankfurt am Main, Germany
marie.hoerova@ecb.int

1. Introduction

The 2007-2009 crisis has intensified the need for indicators of the risk aversion of market participants. It has also become increasingly commonplace to assume that changes in risk appetites are an important determinant of asset prices. Not surprisingly, the behavioral finance literature (see e.g. Baker and Wurgler, 2007) has developed "sentiment indices," and financial institutions have created a wide variety of "risk aversion indicators" (see Coudert and Gex (2008) for a survey).

One simple candidate indicator is the equity variance premium, the difference between the squared VIX index and an estimate of the conditional variance of the stock market. The VIX index is the "risk-neutral" expected stock market variance for the US S&P500 contract and is computed from a panel of options prices. Well-known as a "fear index" (Whaley, 2000) for asset markets, it reflects both stock market uncertainty (the "physical" expected volatility), and a variance risk premium, which is also the expected premium from selling stock market variance in a swap contract. Bollerslev, Tauchen and Zhou (2009) show that an estimate of this variance premium predicts stock returns; Bekaert, Hoerova, Lo Duca (2012) show that there are strong interactions between monetary policy and the variance premium, suggesting that monetary policy may actually affect risk aversion in the market place. The variance premium uses objective financial market information and naturally "cleanses" option-implied volatility from the effect of physical volatility dynamics and uncertainty. We show that in a variety of economic settings, the variance premium is increasing in risk aversion and can therefore serve as an easy-to-compute risk aversion measure.

How to measure the variance premium is not without controversy however, because it relies on an estimate of the conditional variance of stock returns. For example, the measure proposed in Bollerslev, Tauchen and Zhou (2009), BTZ, henceforth, assumes that the conditional variance of stock market returns is a martingale, which is strongly at odds with the data, leading to biased variance premiums.

In this paper, we tackle several measurement issues for the variance premium, assessing a plethora of state-of-the art volatility models and making full use of overlapping daily data, rather than sparse end-of-month data, which is standard.

The conditional variance measure is of interest in its own right. First, there is a long literature on the trade-off between risk, as measured by the conditional variance of stock market returns, and the aggregate risk premium on the market (see e.g. French, Schwert and Stambaugh (1987) for a seminal contribution). This long line of research has mostly failed to uncover a strong positive relationship between risk and return (see Bali, 2008, for a summary). Second, stock market volatility can also be viewed as a market-based measure of economic uncertainty. For example, Bloom (2009) shows that heightened "economic uncertainty" decreases employment and output. Interestingly, he uses the VIX index to measure uncertainty, so that his results may actually be driven by the variance premium rather than uncertainty per se.

Using more plausible estimates of the variance premium and stock market volatility, we then assess whether they predict stock returns and economic activity. We find that the well-known results in BTZ exaggerate the predictive power of the variance premium for stock returns. However, the equity variance risk premium remains a reliable predictor of stock returns, with its predictive power even stronger when the recent crisis period is added to the sample. Stock market volatility does not predict the stock market, but it is a much better predictor of economic activity than is the equity variance premium.

The remainder of the paper is organized as follows. Section 2 provides some theoretical background regarding the interpretation of the variance premium as an indicator of risk aversion. Section 3 discusses the econometric framework that we use to forecast volatility and compare different estimates of the conditional variance of stock returns. Section 4 reports the results of our specification analysis and forecasting horse race. Section 5 uses the preferred estimates of the variance premium and stock market volatility to predict stock returns and economic activity. Section 6 concludes.

2. The VIX, Uncertainty and Risk Aversion

To obtain intuition on how the VIX is related to the actual ("physical") expected variance of stock returns and to risk preferences, we analyze a one-period discrete state economy. Imagine a stock return distribution with three different states x_i, as follows:

Good state: $x_g = \mu + a$ with probability $(1-p)/2$,

Bad state : $x_b = \mu - a$ with probability $(1-p)/2$,

Crash state: $x_c = c < 0$ with probability p,

where $\mu > 0$, $a > 0$ and $p > 0$ are parameters to be determined. We set them to match moments of US stock returns - the mean ($\overline{X}$), the variance (V) and the skewness (Sk) - while fixing the crash return at an empirically plausible number**.** These moments are readily computed from the state values and probabilities.

Consider an investor with power utility over wealth in a one-period world, so that in equilibrium she invests her entire wealth in the stock market:

$$U\left(\tilde{W}\right) = E\left[\frac{\left(W_0 \tilde{R}\right)^{1-\gamma}}{1-\gamma}\right], \tag{1}$$

where $\tilde{R}$ is the gross return on the stock market, W_0 is initial wealth and γ is the coefficient of relative risk aversion. The "pricing kernel" in this economy is given by marginal utility, denoted by m, and is proportional to $\tilde{R}^{-\gamma}$. Hence, the stochastic part of the pricing kernel moves inversely with the return on the stock market.

The physical variance of the stock market is exogenous in this economy, and is simply given by *V*. This variance is computed using the actual probabilities. The (squared) *VIX* represents the "risk-neutral" conditional variance. It is computed using the so-called "risk-neutral probabilities," which are simply probabilities adjusted for risk. In particular, for a general state probability π_i for state i, the risk-neutral probability is:

$$\pi_i^{RN} = \pi_i \frac{m_i}{E[m]} = \pi_i \frac{R_i^{-\gamma}}{E[m]}. \tag{2}$$

So, for a given γ, we can easily compute the risk-neutral probabilities since $R_i = x_i + 1$. For an economy with K states, the risk-neutral variance is then given by:

$$VIX^2 = \sum_{i=1}^{K} \pi_i^{RN} \left(x_i - \overline{X}^{RN}\right)^2 \tag{3}$$

where $\overline{X}^{RN} = \sum_{i=1}^{K} \pi_i^{RN} x_i$ is the risk-neutral mean. The variance premium is:

$$VP = VIX^2 - V = \sum_{i=1}^{K} \pi_i^{RN} \left(x_i - \overline{X}^{RN}\right)^2 - \sum_{i=1}^{K} \pi_i \left(x_i - \overline{X}\right)^2 \tag{4}$$

where $\pi_g^{RN} = \frac{1-p}{2} \frac{(\mu + a + 1)^{-\gamma}}{E[m]}$, $\pi_b^{RN} = \frac{1-p}{2} \frac{(\mu - a + 1)^{-\gamma}}{E[m]}$ and $\pi_c^{RN} = p \frac{(c+1)^{-\gamma}}{E[m]}$.

In our economy, the risk-neutral probability puts more weight on the crash state. The higher is risk aversion, the more weight the crash state gets. As the crash state induces plenty of additional variance, the variance premium is positive.

Numerical Examples

Suppose the statistics to match are as follows: $\overline{X} = 10\%$, $\sigma = 15\%$, both on an annualized basis; and $Sk = -1$ on a monthly basis. These numbers roughly match statistics for the aggregate U.S. stock market. We set $c = -25\%$ (a monthly number). This crash return is in line with the stock market collapses in October 1987 and October 2008. The implied crash probability to match the skewness coefficient of -1 is given by $p = 0.5\%$. With a monthly investment horizon, the crash probability implies a crash every 200 months, or roughly once every 15 years. Panel A of Table 1 provides, for different values of the coefficient of relative risk aversion γ, the values for the *VIX* on an annualized basis in percent (*VIX*) and the annualized variance premium (VP). Note that the variance premium is increasing in the coefficient of relative risk aversion γ.

In structural models, γ is typically assumed to be time-invariant, and the time variation in the variance premium is generated through different mechanisms. For example, in Drechsler and Yaron (2011), who formulate a consumption-based asset pricing model with recursive preferences, the variance premium is directly linked to the probability of a "negative jump" to expected consumption growth. The analogous mechanism in our simple economy would be to decrease the skewness of the return distribution by increasing the crash probability p. This obviously represents "risk" instead of "risk aversion". Yet, it is the interaction of risk aversion and skewness that gives rise to large readings in our risk aversion proxy. To illustrate, let us consider an example with lower skewness. Setting skewness equal to -2 requires a higher crash probability of $p = 1\%$. Panel B of Table 1 shows that the *VIX* increases, and increases more the higher the coefficient of relative risk aversion, both in absolute and in relative terms. The variance premium roughly doubles for all γ levels.

In Bekaert and Engstrom (2010), when a recession becomes more likely, the representative agent also becomes more risk averse through a Campbell-Cochrane (1999)-like external habit formulation. The recession fear then induces high levels of the *VIX*. We can informally illustrate such a mechanism in our one-period model. Imagine that the utility function is over wealth relative to an exogenous benchmark wealth level W_{bm}. Normalizing the initial wealth W_0 to 1, the pricing kernel is now given by $\left(\tilde{R} - W_{bm}\right)^{-\gamma}$, and the coefficient of relative risk aversion is $\gamma \tilde{R} / \left(\tilde{R} - W_{bm}\right)$. Consequently, risk aversion is state dependent and increases as $\tilde{R}$ decreases towards the benchmark level. It is easy to see how a dynamic version of this economy, for instance with a slow-moving W_{bm}, could generate risk aversion that is changing over time as return realizations change the distance between actual wealth and the benchmark wealth level.

To illustrate this mechanism, Panel C considers three different benchmark levels for W_{bm} (0.05, 0.25 and 0.5) with γ fixed at 4 and $Sk = -1$, implying $p = 0.5\,\%$. The second column shows expected relative risk aversion in the economy (RRA), weighting the three possible realizations for risk aversion with the actual state probabilities. The other columns are as in the panels above. Clearly, for $W_{bm} = 0$, RRA = 4 and we replicate the values in Panel A for $\gamma = 4$. Keeping γ fixed and increasing W_{bm}, effective risk aversion increases. For example, RRA increases from 5.323 to 7.968 as W_{bm} increases from 0.25 to 0.5. The *VIX* increases from 19.059 to 26.010 and the variance premium more than triples from 0.014 to 0.045.

3. Econometric Framework

Introduction

We define the variance risk premium as:

$$VP_t = VIX_t^2 - E_t\left[RV_{t+1}^{(22)}\right] \tag{5}$$

Here the *VIX* is the implied option volatility for contracts with a maturity of one month, and $RV_{t+1}^{(22)}$ is the realized variance measured over the next month (22 trading days) using 5 minute returns. Note that $RV_{t+1}^{(22)} - VIX_t^2$ is the return to buying variance in a variance swap contract. Therefore, technically speaking, the variance risk premium refers to the negative of VP. Since that number is always negative, we prefer to define it as we did in Equation (5). The unconditional mean of the variance premium is easy to compute by simply computing the average of $VIX_t^2 - RV_{t+1}^{(22)}$. However, we are interested in the conditional variance premium as described in Equation (5), which relies on the physical conditional expected value of the future realized variance. The common approach to estimate this uses empirical projections of the realized variance on variables in the information set, and subtracts this estimated expected variance from the VIX^2 to arrive at VP. Hence, the problem is reduced to one of variance forecasting.

Variance forecasting

There is an extensive econometric literature on volatility forecasting.[1] It is now generally accepted that models based on high frequency realized variances dominate standard models in the GARCH class (see e.g. Chen and Ghysels, 2012) and we therefore examine the state-of-the-art models in that class. These models stress the importance of

[1] Fernandes, Medeiros and Scharth (2009) forecast the *VIX* index instead.

persistence (using lagged realized variances as predictors), additional information content in the most recent return variances (Corsi, 2009), asymmetry between positive and negative return shocks (the classic volatility asymmetry, see e.g. Engle and Ng, 1993) and potentially differing predictive information present in jump versus continuous volatility components (Andersen, Bollerslev, and Diebold, 2007). We accommodate all of these elements in our model.

In the finance literature, it has been pointed out as early as in Christensen and Prabhala (1998) that option prices as reflected in implied volatility should have information about future stock market volatility. This motivates using the *VIX* as a predictive variable, a variable curiously absent in the econometrics literature.[2] Of course, because the *VIX* also embeds a risk premium, it will not be an unbiased predictor of future realized volatility. Chernov (2007) argues that spot volatility is likely to have additional information about future volatility. Finally, it is well-known that estimation noise hurts out-of-sample forecasting performance. Simple models such as the martingale model may therefore outperform more complex models. We therefore also consider a number of non-estimated models that are special cases of our general framework.

Our most general forecasting model can be represented as follows:

$$\begin{aligned} RV_t^{(22)} &= c + \alpha\, VIX_{t-22}^2 + \beta^m RV_{t-22}^{(22)} + \beta^w RV_{t-22}^{(5)} + \beta^d RV_{t-22}^{(1)} \\ &\quad + \gamma^m J_{t-22}^{(22)} + \gamma^w J_{t-22}^{(5)} + \gamma^d J_{t-22}^{(1)} + \delta^m r_{t-22}^{(22)-} + \delta^w r_{t-22}^{(5)-} + \delta^d r_{t-22}^{(1)-} + \varepsilon_t \end{aligned} \tag{6}$$

As explained below, we want to forecast the monthly (22 trading days) realized variance, denoted by $RV_t^{(22)}$. Our first independent variable is the VIX^2, and we expect α to be positive. The next three variables (realized variances at the monthly, weekly and daily

[2] Exceptions are Busch, Christensen and Nielsen (2011) who examine a number of variance forecasting models embedding option-implied volatility for bond, currency and stock markets and Andersen and Bondarenko (2007), who mostly focus on measurement issues with the officially published *VIX* index.

frequencies) reflect the HAR (Heterogeneous Autoregressive) specification of Corsi (2009), incorporating volatility persistence and the idea that more recent variances may be especially informative.

The third set of variables comprises jumps, denoted as J_t, at the monthly, weekly and daily frequencies. To isolate the jumps contribution to daily quadratic variation, we use bipower variation (Barndorff-Nielsen and Shephard, 2004). Weekly (h=5) and monthly (h=22) jumps are defined as follows: $J_t^{(h)} = \sum_{j=1}^{h} J_{t-j+1}$. The separation of the quadratic variation in a continuous (captured by the *RV*-variables) and a discontinuous ("jump") component follows Andersen, Bollerslev and Diebold (2007).

Finally, following Corsi and Renò (2012), we add negative returns over the past day, week and month, to incorporate a potential leverage effect (see Campbell and Hentschel, 1991; Bekaert and Wu, 2000). To model the leverage effect at different frequencies, they define $r_t^{(h)-} = \min\left(r_t^{(h)}, 0\right)$ where $r_t^{(h)} = \frac{1}{h}\sum_{j=1}^{h} r_{t-j+1}$.

Initial specification analysis and models

Our data start on January 02, 1990 (the start of the model-free *VIX* series)[3] and covers the period until October 01, 2010. The recent crisis period presents special challenges as stock market volatilities peaked at unprecedented levels, but at the same time the crisis represents an informative period during which risk aversion may have been particularly pronounced. For that reason, our main analysis retains the crisis period but also considers a winsorized sample. In addition, we consider models that predict the logarithm of

[3] The CBOE changed the methodology for calculating the *VIX*, initially measuring implied volatility for the S&P100 index, to be measured in a model–free manner from a panel of option prices (see Bakshi, Madan and Kapadia, 2003, for details) only in September 2003. It then backdated the new model–free index to 1990 using historical option prices.

realized variances, and we put much emphasis on parameter stability in our model selection procedure.

Our S&P500 realized variance measurements are computed using five-minute intraday returns plus the close-to-open return. As is standard in the literature (see, e.g., Andersen et al., 2007 or Corsi and Renò, 2012), we exclude holidays and other inactive trading days in our regressions (specifically, we exclude days with less than 64 return-observations per day). We are left with a total of 5177 daily, overlapping observations.

As a first step in our model selection, we use the full sample and estimate the model in Equation (6). To avoid poorly estimated coefficients, we "pare down" the regression using a model selection procedure akin to the one proposed by Campos, Hendry, and Krolzig (2003), among many others. Step 1 conducts a joint estimation using all (remaining) independent variables. Step 2 collects all variables that have coefficients insignificantly different from zero at the 10% level (using a two-sided test). If this set is empty, the final specification is reached. Step 3 performs an F-test in that specification on the non-significant variables. If the test does not reject at the 10% level, all these variables are dropped from the specification and we go back to step 1. If the test rejects, we only drop the variable with the lowest t-statistic (in absolute value) and go to step 1.

We conduct this model selection procedure, which we refer to as the "Hendry-approach," on the original data and three alternative, modified data sets. The first alternative data set uses winsorized data to mitigate the impact of the extreme observations encountered during the crisis.[4] While the crisis observations constitute less than 4% of the sample, their relative contribution to the variance of $RV^{(22)}$ is 81%! We

[4] We define the crisis as the period from September 14, 2008 (collapse of Lehman Brothers) to June 2009 (end of the NBER recession).

select the level of winsorization to bring this relative contribution to below 50%. This leads us to winsorizing the top 1.5% of the $RV^{(22)}$ data. We apply the same winsorization cutoff to all of our variables (winsorizing the bottom 1.5% in case of the negative return variable). The third and fourth data sets use logarithms of all variables in the winsorized and non-winsorized samples except negative returns (for the jump variables, we take the logarithm of $1+J_t^{(h)}$). Because variances have right-skewed distributions, but logarithmic variances tend to have near Gaussian distributions, it may be easier to predict logarithmic variances with linear models. However, ultimately, we still need to identify the model that best forecasts the level of the realized variance. To this end, when we consider a logarithmic model, we assume log-normality to predict levels of monthly realized variances:

$$E_t\left[RV_{t+1}^{(22)}\right] = \exp\left(E_t\left[rv_{t+1}^{(22)}\right] + \frac{1}{2}\mathrm{var}\left[rv_{t+1}^{(22)}\right]\right) \tag{7}$$

where $rv_{t+1}^{(22)} = \ln\left(RV_{t+1}^{(22)}\right)$. We use the logarithmic model to compute the conditional expectation of $rv_{t+1}^{(22)}$ and the sample variance of $rv_{t+1}^{(22)}$ to compute the variance term.

The model specification analysis delivers the model with the best in sample fit over the full sample. This may not be the best model as simpler models may give more robust, stable and precise forecasts going forward. Therefore our second step in model selection consists of an out-of-sample forecasting horse race between various special cases of the encompassing model in Equation (6), including some non-estimated models.

Table 2 summarizes the models we consider. We estimate the models by projecting the current realized variance onto the past values of the explanatory variables using OLS. The first model is to simply use the VIX^2 as a predictor. The other benchmark models are the purely econometric models, either using the realized monthly variance only (model 2) or the realized variance at three frequencies (monthly, weekly and daily) in model 9. We can then add jumps (models 5 and 11), or jumps and negative returns (models 7 and 13). Then we also consider all these models with the VIX^2 as an additional predictor, yielding a total of 13 models (models 3, 4, 6, 8, 10 and 12). Model 14 is the model selected by the Hendry analysis, which we discuss below. The three non-estimated models are: the lagged squared *VIX*; the lagged realized variance (this is the model used in BTZ); and 0.5 times the lagged squared *VIX* plus 0.5 times the lagged realized variance. Finally, we follow the exact same procedure for the logarithmic model, yielding another 14 models. Thus, we consider 31 models in total.

Forecast accuracy and stability analysis

To select among the various models, we verify their out-of-sample forecasting performance and examine their stability. We estimate the models using data between January 1, 1990 and July 15, 2005 (representing 75% of the full sample) and use the rest of the sample (till October 01, 2010) to measure forecasting performance. The parameters are not updated. We examine five different criteria.

We compute the mean-squared error (MSE), mean absolute error (MAE), and mean absolute percentage error (MAPE; the absolute error in percent of the actual realized variance). We evaluate whether the forecast error measures are significantly different among competing forecasting models through the Diebold and Mariano (1995) test (with standard errors computed using 44 Newey-West lags), using a 10% significance level. We also compute the R^2 of Mincer–Zarnowitz (1969) forecasting regressions, that is, we compute the R^2 in a regression of actual data on their forecasted values.[5] The final criterion we examine is a simple joint Chow test for parameter stability over the last part of the sample versus the estimation part of the sample. We also produce the average correlation of the forecasts produced by a particular model with the forecasts produced by the winning model on each of the first 4 criteria. This gives a sense of how close different forecast models are economically.

4. Model Selection Results

Full sample model selection

In Table 3, we discuss the models chosen by the Hendry analysis over the four different data sets. The standard errors below the parameters are computed using 44 Newey-West (1987) lags. The model selection always yields models with weekly and daily realized variances. The monthly realized variance is highly significant in three out of four cases, but it would be dropped in the case of the level regression using the original data. We nevertheless still keep the monthly variance as part of the winning model since the monthly jumps and negative returns, meant to complement its forecast, are significant at the 5% level, and the three variables are jointly significant at the 1% level.

The *VIX* is chosen whenever the monthly realized variance would be, i.e., in all cases except for the non-winsorized levels data, where the Hendry analysis, apart from the *VIX*, also eliminates the weekly jumps and negative returns. In the non-winsorized data with the log regression, the *VIX* enters, but jumps at all frequencies and monthly negative returns are eliminated. With winsorized data, both level and log regressions select the

[5] We also computed two other statistics; the heteroskedasticity adjusted root mean square error suggested in Bollerslev and Ghysels (1996) and the QLIKE loss function (see Patton, 2011). However, these statistics produce rankings very similar to the MAPE-criterion, so we do not discuss them further.

same model. The winning model includes daily jumps and negative returns, in addition to the squared *VIX* and Corsi's realized variances at three frequencies.

In comparison with the other models, the Hendry-chosen models rank among the top three in terms of the Akaike and Bayesian Information Criteria (AIC and BIC henceforth) in each of the four data samples we considered (winsorized/not winsorized; logged/not-logged).

Forecasting horse race

Table 4 produces the ranks of our 31 considered models according to the five criteria discussed above for the non-winsorized data. Results for the winsorized sample are quite similar and are relegated to an online Appendix. Recall that for the logarithmic models, we are predicting the level of the realized variance as discussed in Section 3. For the first three criteria, we test for each model whether it generates a statistic significantly different from the statistic generated by the best ranked model. When such a test fails to reject, the rank is bolded. We view such tests as critical in model selection. A model may rank relatively low, but the criterion may have little power to distinguish different models and generate very similar forecast errors. For example, a quick glance at the table reveals that the MSE criterion has little power to distinguish alternative models. For the R^2 criterion, we view a difference of more than 5% with the winning model as a significant difference in economic terms. Models similar in R^2 to the winning model are bolded. The 5^{th} column produces an average correlation, averaging the correlation of each model's forecasts with the forecasts of the winning models in all four categories. If a model were to be the top model on each criterion, it would get a correlation of 1. Finally, the 6^{th} column reports whether the model is stable or not according to the Chow test (using a 10% significance level). Stable models are bolded.

Using this information, we winnow down our set of models by requiring a good model to be bolded in at least 3 out of 4 criteria, in both non-winsorized and winsorized samples. This leaves us with only 6 models: models 1, 4, 8, 10, 12 and 14. However, the simplest models, models 1 and 4, as well as model 12, are not always stable. We therefore select models 8, 10 and 14 as the winning models. Model 8 is Corsi's HAR model, supplemented with the squared *VIX*. Model 10 adds jump terms to this model, whereas Model 14 is the model shown in Table 3, picked by the full sample model selection procedure.

In Table 5, we show the actual statistics for these 6 models, again focusing on non-winsorized data, but also for some popular simple models used in the literature: the squared *VIX* – realized variance model used in Bekaert, Hoerova, Lo Duca (2013) (our model 3), the martingale model of BTZ (model 16) and the AR(1) model of Londono (2011) (model 2). In the last column, we also produce the average ranks of these models in the forecasting horse race over the 4 different criteria. The MAPE criterion is the most distinguishing one. Model 4 is the best performing model according to the MAPE criterion, but the other selected models produce MAPE statistics that are quite close with the exception of Model 14. Compared to the top models, the martingale and simple autoregressive models perform an order of magnitude worse but the squared *VIX* – realized variance model delivers quite similar performance. The MAE and MSE criteria have much less power to reject models. Simply investigating the magnitudes of the average forecast errors, the presence of realized variances at all three frequencies is important in delivering lower error statistics. Of the simpler models, the squared *VIX* –

realized variance model again performs best. In terms of R^2, models 8, 10, 12 and 14 yield substantial higher values than the other models, with the squared *VIX* – realized variance model one of the best behind these top models. When computing a simple average ranking across the four criteria, our selected models have very low average ranking scores. We also computed the AIC and the BIC criteria for all these models over the "in sample" period. These criteria favor more complex models, with Model 12 minimizing these criteria in both the non-winsorized and winsorized data.

Additional exercises

We performed two alternative exercises. First, a more dramatic way to deal with the crisis is to leave it out of the analysis entirely. We performed our analysis using September 14, 2008 as the end of the sample, thus excluding crisis data points. For this sample, models 1, 4 and 10 of the level models, and models 1, 4 and 6 of the log models yield at least 3 bolded criteria. Hence, three of our previously identified models overlap.[6] Moreover, our preferred models (models 8, 10 and 14) display an average correlation of about 96% with the winning models over this sample period.

Second, we re-consider our forecasting exercise with end-of-month data. In most of the existing articles (including BTZ, Londono, 2011, and Busch, Christensen and Nielsen, 2011), end-of-month data are used to estimate conditional variance models. The use of daily data should lead to more efficient estimates, but the correlation between daily and monthly data induced by the overlapping data structure, may make the increase in efficiency minor. We estimated all our models using end-of-the-month data till mid-2005, mimicking the time span used in our forecasting exercise with daily data. We then use the obtained regression coefficients to construct daily out-of-sample realized variance

[6] Note that none of these models is stable.

forecasts for the remainder of the sample. Computing the usual criteria, we check whether we can accept the various models by looking for three "bolds" (fails to reject) on our four quantitative criteria (MAPE, MAE, MSE, and R^2), and verify the stability criterion.

Not surprisingly, monthly models do not "win" on any of the criteria. Yet, models 1 and 8 still get 3 bolds out of 4. However, only model 8 is also stable. Interestingly, model 8, based on monthly estimates, does well relative to the best models based on the daily information. The monthly model puts less weight on the squared *VIX* and more weight on $RV^{(1)}$ than the daily model does. The more complex models, like our previously winning models 10 and 14, which include jumps and/or asymmetric volatility, are, not surprisingly, more difficult to identify with monthly data. The use of monthly estimation samples should therefore best be restricted to relatively simple models, where the loss of efficiency is not very costly.

5. Economics and Predictability

Risk and risk aversion

In Figure 1, we plot the daily series for the variance risk premium (VP henceforth; displayed in Panel A), which may potentially serve as a proxy for risk aversion, and the conditional (physical) variance of the stock market (CV henceforth; in Panel B), which may potentially serve as a measure of economic uncertainty. We use the non-winsorized sample to do so and show the three series obtained from the winning models 8, 10 and 14 on one graph. The VP and CV series display peaks at the expected times. The largest peaks for CV are observed during the Lehman aftermath in the recent crisis and at the time of the corporate scandals following the Enron debacle. Interestingly, the 1998 Russian crisis and the Gulf war did not generate much uncertainty, but these events do feature substantially elevated levels of VP. The Lehman event seems to have caused both massive uncertainty and massive risk aversion. As the three series are generally highly correlated, they tend to produce similar peaks and valleys; yet, it is visible that model 8, which does not feature jumps and volatility asymmetry, is less jagged than the other models, especially for the VP series.

When realized variances show extreme peaks, the VP series can become negative, which happens more for models 10 and 14 than for model 8. This is a disadvantage of all these models. It is unlikely that during these periods of stress, there was a sudden increase in risk appetite. The more mundane explanation is that realized variances likely have different components with different levels of mean reversion. In a massive crisis, some of the realized variance movements should probably be allowed to mean-revert more quickly and not affect the conditional variance as much as they do now. The models with jumps could theoretically capture this by having negative coefficients on the jump terms. However, when using non-winsorized data, models 10 and 14 put a very large positive coefficient on the monthly jump component, and a negative one on the daily jump component. With the winsorized data, this problem is less prevalent, partially because of the milder data, but also because model 14 now does not feature a monthly jump component, and the daily jump term gets a negative coefficient. Overall, it is likely that a non-linear model may be better equipped to capture the behavior of CV and VP in severe crises.

Contemporaneous correlations

In Table 6, we report correlation matrices for the full sample. We show the three VP and CV measures from the winning models, excess stock returns (the S&P500 return in excess of the Fama-French one-month rate; denoted Ret), industrial production growth (the log-difference of the total industrial production index; denoted IP) and two financial stress indicators, one created by the Kansas Fed (denoted FS Fed), and one created by the ECB (called CISS, see Hollo, Kremer and Lo Duca, 2012; denoted FS ECB). The Kansas Fed indicator combines a large number of interest rate variables such as the TED spread and the off/on-the-run-Treasury spread; a number of corporate yield spreads, risk indicators drawn from banking stock returns, but also the stock-bond return correlation and the *VIX* itself (see Hakkio and Keeton, 2009, for details). The series starts in February 1990. The ECB indicator is based on European Monetary Union data, combining information from the money, equity, bond, and foreign exchange markets, and some financial intermediaries-related information. The indicators mostly comprise realized volatilities for various return, currency or interest rate measures.

The different CV measures are very highly correlated (correlations in excess of 94%), but Model 8's VP measure shows correlation lower than 90% with the other two VP measures. Before the crisis, these correlations (unreported) were above 90%. The VP measures of Models 10 and 14 are 96% correlated. The VP and the CV measures display correlations roughly in the 25-45% range. In a pre-crisis sample (unreported), these correlations would be 20% higher.

The financial instability indicators are relatively more correlated with the CV measures, than with the VP measures. The Kansas Fed measure shows a stronger association with the *VIX* components than does the ECB measure, which is not surprising given that it is based on US data and includes the *VIX* index itself. Excess returns not surprisingly correlate negatively with both *VIX* components, but the correlations are higher (in absolute magnitude) for the CV measure. Industrial Production growth is negatively correlated with both the CV and VP measures, with the VP – industrial production growth correlations being generally lower in absolute magnitude.[7]

Finally, while not reported, we should point out that when BTZ's martingale model is used, it produces the relatively lowest correlation between VP and CV, and it produces surprisingly positive (but very small) correlations between the variance premium and both industrial production and excess returns. One reason for this is that the martingale model generates negative variance premiums during the crisis period more often than our preferred models.

Predicting stock market returns

The two components of the squared *VIX* index have been considered as separate potential predictors of stock market returns. Starting with French, Schwert and Stambaugh (1987), a large literature focuses on the relationship between aggregate stock market returns and their conditional variance. In a simple static CAPM model, the coefficient on the conditional stock market variance would be the wealth weighted risk aversion coefficient, but such a relationship need not hold perfectly in a dynamic model. In the literature on the risk–return relationship, estimates vary from positive to negative and the relationship is often insignificant. Lundblad (2007) suggests that the samples typically used are too short to uncover a relationship that is robustly and statistically significantly positive in the sample of over 150 years that he considers. Yet, the measurement of the conditional variance of stock returns may matter too. The bulk of the extant literature has considered GARCH-in-mean models to measure the conditional

[7] The correlation matrices for winsorized data look similar to the ones discussed here and are therefore not reported.

stock market variance, which likely induces substantial measurement error in the regression. Ghysels, Santa Clara and Valkanov (2005) recover a positive risk–return trade-off measuring the conditional variance with a flexible function of past returns, applying MIDAS modeling.

BTZ recently showed that the variance risk premium has predictive power for future stock returns, which is logical since it harbors information about aggregate risk aversion. As we showed above, their measure implicitly uses a volatility model that is strongly rejected by the data. We therefore reconsider the predictive power of both the equity variance risk premium ("risk") and the conditional variance of the stock market ("uncertainty"), using our improved measures of the conditional variance of stock market returns.

Given the importance of the BTZ-paper, we start by replicating their results using their sample period, which ends in December 2007 and therefore conveniently excludes the crisis period. We also rely on end-of-the-month observations but we consider various estimates of the variance premium as a predictor of equity returns. Table 7 contains the results. The left hand side variable is always Ret, as described above. We use three different horizons, monthly, quarterly and annual (denoted by 1, 3 and 12, respectively). The overlap in the monthly data creates serial correlation in the error term that must be corrected for in creating standard errors. We use a relatively large number of Newey-West lags, namely max{3, 2*horizon}, to do so, rather than create standard errors under the null of no predictability, as in Hodrick (1992). While the Hodrick estimator has very good size properties, selecting a large number of lags may improve power (see Sun, Phillips, and Jin, 2008).

In the last specification, we show that the squared *VIX* itself fails to predict stock returns. Just above, we repeat the BTZ specification that uses the past realized variance as the estimate of the conditional variance of stock market returns. The resulting variance premium proxy predicts stock market returns at all three horizons with the predictive power strongest at the quarterly horizon, both in terms of statistical significance of the coefficient and the adjusted R^2. These results confirm the results in BTZ.[8] Compared to the predictability results when using the three best models - models 8, 10 and 14 - to estimate the variance premium, BTZ's martingale model maximizes the predictive power of the variance premium for returns. For the best models, there is only statistical significant predictive power at the quarterly frequency and the R^2 drops from 7% to somewhere between 3 and 5%.[9]

In unreported results, we find that various estimates of the conditional variance do not predict future stock market returns for this sample. With the exception of model 14, where the conditional variance predicts future stock market returns next month, we do no record any significant coefficients.

This generates somewhat of a puzzle regarding the origin of the strong predictive power of the BTZ-variance premium. If the *VIX* itself does not predict stock market returns and aggregate realized variance does not either, why does their difference provide strong predictive power? The coefficient on the variance premium can be decomposed as follows:

[8] Our results differ slightly from BTZ estimates because our monthly realized variance measure is based on the sum of squared returns over 22 trading days, whereas the BTZ measure is based on the sum of squared returns over a calendar month (which is mostly, but not always, 22 trading days).

[9] One possibility is that because we pre-estimate the conditional variance and BTZ do not, measurement noise affects our estimates. However, our measurement provides proxies for the variance premium and the conditional stock market variance closer to the true economic concepts.

$$\beta_{VP} = \beta_{VIX^2} \frac{\text{var}[VIX^2]}{\text{var}[VP]} - \beta_{CV} \frac{\text{var}[RV]}{\text{var}[VP]} \qquad (8)$$

It turns out that the variance of the squared *VIX* is higher than the variance of *RV*, which is itself rather similar to the variance of the variance risk premium. Therefore, the variance premium coefficient at the quarterly frequency scales up the positive coefficient on the *VIX*, and gets an additional small contribution from the coefficient on stock market volatility which is negative at the quarterly horizon.

Economically, it does appear that the variance risk premium uncovers a component in the *VIX* index that is related to future stock market returns, but the statistical evidence is not very strong. Apart from the small sample, one possible reason for this is the well-known fact that equity risk premiums are likely driven by multiple state variables (see Ang and Bekaert; 2007, Menzly, Santos and Veronesi, 2004) so that the univariate regressions are necessarily mis-specified. In the consumption-based asset pricing model of Bekaert, Engstrom and Xing (2009), risk aversion and uncertainty are the two state variables driving time-variation in the equity risk premium.[10]

Switching to the full sample, we therefore investigate bivariate regressions using both the variance premium and the conditional variance as predictors, a specification not considered in BTZ. We also performed tests allowing for dummies to capture potential coefficient changes during the crisis. However, we do not report these regressions with dummies as the evidence for changes during the crisis is overall econometrically weak. When there are interesting changes, we simply note them in the text.

The bivariate regression results are in Panel A of Table 8. The VP remains overall the

[10] Anderson, Ghysels, and Juergens (2009) also examine the impact of "risk" and "uncertainty", but in their paper risk represents physical volatility and uncertainty disagreement among forecasters.

stronger predictor, with its coefficients and t-stats increasing relative to the shorter sample. It is now statistically significant at both the quarterly and annual horizons for all three preferred measures. The CV coefficients are always negative and sometimes significantly so for models 8 and 10.

These results have implications for the consumption–based asset pricing literature, where there is a persistent debate about what economic mechanism generates a large equity premium, volatile stock market and long–horizon stock return predictability. In the Bansal–Yaron (2004) long-run risk model, time–variation in the equity premium comes from time–variation in economic uncertainty. Recent versions of the model (see e.g. Bansal, Kiku, Yaron, 2012) put more and more emphasis on the role of volatility and argue that substantial persistence in volatility is necessary to make the models fit the salient asset return features. However, our empirical results cast doubt on this economic mechanism. The persistence of the conditional variance varies between 0.67 and 0.73 across models. Moreover, the time-varying risk premium component in equity returns comes predominantly from the variance risk premium, not from time-varying economic uncertainty. The effects of economic uncertainty on risk premiums we do document seem short-lived. This suggests that the alternative class of models (see Campbell and Cochrane, 1999), which relies on counter-cyclical changes in risk aversion to generate variation in risk premiums, has more chance of being the true economic mechanism explaining time-variation in equity risk premiums.

In Panel B, we consider a multivariate regression including other well-known predictor-variables, namely the real 3-month rate (the three-month T-bill minus CPI inflation, denoted 3MTB), the log dividend yield (denoted Log(DY)), the credit spread (the difference between Moody's BAA and AAA bond yield indices, denoted CS) and the term spread (the difference between the 10-year and the 3-month Treasury yields, denoted TS). The addition of the other variables strengthens the predictive power of the variance premium for equity returns, with the coefficients uniformly increasing slightly. The evidence for the predictive power of CV does change considerably. The uncertainty coefficients are now mostly small and insignificantly different from zero.

As to the other variables, the term structure variables are never significant. Both the real rate and the term spread have consistently positive coefficients, reversing a pattern observed when crisis data are excluded, where they have negative coefficients at the monthly and quarterly horizons, but positive ones at the annual horizon (which are also not significant). The credit spread obtains a large negative coefficient that is not significantly different from zero, and the dividend yield is at best significant at the 10% level, mostly at the longer horizons. Here, the crisis adversely affected the predictive power of these variables. Excluding crisis data, the dividend yield and the credit spread were highly statistically significantly different from zero at all horizons for all specifications, with the dividend yield having the expected positive coefficient, but the credit spread negatively affecting the equity premium.[11]

The adjusted R^2's remains small at the one month horizon, but now becomes quite large at the quarterly (12 to 20% range) and annual horizons (around 27%). It is likely that this high explanatory power may partially reflect statistical bias (see Boudoukh,

[11] There is no issue of multi-collinearity in the regression as the dividend yield–credit spread correlation is in fact close to zero. While the negative credit spread coefficient may surprise some readers, BTZ also report negative coefficients for the credit spread in univariate excess return regressions. It is conceivable that the credit spread is a good indicator of economic prospects (for example, it is relatively highly correlated with economic uncertainty) and therefore helps cleanse the dividend yield from variation driven by cash flows, rather than risk premiums (see Golez, 2012 for a recent interesting attempt to cleanse the dividend yield of cash flow effects in a predictability regression).

Richardson and Whitelaw, 2007).

We conducted the full sample analysis using winsorized data, finding that our results are unchanged. We therefore omit the results.

Predicting the Real Economy

In Table 9, we examine the predictive power of the variance risk premium, and stock market volatility (using our three preferred measures) for economic activity as measured by industrial production growth. Bloom (2009) shows that uncertainty shocks lead to a rapid drop and rebound in aggregate output and employment. In a model with adjustment costs to labour and capital, this occurs because higher uncertainty causes firms to temporarily pause their investment and hiring. In some of his empirical work, Bloom actually uses the *VIX* to help measure uncertainty shocks. Here, we investigate whether the *VIX* and/or its two components predict economic activity in a simple regression framework.

The last specification shows that the squared *VIX* itself predicts economic activity with a negative sign at all horizons (significant at the 1% level). The bivariate regressions with its two components show that whatever predictive power the *VIX* has for future output, is coming from the uncertainty component. The coefficient on VP is negative at monthly and quarterly horizons, but it is always statistically insignificantly different from zero. The coefficient on CV is always negative, and statistically significant at the 1% level for all three horizons. We conclude that CV is a robust and significant predictor of economic activity.

6. Conclusions

We decompose the squared *VIX*, the risk neutral expected stock market variance, into two components, the conditional (physical) variance of the stock market (CV) and the equity variance premium (VP), which is the difference between the two (VP=VIX^2-CV). Because this decomposition critically depends on the accuracy of the model for CV, we first conduct an extensive analysis of state-of-the-art variance forecasting models, where we make sure to also consider the squared *VIX* itself as a potential predictor. We find that the winning models on a number of forecasting criteria and specification tests always include the *VIX*. Of the two components, the conditional variance is more strongly correlated with some recent financial stress indicators, tracked by the Fed and the ECB, than is the variance premium.

We use these models to re-examine and expand the evidence on the predictive power of VP and CV for stock returns and economic activity, as measured by industrial production. We find that the variance premium is a significant predictor of stock returns, but the conditional variance mostly is not. However, CV robustly and significantly predicts economic activity with a negative sign, whereas VP has no predictive power for future output growth.

Acknowledgements

Carlos Garcia de Andoain Hidalgo provided excellent research assistance. The views expressed do not necessarily reflect those of the European Central Bank or the Eurosystem.

REFERENCES

Andersen, T.G., and O. Bondarenko (2007). "Construction and Interpretation of Model-Free Implied Volatility," in I. Nelken (Ed.), Volatility as an Asset Class, Risk Books, London, UK (2007), 141–181.

Andersen, T. G., T. Bollerslev and F. X. Diebold (2007). "Roughing It Up: Including Jump Components in the Measurement, Modeling, and Forecasting of Return Volatility," Review of Economics and Statistics 89(4), 701-720.

Anderson, E., E. Ghysels, and J.L. Juergens (2009). "The Impact of Risk and Uncertainty on Expected Returns," Journal of Financial Economics, 94(2), 233-263.

Ang, A. and G. Bekaert (2007). "Stock Return Predictability: Is it There?" Review of Financial Studies, 20(3), 651-707.

Baker, M. and J. Wurgler (2007). "Investor sentiment in the stock market," Journal of Economic Perspectives 21(2), 129-152.

Bakshi, G., N. Kapadia and D. Madan (2003). "Stock Return Characteristics, Skew Laws, and Differential Pricing of Individual Equity Options," Review of Financial Studies 16(1), 101-143.

Bali, T.G. (2008). "The Intertemporal Relation between Expected Returns and Risk," Journal of Financial Economics 87(1), 101-131.

Bansal R. and A. Yaron (2004). "Risks for the Long-Run: A Potential resolution of Asset Pricing Puzzles," Journal of Finance 59(4), 1481-1509.

Bansal R., D. Kiku and A. Yaron (2012). "An Empirical Evaluation of the Long-Run Risks Model for Asset Prices," Critical Finance Review 1, 183-221.

Barndorff-Nielsen, O., and M. Shepherd (2004). "Power and Bi-Power Variation with Stochastic Volatility and Jumps," Journal of Econometrics 2(1), 1-48.

Bekaert, G., M. Hoerova and M. Lo Duca (2012). "Risk, Uncertainty, and Monetary Policy," working paper, Columbia GSB.

Bekaert, G. and E. Engstrom (2010). "Asset Return Dynamics under Bad Environment-Good Environment Fundamentals," working paper, Columbia GSB.

Bekaert, G., E. Engstrom, and Y. Xing (2009). "Risk, Uncertainty, and Asset Prices," Journal of Financial Economics 91, 59-82.

Bekaert, G. and G. Wu, (2000), "Asymmetric Volatility and Risk in Equity Markets," Review of Financial Studies 13(1), 1-42.

Bloom, N. (2009). "The Impact of Uncertainty Shocks," Econometrica 77(3), 623-685.

Bollerslev, T. and E. Ghysels (1996). "Periodic Autoregressive Conditional Heteroscedasticity," Journal of Business & Economic Statistics 14(2), 139-51.

Bollerslev, T., G. Tauchen and H. Zhou (2009). "Expected Stock Returns and Variance Risk Premia," Review of Financial Studies 22(11), 4463-4492.

Boudoukh, J., M. Richardson, and R.Whitelaw (2007). "The Myth of Long-Horizon Predictability," Review of Financial Studies 24(4), 1577-1605.

Britten-Jones, M. and A. Neuberger (2000). "Option Prices, Implied Price Processes, and Stochastic Volatility," Journal of Finance 55(2), 839-866.

Busch, T., B.J. Christensen, and M.O. Nielsen (2011). "The role of implied volatility in forecasting future realized volatility and jumps in foreign exchange, stock, and bond markets," Journal of Econometrics 160(1), 48-57.

Campbell, J. Y and J. Cochrane (1999). "By Force of Habit: A Consumption Based Explanation of Aggregate Stock Market Behavior," Journal of Political Economy 107(2), 205-251.

Campbell, John Y., and Ludger Hentschel (1992). "No News is Good News: An Asymmetric Model of Changing Volatility in Stock Returns," Journal of Financial Economics 31(3), 281-318.

Campos, J., D.F. Hendry, and H.M. Krolzig (2003). "Consistent Model Selection by an Automatic Gets Approach," Oxford Bulletin of Economics and Statistics 65, 803-819.

Carr, P. and L. Wu (2009). "Variance Risk Premiums," Review of Financial Studies 22(3), 1311-1341.

Chen, X. and E. Ghysels (2012). "News – Good or Bad – and its Impact on Volatility Predictions over Multiple Horizons," Review of Financial Studies 24(1), 46-81.

Chernov, M (2007). "On the Role of Risk Premia in Volatility Forecasting," Journal of Business & Economic Statistics 25(4), 411-426.

Chicago Board Options Exchange (2004). "VIX CBOE Volatility Index," White Paper.

Christensen, B.J. and N.R. Prabhala (1998) "The Relation between Implied and Realized Volatility," Journal of Financial Economics 50(2), 125-150.

Corsi F. (2009). "A Simple Approximate Long Memory Model of Realized Volatility," Journal of Financial Econometrics 7(2), 174-196.

Corsi F. and R. Renò (2012). "Discrete-time Volatility Forecasting with Persistent Leverage Effect and the Link with Continuous-time Volatility Modeling," Journal of Business & Economic Statistics, forthcoming.

Corsi F., D. Pirino and R. Renò (2010). "Threshold Bipower Variation and the Impact of Jumps on Volatility Forecasting," Journal of Econometrics 159(2), 276-288.

Coudert, V. and M. Gex (2008). "Does Risk Aversion Drive Financial Crises? Testing the Predictive Power of Empirical Indicators," Journal of Empirical Finance 15, 167-184.

Diebold, F. X. and R.S. Mariano (1995). "Comparing Predictive Accuracy," Journal of Business & Economic Statistics 13(3), 253-63.

Drechsler, I. (2009). “Uncertainty, Time-Varying Fear, and Asset Prices,” working paper, Wharton School.

Drechsler, I. and A. Yaron (2011). “What’s Vol Got to Do with It,” Review of Financial Studies 24(1), 1-45.

Engle, R. and V. Ng (1993). “Measuring and Testing the Impact of News and Volatility,” Journal of Finance 48(5), 1749-1778.

Fernandes, M., M. C. Medeiros, and M. Scharth (2009). “Modeling and Predicting the CBOE Market Volatility Index,” working paper.

French, K., W. Schwert and R. Stambaugh (1987), “Expected Stock Returns and Volatility,” Journal of Financial Economics 19, 3-29.

Ghysels, E., P. Santa-Clara, and R. Valkanov (2005). “There is a risk-return trade-off after all,” Journal of Financial Economics 76(3), 509-548,

Golez, B. (2012). “Expected Returns and Dividend Growth Rates Implied in Derivative Markets,” Working paper, University of Notre Dame.

Hakkio, C. S. and W. R. Keeton (2009). “Financial Stress: What Is It, How Can It Be Measured, and Why Does It Matter?” Kansas Fed Economic Review, second quarter 2009.

Hodrick, R.J. (1992). “Dividend yields and expected stock returns: alternative procedures for inference and measurement”, Review of Financial Studies 5(3), 357-386.

Holló, D., M. Kremer, and M. Lo Duca (2012). “CISS - a composite indicator of systemic stress in the financial system,” ECB Working Paper No. 1426.

Londono, J.M. (2011). “The Variance Risk Premium around the World,” IFDP Working Paper.

Lundblad, C. (2007). “The Risk Return Tradeoff in the Long-Run: 1836-2003,” Journal of Financial Economics 85, 123-150.

Menzly, L., T. Santos, and P. Veronesi (2004), “Understanding Predictability”, Journal of Political Economy 112, 1-47.

Mincer, J. and V. Zarnowitz (1969). “The Evaluation of Economic Forecasts,” In: J. Mincer (Ed.), Economic Forecasts and Expectations, NBER, New York, 3–46.

Newey, W. and K. West (1987). “A Simple, Positive Semi-definite, Heteroskedasticity and Autocorrelation Consistent Covariance Matrix,” Econometrica 55(3), 703-708.

Patton, A. (2011). “Volatility Forecast Comparison using Imperfect Volatility Proxies,” 2011, Journal of Econometrics 160(1), 246-256.

Sun, Y., P.C. B. Phillips and S. Jin (2008). “Optimal Bandwidth Selection in Heteroskedasticity-Autocorrelation Robust Testing,” Econometrica 76(1), 175-194.

Whaley, R. E. (2000). “The Investor Fear Gauge,” Journal of Portfolio Management, Spring, 12-17.

Table 1: The VIX and variance premium

Parameters	RRA	*VIX*	VP
Panel A: Varying γ , $Sk=-1$, $p=0.5\%$			
Parameters		*VIX*	VP
$Sk=-1, \gamma=2$		15.925	0.003
$Sk=-1, \gamma=4$		17.342	0.008
$Sk=-1, \gamma=6$		19.472	0.015
Panel B: Varying γ , $Sk=-2$, $p=1\%$			
Parameters		*VIX*	VP
$Sk=-2,\ \gamma=2$		16.838	0.006
$Sk=-2,\ \gamma=4$		19.516	0.016
$Sk=-2,\ \gamma=6$		23.194	0.031
Panel C: Varying W_{bm} , $\gamma=4$, $Sk=-1$, $p=0.5\%$			
Parameters	RRA	*VIX*	VP
$\gamma=4, W_{bm}=0$	4.000	17.342	0.008
$\gamma=4, W_{bm}=0.25$	5.323	19.059	0.014
$\gamma=4, W_{bm}=0.50$	7.968	26.010	0.045

Notes: Values of the *VIX* on an annualized basis in percent (*VIX*) and the annualized variance premium (VP) for different values of the underlying parameters, while keeping the crash return *c* fixed at -25%. In Panel A, the varying parameter is the coefficient of relative risk aversion γ while skewness *Sk* is fixed at -1. In Panel B, *Sk* is fixed at -2. Panel C computes, for γ fixed at 4 and *Sk* fixed at -1, expected relative risk aversion (RRA), *VIX* and VP for different values of the benchmark wealth level W_{bm}.

Table 2: Models considered

Variables	VIX^2	$RV^{(22)}$	$RV^{(5)}, RV^{(1)}$	$J^{(22)}$	$J^{(5)}, J^{(1)}$	$r^{(22)-}$	$r^{(5)-}, r^{(1)-}$
Estimated models							
Model 1	X						
Model 2		X					
Model 3	X	X					
Model 4	X	X		X			
Model 5		X		X			
Model 6	X	X		X		X	
Model 7		X		X		X	
Model 8	X	X	X				
Model 9		X	X				
Model 10	X	X	X	X	X		
Model 11		X	X	X	X		
Model 12	X	X	X	X	X	X	X
Model 13		X	X	X	X	X	X
Model 14	Hendry-chosen model - see Table 3						
Non-estimated models							
Model 15	X						
Model 16		X					
Model 17	0.5* X	0.5*X					

Notes: Summary of variables included in estimated and non-estimated models.

Table 3: Hendry analysis

Model	(1)	(2)	(3)	(4)
VIX^2		0.203***	0.162**	0.235***
		[0.074]	[0.076]	[0.069]
$RV^{(22)}$	-0.255	0.359***	0.233***	0.349***
	[0.276]	[0.054]	[0.077]	[0.052]
$RV^{(5)}$	0.216**	0.241***	0.204**	0.228***
	[0.103]	[0.052]	[0.087]	[0.042]
$RV^{(1)}$	0.141***	0.101***	0.165***	0.122***
	[0.033]	[0.014]	[0.035]	[0.017]
$J^{(22)}$	4.954**			
	[2.352]			
$J^{(5)}$				
$J^{(1)}$	-0.308**		-0.252**	-61.924**
	[0.145]		[0.112]	[28.381]
$r^{(22)-}$	-0.013**			
	[0.006]			
$r^{(5)-}$		-0.356*		
		[0.195]		
$r^{(1)-}$	-0.002***	-0.231***	-0.0009***	-0.283***
	[0.0009]	[0.060]	[0.0003]	[0.049]
Constant	0.0003**	-0.810***	0.0002	-0.621***
	[0.0001]	[0.186]	0.0001	[0.191]
# daily observations	5155	5155	5155	5155

Notes: The table reports the OLS estimates for monthly variance forecast regressions. Columns 1-2 use the daily non-winsorized sample and Columns 3-4 use the daily winsorized sample. Columns 2 and 4 use the data in logarithms. The standard errors reported in brackets are computed using 44 Newey-West lags. ***, **, * denote significance at the 0.01, 0.05 and 0.10-level.

Table 4: Model ranking, non-winsorized sample

Model	MAPE rank	MAE rank	MSE rank	R^2 rank	Corr w/ #1 models	Stable
Estimated level models						
Model 1	**2**	**12**	**22**	22	0.952	N
Model 2	18	15	**14**	14	0.963	**Y**
Model 3	5	**9**	**13**	13	0.977	**Y**
Model 4	**1**	**10**	**19**	19	0.974	**Y**
Model 5	17	14	**15**	15	0.963	N
Model 6	6	**7**	**11**	11	0.983	**Y**
Model 7	12	11	**9**	9	0.977	**Y**
Model 8	**3**	**3**	**8**	**8**	0.986	**Y**
Model 9	11	8	**4**	**4**	0.981	**Y**
Model 10	**4**	**1**	**7**	**7**	0.989	**Y**
Model 11	13	6	**3**	**3**	0.984	**Y**
Model 12	7	**2**	**5**	**5**	0.991	N
Model 13	9	**4**	**2**	**2**	0.989	N
Model 14	10	**5**	**1**	**1**	0.989	**Y**
Non-estimated models						
Model 15	31	31	**25**	25	0.952	N/A
Model 16	20	26	**23**	23	0.963	N/A
Model 17	30	30	**20**	20	0.977	N/A
Estimated log models						
Log Model 1	8	13	**17**	17	0.955	N
Log Model 2	29	24	12	12	0.964	**Y**
Log Model 3	21	20	**18**	18	0.977	N
Log Model 4	14	**28**	**30**	30	0.719	N
Log Model 5	27	29	**31**	31	0.702	**Y**
Log Model 6	15	**23**	**29**	29	0.793	N
Log Model 7	26	27	**27**	27	0.830	**Y**
Log Model 8	22	17	**10**	10	0.988	N
Log Model 9	28	16	**6**	**6**	0.988	N
Log Model 10	16	**22**	**28**	28	0.818	N
Log Model 11	25	25	**26**	26	0.842	N
Log Model 12	19	19	**24**	24	0.835	N
Log Model 13	23	18	**21**	21	0.842	N
Log Model 14	24	21	**16**	16	0.973	N

Notes: Model ranking based on the out-of-sample performance, non-winsorized sample. Parameters are estimated using data between January 1, 1990 and July 15, 2005 and the rest of the sample (till October 01, 2010) is used to assess forecasting performance. The first three columns produce ranking according to the mean absolute percentage error (MAPE), mean absolute error (MAE) and mean-squared error (MSE). The rank is bolded if the Diebold-Mariano test fails to reject (at 10% level) the null of no significant difference from the best ranked model. The 4^{th} column produces ranking according to the Mincer-Zarnowitz R^2, with the rank bolded if the difference with the winning model is less than 5%. The 5^{th} column produces the average correlation of each model with the winning models in the 4 categories. The 6^{th} column reports whether the model is stable or not according to the Chow test (using 10% significance level); stable models are bolded.

Table 5: Selected model statistics, non-winsorized sample

Model	MAPE	MAE	MSE	R^2	Average score
Model 1	**0.342**	**1.878E-03**	**2.840E-05**	0.410	14.50
Model 2	0.494	1.914E-03	**2.490E-05**	0.483	15.25
Model 3	0.354	**1.775E-03**	**2.460E-05**	0.489	10.00
Model 4	**0.338**	**1.784E-03**	**2.570E-05**	0.466	12.25
Model 8	**0.346**	**1.693E-03**	**2.150E-05**	**0.553**	5.50
Model 10	**0.353**	**1.681E-03**	**2.120E-05**	**0.559**	4.75
Model 12	0.374	**1.682E-03**	**2.090E-05**	**0.565**	4.75
Model 14	0.437	**1.711E-03**	**1.990E-05**	**0.587**	4.25
Model 16	0.501	2.228E-03	**2.850E-05**	0.407	23.00

Notes: Selected model statistics based on the out-of-sample performance. The first three columns show the mean absolute percentage error (MAPE), mean absolute error (MAE) and mean-squared error (MSE). The statistic is bolded if the Diebold-Mariano test fails to reject (at 10% level) the null of no significant difference from the best ranked model. The 4th column reports Mincer-Zarnowitz R^2, with the statistic bolded if the difference with the winning model is less than 5%. The 5th column produces the average ranking score of each model in the 4 categories.

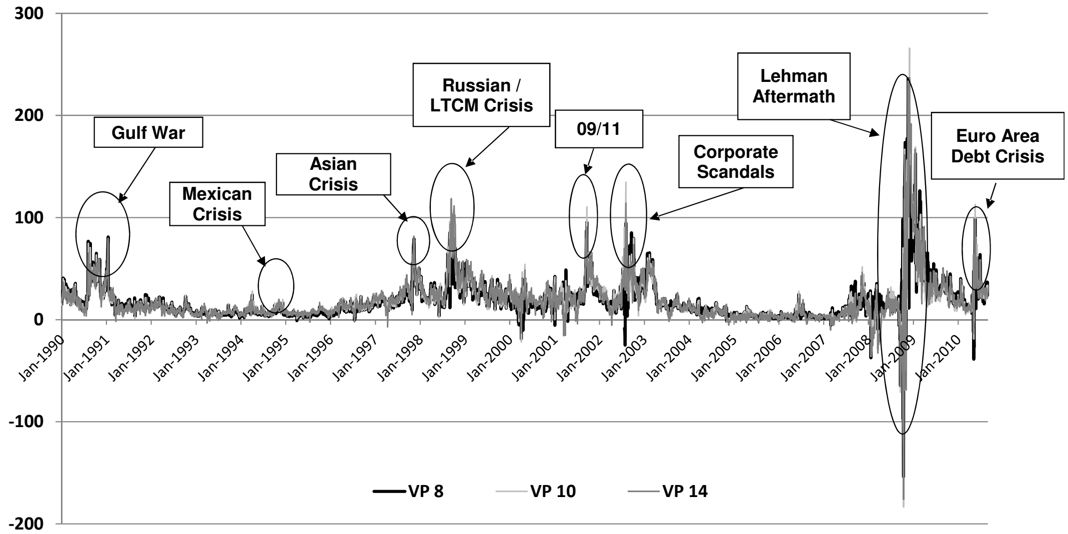
Table 6: Correlation matrix

Correlations										
	VP 8	VP 10	VP 14	CV 8	CV 10	CV 14	FS ECB	FS Fed	Ret	IP
VP 8	1									
VP 10	0.871	1								
VP 14	0.838	0.963	1							
CV 8	0.375	0.443	0.414	1						
CV 10	0.426	0.343	0.318	0.955	1					
CV 14	0.418	0.336	0.266	0.942	0.986	1				
FS ECB	0.457	0.363	0.360	0.670	0.718	0.696	1			
FS Fed	0.584	0.508	0.518	0.772	0.807	0.772	0.837	1		
Ret	-0.172	-0.221	-0.050	-0.456	-0.425	-0.515	-0.237	-0.238	1	
IP	-0.211	-0.096	-0.147	-0.234	-0.299	-0.260	-0.371	-0.479	-0.004	1

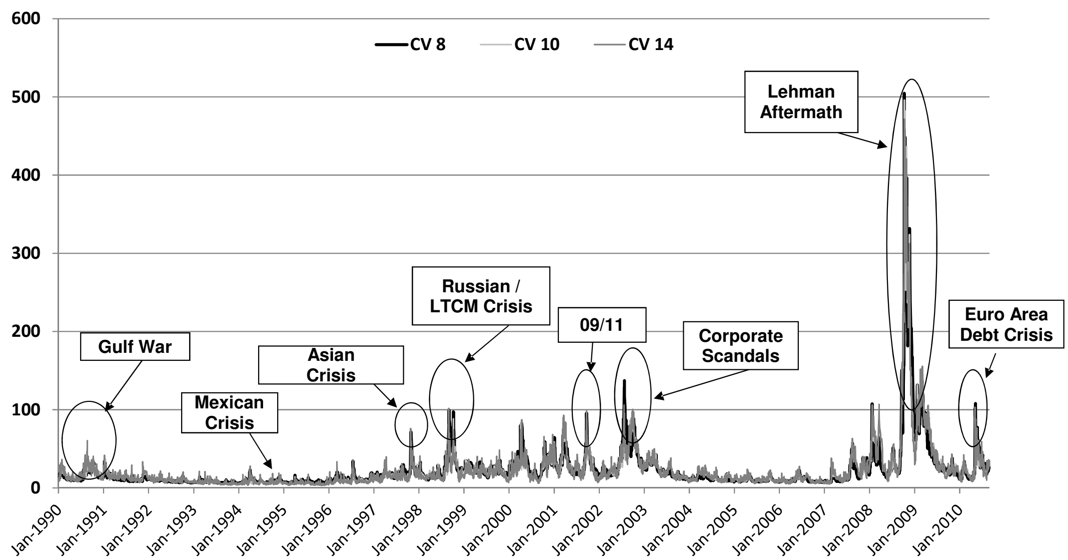
Notes: Sample period February 1990 – September 2010, monthly observations. Correlations for the three VP and CV measures from the winning models 8, 10 and 14; two financial stress indicators, one created by the Kansas Fed (FS Fed) and one created by the ECB (FS ECB); excess stock returns (Ret), and industrial production growth (IP).

Figure 1: Variance premium and conditional variance

Panel A: Variance premia



Panel B: Conditional variances



Notes: Daily series for the variance premium (VP) and the conditional variance (CV) from the winning models 8, 10 and 14 (full non-winsorized sample).

Table 7: Stock return regressions, sample till 2007

Monthly, quarterly and annual regressions with variance premium															
Horizon	1	3	12	1	3	12	1	3	12	1	3	12	1	3	12
VP 8	0.381	0.462***	0.116												
	[0.237]	[0.171]	[0.140]												
VP 10				0.232	0.382**	0.0643									
				[0.246]	[0.174]	[0.143]									
VP 14							0.123	0.351**	0.0592						
							[0.266]	[0.172]	[0.139]						
VP 16										0.363*	0.501***	0.188**			
										[0.189]	[0.097]	[0.086]			
VIX^2													0.192	0.188	-0.007
													[0.134]	[0.122]	[0.100]
constant	0.231	-1.423	3.626	2.496	-0.356	4.422	4.286	0.133	4.501	-0.0402	-2.826	2.184	-0.0439	-0.262	5.732
	[4.117]	[4.012]	[3.914]	[4.180]	[3.829]	[3.441]	[4.393]	[3.813]	[3.543]	[3.843]	[3.507]	[4.062]	[4.426]	[4.424]	[3.779]
Adj. R^2	0.007	0.049	0.006	0.000	0.035	-0.001	-0.003	0.029	-0.002	0.008	0.070	0.028	0.005	0.023	-0.005

Notes: Sample period January 1990 – December 2007. All regressions are based on monthly observations. The standard errors reported in brackets are computed using max{3, 2*horizon} Newey-West lags. ***, **, * denote significance at the 0.01, 0.05 and 0.10-level.

Table 8, Panel A: Stock return regressions, full sample

Panel A: Monthly, quarterly and annual regressions with variance premium and conditional variance															
Horizon	1	3	12	1	3	12	1	3	12	1	3	12	1	3	12
VP 8	0.487	0.705***	0.245**												
	[0.347]	[0.134]	[0.095]												
CV 8	-0.269*	-0.294***	-0.025												
	[0.139]	[0.050]	[0.059]												
VP 10				0.393	0.485***	0.254**									
				[0.327]	[0.133]	[0.119]									
CV 10				-0.239	-0.209*	-0.034									
				[0.231]	[0.119]	[0.066]									
VP 14							0.317	0.410***	0.260**						
							[0.311]	[0.143]	[0.118]						
CV 14							-0.200	-0.170	-0.035						
							[0.229]	[0.123]	[0.060]						
VP 16										0.459**	0.572***	0.204**			
										[0.182]	[0.109]	[0.084]			
CV 16										-0.057	-0.010	0.053			
										[0.080]	[0.054]	[0.062]			
VIX^2													-0.036	0.014	0.059
													[0.167]	[0.140]	[0.047]
constant	1.548	-1.772	1.175	2.604	0.358	1.228	3.068	0.802	1.111	-2.274	-5.239	0.326	5.913	3.993	2.733
	[5.157]	[4.072]	[5.098]	[4.624]	[3.857]	[4.999]	[4.794]	[4.123]	[5.056]	[4.574]	[4.071]	[5.257]	[5.563]	[4.689]	[4.452]
Adj. R^2	0.017	0.111	0.034	0.012	0.056	0.042	0.008	0.045	0.050	0.032	0.132	0.037	-0.003	-0.004	0.010

Notes: Sample period January 1990 – September 2010. All regressions are based on monthly observations. The standard errors reported in brackets are computed using max{3, 2*horizon} Newey-West lags. ***, **, * denote significance at the 0.01, 0.05 and 0.10-level.

Table 8, Panel B: Stock return regressions, full sample

Panel B: Monthly, quarterly and annual regressions with variance premium, conditional variance and other predictors

Horizon	1	3	12	1	3	12	1	3	12	1	3	12
3MTB	4.146	3.472	3.716	4.073	3.654	3.532	3.961	3.456	3.314	3.897	3.450	3.726
	[3.351]	[3.426]	[3.205]	[3.157]	[3.660]	[3.119]	[3.103]	[3.696]	[3.068]	[3.452]	[3.506]	[3.209]
Log(DY)	19.19	21.67*	19.20*	19.60	21.72	19.63*	19.96	22.26*	20.11*	19.38	21.46	19.13*
	[14.54]	[13.08]	[10.69]	[14.40]	[13.36]	[10.69]	[14.38]	[13.37]	[10.67]	[14.48]	[13.04]	[10.67]
CS	-13.92	-16.40	-6.016	-12.27	-14.53	-5.253	-14.82	-17.53	-6.496	-10.28	-12.23	-5.138
	[16.98]	[12.45]	[4.890]	[18.50]	[14.59]	[5.141]	[17.77]	[14.89]	[4.922]	[16.47]	[12.59]	[4.976]
TS	2.420	2.297	4.224	2.357	2.431	4.076	2.270	2.277	3.905	2.123	2.161	4.207
	[4.022]	[4.369]	[3.671]	[3.987]	[4.505]	[3.618]	[3.953]	[4.490]	[3.575]	[4.033]	[4.365]	[3.662]
VP 8	0.562**	0.809***	0.274***									
	[0.283]	[0.141]	[0.102]									
CV 8	-0.078	-0.090	0.076									
	[0.210]	[0.091]	[0.050]									
VP 10				0.421	0.539***	0.271***						
				[0.312]	[0.166]	[0.079]						
CV 10				-0.045	-0.002	0.062						
				[0.290]	[0.118]	[0.057]						
VP 14							0.388	0.514***	0.291***			
							[0.304]	[0.188]	[0.089]			
CV 14							0.005	0.050	0.069			
							[0.261]	[0.116]	[0.052]			
VP 16										0.513***	0.646***	0.233***
										[0.177]	[0.107]	[0.078]
CV 16										0.069	0.130	0.125**
										[0.126]	[0.096]	[0.054]
constant	-12.65	-15.25	-20.39	-12.49	-14.52	-20.63	-10.58	-12.31	-19.79	-17.54	-20.37	-21.44
	[16.55]	[13.00]	[13.01]	[17.50]	[14.03]	[12.76]	[16.87]	[14.16]	[12.63]	[16.22]	[12.98]	[13.03]
Adj. R^2	0.034	0.189	0.270	0.027	0.131	0.272	0.025	0.126	0.278	0.045	0.199	0.270

Notes: Sample period January 1990 – September 2010. All regressions are based on monthly observations. The standard errors reported in brackets are computed using max{3, 2*horizon} Newey-West lags. ***, **, * denote significance at the 0.01, 0.05 and 0.10-level.

Table 9: Industrial production regressions

Monthly, quarterly and annual regressions with variance premium and conditional variance															
Horizon	1	3	12	1	3	12	1	3	12	1	3	12	1	3	12
VP 8	-0.043	-0.027	0.020												
	[0.043]	[0.041]	[0.024]												
CV 8	-0.097***	-0.109***	-0.053***												
	[0.019]	[0.007]	[0.011]												
VP 10				-0.042	-0.015	0.038									
				[0.060]	[0.038]	[0.023]									
CV 10				-0.098***	-0.116***	-0.063***									
				[0.028]	[0.008]	[0.016]									
VP 14							-0.053	-0.021	0.038						
							[0.059]	[0.038]	[0.024]						
CV 14							-0.092***	-0.113***	-0.063***						
							[0.023]	[0.008]	[0.015]						
VP 16										-0.036	-0.027	0.014			
										[0.033]	[0.030]	[0.019]			
CV 16										-0.082***	-0.086***	-0.033***			
										[0.019]	[0.010]	[0.011]			
VIX^2													-0.080***	-0.084***	-0.031***
													[0.020]	[0.018]	[0.009]
constant	4.792***	4.728***	2.724**	4.808***	4.671***	2.617**	4.887***	4.695***	2.591**	4.370***	4.263***	2.411**	5.103***	5.205***	3.145***
	[0.846]	[0.694]	[1.118]	[0.885]	[0.659]	[1.133]	[0.951]	[0.711]	[1.139]	[0.849]	[0.701]	[1.178]	[0.820]	[0.678]	[0.959]
Adj. R^2	0.123	0.278	0.083	0.124	0.294	0.118	0.122	0.293	0.129	0.131	0.296	0.095	0.122	0.257	0.056

Notes: Sample period January 1990 – September 2010. All regressions are based on monthly observations. The standard errors reported in brackets are computed using max{3, 2*horizon} Newey-West lags. ***, **, * denote significance at the 0.01, 0.05 and 0.10-level.