Forecasting International Index Returns using Option-implied Variables

Marie-Hélène Gagnon
Gabriel J. Power
Dominique Toupin

Avril / April 2018

**Centre de recherche sur les risques
les enjeux économiques et les politiques publiques**

www.crrep.ca

Abstract

This paper investigates international index return predictability using option-implied information. We document the significant predictive power of the variance risk premium (VRP), Foster-Hart risk (FH), and higher-order moments for horizons ranging from 1 to 250 days. Our results from predictive regressions show that these four risk-neutral metrics, which have the advantage of daily updating, perform well internationally. VRP and FH risk are significant predictors for several horizons, including less than one month (VRP) and longer horizons (FH). Risk-neutral skewness and kurtosis are significant for several countries across multiple horizons. Out-of-sample forecasts and utility gain calculations confirm the statistical and economic significance of these risk-neutral variables internationally.

Marie-Hélène Gagnon : FSA ULaval, CRREP, IHEIQ, CEPCI, Marie-Helene.Gagnon@fsa.ulaval.ca
Gabriel J. Power : FSA ULaval, CRREP, CRIB, Gabriel.Power@fsa.ulaval.ca
Dominique Toupin : Université Laval, dominique.toupin.1@ulaval.ca

1. – Introduction

A large literature has explored long-horizon index return predictability using predictive regressions (Rapach and Zhou, 2013). In particular, recent evidence suggests that the variance risk premium (VRP), the difference between option-implied and realized variances, provides a parsimonious way to outperform traditional, macroeconomic predictive variables such as the consumption/wealth ratio (e.g., Bollerslev, Tauchen, and Zhou, 2009). While the relevance of VRP to predict future index returns is recognized, the literature is scarce when it comes to using information from the risk-neutral distribution-- aside from variance--to improve forecasts.

Predictive variables constructed from options markets are appealing for forecasting purposes relative to variables that reflect realizations, as they are forward-looking and capture the risk-adjusted expectations of market participants (Constantinides, Jackwerth, and Perrakis, 2007). Moreover, theory suggests that these variables should be relevant, as substantial price discovery occurs in options markets (Easley, O'Hara and Srinivas, 1998). The relevance of higher moments is further underscored by the non-normality of the risk-neutral index return distributions (e.g. Jackwerth and Rubinstein, 1996). Higher-order moments can also be estimated more reliably using the risk-neutral distribution compared to the historical distribution (Conrad, Dittmar and Ghysels, 2013). Thus, it is appealing to develop additional option-implied metrics that can capture this information for use in predictive regressions.

This paper investigates the predictive power of option-implied information for stock market index returns using data for five major international equity indexes (S&P500, DAX, FTSE, CAC, and SMI) at horizons ranging from one day ahead to one year ahead. Our motivation is to study together for the first time several distinct option-implied measures of risk to forecast index returns. Our evidence relies on both predictive regressions and out-of-sample forecasts, presented as complementary evidence in support

of return predictability. We retain four option-implied predictive variables.[1] The VRP is included to build on the existing literature on predictive regressions (Bollerslev, Marrone, Xu and Zhou, 2014). The other variables, however, have not been studied using predictive regressions. They are risk-neutral skewness (RNS) (Stilger, Kostakis and Poon, 2017), risk-neutral kurtosis (RNK) (Bakshi and Cao, 2003), and Foster-Hart riskiness (FH) (Foster and Hart, 2009; Bali, Cakici and Chabi-Yo, 2011), a measure of bankruptcy risk. We argue that each variable captures a particular aspect of risk that is non-redundant to the investor, and thus should be reflected in future returns. These four variables are obtained from the risk-neutral distribution, and have been previously studied (separately) for the U.S. market. Given that most of the evidence from option-implied variables pertains to the cross-sectional domestic setting, we provide novel international evidence using predictive regressions and out-of-sample frameworks.

The paper makes two contributions to the literature. First, we provide international evidence of the predictive performance of several option-implied risk variables beyond the VRP in a joint setting. While previously such evidence was limited to the VRP using predictive regressions, we present evidence for three additional option-implied predictors and by means of out-of-sample forecasts in addition to predictive regressions. These results suggest both statistical and economic significance. Second, we provide new evidence of predictability using option-implied variables at horizons of less than one month, as well as longer horizons than the previously-documented quarterly peak. To our knowledge, ours is the first study to provide such evidence.

We investigate these predictive variables using a forecasting methodology that is benchmarked with the literature (Clark and McCracken, 2013; Chen, 2009; Ludvigson and Ng, 2007; Maio and Santa

---

[1] Other measures could be considered, including downside risk metrics (such as value-at-risk or expected shortfall), or alternative measures of dispersion such as interquartile ranges. However, their high correlations with VRP suggest that they are redundant for our analysis. Moreover, the VIX or realized variance alone have no predictive power (Bollerslev, Tauchen and Zhou, 2009) and are thus omitted.

Clara, 2015; Rapach, Strauss and Zhou, 2013). First, to allow for higher-frequency return forecasts than previously shown in the literature, we estimate constant-horizon risk-neutral distributions (RNDs) for each equity index using daily options data. Each of the four predictive variables is then measured using the estimated daily RNDs. To evaluate their respective predictive power, we first run predictive regressions and compute the regression R^2 as well as the significance of each variable's slope coefficient. In accordance with the literature, simulations are performed on the critical points of our tests to account for persistence in cumulative returns (Bollerslev et al., 2014; Rapach and Zhou, 2013). Then, we conduct out-of-sample forecasts, compute the R^2_{OOS} , and use forecasting tests to assess the significance of the forecast improvements (Clark and MacCracken, 2005). As with the predictive regressions, our data are sampled at a daily frequency, and several horizons ranging from one to 250 days are studied.

This paper finds, first, that across countries in our sample VRP has significant predictive power for future returns for a wider range of horizons than previously shown. More specifically, we document stock index return predictability below one month, which matters to investors. The additional proposed variables are relevant to describe risk beyond the variance risk premium. We show that FH riskiness contributes predictive power, and typically at longer horizons than does VRP. The out-of-sample forecast evidence supports the results obtained from predictive regressions, as VRP and FH are complementary for forecasting purposes. Risk-neutral skewness and kurtosis are distinct from FH risk and contribute incremental predictive power in addition to the inclusion of FH risk. Internationally, we find out-of-sample evidence of return predictability. For all countries at least at some horizons, the out-of-sample performance is statistically and--using utility gain computations--economically significant. This is a novel result for our option-implied risk variables, and a significant one because predictability is considerably more difficult to obtain from out-of-sample forecasts than from predictive regressions (e.g. Goyal and Welch, 2007).

1.1 – REVIEW OF THE LITERATURE

The first contribution of this paper to the existing literature concerns the international predictive power of option-implied variables beyond the VRP. We assess, in a joint setting, the performance of VRP, Foster-Hart risk, and risk-neutral skewness and kurtosis using both predictive regressions and out-of-sample forecast tests. This is a meaningful contribution because the prior literature in this area has focused only on VRP and predictive regressions. We now summarize the current state of the literature for each chosen risk-neutral variable.

VRP: VRP has been interpreted as a fear gauge or as the price investors pay to hold assets that will pay off higher returns specifically in high-volatility states [Bollerslev, Tauchen and Zhou (2009), Carr and Wu (2009, 2016)]. The literature has found that VRP is a concise and economically meaningful way to link option-implied information to an asset's future returns. There is strong evidence that the stock market VRP has explanatory power for U.S. stock returns [Bali and Zhou (2016), Bekaert and Hoerova (2014), Carr and Wu (2009, 2016)]. Bollerslev, Tauchen and Zhou (2009) find that VRP predicts S&P index returns particularly well at a quarterly horizon, and that it outperforms traditional variables such as the Price/Earnings (P/E) ratio or the consumption-wealth ratio. International evidence for VRP's usefulness is limited, however. Bollerslev, Marrone, Xu and Zhou (2014) run predictive regressions using VRP for international equity indexes and find results that are similar to the S&P for most countries in their sample (see also Londono, 2015). Both U.S. and international studies, however, use monthly-sampled data.

FH: Foster and Hart (2009) introduce a measure of riskiness (FH) representing the critical wealth level above which it becomes safe for an investor to accept a particular gamble [see also Aumann and Serrano (2008)]. The measure is equivalent to a strategy that guarantees the investor will avoid bankruptcy. The existing literature on the link between Foster-Hart risk and financial returns is recent

and developing. Recent research has found that a generalized version of FH risk has predictive power for the risk-adjusted returns of individual U.S. stocks, and can be useful to improve portfolio allocation [Bali, Cakici, and Chabi-Yo (2011); Leiss and Nax (2018)] and portfolio performance (Anand, Li, Kurosaki and Kim, 2016). However, little is known about the predictive power of FH risk for index returns, how this measure relates to other option-implied variables, or how it performs internationally or out-of-sample. Our study expands on this recent literature on several fronts. First, we look at U.S. and international equity indexes rather than only U.S. stocks. Second, we study FH risk jointly with VRP and risk-neutral higher moments, in order to better assess its incremental predictive power. Finally, we perform out-of-sample forecasting tests rather than rely only on the evidence from predictive regressions.

Risk-neutral skewness and kurtosis:

The prior literature relates RNS to negative asymmetric returns in asset markets (Conrad, Dittmar and Ghysels, 2013) and is can be interpreted as "crash-phobia" (Rubinstein, 1994) or as the cost of insuring against crash risk [Brunnermeier, Nagel and Pedersen (2008); Doran, Carson and Peterson (2006)]. RNK can be interpreted as a tail risk proxy (Bakshi and Cao, 2003). Several papers document the relevance of RNS for the cross-section of expected returns [Chang, Christoffersen and Jacobs (2013); Bali and Murray (2013); Dennis and Mayhew (2002); Conrad, Dittmar and Ghysels (2013)]. Relatedly, Diavatopoulos, Doran, Fodor and Peterson (2012) show for individual U.S. equities that risk-neutral higher moments anticipate earnings announcements and thus help explain equity returns.

In contrast to cross-sectional studies, however, there is very little evidence on higher moments in predictive regressions. The existing literature focuses on U.S. data. Stilger, Kostakis and Poon (2017) show that risk-neutral skewness and kurtosis have predictive power for individual U.S. stock returns. Their results for individual equities suggest that RNS and RNK contribute distinct information. Similarly, Rehman and Vilkov (2012) find that individual U.S. stock option risk-neutral skewness has

7

predictive power for equity returns. This recent U.S. evidence of the predictive power of higher-order moments supports our thorough examination of the question, as international evidence in this area is scarce.

Our second contribution is to provide evidence of return predictability at horizons of less than one month (see e.g. Cremers and Weinbaum, 2010; Diether, Lee and Werner, 2008) as well as for longer horizons. While the prior literature has focused on a monthly frequency, we construct constant-maturity RNDs, allowing us to obtain the predictive variables at a daily frequency. Our results show evidence of return predictability under one month for several markets both using predictive regressions as well as out-of-sample tests. The existing literature has focused on long-horizon regressions because the macroeconomic variables that are typically used to predict returns have a lower sampling frequency. This literature has usually adopted a monthly frequency and has reported a peak in predictability at the quarterly horizon (Bollerslev, Tauchen and Zhou, 2009; Bollerslev et al., 2014). Shorter horizons are, however, of considerable interest, as risk management practices are often focused on short-run measures [e.g., Brown (2001); Brownlees, Engle and Kelly (2014); Christoffersen and Diebold (2000)]. Unfortunately, it is not feasible to sample traditional predictive variables at a less-than-monthly frequency. In addition, it is well known than option-implied data reflect information not found in realized asset returns (Chang, Christoffersen and Jacobs, 2013). Thus, it is more sensible to use risk-neutral distributions to construct high-frequency predictor variables. It is therefore fitting to investigate the forecasting power of VRP and our other risk variables at a higher frequency than monthly.

2. – Methodology

2.1 – EMPIRICAL RISK-NEUTRAL DISTRIBUTIONS

To obtain the explanatory variables of interest, we first have to recover the empirical risk-neutral distribution from a set of option prices. We follow the method described in Birru and Figlewski (2012)

by first calculating the empirical RND and then fitting each missing tail to a Generalized Extreme Value (GEV) distribution. This methodology is motivated by the goal of reliably measuring higher-order risk-neutral moments as well as Foster-Hart risk, which depend on the full distribution. For instance, using the Bakshi, Kapadia and Madan (2003) method (see also Britten-Jones and Neuberger, 2000) to recover the moments of the empirical risk-neutral distribution limits how one can extrapolate outside of available strike prices and can affect how the tails are modeled (Markose and Alentorn, 2011; Elliott and Timmermann, 2013). Details are provided in Appendix I.

2.2 – RISK-NEUTRAL MOMENTS AND THE VARIANCE RISK PREMIUM

Once the risk-neutral distribution $f(x)$ is computed for a given day, its central moments are given by the usual mass function formulas:

$$M_n = \int_{-\infty}^{\infty}(x - c)^n f(x)\, dx \tag{1}$$

$$Var = \frac{M_2}{T}$$

$$Skew = \frac{M_3}{M_2^{3/2}}$$

$$Kurt = \frac{M_4}{M_2^{4/2}}$$

The variance risk premium (VRP) is obtained each day by subtracting the realized variance computed from intraday returns (see section 3.1) from the risk-neutral variance of eq. (1).

2.3 – FOSTER-HART MEASURE OF RISKINESS

Foster-Hart risk provides a measure of bankruptcy risk. More generally, given a lottery represented by the random variable X, the Foster and Hart (2009) measure of riskiness ρ is the solution to this equation:

$$\mathrm{E}\left[\log\left(1 + \frac{X}{\rho}\right)\right] = 0 \tag{2}$$

9

The measure of riskiness ρ represents the critical wealth level below which the investor will reject the gamble in order to avoid the risk of bankruptcy. Bali, Cakici and Chabi-Yo (2011) generalize this measure whereby it solves the following equation:

$$\mathrm{E}\left[\frac{\left(1+\frac{X}{\rho}\right)^{\delta}-1}{\delta}\right] = 0 \tag{3}$$

where $\delta = (1 - \gamma)$ for γ the coefficient of relative risk aversion for a representative investor with CRRA utility. In a financial setting, the random variable X is the return on an index over a certain horizon. As noted by Bali, Cakici and Chabi-Yo (2011), the solution ρ of eq. (3) should be restricted to values that are greater than the maximum loss of the gamble. With a stock index, the lowest possible return is -100%. If the resulting measure of riskiness is limited by this restriction, it might not be a very informative signal as it could take exactly the same minimal value for many consecutive days. This can be prevented by using a sufficiently low value for δ. In our case, $\delta = -4$. The riskiness measure is scaled for use in the linear regressions by using the transformation: $FH_t = \ln(\rho_t)$.

2.4 – PREDICTIVE REGRESSIONS

This paper investigates the predictive power of several variables implied from options on equity indexes. We consider a linear model where excess index returns from time t to $t+h$ are explained by the following variables at time t:

$$R_{t,t+h} = b_0 + b_1 \cdot VRP_t + b_2 \cdot FH_t + b_3 \cdot RNS_t + b_4 \cdot RNK_t + u_t \tag{4}$$

where VRP is the variance risk premium defined as risk-neutral variance minus realized variance from intraday returns; RNS_t and RNK_t are respectively the third and fourth moments of the risk-neutral distribution at time t; and FH_t is the Foster-Hart measure of riskiness computed from the risk-neutral distribution. Following Bollerslev et al. (2014), we use bootstrapped critical values to evaluate standard errors. We use a simulation design based on the VAR (1) GARCH(1,1) DCC model (Engle, 2002)

suggested in Bollerslev et al. (2014) to compute bootstrapped critical values for the Newey-West t-statistics for each of the four predictive variables in the regression analysis. The bootstrap design is presented in Appendix II.

2.5 – OUT-OF-SAMPLE RETURN FORECASTS

To build additional evidence on return predictability, we conduct out-of-sample forecast tests (for a survey, see Clark and McCracken, 2013). This out-of-sample empirical strategy has been used before in the stock return predictability literature (e.g., Campbell and Thompson, 2008; Polk, Thompson and Vuolteenaho, 2006; Rapach, Strauss and Zhou, 2013), but not to investigate option-implied predictive variables. The linear model described in Equation (4) is therefore used to perform out-of-sample forecasting on the last 30% of the sample of index returns. This model is calibrated using the first 70% of the sample to make one h-day-ahead prediction, for h varying from one to 250 days. The next day, the model is recalibrated using the same, fixed length of past observations (70% of the sample), and as a result one new h-day-ahead prediction is made. This computation is repeated until the end of the sample. The prediction errors from the forecast are then compared to those of restricted specifications of model (4) to better identify the contribution of each of the different predictive variables.

Then, three measures of statistical significance are reported. First, we use a modified Diebold-Mariano (1995) test on the squared prediction errors that takes into account the effect of the overlapping periods of returns (Harvey, Leybourne, and Newbold, 1997). The null hypothesis of this test is that of equal forecast performance for two models. It is based on the series d_t of differences between the squared forecast errors (u^2) of two models at each period t. Given that this type of test is one-sided to the right, we reject the null hypothesis of equal forecasting performance if the test statistic is larger than the critical value. We use bootstrapped critical values based on recommendations from the literature regarding direct multistep predictions from nested regression models (see e.g., Clark and McCracken,

2005). The bootstrap design is similar to the one used to evaluate the predictive regressions above and is also presented in Appendix II.

The second test is also a test of forecast accuracy, but it is based on an F-type test of equal mean squared error (MSE) proposed in Clark and McCracken (2005). This test statistic is given by:

$$MSEF = n \times \frac{\bar{d}}{MSE_2} \tag{5}$$

where $\bar{d}$ is the mean of the n observations in d_t and MSE_2 is the mean squared error of the unrestricted model.

The third test considers forecast encompassing rather than accuracy. The purpose is to assess whether the information content of one set of forecasts dominates the other. Following Clark and McCracken (2005), an F-test is used which is based on the covariance between $u_{1,t}^2$ and $\left(u_{1,t}^2 - u_{2,t}^2\right)$. This test statistic, denoted ENCF, is given by:

$$ENCF = n \times \frac{\bar{c}}{MSE_2} \tag{6}$$

$$c_t = u_{1,t}^2 \times \left(u_{1,t}^2 - u_{2,t}^2\right) \tag{7}$$

$$\bar{c} = \frac{1}{n}\sum_{t=1}^{n} c_t \tag{8}$$

The covariance term c_t is positive when the additional variables in the unrestricted model 2 have predictive power. This test is also one-sided to the right. Clark and McCracken (2005) find that for these test statistics, a bootstrap approach provides more reliable results than their alternative approach, based on asymptotic critical values.[2] Thus, in our analysis, the MDM, MSEF and ENCF test statistics are compared to critical values constructed from the same bootstrapped series presented in Appendix II.

---

[2] Clark and McCracken (2005) show that asymptotic critical values are oversized at long horizons, while bootstrap critical values yield better-sized results. This difference is even greater at longer horizons, where they find that: "Once the forecast horizon increases beyond a few periods, neither a standard normal approximation nor our asymptotic distribution yields reliable inference in finite samples; bootstrap methods are much more reliable" (p.390).

The significance of this forecast improvement is assessed using an out-of-sample R_{OOS}^2. For a given test, an unrestricted forecasting model is compared to a restricted model where one or more variables are excluded. Following Campbell and Thompson (2008), we compute $R_{OOS}^2 = \left(1 - \frac{MSFE_{unrestricted}}{MSFE_{restricted}}\right)$ where MSFE is the mean squared forecast error.

Finally, a utility-based metric is used to quantify the economic significance of the out-of-sample results. As presented in Rapach and Zhou (2013), this metric represents the utility gained by a mean-variance investor with a relative risk-aversion coefficient γ who allocates his portfolio between stocks and a risk-free asset according to the forecasted equity premium.[3] This utility-based metric has been used, for example, in Rapach, Strauss and Zhou (2010) and Dangl and Halling (2012). For period t to $t+1$, the share of the portfolio invested in equities is:

$$a_{i,t} = \frac{1}{\gamma} \times \frac{\hat{r}_{i,t+1}}{\hat{\sigma}_{t+1}^2} \tag{9}$$

where $\hat{r}_{i,t+1}$ and $\hat{\sigma}_{t+1}^2$ are the return and variance forecasts according to a specified model.[4] The average utility realized by this investor over the forecasting period is given by:

$$\hat{v}_i = \hat{\mu}_i - 0.5\gamma\hat{\sigma}_i^2 \tag{10}$$

where $\hat{\mu}_i$ and $\hat{\sigma}_i^2$ are the mean and variance of the portfolio returns over the forecasting period. The difference in realized utility between the two nested models represents the utility gained by the investor from using the additional information revealed by the unrestricted model over the period studied. This utility gain can be interpreted as the management fee the investor would be willing to pay to access this additional information.

---

[3] Following Rapach, Strauss and Zhou (2010), we report results for $\gamma = 3$ and portfolio weights $a_{i,t}$ that are restricted to [0%, 150%]. We have also performed robustness checks using $\gamma = 5$, and the results concur.

[4] The return forecast is computed using Equation (5) while the variance forecast is based on the realized volatility of the past month (the same measure used in computing the VRP in section 3.1) and scaled for the appropriate horizon.

3. – Data

3.1 – REALIZED VARIANCE, INDEX RETURNS AND RISK-FREE RATES

Equity index excess returns are computed including dividends using data from Bloomberg. Country risk-free rates index are obtained from the FRED website of the Federal Reserve Bank of St. Louis. Realized variances computed as the sum of the 5-minute realized variance and squared overnight returns of the past month for each index are obtained from the Oxford-Man Institute's realized library's website (Gerd, Lunde, Shephard and Sheppard, 2009).

3.2 – OPTION DATA

Index option data are obtained from OptionMetrics Ivy DB U.S. and Europe. The selected indexes are those with the most options available among the markets covered by this database. They are the S&P500 index for the United States, the DAX index for Germany, the SMI index for Switzerland, the CAC index for France, and the FTSE index for the United Kingdom. These markets are significant for the Eurozone, and have active index options markets. Strike prices, maturities, and implied volatilities are extracted at a daily frequency for all available options on the selected indexes. Table I presents descriptive statistics for the raw data on options.

3.3 – DESCRIPTIVE STATISTICS OF THE TIME SERIES

Table II presents the descriptive statistics for the time series of predictive variables computed from the risk-neutral distribution. Judging from their low correlation coefficients, the selected variables represent different components of the risk-neutral distribution.[5] For all indexes, VRP and FH increase substantially during the financial crisis in early 2009.

---

[5] Since the correlation between RNS and RNK for SMI is .68, we also ran regressions containing only one of these variables. Results are stable across the sample and horizons and largely similar to the ones reported, and thus are omitted for brevity.

4. – Results

4.1 – PREDICTIVE REGRESSIONS OF FUTURE INDEX RETURNS

Future returns for all indexes at different horizons h (in business days) are regressed on the explanatory risk-neutral variables as presented in eq. (4). Figure 1 shows the results of two-sided individual significance tests for different return horizons h. The results are shown as Newey-West corrected t-statistics computed from 10,000 bootstrapped series. In the discussion below, we use the 10% level of significance, which is customary in this literature (e.g., Bollerslev et al., 2014; Rapach and Zhou, 2013), but other levels are reported in the figure and further confirm our interpretation of the findings.

4.1.a – Assessing the Evidence of Predictability using the Adjusted R^2

We first want to assess the performance of risk-neutral variables in predictive regressions. Figure 2 (panels A through E) presents graphically the regression R^2s against different horizons, separately for each index. The pattern is similar across indexes, with the R^2 being lowest at the shortest horizons and increasing about monotonically with the horizon. As shown by Kirby (1997), regressions using overlapping return data can generate R^2s that increase mechanically with the horizon even if the variables have no predictive power. To address this, we compute and report graphically the percentiles of R^2 for regressions on the bootstrapped time series under the null of no informative content. Thus, significance is achieved when the regression's actual R^2 is greater than the simulated R^2 for a given horizon.

Nearly in all cases, the R^2 exceeds the simulated R^2. For the S&P 500 index, R^2 increases with the horizon from about 1% to 18%, peaking at 250 days, and always exceeds the simulated R^2. For DAX, R^2 increases from less than 1% to about 9%, plateauing around 60 days (3 months), and is significant for all horizons above 25 days. For SMI, R^2 increases from about 1% to 10%, peaking at about 60 days, and is always significant. For CAC, R^2 increases from about 1% to 16% peaking at about

60 days and is always significant. For FTSE, R^2 increases from less than 1% to about 12% reaching a plateau around 80 days (4 months), and is significant for horizons above 12 days. These results are economically significant, as they compare favorably to the R^2 for different predictive variables reported by Bollerslev, Tauchen and Zhou (2009) which range from 0-4% at a monthly horizon. Since the model is significant for all the countries considered, we now assess the marginal contribution of each variable in each case.

4.1.b – Variance Risk Premium

Using monthly returns, Bollerslev et al. (2014) find a peak of predictability for VRP at a quarterly horizon. Using our daily sampled series, we show that VRP also has predictive power for shorter horizons (under one month) and longer horizons (up to 6 months). For the S&P index, we find predictability for horizons of about 3-100 days, that is, up to 5 months. For SMI, it is significant at horizons of 1-35 days and 50-100 days. For CAC, it is significant over horizons of 10-125 days. For FTSE, we find predictability for 15-20 days and 75-125 days. Finally for DAX, it is significant for 50-90 days, which is roughly the quarterly horizon. Thus, VRP is relevant and positively related to future returns, in line with Bollerslev et al. (2014). We corroborate their finding of predictability at the quarterly horizon and also document new evidence of short-horizon predictability.

4.1.c – Foster-Hart Risk

Foster-Hart riskiness serves to capture the likelihood of large negative returns and the risk of bankruptcy. Financial theory suggests there should be a positive relation between FH and future returns, as the investor requires higher expected returns when his investment is at greater risk of large losses. Our results present the first international evidence related to FH risk and the first evidence of FH risk performance in predictive regressions. The results suggest that FH risk is relevant to predict future returns internationally. Our regressions show that FH offers return predictability at the international level

with a positive slope, as expected. FH is typically significant for horizons of 20 to 120 business days (i.e., 1 to 6 months), with some variation between indexes. This return predictability extends to 200 days for SMI and to 250 days for the S&P. Based on these results, we argue that the longer-horizon predictability of FH is complementary to information contained in VRP.

4.1.d – Risk-neutral Skewness

Risk-neutral skewness is significantly and positively related to future returns for all indexes except DAX, for which it is not significant. It is significant at short horizons of less than one week (S&P, FTSE and CAC), one month (S&P, FTSE and CAC), and longer than 50 days (FTSE, CAC) or 100 days (S&P, SMI). Although there is not a clear pattern of predictability, the variable is useful at several different horizons for all indexes but one.

   The positive coefficient on risk-neutral skewness means that when its level is higher, buying the index leads to higher future returns. Although this result differs from the cross-sectional evidence on realized skewness or risk-neutral skewness [e.g., Amaya, Christoffersen, Jacobs and Vasquez (2015); Conrad, Dittmar and Ghysels (2013)], it is consistent with the evidence for individual stocks in Stilger, Kostakis and Poon (2017). They find that a quintile portfolio of low skewness stocks significantly underperforms the quintile portfolio of high skewness stocks. According to their analysis, this underperformance comes from the highly negative risk-neutral skewness of stocks that are "overpriced but hard to sell short" according to investors. A similar positive relation between risk-neutral skewness and future returns of individual stocks is found by Rehman and Vilkov (2012). While these papers confirm our finding in other settings, the advantage of our skewness measure relies in its construction from the complete risk-neutral distribution using GEV tails, preventing the over-importance of the center of the distribution in the computations.

4.1.e – Risk-neutral Kurtosis

Risk-neutral kurtosis is usually interpreted as risk-adjusted "tail risk", so we expect a higher measure of tail risk today to lead to higher future returns. The positive relation between kurtosis and expected returns in the cross-section is documented in Conrad, Dittmar and Ghysels (2013) and Amaya et al. (2015). This study confirms the result, as the kurtosis slope coefficient is positive when significant. We find a significant and positive relation for S&P, CAC and FTSE at horizons above 20 days, and for all horizons in the case of the S&P. Kurtosis is not significant for DAX or SMI.[6] To our knowledge, these are the first findings of index return predictability using risk-neutral kurtosis and in particular, the first evidence of return predictability under one month using kurtosis (for S&P500).

4.2 – OUT-OF-SAMPLE FORECAST ANALYSIS

We now examine the out-of-sample performance of forecasting returns using different variations of the linear model in (4). Our objectives are threefold: 1) to build evidence regarding the predictive power of our option-implied variables with out-of-sample results and to provide a more thorough evaluation of option-implied information than what is typically reported, 2) to explore the forecasting power of each variable separately, and 3) to benchmark our results against a traditional constant model (i.e., forecasting with the historical mean). This is done using the three test statistics (MDM, MSEF, ENCF). In each case, a significantly positive test statistic implies that the unrestricted model performs significantly better than does the restricted model and thus, that the additional variables have forecasting power over future returns. For each index, the first 70% of the time series observations are used to construct a rolling window and to calibrate each day a new forecasting model. The predictions made on the last 30% of the data are used to evaluate its performance.

---

[6] For SMI it is significant and negative for only two specific longer horizons.

The R_{oos}^2 for the return forecasts can be interpreted as the fraction of the benchmark model's forecast error that can be explained by the unrestricted model's additional variable(s). For a given horizon, we compare the R_{oos}^2 for different models (Campbell and Thompson, 2008). This measure indicates the economic importance of each explanatory variable for a given return horizon. The results are provided in table III, panels A through E, for each of the five equity indexes. In each panel, each row reports a different case corresponding to a comparison between two models. As such, rows 1-4 in each panel examine the contribution of each individual variable against the benchmark of using only a constant. Rows 6-9 report the marginal contribution of a given variable by comparing the full model against the same model omitting one variable. Row 5 compares the full model against the benchmark.

4.2.a – Variance Risk Premium

Row 1 shows that VRP significantly improves on the constant-only model across most indexes for short return horizons of 1 to 5 business days. At a horizon of 5 business days, the model including VRP is significantly better according to the three test statistics considered for several indexes (S&P500, DAX, CAC), and for two of three test statistics for FTSE. VRP is particularly informative for the S&P500, where it improves on the base model for all horizons. Internationally, we find predictive power at the longest horizon of 250 days for the DAX and FTSE.

Row 6 provides an alternative assessment that VRP is not a redundant variable, by comparing the full model to the full model omitting VRP. Moreover, this novel evidence of short-run predictability suggests once more that the "fear gauge" interpretation of VRP is relevant at shorter horizons than previously understood. Indeed, in the out-of-sample forecasts across indexes, the informative content of VRP is generally significant for horizons under 1 month. This finding represents new evidence obtained from daily frequency data that provides additional support for the relevance of VRP in international equity markets as well as in the U.S.

4.2.b – Foster-Hart Risk

The contribution of Foster-Hart risk is assessed in rows 2 and 7. Row 2 shows that out-of-sample, FH risk is more important for longer return horizons of 20 to 250 business days (S&P500, DAX and CAC). The evidence is strongest for the S&P500 where FH is significant according to all test statistics for horizons of 20 to 250 business days. For the SMI, FH improves upon forecasts at an annual horizon. FH is not meaningfully significant out-of-sample for FTSE. Overall, we observe that FH has a higher explanatory power (R^2_{oos}) for horizons above 20 business days, while VRP is more important for horizons under 20 days. Row 7 confirms that those results hold when the marginal contribution of FH is investigated with respect to the other option-implied variables.

4.2.c – Risk-neutral Skewness

Rows 3 and 8 of each panel document the contribution of risk-neutral skewness to predicting out-of-sample index returns. For the S&P index, row 3 shows a significant improvement at the 1- and 2-day (MDM), 20-day (MDM and MSEF) and 250-day-ahead horizons (MDM, ENCF and MSEF). However, the results do not remain significant in row 8, when the marginal contribution of skewness is assessed. For CAC, row 3 reports significance for horizons of 5-100 days (MDM, ENCF and MSEF) and 100-250 days (ENCF and MSEF). When compared against the other option-implied variables (row 8), the results are broadly similar, with significance for 5-100 days (MDM) or 50-250 days (ENCF and MSEF). For DAX, skewness is significant against the benchmark for horizons of 2-5 days (MDM) and 5-20 days (MDM, ENCF and MSEF). It remains significant for horizons of 1 day (MDM) and 20 days (MDM, ENCF and MSEF) when all variables are included. Skewness is not significant for SMI or FTSE.

4.2.d – Risk-neutral Kurtosis

Evidence for the informative content of risk-neutral kurtosis in out-of-sample forecasts is presented in rows 4 and 9. Overall, the evidence is weaker than for the other variables. For CAC, row 9 reports that

kurtosis is significant at the 1- and 2-day (MDM and MSEF) and 100-250 day horizons (ENCF and MSEF). Moreover, for CAC the 250-day R^2_{oos} is 4.2%. For FTSE, it is significant at the 100-day horizon (ENCF and MSEF), with an R^2_{oos} of 1.4%. However, kurtosis is not significant for S&P, DAX or SMI.

4.2.e – Economic Significance

Table III also reports annualized utility gains and shows that nearly all statistically significant results are also economically significant. For example, at a horizon of 20 business days on the S&P500, a mean-variance investor with relative risk-aversion of 3 would be willing to pay an annual management fee of up to 2.5% to have access to the information represented by the four studied variables.[7]

Indeed, as observed by Rapach and Zhou (2013), even when statistical tests fail to detect out-of-sample improvements, a utility-based metric can indicate clear economic significance. This is the case for skewness (row 8 in each panel), which is economically significant at some horizons for most indexes. There is also evidence that kurtosis (row 9) is economically significant for S&P500 and CAC. Moreover, although the statistical evidence for Foster-Hart risk is weaker for the SMI and FTSE indexes, this variable is economically significant for all indexes (rows 2 and 7). Finally, the complementarity of VRP and FH is also found in the utility gains, where for all indexes VRP has greater economic significance at horizons of one month or less, while FH is more important at longer horizons. Internationally, this analysis of economic significance confirms the relevance of risk-neutral variables beyond the U.S case as well as heterogeneity in performance in terms of horizons. Our results underscore the importance of investigating model performance for specific countries and horizons of investment.

5. – Conclusion

Option data provide a forward-looking view of market risk anticipations (Christoffersen, Jacobs and

---

[7] The reported results on utility gains are consistent with the findings of Rapach and Zhou (2013) for the out-of-sample predictability on monthly excess returns of the S&P500 index. They find annualized utility gains of 0.24% to 2.47% for various models combining commonly used economic variables.

Chang, 2013; Cremers and Weinbaum, 2010). These data allow for daily estimations of forward-looking measures of risk, which we exploit in this paper. We show that these risk measures translate into significant index return predictability and out-of-sample performance. Our results extend previous evidence which ascribed a predictive role mainly to the U.S. S&P index (e.g. Pástor and Stambaugh, 2009; Rapach, Strauss, and Zhou, 2013). Indeed, we show that each market's own index options contain significant predictive information.

We have three main findings. First, risk-neutral metrics matter beyond the variance risk premium. In fact, we find predictability at the international level for several horizons for all metrics considered (VRP, FH, RNS and RNK). Foster-Hart risk and risk-neutral skewness and kurtosis each contribute distinct information that is relevant for expected returns. In particular, FH risk explains future returns at longer horizons than does VRP. Out of sample, we find that a combination of VRP and FH risk fares well in several countries. The out-of-sample performance of the option-implied variables further underscores their economic significance.

Second, we document predictability for several horizons that are important for asset managers and which were previously undocumented. In particular, VRP has predictive power at horizons less than one month as well as longer-run horizons, in addition to the previously reported quarterly horizon. Using the theoretical model in Bollerslev, Tauchen and Zhou (2009), this result suggests that investors care about the "volatility-of-volatility" at fairly short horizons. We also find that FH risk contributes more to return predictability at longer horizons. Under typical market conditions, it is intuitive that the risk of bankruptcy (large negative returns) would be associated with more distant forecasts and perceived as highly unlikely in the very short run.

Third, the international evidence presented for several metrics is novel to the literature. We show that option markets for the DAX, FTSE, CAC and SMI indexes contain significant predictive power not

only as captured by VRP, but also in other option-implied variables. Our main results indeed hold for these international indexes, and also show cross-country variation in the performance of the models. Thus, finding relevant predictors internationally for stock returns is as challenging as ever, especially in the out-of-sample context.

Appendix I: Obtaining the Risk-Neutral Distribution

The first step toward obtaining moments of the RND is to construct a continuous volatility surface. Standard to the empirical option literature, options are excluded if they expire in five days or less or if they have implied volatility that is negative or above 100%. Only out-of-the-money options are used. For all remaining options, implied volatilities are recovered from the data source (OptionMetrics). This step provides, for each day, a surface of points in implied volatility-maturity-strike space. For a chosen maturity, the next steps are to convert the corresponding IVs to call prices using Black-Scholes, fit the points to a continuous form such as a spline, and apply Breeden and Litzenberger's (1978) formula to obtain the risk-neutral distribution for a given date and maturity.

Our objective is to construct time series of option-implied risk variables at a daily frequency, which requires a constant maturity for each daily observation. However, actual option prices have maturities that change each day. For this reason, the set of available points is transformed into a continuous surface. This is done using cubic spline interpolation, where each node is an actual observation. Afterward, a "slice" of this surface is taken at a maturity of one month and is corrected to avoid arbitrage possibilities, following Aït-Sahalia and Duarte (2003). This step corrects possible pricing errors introduced by interpolating without having to make strong parametric assumptions about the IV surface.

Following Birru and Figlewski (2012), a fourth-degree smoothing spline is applied to the arbitrage-free interpolated IVs for a given maturity in order to prevent sharp spikes in the fitted density.

Once the smoothed IV curve is converted back to a call option price function using Black-Scholes, the empirical risk-neutral distribution function is obtained from Breeden and Litzenberger (1978). Then, tails are fitted using a Generalized Extreme Value (GEV) distribution to each missing tail (Birru and Figlewski, 2012). This assumption is more appropriate than is assuming normality and ensures that, for example, risk-neutral variables computed on dates when the RND is missing a considerable part of its lower tail are closer to those computed on dates when these data are available. The Generalized Extreme Value (GEV) cumulative distribution function used for each missing tail is given by:

$$F(X) = \exp\left[-\left(1 + \xi\frac{X-\mu}{\sigma}\right)^{-\frac{1}{\xi}}\right] \tag{11}$$

For each tail, the three parameters (μ, σ, ξ) of the GEV distribution are fitted such that the total probability in the fitted tail is equal to the missing total probability in this tail of the empirical risk-neutral distribution, and must connect with the empirical RND at the 2nd and 5th or 95th and 98th percentiles. The objective function is to minimize the sum of the squared distances between the empirical RND and the GEV distribution on the domain between each pair of connection points. The GEV approach should then generate risk-neutral variables that better reflect the information contained in the options near the tails. It is documented in the literature that the normality assumption underestimates the prices of options in the distributional tails (Birru and Figlewski, 2012; Markose and Alenthorn, 2011). Moreover, this method addresses the bias in risk-neutral skewness discussed in Dennis and Mayhew (2002) linked to using an incomplete distribution.

Appendix II: Bootstrap Procedure

This bootstrap algorithm is based on the out-of-sample methodology presented in Clark and McCracken (2005) and Bollerslev et al. (2014) for possible GARCH effects. A 5-dimensional vector autoregressive (VAR) model is built with equations for the daily returns Y as well as the four explanatory variables X

(VRP, FH, RNS, RNK). Each possible combination of lags from 0 to 3 for each variable is estimated. The model is chosen for each country according to the Bayesian Information Criterion. For example, for the S&P500, the best model has two lags in the VRP equation, three for FH, three for RNS and two for RNK. The appropriate model for each index is further simplified by dropping the terms that are not significant. For parsimony, coefficients that have absolute values inferior to 1.6 times their standard errors are dropped. This is roughly equivalent to a 10% significance level. This model is estimated using the full sample of observations and the residuals are stored in order to construct the simulated series. The model used to generate the 1000 simulated series is restricted to impose the null hypothesis that variable X has no predictive power over future values of Y. In addition, we account for possible GARCH effects in the VRP and in the returns (see Bollerslev et al., 2014) by using a GARCH-DCC model to generate the simulated residuals. We use GARCH-DCC code provided by Sheppard (2013) to implement this part. The VRP and the index returns thus follow a GARCH(1,1). For return horizons longer than one day, the overlapping returns time series are constructed from this same simulated series of daily returns. This is done to ensure that the bootstrapped series have the same overlapping nature as do the actual observations. An alternative bootstrap technique was also implemented, with similar results. The alternative design is an unconstrained VAR, as developed in Clark and McCracken (2005). Given the available evidence regarding the VRP found in Bollerslev et al. (2014), for the sake of brevity we report results for the GARCH-DCC design only.

6. – References

Aït-Sahalia, Y. and Duarte, J. (2003). "Nonparametric Option Pricing under Shape Restrictions." Journal of Econometrics, 116, 9-47.

Anand, A., Li, T., Kurosaki, T., and Kim, Y. S. (2016). "Foster-Hart Optimal Portfolios." Journal of Banking and Finance, 68, 117–130.

Amaya, D., Christoffersen, P., Jacobs, K., and Vasquez, A. (2015). "Does Realized Skewness Predict the Cross-Section of Equity Returns?" Journal of Financial Economics, 118(1), 135-167.

Aumann, R. J., and Serrano, R. (2008). An Economic Index of Riskiness. Journal of Political Economy, 116(5), 810-836.

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

Bakshi, G., and Cao, C. (2003). "Risk-Neutral Kurtosis, Jumps, and Option Pricing: Evidence from 100 Most Actively Traded Firms on the CBOE." Manuscript, University of Maryland.

Bali, T. G., Cakici, N., and Chabi-Yo, F. (2011). "A Generalized Measure of Riskiness." Management Science, 57(8), 1406–1423.

Bali, T. G., and Zhou, H. (2016). "Risk, Uncertainty, and Expected Returns." Journal of Financial and Quantitative Analysis, 51(3), 707-735.

Bekaert, G. and Hoerova, M. (2014). "The VIX, the Variance Premium and Stock Market Volatility." Journal of Econometrics, 183, 181-192.

Birru, J. and Figlewski, S. (2012). "Anatomy of a Meltdown: The Risk-Neutral Density for the S&P 500 in the Fall of 2008." Journal of Financial Markets 15, 2, 151-180.

Bollerslev, T., Marrone, J., Xu, L., and Zhou, H. (2014). "Stock Return Predictability and Variance Risk Premia: Statistical Inference and International Evidence." Journal of Financial and Quantitative Analysis, 49(3), 633–661.

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

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

Brown, G. (2001). "Managing Foreign Exchange Risk with Derivatives." Journal of Financial
Economics, 60, 401–448.

Brunnermeier, M. K., Nagel, S., and Pedersen, L. H. (2008). "Carry Trades and Currency Crashes."
NBER Macroeconomics Annual, 23(1), 313-348.

Campbell, J. Y., and Thompson, S. B. (2007). "Predicting Excess Stock Returns Out of Sample: Can
Anything Beat the Historical Average?" Review of Financial Studies, 21(4), 1509-1531.

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

Carr, P., and Wu, L. (2016). "Analyzing Volatility Risk and Risk Premium in Option Contracts: A New
Theory." Journal of Financial Economics, 120(1), 1-20.

Chang, B. Y., Christoffersen, P., and Jacobs, K. (2013). "Market Skewness Risk and the Cross Section of
Stock Returns." Journal of Financial Economics, 107(1), 46-68.

Chen, L. (2009). "On the Reversal of Return and Dividend Growth Predictability: A Tale of Two
Periods." Journal of Financial Economics, 92(1), 128-151.

Christoffersen, P. F., and Diebold, F. X. (2000). "How Relevant Is Volatility Forecasting for Financial
Risk Management?" Review of Economics and Statistics, 82(1), 12-22.

Christoffersen, P., Jacobs, K., and Chang, B. Y. (2013). "Forecasting with Option-Implied Information."
Chap. 10 in: Elliott, G., and Timmermann, A., ed., Handbook of Economic Forecasting, vol.2 (part
A), 581-656.

Clark, T. E., and McCracken, M.W. (2005). "Evaluating Direct Multistep Forecasts." Econometric
Reviews, 24(4), 369–404.

Clark, T.E. and McCracken, M.W. (2013). "Advances in Forecast Evaluation." Chap. 20 in: Elliott, G.,
and Timmermann, A., ed., Handbook of Economic Forecasting, vol. 2B, 1107-1201.

Conrad, J., Dittmar, R., and Ghysels, E. (2013). "Ex-ante Skewness and Expected Stock Returns."
Journal of Finance, 68, 1, 85-124.

Constantinides, G. M., Jackwerth, J. C., and Perrakis, S. (2007). "Option Pricing: Real and Risk-Neutral
Distributions." In: Birge, J.R. and Linetsky, V., ed., Handbooks in Operations Research and
Management Science, vol. 15, 565-591.

Cremers, M., and Weinbaum, D. (2010). "Deviations from Put-Call Parity and Stock Return
Predictability." Journal of Financial and Quantitative Analysis, 45(2), 335-367.

Dangl, T., and Halling, M. (2012). "Predictive Regressions with Time-Varying Coefficients." Journal of
Financial Economics, 106(1), 157–181.

Diavatopoulos, D., Doran, J.S., Fodor, A., and Peterson, D.R. (2012). "The Information Content of
Implied Skewness and Kurtosis Changes Prior to Earnings Announcements for Stock and Option
Returns." Journal of Banking & Finance, 36(3), 786-802.

Diebold, F.X., and Mariano, R.S. (1995). "Comparing Predictive Accuracy." Journal of Business and
Economic Statistics, 13, 253–265.

Diether, K. B., Lee, K. H., and Werner, I. M. (2008). "Short-Sale Strategies and Return Predictability."
Review of Financial Studies, 22(2), 575-607.

Doran, J., Carson, J.M., and Peterson, D.R. (2006). "Market Crash Risk and Implied Volatility
Skewness: Evidence and Implications for Insurer Investments." Working paper available at SSRN no.
897401.

Easley, D., O'Hara, M., and Srinivas, P. S. (1998). "Option Volume and Stock Prices: Evidence on
Where Informed Traders Trade." Journal of Finance, 53(2), 431-465.

Elliott, G., and Timmermann, A. (2013). Handbook of Economic Forecasting, volume 2. NY: Elsevier.

Engle, R.F. (2002). "Dynamic Conditional Correlation: A Simple Class of Multivariate GARCH Models." Journal of Business and Economic Statistics, 20, 339–350.

Foster, D. P., and Hart, S. (2009). "An Operational Measure of Riskiness." Journal of Political Economy, 117(5), 785–814.

Gerd, H., Lunde, A., Shephard, N. and Sheppard, K. (2009). Oxford-Man Institute's realized library v. 0.2, Oxford-Man Institute, University of Oxford.

Giacomini, R., and White, H. (2006). "Tests of Conditional Predictive Ability." Econometrica, 74(6), 1545-1578.

Goyal, A., and Welch, I. (2007). "A Comprehensive Look at the Empirical Performance of Equity Premium Prediction." Review of Financial Studies, 21(4), 1455-1508.

Harvey, D., Leybourne, S., and Newbold, P. (1997). "Testing the Equality of Prediction Mean Squared Errors." International Journal of Forecasting, 13(2), 281–291.

Jackwerth, J. C., and Rubinstein, M. (1996). "Recovering Probability Distributions from Option Prices." Journal of Finance, 51(5), 1611-1631.

Kirby, C. (1997). "Measuring the Predictable Variation in Stock and Bond Returns." Review of Financial Studies, 10(3), 579–630.

Leiss, M., and Nax, H. H. (2018). "Option-Implied Objective Measures of Market Risk." Journal of Banking & Finance, 88, 241-249.

Londono, J. (2015). "The Variance Risk Premium around the World." Working Paper, Federal Reserve Board, Washington, DC and SSRN paper 2517020.

Ludvigson, S. C., and Ng, S. (2007). "The Empirical Risk–Return Relation: A Factor Analysis Approach." Journal of Financial Economics, 83(1), 171-222.

Maio, P., and Santa-Clara, P. (2015). "Dividend Yields, Dividend Growth, and Return Predictability in the Cross Section of Stocks." Journal of Financial and Quantitative Analysis, 50(1-2), 33-60.

Markose, S., and Alentorn, A. (2011). "The Generalized Extreme Value Distribution, Implied Tail Index, and Option Pricing." Journal of Derivatives, 18(3), 35-60.

Pástor, Ľ., and Stambaugh, R. F. (2009). "Predictive Systems: Living with Imperfect Predictors." Journal of Finance, 64(4), 1583-1628.

Polk, C., Thompson, S., and Vuolteenaho, T. (2006). "Cross-sectional Forecasts of the Equity Premium." Journal of Financial Economics, 81(1), 101-141.

Rapach, D. E., Strauss, J. K., and Zhou, G. (2010). "Out-Of-Sample Equity Premium Prediction: Combination Forecasts and Links to the Real Economy." Review of Financial Studies, 23(2), 821–862.

Rapach, D. E., Strauss, J. K., and Zhou, G. (2013). "International Stock Return Predictability: What is the Role of the United States?" Journal of Finance, 68(4), 1633-1662.

Rapach, D. E., and Zhou, G. (2013). "Forecasting Stock Returns." In: Elliott, G., and Timmermann, A., ed., Handbook of Economic Forecasting, vol. 2(Part A), 328-383.

Rehman, Z., and Vilkov, G. (2012). "Risk-Neutral Skewness: Return Predictability and Its Sources." Available at SSRN: http://doi.org/10.2139/ssrn.1301648

Rubinstein, M. (1994). "Implied Binomial Trees." Journal of Finance, 49(3), 771-818.

Sheppard, K. (2013). "MFE Toolbox." University of Oxford. Available at: https://www.kevinsheppard.com/MFE_Toolbox#Last_Updated

Stilger, P.S., Kostakis, A., and Poon, S.-H. (2016). "What Does Risk-Neutral Skewness Tell Us About Future Stock Returns?" Management Science, 63,6: 1814-1834.

Table I: Descriptive statistics for the raw option data, January 2, 2000 -December 31, 2013.

	S&P	DAX	CAC	FTSE	SMI
Dates	2000/01/02 to 2013/12/31 3356 obs.	2002/01/02 to 2013/12/31 2906 obs.	2003/04/14 to 2013/12/31 2523 obs.	2002/01/02 to 2013/12/31 2848 obs.	2002/01/02 to 2013/12/31 2872 obs.
Number of options per day	546.71	627.76	362.68	365.44	452.51
Implied volatility	.26	.28	.24	.23	.22
Strike prices	1172.42	5447.45	4044.97	5140.31	6388.85
Number of strikes per day	118.16	103.54	57.70	77.98	83.87
Time to maturity (days)	202.59	358.45	449.53	233.18	335.56
Number of maturities per day	10.98	13.59	13.26	9.85	11.74

This table reports, for each equity index, the mean of each of the variables relating to characteristics of the raw option data. The source of these data is Optionmetrics Ivy DB USA and Europe.

Table II: Descriptive statistics for the time series of option-implied moments

	SPX				DAX				SMI			
	VRP	FH	RNS	RNK	VRP	FH	RNS	RNK	VRP	FH	RNS	RNK
Mean	.01	1.64	-1.05	11.73	.00	1.82	-.71	7.75	.00	1.89	-0.28	13.40
Std. Dev.	.03	1.10	.96	11.67	.05	.98	.80	16.06	.03	1.07	1.26	35.69
Autocorr(1)	.85	.91	.46	.38	.93	.87	.34	.18	.91	.74	.41	.11
P.Perron(10)	.00	.00	.00	.00	.00	.00	.00	.00	.00	.00	.00	.00
Correlations												
VRP	1	.08	-.05	.06	1	.01	-.03	.02	1	.00	.10	.05
FH		1	.31	-.22		1	.03	-.11		1	.03	-.03
SK			1	-.51			1	.45			1	.68
Kurt				1				1				1

	CAC				FTSE			
	VRP	FH	RNS	RNK	VRP	FH	RNS	RNK
Mean	.00	1.62	-0.74	10.27	.01	1.19	-1.10	14.98
Std. Dev.	.04	.99	.99	13.09	.03	1.16	1.61	23.51
Autocorr(1)	.92	.81	.26	.34	.90	.88	.19	0.32
P.Perron(10)	.00	.00	.00	.00	.00	.00	.00	.00
Correlations								
VRP	1	.08	.00	.09	1	.02	-.11	.14
FH		1	.12	-.21		1	.21	-.29
SK			1	-.19			1	-.46
Kurt				1				1

This table reports descriptive statistics for the daily time series of risk variables obtained from the risk-neutral distribution implied by options data on each of the equity indexes studied in the paper. VRP is the variance risk premium defined as the difference between model-free risk-neutral variance and realized variance. FH is the Foster-Hart generalized measure of riskiness taken on the risk-neutral distribution each day in the sample. RNS and RNK are the skewness and kurtosis of the risk-neutral distribution using GEV tails. *Autocorr(1)* is the autocorrelation of the series at lag 1 and *P.Perron(10) p-value* is the p-value of a Phillips–Perron unit-root test with 10 lags.

Table III: This table (panels A through E) reports test statistics for the out-of-sample index return forecasting performance of unrestricted models (first column) compared to a (nested) restricted model (second column). VRP is variance risk premium; FH is Foster-Hart risk; RNS is risk-neutral skewness; and RNK is risk-neutral kurtosis. MDM is the modified Diebold-Mariano test statistic on the squared prediction error. MSEF is a test statistic of forecast accuracy. ENCF is a test statistic of forecast encompassing. MDM, MSEF and ENCF are compared to 1000 bootstrapped values. Test statistics that are positive indicate that the unrestricted model performs better than does the restricted version. When the test statistic is significantly positive, the asterisks (*) correspond to the usual significance levels (*:10%, **:5%, ***:1%). When the unrestricted model underperforms the restricted model, R^2_{oos} is negative and omitted from the table. UG is the average annualized utility gain for a mean-variance investor with a relative risk aversion of $\gamma = 3$ using the unrestricted model, compared to using the restricted model. UG is omitted when it is negative.

				Return horizon (in business days)						
Table III: Out-of-Sample Performance										
Panel A: S&P 500										
No.	Unrest.	Rest.	Test Stat.	1	2	5	20	50	100	250
1	VRP	CST only	MDM	1.69**	1.76**	2.31***	2.12***	1.66**	1.45*	1.14
			ENCF	10.84***	12.93***	19.67***	34.86***	46.19***	38.04**	33.79**
			MSEF	16.30***	19.16***	30.38***	55.12***	67.30***	60.31***	59.44***
			R^2_{oos}	1.6%	1.9%	2.9%	5.2%	6.4%	5.8%	6.0%
			UG	0.6%	1.3%	1.6%	1.0%	1.1%	0.3%	0.4%
2	FH	CST only	MDM	-1.81	-1.99	0.95*	2.31***	2.28***	1.10*	4.90***
			ENCF	-1.01	-1.71	1.24	21.91**	48.18**	57.34**	210.77***
			MSEF	-2.25	-3.65	2.19	38.26**	84.33**	88.19**	355.35***
			R^2_{oos}	-	-	0.2%	3.7%	7.8%	8.3%	27.6%
			UG	-	-	0.6%	1.9%	1.6%	1.4%	1.6%
3	RNS	CST only	MDM	1.51**	1.67**	-1.79	2.29***	-1.23	0.02	1.41*
			ENCF	0.38	0.50	-0.31	3.22	-2.57	0.31	30.18***
			MSEF	0.68	0.93	-0.64	6.03*	-5.24	0.22	57.96***
			R^2_{oos}	0.1%	0.1%	-	0.6%	-	0.0%	5.9%
			UG	0.5%	0.3%	-	0.3%	-	0.2%	0.5%
4	RNK	CST only	MDM	-1.86	-1.68	-1.67	-3.50	-2.49	-1.75	-1.39
			ENCF	-0.61	-0.81	-1.46	-4.61	-9.24	-15.15	-12.42
			MSEF	-1.37	-1.83	-3.31	-9.86	-19.47	-31.27	-25.46
			R^2_{oos}	-	-	-	-	-	-	-
			UG	-	-	-	-	-	-	-
5	All	CST only	MDM	1.42**	1.29**	2.04***	2.20***	1.90**	1.56**	2.77**
			ENCF	9.94***	10.81***	18.86***	53.39***	83.98***	84.89**	242.55***
			MSEF	13.17***	13.61***	27.15***	77.91***	121.50***	117.29**	392.02***
			R^2	1.3%	1.3%	2.6%	7.2%	10.9%	10.7%	29.6%
			UG	2.1%	1.6%	1.6%	2.5%	1.9%	1.3%	1.8%
6	All	All ex. VRP	MDM	1.76**	1.77***	2.15***	1.77**	1.23*	1.00	1.01
			ENCF	11.05***	12.52***	17.55***	28.01***	32.93**	27.08**	17.44*
			MSEF	16.83***	18.74***	26.80***	43.02***	43.95**	41.32**	29.09*
			R^2_{oos}	1.6%	1.8%	2.6%	4.1%	4.2%	4.1%	3.0%
			UG	1.8%	1.6%	1.1%	0.8%	0.4%	-	0.1%
7	All	All ex. FH	MDM	-0.95	-1.28	-1.27	1.59**	2.16***	1.09*	9.78***
			ENCF	-0.69	-1.90	-0.74	11.13*	35.38**	46.14**	168.47***
			MSEF	-2.33	-4.90	-1.54	19.35*	62.46**	71.78**	277.50***
			R^2_{oos}	-	-	-	1.9%	5.9%	6.8%	22.9%
			UG	0.4%	-	-	1.0%	0.8%	1.0%	1.3%
8	All	All ex. RNS	MDM	-0.03	-0.02	-0.07	0.35	-0.98	-0.62	0.04
			ENCF	0.76	0.84	0.23	2.50	-1.80	-2.74	5.99**
			MSEF	-0.08	-0.07	-0.15	1.90	-4.72	-8.28	1.36
			R^2_{oos}	-	-	-	0.2%	-	-	0.1%
			UG	1.5%	0.6%	0.3%	0.1%	-	-	0.0%
9	All	All ex. RNK	MDM	-0.94	-0.75	-0.45	-0.55	-0.35	-0.32	-0.50
			ENCF	-0.33	-0.26	-0.08	0.23	0.49	0.00	1.98
			MSEF	-1.38	-1.37	-1.27	-4.28	-4.03	-8.26	-11.48
			R^2_{oos}	-	-	-	-	-	-	-
			UG	0.5%	0.1%	0.1%	-	0.1%	0.1%	0.0%

Table III: Out-of-Sample Performance
Panel B: DAX

No.	Unrest.	Rest.	Test Stat.	Return horizon (in business days)						
				1	2	5	20	50	100	250
1	VRP	CST only	MDM	1.15**	1.39**	1.33**	0.11	-1.33	-0.76	3.39**
			ENCF	1.74**	4.41**	5.94**	2.00	-8.33	-8.99	2.02
			MSEF	2.92**	7.63***	10.17**	0.83	-22.45	-21.60	3.04
			R^2_{oos}	0.3%	0.9%	1.2%	0.1%	-	-	0.4%
			UG	0.7%	1.2%	0.9%	-	-	-	0.0%
2	FH	CST only	MDM	-0.40	-1.39	-0.37	0.55	0.75	0.80	0.47
			ENCF	-0.08	-0.74	-0.25	9.25	38.68**	92.58***	176.29***
			MSEF	-0.38	-1.63	-1.33	10.58	38.98**	76.49**	179.45***
			R^2_{oos}	-	-	-	1.2%	4.4%	8.3%	18.4%
			UG	0.7%	-	0.5%	1.2%	1.4%	1.1%	1.1%
3	RNS	CST only	MDM	0.04	1.52**	1.37**	1.24**	0.22	0.29	0.62
			ENCF	0.02	0.33	0.85	2.23*	0.39	0.85	0.92
			MSEF	0.01	0.58	1.48	3.72**	0.47	1.29	1.62
			R^2_{oos}	0.0%	0.1%	0.2%	0.4%	0.1%	0.2%	0.2%
			UG	0.2%	0.5%	0.4%	0.2%	0.1%	0.1%	0.0%
4	RNK	CST only	MDM	-1.25	-0.98	-1.35	-2.48	-0.18	-0.31	0.33
			ENCF	-0.22	-0.15	-0.29	-0.45	0.12	0.32	0.52
			MSEF	-0.48	-0.35	-1.70	-0.98	-0.16	-0.97	1.02
			R^2_{oos}	-	-	-	-	-	-	0.1%
			UG	-	-	-	-	0.0%	-	0.0%
5	All	CST only	MDM	0.89**	1.03**	0.96**	0.75*	0.35	0.58	0.49
			ENCF	1.55	3.56*	5.51*	12.61	28.08*	80.39***	187.38***
			MSEF	2.33**	5.67**	7.34*	14.06*	16.67	56.06**	191.61***
			R^2_{oos}	0.3%	0.6%	0.8%	1.6%	1.9%	6.2%	19.4%
			UG	1.0%	1.1%	1.0%	0.8%	1.0%	1.1%	1.1%
6	All	All ex. VRP	MDM	1.20**	1.39**	1.34**	0.26	-1.53	-0.91	1.12
			ENCF	1.84**	4.52***	6.10**	2.64	-6.83	-6.79	4.56
			MSEF	3.10**	7.80***	10.49**	2.12	-19.01	-16.55	7.93
			R^2_{oos}	0.4%	0.9%	1.2%	0.2%	-	-	1.0%
			UG	0.2%	1.0%	0.5%	-	-	-	0.0%
7	All	All ex. FH	MDM	-0.68	-1.63	-0.51	0.46	0.75	0.77	0.48
			ENCF	-0.19	-0.77	-0.49	7.74	36.90**	90.51***	180.89***
			MSEF	-0.53	-1.68	-1.80	8.30	37.95**	75.18**	182.39***
			R^2_{oos}	-	-	-	0.9%	4.2%	8.2%	18.6%
			UG	0.0%	-	0.0%	0.8%	1.2%	1.2%	1.1%
8	All	All ex. RNS	MDM	1.10**	-0.07	0.31	1.32**	-0.78	-0.15	0.80
			ENCF	0.14	0.00	0.13	2.02*	-0.70	-0.14	1.34
			MSEF	0.26	-0.02	0.20	3.64*	-1.72	-0.67	2.38
			R^2_{oos}	0.0%	-	0.0%	0.4%	-	-	0.3%
			UG	0.3%	0.2%	-	-	0.0%	0.0%	0.0%
9	All	All ex. RNK	MDM	-0.93	-1.97	-2.08	-2.59	-1.31	-1.86	-1.35
			ENCF	-0.11	-0.43	-1.05	-0.79	-0.38	-1.98	-1.11
			MSEF	-0.25	-0.93	-2.90	-1.63	-0.88	-4.03	-2.26
			R^2_{oos}	-	-	-	-	-	-	-
			UG	0.1%	-	-	-	-	-	-

Table III: Out-of-Sample Performance
Panel C: SMI

No.	Unrest.	Rest.	Test Stat.	Return horizon (in business days)						
				1	2	5	20	50	100	250
1	VRP	CST only	MDM	-0.17	0.16	0.26	-0.38	-0.56	-0.57	-2.32
			ENCF	0.62	2.69**	4.46*	0.15	-2.76	-8.43	-5.77
			MSEF	-0.96	1.64*	2.74	-6.82	-16.11	-28.71	-13.91
			R^2_{oos}	-	0.2%	0.3%	-	-	-	-
			UG	-	-	0.2%	0.1%	0.3%	0.0%	-
2	FH	CST only	MDM	-0.44	-0.98	-0.94	-0.79	-0.69	-0.67	0.19
			ENCF	0.81*	-0.44	-0.65	-3.30	4.00	5.56	70.56***
			MSEF	-1.60	-4.74	-10.04	-27.32	-31.87	-40.56	62.85***
			R^2_{oos}	-	-	-	-	-	-	7.4%
			UG	-	-	-	-	0.0%	0.3%	1.0%
3	RNS	CST only	MDM	-0.46	0.00	-0.60	-0.59	-1.96	-0.50	-0.05
			ENCF	-0.04	0.01	-0.14	-0.79	-3.25	-3.00	2.00
			MSEF	-0.09	0.00	-0.32	-2.02	-6.73	-9.22	-2.13
			R^2_{oos}	-	0.0%	-	-	-	-	-
			UG	-	0.0%	-	0.0%	-	-	-
4	RNK	CST only	MDM	-0.24	-0.31	-0.13	-0.05	-0.52	0.04	-0.42
			ENCF	-0.01	-0.01	-0.01	-0.01	-0.30	0.16	-0.13
			MSEF	-0.01	-0.02	-0.01	-0.03	-0.60	0.21	-0.39
			R^2_{oos}	-	-	-	-	-	0.0%	-
			UG	-	-	0.0%	0.0%	-	0.0%	-
5	All	CST only	MDM	-0.49	-0.35	-0.56	-1.01	-1.06	-1.14	-0.07
			ENCF	1.28	2.17	3.51	-4.83	-4.20	-12.95	50.84**
			MSEF	-3.19	-3.85	-7.87	-42.11	-61.68	-104.09	-26.49
			R^2_{oos}	-	-	-	-	-	-	-
			UG	-	-	-	-	-	0.1%	0.9%
6	All	All ex. VRP	MDM	-0.23	0.13	0.27	-0.50	-0.78	-0.83	-7.93
			ENCF	0.40	2.33*	4.33*	-0.81	-3.98	-9.84	-7.31
			MSEF	-1.21	1.27	2.91	-7.59	-16.90	-28.33	-16.12
			R^2_{oos}	-	0.1%	0.3%	-	-	-	-
			UG	0.5%	0.7%	0.8%	0.2%	0.1%	-	-
7	All	All ex. FH	MDM	-0.64	-1.22	-1.07	-0.96	-0.81	-1.11	-0.03
			ENCF	0.39	-1.04	-1.34	-5.90	1.04	-3.35	40.71***
			MSEF	-2.24	-5.64	-10.80	-32.06	-37.55	-65.19	-10.63
			R^2_{oos}	-	-	-	-	-	-	-
			UG	0.6%	0.1%	-	-	-	0.2%	0.9%
8	All	All ex. RNS	MDM	0.31	0.43	0.81*	-1.52	-1.53	-2.32	-0.71
			ENCF	0.08	0.14	0.70	-2.39	-4.43	-13.10	-19.91
			MSEF	0.12	0.23	1.13	-5.20	-9.55	-30.22	-55.39
			R^2_{oos}	0.0%	0.0%	0.1%	-	-	-	-
			UG	-	-	-	-	-	-	-
9	All	All ex. RNK	MDM	-0.10	0.28	0.53	-0.79	-1.58	-2.68	-1.05
			ENCF	0.00	0.03	0.13	-0.34	-1.43	-2.48	-8.00
			MSEF	-0.01	0.06	0.23	-0.71	-2.95	-5.09	-16.93
			R^2_{oos}	-	0.0%	0.0%	-	-	-	-
			UG	-	-	0.0%	-	-	-	-

				Return horizon (in business days)						
Table III: Out-of-Sample Performance										
Panel D: CAC										
No.	Unrest.	Rest.	Test Stat.	1	2	5	20	50	100	250
1	VRP	CST only	MDM	1.81**	1.86**	2.12***	1.01*	-0.17	-0.12	-1.87
			ENCF	2.86**	5.10**	7.69**	10.20	0.82	1.93	-3.44
			MSEF	4.90***	9.02***	13.54**	15.38*	-3.66	-5.25	-7.32
			R^2_{oos}	0.6%	1.2%	1.8%	2.0%	-	-	-
			UG	2.0%	2.1%	2.2%	1.2%	0.6%	0.2%	0.0%
2	FH	CST only	MDM	0.97**	0.46	0.52	1.24**	1.68**	1.43*	0.48
			ENCF	0.79*	0.95	3.33*	34.72***	72.30***	184.58***	282.40***
			MSEF	1.17**	1.07	3.45*	44.06***	100.57***	233.22***	326.55***
			R^2_{oos}	0.2%	0.1%	0.5%	5.5%	11.9%	24.3%	32.4%
			UG	2.4%	2.1%	2.6%	4.7%	4.2%	3.7%	1.5%
3	RNS	CST only	MDM	0.18	0.43	1.23**	1.43**	3.51***	2.12**	0.90
			ENCF	0.20	0.72	2.25**	6.18***	14.84***	16.36***	31.93***
			MSEF	0.14	0.72	3.45**	10.54***	25.67***	29.54***	54.09***
			R^2_{oos}	0.0%	0.1%	0.5%	1.4%	3.3%	3.9%	7.3%
			UG	1.4%	1.5%	1.6%	1.1%	0.7%	0.1%	0.0%
4	RNK	CST only	MDM	0.99**	1.27**	-0.73	-1.16	-1.49	-1.89	0.17
			ENCF	0.26	0.61	-0.22	-2.01	-4.75	-8.50	3.57*
			MSEF	0.42	1.00	-0.53	-4.42	-10.49	-19.31	2.63
			R^2_{oos}	0.1%	0.1%	-	-	-	-	0.4%
			UG	0.6%	1.0%	-	-	-	-	-
5	All	CST only	MDM	1.99***	2.02***	1.74***	1.31**	1.36**	1.15*	0.54
			ENCF	3.69**	6.65**	11.30**	43.46***	71.73***	177.79***	293.62***
			MSEF	6.01***	10.91***	16.63***	54.34***	89.57***	209.88***	364.74***
			R^2_{oos}	0.8%	1.4%	2.2%	6.7%	10.8%	22.4%	34.8%
			UG	4.1%	3.8%	3.5%	4.5%	4.1%	3.5%	1.4%
6	All	All ex. VRP	MDM	1.77**	1.82**	1.95**	0.52	-0.75	-0.82	-3.74
			ENCF	2.73**	5.02**	6.76*	6.01	-4.98	-10.70	-4.28
			MSEF	4.70**	8.90***	11.92**	7.84	-14.43	-28.77	-9.01
			R^2_{oos}	0.6%	1.2%	1.6%	1.0%	-	-	-
			UG	1.4%	1.2%	0.6%	-	-	-	-
7	All	All ex. FH	MDM	0.19	-0.45	0.00	0.97*	1.45**	1.47**	0.54
			ENCF	0.12	-0.15	0.94	26.13***	55.33***	158.39***	271.33***
			MSEF	0.12	-0.54	0.00	31.26**	75.55***	198.93***	316.36***
			R^2_{oos}	0.0%	-	0.0%	4.0%	9.2%	21.5%	31.7%
			UG	-	-	0.1%	2.8%	3.0%	3.3%	1.5%
8	All	All ex. RNS	MDM	0.01	0.38	1.12**	1.10**	1.32*	1.17*	0.65
			ENCF	0.04	0.32	1.22	2.47*	6.94***	9.18***	13.45***
			MSEF	0.01	0.37	1.94*	4.45**	11.80***	16.53***	17.52***
			R^2_{oos}	0.0%	0.0%	0.3%	0.6%	1.6%	2.2%	2.5%
			UG	0.8%	0.5%	0.4%	0.1%	0.2%	0.1%	-
9	All	All ex. RNK	MDM	1.22**	1.55**	-0.51	-0.26	-0.26	0.71	0.96
			ENCF	0.45	0.98	-0.14	-0.23	-0.42	4.50**	18.70***
			MSEF	0.73*	1.59*	-0.35	-0.84	-1.78	5.35**	29.99***
			R^2_{oos}	0.1%	0.2%	-	-	-	0.7%	4.2%
			UG	1.3%	0.9%	-	-	-	-	-

Table III: Out-of-Sample Performance
Panel E : FTSE

No.	Unrest.	Rest.	Test Stat.	Return horizon (in business days)						
				1	2	5	20	50	100	250
1	VRP	CST only	MDM	1.30*	1.61**	2.34***	1.45**	0.60	0.52	0.71
			ENCF	1.66*	3.17*	3.69	8.86	5.13	9.58	21.39*
			MSEF	2.88**	5.75**	6.92*	14.52*	5.63	8.87	35.88*
			R^2_{oos}	0.3%	0.7%	0.8%	1.7%	0.7%	1.1%	4.4%
			UG	0.3%	0.5%	0.7%	0.4%	0.2%	0.1%	0.1%
2	FH	CST only	MDM	-0.64	-0.56	-0.64	-0.40	0.05	-0.10	-0.76
			ENCF	-0.14	-0.30	-1.03	2.35	25.18*	72.49**	-65.66
			MSEF	-0.36	-0.99	-4.54	-12.59	4.10	-18.95	-491.30
			R^2_{oos}	-	-	-	-	0.5%	-	-
			UG	-	-	-	0.2%	0.6%	0.6%	0.9%
3	RNS	CST only	MDM	-1.26	-0.78	0.40	-0.20	-0.15	0.19	-2.18
			ENCF	-0.16	-0.22	0.23	-0.05	0.10	0.36	-16.55
			MSEF	-0.35	-0.54	0.34	-0.33	-0.23	0.15	-37.85
			R^2_{oos}	-	-	0.0%	-	-	0.0%	-
			UG	-	0.0%	0.1%	-	0.0%	-	-
4	RNK	CST only	MDM	-0.76	0.03	-0.71	-0.79	-0.61	-0.47	-0.82
			ENCF	-0.04	0.00	-0.15	-0.75	-2.12	-3.80	-37.92
			MSEF	-0.09	0.00	-0.32	-1.60	-4.66	-9.53	-85.72
			R^2_{oos}	-	0.0%	-	-	-	-	-
			UG	-	0.1%	-	-	-	-	-
5	All	CST only	MDM	0.90*	0.96**	0.19	-0.10	-0.01	-0.12	-0.76
			ENCF	1.34	2.58	2.50	9.46	26.75	82.20**	-68.32
			MSEF	2.10*	4.00*	1.53	-3.68	-0.81	-24.67	-482.62
			R^2_{oos}	0.2%	0.5%	0.2%	-	-	-	-
			UG	0.2%	0.2%	-	0.3%	0.5%	0.7%	0.9%
6	All	All ex. VRP	MDM	1.34*	1.63**	2.32***	1.17**	-0.09	-0.92	-0.60
			ENCF	1.68*	3.23*	3.43	7.13	1.29	-3.21	-1.79
			MSEF	2.94**	5.89**	6.45*	11.75*	-0.58	-13.20	-6.27
			R^2_{oos}	0.3%	0.7%	0.8%	1.4%	-	-	-
			UG	0.6%	0.6%	0.7%	0.2%	-	-	-
7	All	All ex. FH	MDM	-1.42	-1.07	-0.84	-0.54	-0.02	-0.14	-0.73
			ENCF	-0.28	-0.63	-1.77	-0.19	22.92*	77.25**	-50.71
			MSEF	-0.60	-1.49	-5.79	-16.86	-1.90	-26.55	-444.21
			R^2_{oos}	-	-	-	-	-	-	-
			UG	-	-	-	0.0%	0.5%	0.8%	1.3%
8	All	All ex. RNS	MDM	-1.08	-0.86	0.42	0.18	-0.41	1.03	-0.60
			ENCF	-0.17	-0.29	0.33	0.38	-0.13	0.80	-0.70
			MSEF	-0.36	-0.71	0.47	0.37	-0.79	1.30	-2.69
			R^2_{oos}	-	-	0.1%	0.0%	-	0.2%	-
			UG	-	-	0.1%	-	0.0%	0.0%	-
9	All	All ex. RNK	MDM	0.33	0.22	-0.66	-0.77	-0.39	0.85	-0.60
			ENCF	0.03	0.01	-0.20	-0.62	-0.04	8.43***	-3.24
			MSEF	0.05	0.02	-0.50	-1.83	-1.89	11.49***	-9.42
			R^2_{oos}	0.0%	0.0%	-	-	-	1.4%	-
			UG	-	0.0%	-	-	0.0%	0.1%	-

**Figure 1: Predictive regressions using option-implied risk variables**

The following panels show results, for each equity index, of two-sided individual significance tests for the coefficients of regression (1) for different return horizons *h* measured in days. The results are shown as Newey-West corrected t-statistics, but the critical values for the usual levels of significance are computed from 10,000 bootstrapped series and plotted as gray lines (10%: dots, 5%: dash-dot, 1%: dashes). The null hypothesis is that a coefficient is equal to zero. The alternative hypothesis is that the coefficient is not equal to zero, meaning that this variable has significant explanatory power over future returns. VRP is the variance risk premium defined as the difference between model-free risk-neutral variance and realized variance. FH is the Foster-Hart generalized measure of riskiness taken on the risk-neutral distribution each day in the sample. RNSkew and RNKurt are the skewness and kurtosis of the risk-neutral distribution using GEV tails. The horizontal axis is in log scale.
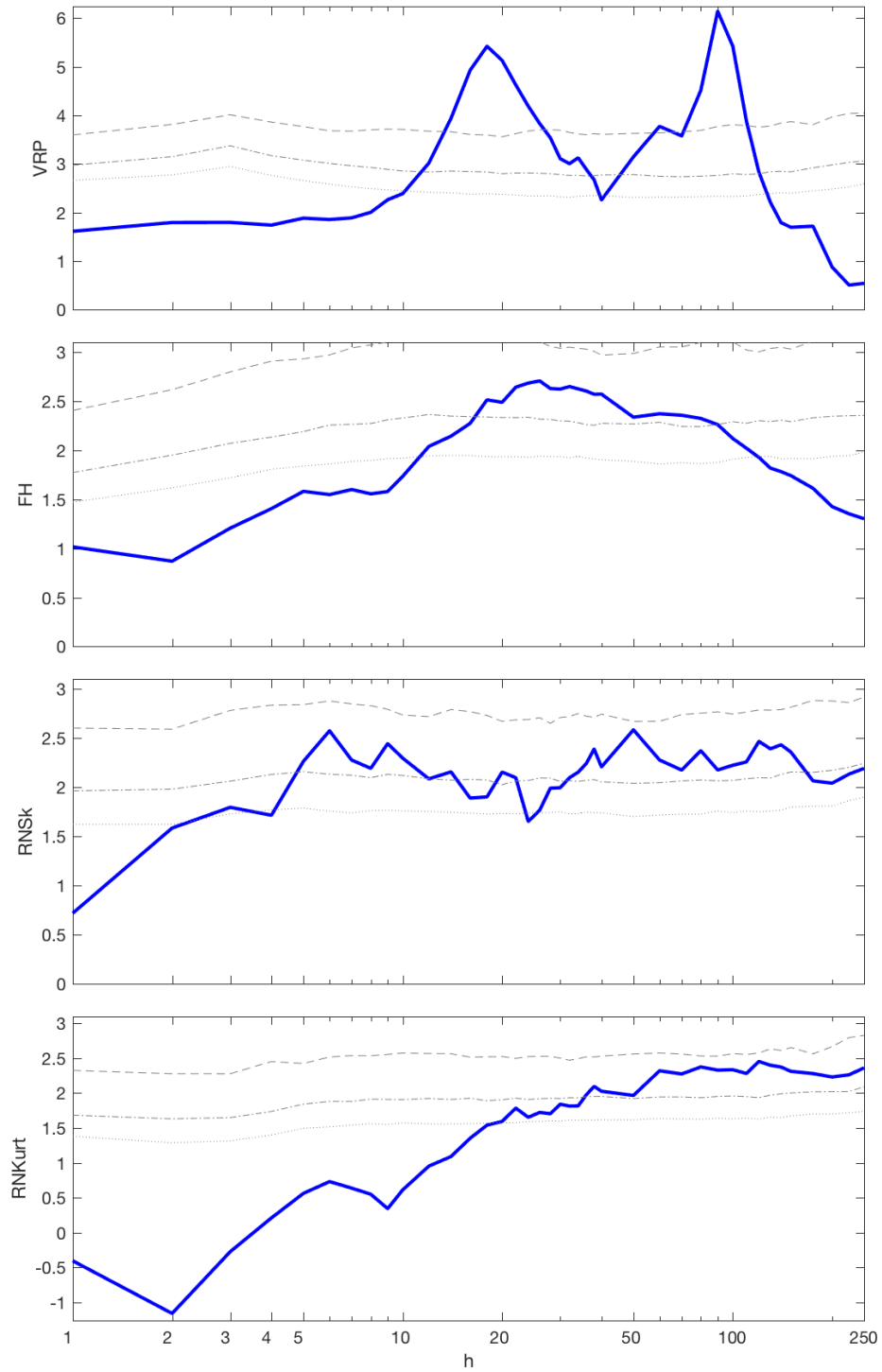
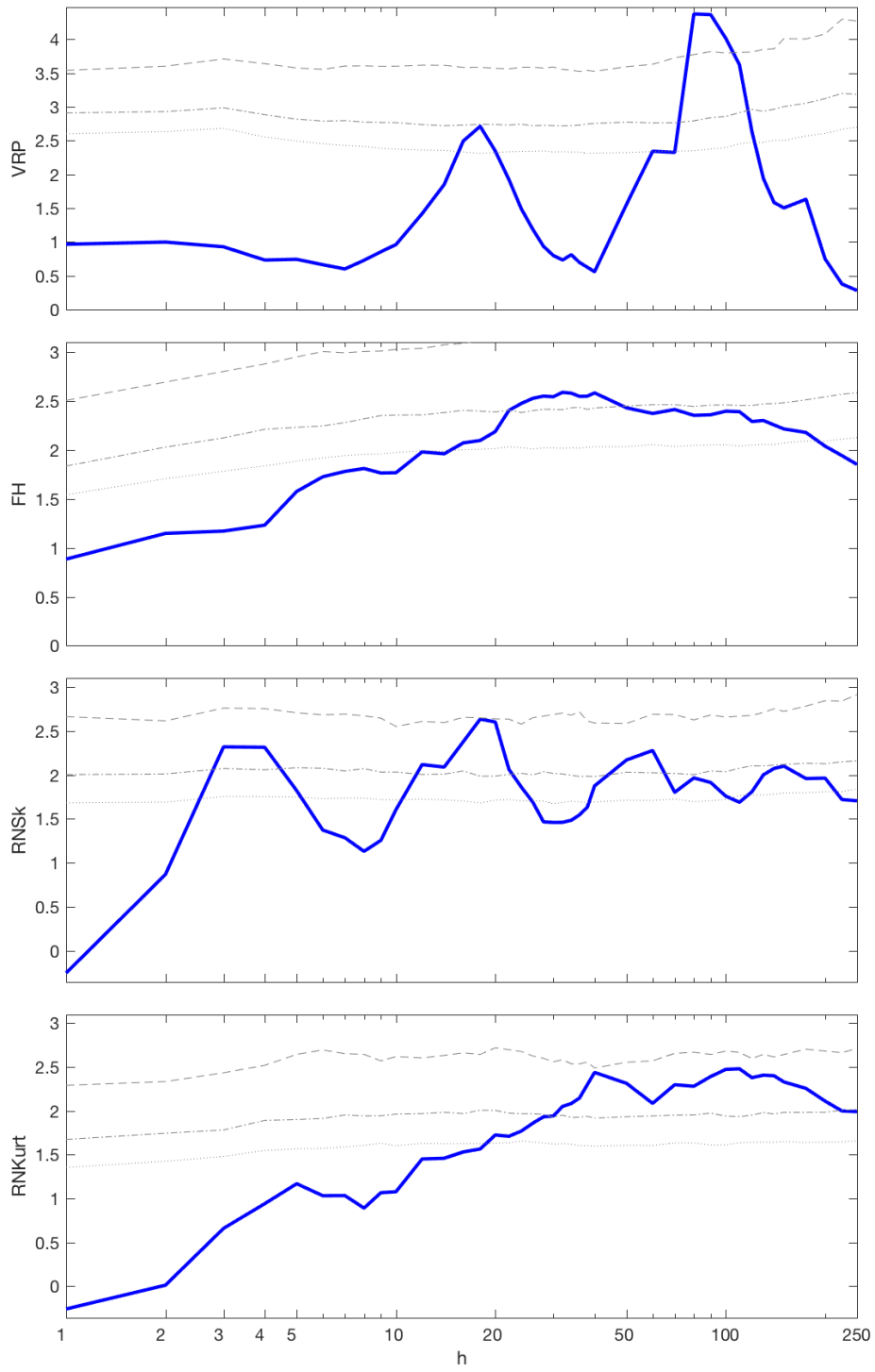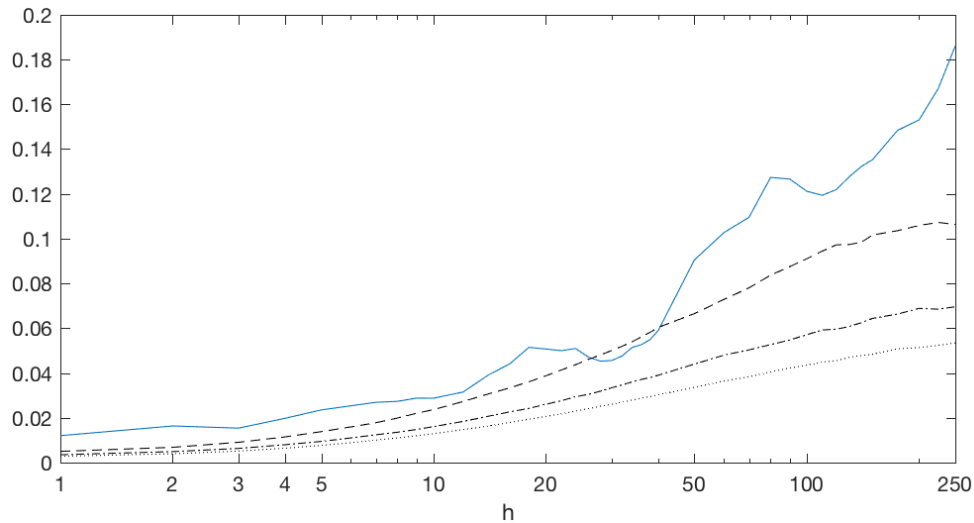**Panel A: S&P 500 index**
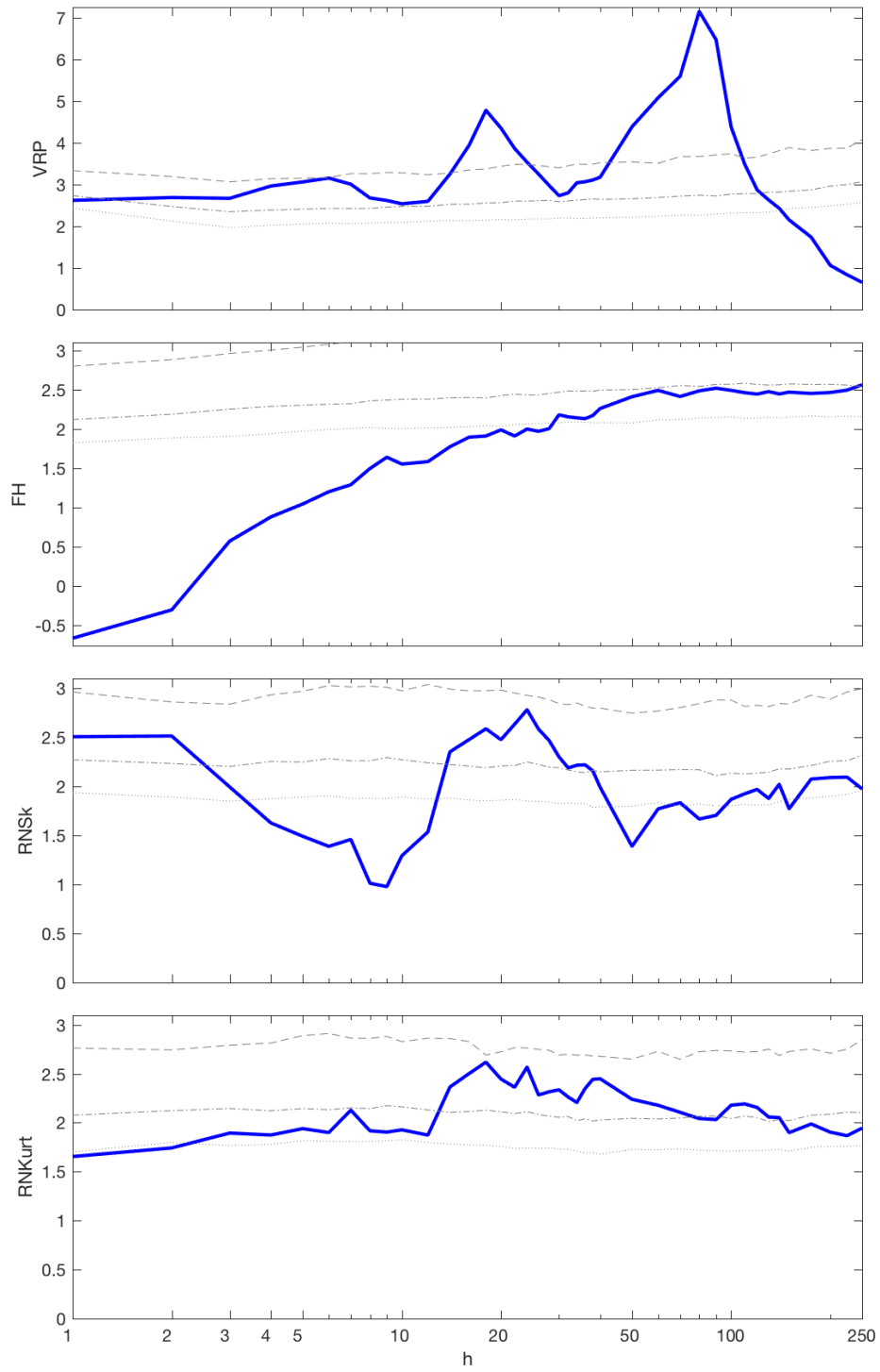
Panel B: DAX
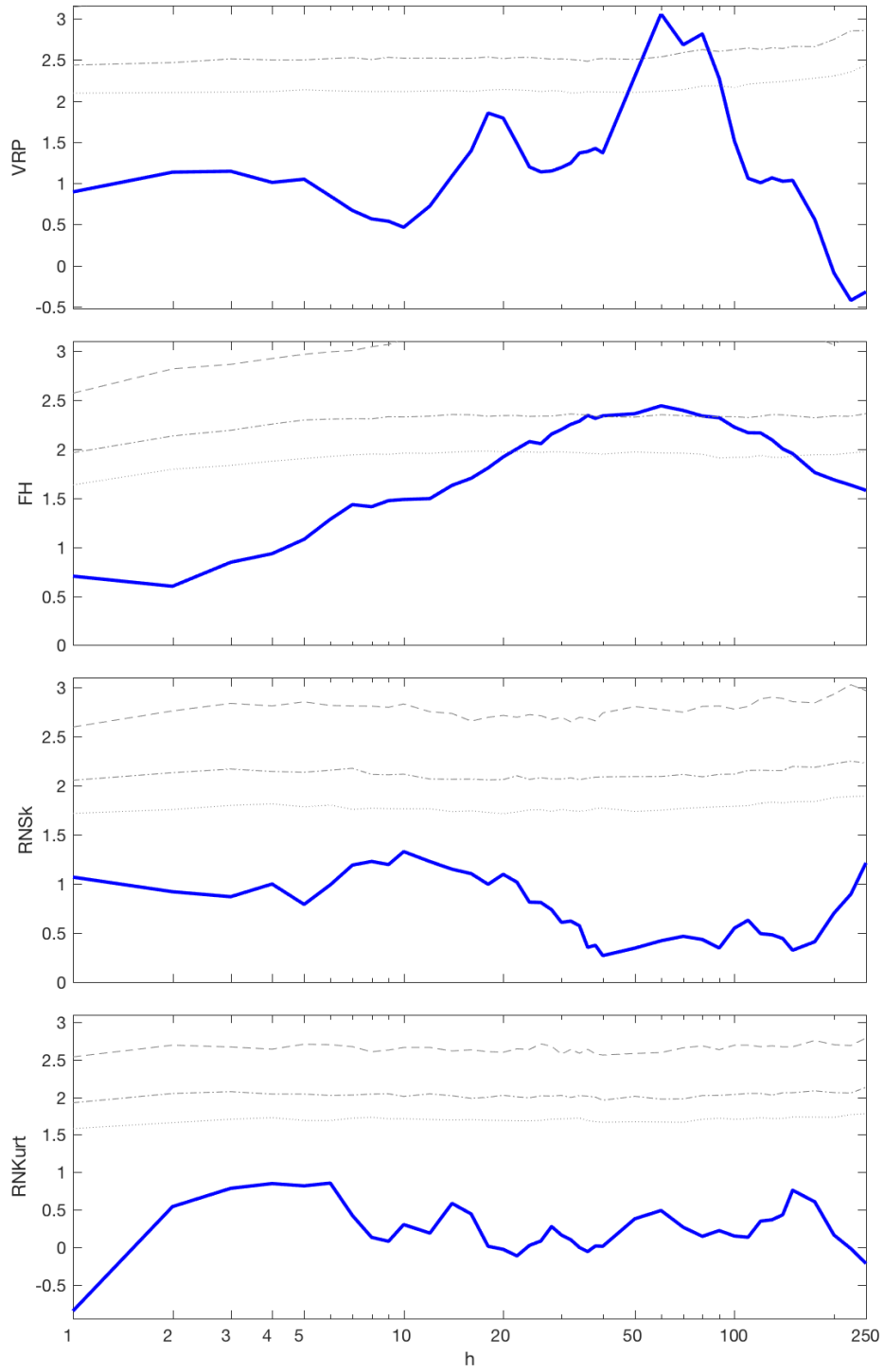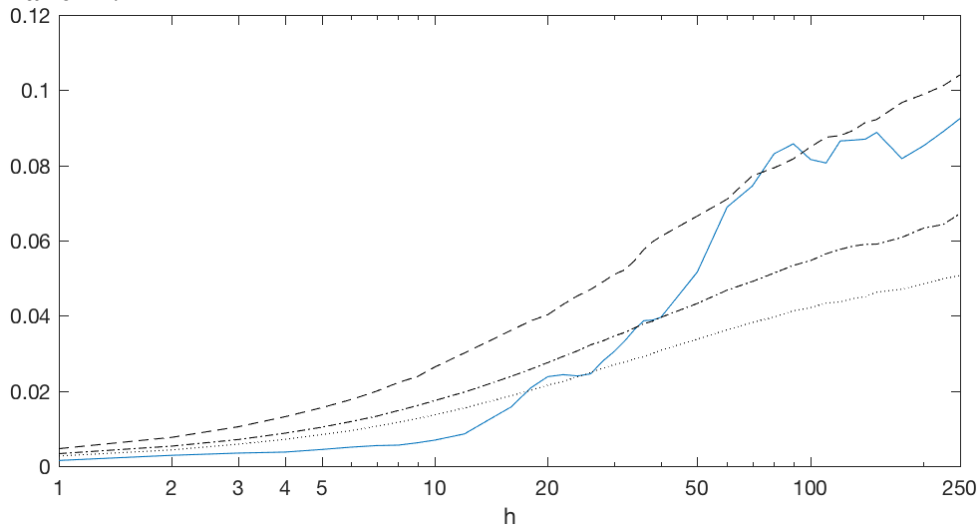
Panel C: SMI
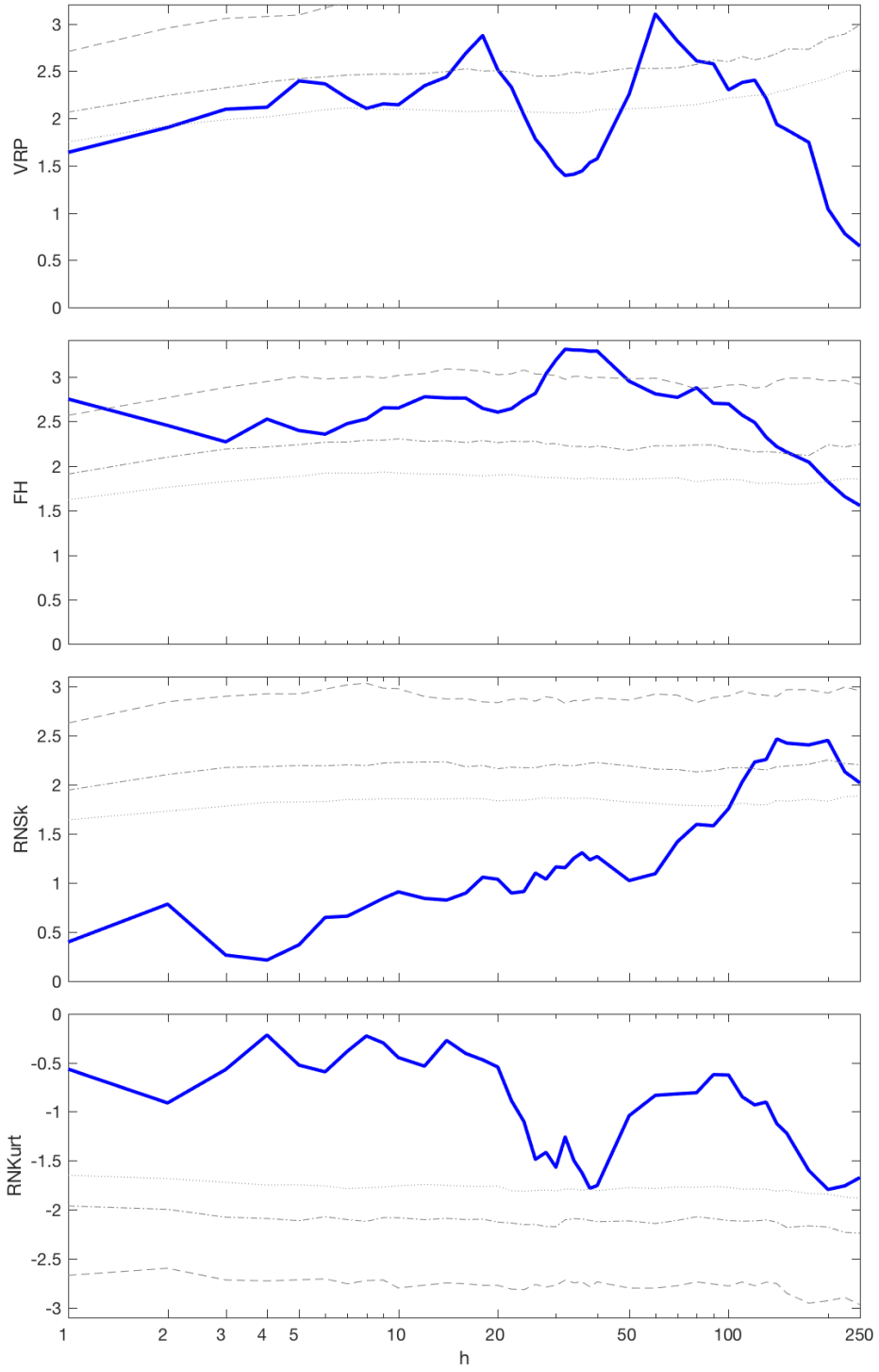
Panel D: CAC

Panel E: FTSE

Figure 2: Adjusted R^2 for predictive regressions over several horizons

This figure (panels A through E) reports the adjusted R^2 for the regression described in Equation (4) for future returns of each market index at different horizons h (measured in days). The Adjusted R^2 are compared to the percentiles of those from 10,000 bootstrapped series that are plotted as gray lines (90%: dots, 95%: dash-dot, 99%: dashes). The horizontal axis is in log scale.
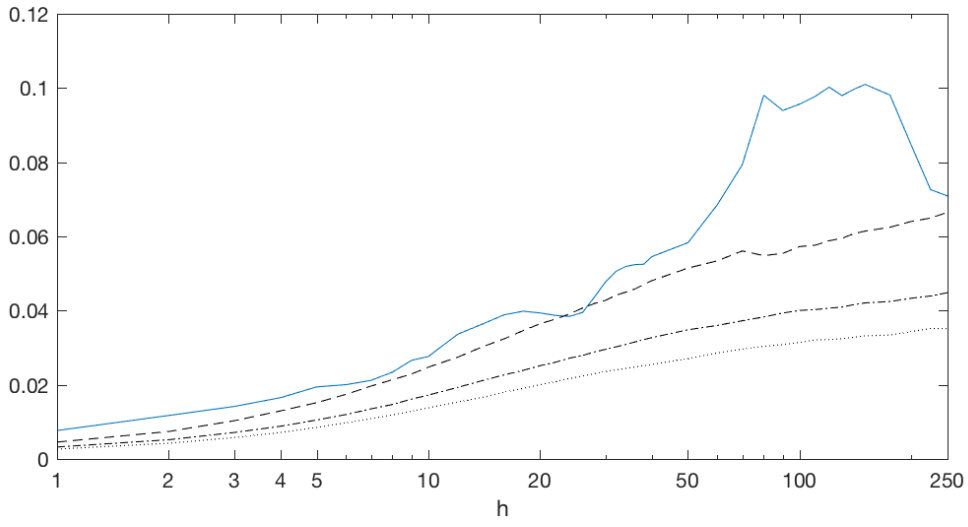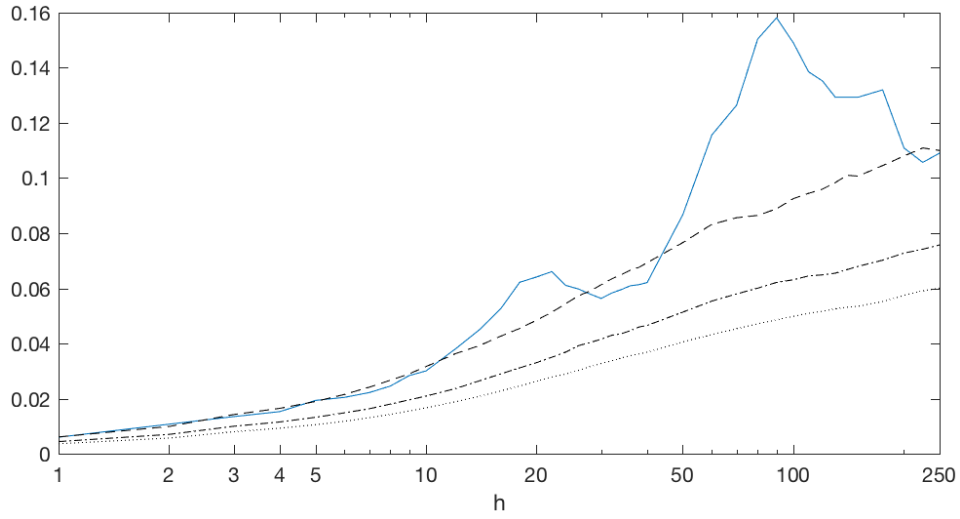
Panel A: S&P 500

Panel B: DAX

Panel C: SMI

Panel D– CAC

46

Panel E- FTSE