Forecasting International Stock Market Variances*

Geert Bekaert† Nancy R. Xu‡ Tiange Ye§

July, 2025

Abstract

We first examine 320 different forecasting models for international monthly stock return volatilities, using high frequency realized variances and the implied option variance as the predictor variables. We evaluate linear and non-linear models, and logarithmic transformed and weighted least squares estimation approaches. A logarithmically transformed Corsi (2009) model combined with the option implied variance ("lm4_log") is robustly, across countries and time, among the best forecasting models. This conclusion survives when estimating a multitude of alternative models, including panel and MIDAS models, models embedding quarticity, leverage, downside, and jump risk. While not always the very best, the "lm4_log" model always generates forecasts extremely highly correlated with the very best model for a particular sample.

JEL Classification: C58, F30, G10, G17
Keywords: Realized variance, implied volatility, international stock market, volatility forecasting

*First SSRN draft: May, 2024. All errors are our own.

†Finance Division, Columbia Business School, 3022 Broadway, New York, NY 10027, USA; CEPR; Email: gb241@gsb.columbia.edu.

‡Finance Department, Carroll School of Management, Boston College, 140 Commonwealth Avenue, Chestnut Hill, MA 02467, USA; Email: nancy.xu@bc.edu.

§Finance and Business Economics Department, Marshall School of Business, University of Southern California, 3670 Trousdale Pkwy, Los Angeles, CA 90089, USA; Email: tiangeye@usc.edu.

1 Introduction

The conditional expected stock market variance is a critical variable for risk and asset management, and not surprisingly the subject of a gigantic literature (see e.g. Corsi, Audrino, and Renò, 2012). There is much less work on international stock market variance forecasting however, with most research focusing on the US. While the US stock market constitutes an important part of the global stock market, it is also by far the least volatile market, and great US volatility models may work less well in other markets. In this article, we identify volatility models that forecast stock market variances well for a set of developed countries, which together comprise more than 90% of the developed world market capitalization. We start from the state-of-the-art literature using the future realized variance of stock market returns, computed from high frequency data, as the variable to forecast (Andersen, Bollerslev, Diebold, and Labys, 2003)

Our analysis is two-pronged. In the first part, we consider relatively standard predictor variables, which include realized variances at different aggregation levels as in Corsi (2009) and, importantly, option implied variances (as in Bekaert and Hoerova, 2014). We consider a very wide range of models, examining all possible combinations of these independent variables in linear and non-linear models. In non-linear versions of these models, the predictive coefficients can change with the level of the independent variable. We also examine logarithmic transformations of realized variances and weighted least squares estimates. Non-linear coefficients can help capture sudden changes in mean reversion in crisis times, whereas logarithmic transformations render the resulting volatility distributions more Gaussian, leading to improved linear forecasts. We estimate a total of 320 different models. We use the BIC, RMSE (root mean squared error) and QLIKE (quasi-likelihood) criteria (see Patton, 2011) to measure initial forecasting performance, in a cross-validation and forward-chained validation approach. We ultimately select models that perform well across all countries and criteria and significantly beat simple benchmark models in statistical horse races relative to three easy-to-estimate linear benchmark models. The first benchmark model is the Heterogeneous Autoregressive (HAR) model of

Corsi (2009), which incorporates three realized variances measuring quadratic variation from the past day, week and month respectively. The second model adds the option implied variance to the Corsi model as in Bekaert and Hoerova (2014). The final model uses the past monthly realized variance and the past option implied variance as independent variables, as proposed in Bekaert, Hoerova, and Lo Duca (2013)

Different from most existing econometric analysis on volatility forecasting, which mostly focuses on short forecasting horizons (like one day), we focus on the one-month horizon, which is more relevant for asset management. Time zone differences also complicate the use of daily forecast models in international data. In addition, we cast a particular wide net in terms of models examined.

Our main result is that two fairly simple models provide consistently superior forecasts to simple benchmark models and perform well in all countries across all performance criteria. The first model is simply the logarithmic transformation of the Corsi model, combined with the implied variance (which we label as "lm4_log"); the second model drops the daily realized variance from Model "lm4_log" (which we label as "lm7_log"). We establish this result first using a long sample period for Japan, Germany and the U.S. from 1992 to 2019. We later show it also applies to other countries, including the UK, France, the Netherlands, Switzerland and the Euro area. While alternative models outperform these models in a few settings, the resulting volatility forecasts are highly correlated with the forecasts generated by our proposed best models. Overall, the lm4_log model is slightly better than the lm7_log model and thus provides an easy-to-estimate volatility model that robustly performs well in an international setting.

In the second part of the paper, we consider a variety of alternative models and verify whether the forecasting power of lm4_log model survives.[1] These alternative models include a "global" model including cross-country variables, a panel estimation (inspired by Bollerslev, Hood, Huss, and Pedersen, 2018); a model including (negative) returns in the forecasting equations, to capture the "leverage effect" (see Corsi and Renò, 2012); a model decomposing quadratic variation in continuous and jump variables (Andersen,

[1] We provide the results for the lm7_log model in the Online Appendix.

Bollerslev, and Diebold, 2007; Barndorff-Nielsen and Shephard, 2002); a model accommodating downside risk (that is, differentiating between the variation of positive and negative returns by using realized semi-variances; Chen and Ghysels, 2011; Patton and Sheppard, 2015); embedding quarticity in the lm4_log model (Bollerslev, Patton, and Quaedvlieg, 2016) and a MIDAS model (Ghysels, Plazzi, Valkanov, Rubia, and Dossani, 2019). None of these models consistently outperforms our preferred models, but the global, leverage and "downside risk" models do outperform in a few isolated instances.

We also provide some further economic analysis of the results. The volatility forecasts of our proposed best models and the benchmark models generally show relatively high correlation, but this correlation drops substantially in crisis periods. This is important as many models generate somewhat unrealistic forecast values during crisis periods, which often leads to negative variance risk premiums. The variance premium is the difference between the option implied variance and the "physical" stock market variance. While theoretically it is possible for the variance risk premium to be negative (See Bekaert and Engstrom, 2017; Bekaert, Engstrom, and Ermolov, 2023, for theoretical explanations based on "good" uncertainty, and Cheng, 2019, for an explanation based on dealer hedging demands), there is a strong prior that the variance risk premium should be predominantly positive. In fact, several articles show the variance premium to be an important component of the equity risk premium (see Bollerslev, Tauchen, and Zhou, 2009; Bekaert and Hoerova, 2014). Standard volatility forecasting models tend to generate a large number of negative variance risk premiums. However, we show that the proposed models are particularly effective in reducing the number of negative variance risk premiums, especially during crisis periods. Among the better performing alternative models, the global and downside risk models are particularly effective in producing fewer negative variance risk premiums.

Finally, following Bollerslev et al. (2018), we calculate the utility benefit of using our preferred models, the lm4_log and lm7_log models, to forecast volatility, relative to using a very competitive benchmark model (the lm4 model). We find positive utility benefits for all 7 countries examined , except the general Eura area. For these other countries, the benefits are largest under the forward-chained cross-validation approach, ranging between 40 basis points (bps) for the U.S. and 2.43% for Germany.

While the literature on stock return volatility forecasting is too large to adequately summarize, the literature on forecasting international stock return variances is much smaller. Kourtis, Markellos, and Symeonidis (2016) show that GARCH models underperform the HAR model and/or implied volatility depending on the forecast horizon. Buncic and Gisler (2017) examine the Corsi and Renò (2012) model that adds jump components and leverage effects (using realized returns) to Corsi's HAR model for 18 international equity indices. They find that jump components are not helpful for longer horizons but leverage effects do lead to significant forecast gains. Buncic and Gisler (2016) show that adding US variables leads to forecasting gains with respect to a standard HAR model for 17 international stock markets. The additional information content of cross-national information within HAR models is more generally confirmed in an early note by Taylor (2015) and in Liang, Wei, Lei, and Ma (2020) and Zhang, Ma, and Liao (2020). Finally, Liang, Wei, and Zhang (2020) show that option implied volatility enhances forecasting accuracy for international stock market volatilities relative to a standard HAR model.

Given the results in the extant literature, we do not consider GARCH models.[2] However, our set of alternative models re-evaluates the use of jumps, leverage variables and cross-national variables. This additional information is either less useful with respect to our logarithmically transformed models or the corresponding models generate highly correlated forecasts with our top models.

In sum, we propose two stock market volatility models that are easy to compute and provide highly competitive stock market forecasts across the developed world.[3] The remainder of the paper is organized as follows. Section 2 describes the data and the main models we estimate. Section 3 describes our model selection procedures, and Section 4 the main results for a long sample on Germany, Japan and the US. Before we report on the overall best models, we demonstrate that non-linear models have great

[2] Bekaert, Bergbrant, and Kassa (2025), show that GARCH models perform by far the worst in predicting firm specific volatilities.

[3] We plan on sharing and updating these forecasts on our websites.

potential to improve forecasting accuracy relative to linear models, but also find that log-transformations appear to almost uniformly improve forecasting performance. Section 5 investigates the performance of a large suite of alternative models, showing the overall robust performance of the simple lm4_log model. Section 6 extends the sample to other developed countries but for a shorter sample period. Finally, Section 7 examines robustness to using an alternative cross-validation method.

2 Data and Base Models

We first focus on three countries with a long sample (January 1992 to December 2019): Germany (DE), Japan (JP), and the United States (US). The longest common sample for other developed market variances that we consider later in the paper only starts from January 2000. All variance variables and estimations are at the daily frequency. We obtain our data from standard databases, i.e., the Oxford-Man Institute for realized variances, and Refinitiv DataStream for option implied volatilities. The realized variance statistics use 5 minute returns (see Liu, Patton, and Sheppard, 2015, for evidence on the optimality of the 5 minute interval).

We focus on forecasting the future monthly realized variance (22 trading days) from $t + 1$ to $t + 22$, denoted as $RV^{(22)}_{i,t+22}$. Following the literature, we consider four independent variables. The first three are the recent monthly, weekly, and daily realized variance, denoted as $RV^{(22)}_t$ ($t - 21$ to t), $RV^{(5)}_t$ ($t - 4$ to t), and RV_t ($t - 1$ to t), as first proposed by Corsi (2009). As is typical, realized variances are computed using squared five-minute intraday returns and the squared close-to-open returns. The fourth independent variable is the option implied stock return variance, denoted as IV^2_t. IV represents the option implied volatility index for contracts of approximately one month. These indices are computed using a weighted average of European style call and put options on the index. As is common in this literature, each variance variable is converted into monthly percentages. For instance, the implied volatility is quoted as an annualized number and our IV^2_t variable is constructed as implied volatility squared divided by 12.

The original data sources for the volatility indices are:

Country	Volatility Index	Source	Currency
Germany	VDAX	Deutsche Boerse	Euro
Japan	VXJ	NIKKEI	Japanese Yen
United States	VIX	CBOE	US Dollar

We consider 15 linear models and 65 non-linear models. Furthermore, we have three additional transformations for each model: the log transformation, weighted least squares, and the combination of both. Consequently, we investigate 320 models in total. When all four independent variables are included in a model, it is referred to as a *full* model. Next, we introduce the four full baseline models first: full linear model, full log linear model, full non-linear model, and the full weighted least square model.

Full Linear Model The most basic full linear model (labeled as "lm4") is as follows:

$$\mathbf{E}_t\left(RV_{i,t+22}^{(22)}\right) = \hat{\alpha}_i + \hat{\beta}_i^m RV_{i,t}^{(22)} + \hat{\beta}_i^w RV_{i,t}^{(5)} + \hat{\beta}_i^d RV_{i,t} + \hat{\gamma}_i IV_{i,t}^2. \quad (1)$$

The model is estimated using OLS, using overlapping daily data. There are a total of 15 possible models combining these 4 variables linearly. We list them in Table 1. This specification comprises our three benchmark models as special cases: (1) the seminal Corsi model (our "lm3") which has the three realized variance variables, (2) the full model which also includes the option implied variance (our "lm4"), and (3) the simpler lm2 model. The lm2 model, initially proposed and tested in Bekaert, Hoerova, and Lo Duca (2013), uses the past monthly realized variance and the implied variance. Despite being very simple and parsimonious, they show that lm2 performs very well in out-of-sample forecasting exercises.

[Insert Table 1]

Full Non-linear Model In the full non-linear model (nlm4-1), each coefficient is the typical feedback coefficient multiplied by a logistic function of the independent variable itself. Thus, there are two coefficients to estimate for each independent variable, e.g. β_{m0} and β_{m1}:

$$\begin{aligned}\mathbf{E}_t\left[RV_{i,t+22}^{(22)}\right] = \hat{\alpha}_i + \hat{\beta}_i^{m0}\frac{\exp\left(\hat{\beta}_i^{m1}RV_{i,t}^{(22)}\right)}{\exp\left(\hat{\beta}_i^{m1}RV_{i,t}^{(22)}\right)+1}RV_{i,t}^{(22)} + \hat{\beta}_i^{w0}\frac{\exp\left(\hat{\beta}_i^{w1}RV_{i,t}^{(5)}\right)}{\exp\left(\hat{\beta}_i^{w1}RV_{i,t}^{(5)}\right)+1}RV_{i,t}^{(5)} \\ + \hat{\beta}_i^{d0}\frac{\exp\left(\hat{\beta}_i^{d1}RV_{i,t}\right)}{\exp\left(\hat{\beta}_i^{d1}RV_{i,t}\right)+1}RV_{i,t} + \hat{\gamma}_i^{0}\frac{\exp\left(\hat{\gamma}_i^{1}IV_{i,t}^{2}\right)}{\exp\left(\hat{\gamma}_i^{1}IV_{i,t}^{2}\right)+1}IV_{i,t}^{2}\end{aligned} \tag{2}$$

Economically, such non-linear coefficients help capture sudden changes in mean reversion in crisis times. For example, when a particular month witnesses tremendous volatility, resulting in high realized variances, it is quite likely that such high variance realization does not persist in the same fashion as it does in moderate times. Similarly, an event that makes agents very risk averse causing implied volatility to rise sharply may be expected to revert to less extreme levels more quickly than more moderate increases in risk aversion. Thus, we generally expect the interaction coefficients within the logistic functions to be negative. The logistic function ensures the interaction effect is strictly in the $(0, 1)$ continuous interval.

Moreover, there is a purely econometric justification for this specification, as indicated in Bollerslev et al. (2016). They estimate the Corsi model with some or all of the coefficients interacted with the relevant quarticity measure. Quarticity reflects sums of high frequency returns to the 4th power and is proportional to the asymptotic variance of realized variance measures. Because there is an obvious positive correlation between quadratic variation and quarticity, and quarticity is not defined for implied variance measures, we use the realized variances themselves in the interaction terms.[4]

The nomenclature for the models follows Table 1. For example, nlm4-11 refers to a non-linear model with 4 independent variables but with the first two independent variables ($RV^{(22)}$ and $RV^{(5)}$) entering in a linear instead of non-linear fashion. All nlm3 models refer to versions of the Corsi model, with 7 such models describing different combinations of non-linear and linear independent variables. Table A1 lists the specification for all 15 full non-linear models. That is, each model on this list has the 4 independent variables,

[4]Quarticity may also be less publicly available than is RV. Yet, we do estimate a model embedding quartcity inferred from daily returns in Section 5.

which can be either in linear or non-linear form. Table A2 lists the remaining 50 non-linear models, where variables can also be left out, meaning that models have at least one but fewer than 4 non-linear independent variables.

The estimation is conducted by minimizing the sum of squared residuals:

$$\min_{\left\{\alpha_i,\beta_i^{m0},\beta_i^{m1},\beta_i^{w0},\beta_i^{w1},\beta_i^{d0},\beta_i^{d1},\gamma_i^0,\gamma_i^1\right\}} \sum_t \left(RV_{i,t+22}^{(22)} - \mathbf{E}_t\left[RV_{i,t+22}^{(22)}\right]\right)^2$$

Full Log Linear Model The log transformed models are models that predict the logarithm of the realized variance using the logarithms of the independent variables as predictors. The full log linear model (lm4_log) specification is as follows:

$$\mathbf{E}_t\left[\ln\left(RV_{i,t+22}^{(22)}\right)\right] = \hat{\alpha}_i + \hat{\beta_i^m}\ln\left(RV_{i,t}^{(22)}\right) + \hat{\beta_i^w}\ln\left(RV_{i,t}^{(5)}\right) + \hat{\beta_i^d}\ln(RV_{i,t}) + \hat{\gamma}_i\ln\left(IV_{i,t}^2\right), \quad (3)$$

Analogously, a log non-linear model replaces the independent and dependent variables with their logarithms. The logarithmic transformation renders variance distributions, which are right skewed, more Gaussian. While this may impart better forecasting properties to linear models (which we estimate by OLS), we must still estimate the expected variance. Therefore, when considering a log transformed model, we assume lognormality to predict levels of monthly realized variances as in Equation (4):

$$\mathbf{E}_t\left[RV_{i,t+22}^{(22)}\right] = \exp\left\{\mathbf{E}_t\left[\ln\left(RV_{i,t+22}^{(22)}\right)\right] + \frac{1}{2}\mathrm{Var}\left[\epsilon_{i,t+22}^{(22)}\right]\right\}, \quad (4)$$

where $\epsilon_{i,t+22}^{(22)} = \ln\left(RV_{i,t+22}^{(22)}\right) - \mathbf{E}_t\left[\ln\left(RV_{i,t+22}^{(22)}\right)\right]$. That is, the variance correction uses the sample variance of the (logarithmic) residual for country i.

Weighted Least Squares Model In the weighted least squares (WLS) model, the weight is the reciprocal of the recent monthly realized variance, i.e. $1/RV_{i,t}^{(22)}$. Thus, observations in the right tail of the variance distribution are down weighted. Finally, we also consider WLS estimation of the logarithmic models. Note that we do not consider the martingale model, which is a restricted version of a particular linear model or a constant variance model, as these models have been convincingly rejected in the volatility forecasting literature.

3 Model Selection

Our model selection uses a combination of three popular performance criteria and two validation techniques to identify the overall best model(s), and then employs "horserace" regression methods to test their forecasts relative to those of the three benchmark models, mentioned before (lm2, lm3 and lm4).

3.1 Forecasting Criteria

We use three performance measure criteria: the well-known BIC and $RMSE$ criteria, but also the $QLIKE$ criterion ("Quasi-likelihood", Patton, 2011) as follows:

$$QLIKE = \frac{1}{T}\sum_t \left[\frac{RV_t}{FV_t} - \ln\left(\frac{RV_t}{FV_t}\right) - 1\right],$$

where RV is the realized variance and FV is the predicted variance. Patton shows that the MSE and $QLIKE$ criteria represent loss functions that are robust to noise in the volatility proxy. In addition, they yield inference that is invariant to the choice of units of measurement. Because $QLIKE$ depends on a standardized forecast error, it is centered approximately around 1, regardless of the level of the volatility of returns. Thus, the average $QLIKE$ loss is less affected (generally) by the most extreme observations in the sample. The MSE loss, on the other hand, with the forecast error centered approximately around zero, has a variance that is proportional to the square of the variance of returns, and is thus sensitive to extreme observations and the level of the volatility of returns.

3.2 Cross-Validation and Forward-Chained Validation

To address overfitting and selection bias, we employ the cross-validation methodology. That is, we estimate the coefficients ("trains the model") using one sub-set of the data, use the estimated coefficients to provide forecasts on another part of the data set ("tests the model"), out-of-sample, and repeat it using multiple data subsamples. More specifically, we partition the sample into 7 subsets so that each sub-sample has around

1,000 daily observations. For the first iteration, we use Subsets 1 to 6 as the training sample to estimate the coefficients and Subset 7 as the out-of-sample data for testing the model's performance. In the next iteration, we use Subsets 1-5, as well as Subset 7, to train the model and Subset 6 to test the model performance. There are a total of 7 iterations since each data subset is used once as a test sample. Table 2's panel A illustrates the methodology. For each iteration, we calculate the performance measures based on the out-of-sample prediction results in the test sample. Lastly, we average each performance measure across all 7 iterations to obtain the final cross-validation performance measures for our aforementioned 320 models.

While the cross-validation methodology is powerful to ensure that stable models are retained, six of the seven test samples partially use future information to produce forecasts. Therefore, we further consider the forward-chained methodology, which ensures that the model coefficients are estimated only using past data. For example, when using Subset 6 as the test sample, we use only Subsets 1 to 5 to estimate the model and drop Subset 7 since it contains future information. Panel B of Table 2 illustrates the forward-chain methodology. Because no model coefficients can be obtained for subset-1 without using future data, we now have only six test sub-samples.

[Insert Table 2]

The forward-chaining methodology has a few limitations. First, each test in the forward-chained validation estimates the model with a sample of a different size. In our example, while the first iteration uses 6 data subsets (around 6,000 observations) to estimate the models, the last iteration only uses 1 data subset (about 1,000 observations) to estimate the various models. Short samples may lead to inaccurate estimation of models. Since we average the performance measures across iterations, each test receives the same weight. Therefore, an inaccurate estimation due to a short estimation sample could result in poor overall forward-chained performance. A further consequence of this mechanism is that the forward-chained method tends to favor simple models since they rely less on large estimation samples. The second limitation is that earlier samples are used more heavily than later samples. In our example, subset-1 is used in all six tests, but subset-6 is only used for one test. As a result, the forward-chained method might not accurately reflect model performance over the full sample if a model has difficulty in the early part of the sample. Therefore, in our formal analysis, we use the standard cross-validation methodology as the main validation methodology and the forward-chaining methodology as a robustness check.

3.3 Horserace Regressions

The goal of the horserace regression is to statistically compare the performance of one model with a benchmark model. If a model generates forecasts that are extremely highly correlated with the simpler benchmark model, then it should not be selected, given the principles of parsimony and simplicity. We run the test against three benchmark models: lm2, lm3, and lm4. The horserace regression between model k and the benchmark model is as follows (ignoring country indicators for simplicity):

$$RV_{t+22}^{(22)} = (1-\alpha)\mathbf{E}_{t,BM}\left[RV_{t+22}^{(22)}\right] + \alpha\mathbf{E}_{t,k}\left[RV_{t+22}^{(22)}\right] + \epsilon_{t+22}, \tag{5}$$

where $\mathbf{E}_{t,BM}\left[RV_{t+22}^{(22)}\right]$ is the predicted variance using the benchmark model, $\mathbf{E}_{t,k}\left[RV_{t+22}^{(22)}\right]$ is the predicted variance using model k, and $RV_{t+22}^{(22)}$ is the actual realized variance. Here, α captures the relative explanatory power of model k compared to the benchmark model with $\alpha = 1$ ($\alpha = 0$) indicating model k (the benchmark model) fully explains future realized variances. We report t-statistics testing $\alpha = 0.5$. Rearranging Equation (5), we get equation (6), which can be easily estimated using OLS:

$$RV_{t+22}^{(22)} - \mathbf{E}_{t,BM}\left[RV_{t+22}^{(22)}\right] = \alpha\left(\mathbf{E}_{t,k}\left[RV_{t+22}^{(22)}\right] - \mathbf{E}_{t,BM}\left[RV_{t+22}^{(22)}\right]\right) + \epsilon_{t+22}. \tag{6}$$

In sum, we record three different performance criteria (BIC, $RMSE$, $QLIKE$) over two different validation techniques (standard cross-validation and forward-chained cross-validation) for each of the three countries. We use these results to select models that are "overall" great, across performance criteria, across countries, and across validation techniques.

4 Main Model Selection Results

We present model selection results using our main sample mentioned before (Germany, Japan and the U.S. from 1992 to 2019). We characterize more generally which data/model transformations work well in Section 4.1, and discuss the selection results of the winning models under the main validation techniques in Section 4.2.

4.1 The effect of logarithmic transformation, WLS, and non-linearities

The literature on volatility forecasting for US data is huge, but there is little systematic work on which transformations work best. An exception is Clements and Preve (2021) who conclude that WLS and robust estimations tend to improve on standard HAR models whereas logarithmic transformations work less well. Their sample period is quite short extending from April 1997 to August 2013. We base our analysis on the standard cross-validation results. In Table 3, we report the distribution of performance changes comparing a linear model to its transformed counterpart, using three transformation methods (WLS; logarithmic transformation; and both, i.e., using WLS on logarithmically transformed data). That is, each linear model is compared with its corresponding transformed model, e.g. lm4 versus lm4_log. As we have 15 linear models, we have 15 pairs of comparisons for each transformation; we report the 25th percentile, the average, the median, the 75th percentile, and the maximum of these performance changes. Changes are expressed as the percent differences between the transformed and the base model. To help with interpretation, we take the negative of the percentage change for $RMSE$ and $QLIKE$ so that positive (negative) numbers indicate improvement (deterioration). For BIC, the negative denominator turns a negative percentage change automatically into a positive number, so that a similar interpretation applies.

[Insert Table 3]

Table 3 reports performance change statistics for the three transformations across three performance criteria and three countries. At the median, the logarithmic and WLS transformations are uniformly better than the base linear models, whereas for WLS/log, there are two instances where the base linear model still produces lower forecast errors. The improvements are largest for the QLIKE criterion, exceeding 9% at the median for both the US and Japan. The logarithmic transformation is still uniformly better than linear models at even the 25th percentile of the distribution, suggesting that the base linear model specifications are strictly dominated by logarithmic models.

Next, we perform the same analysis for our 65 non-linear models, relegating the detailed results to Table A3 in the Online Appendix. Here, we discuss the main takeaways. The logarithmic transformation still uniformly dominates the non-transformed models for the BIC criterion. However, there are a few exceptions for the QLIKE and RMSE criteria, perhaps because the non-linear coefficients may also serve to dampen the impact of large realized variances or implied variance realizations. WLS works even better than logarithmic transformation for the non-linear models, with uniform improvement at the median and the mean (but not at the 25th percentile). Not suprisingly, the percent improvements are more modest than in the case of linear models.

We next investigate whether non-linearities help forecasting performance relative to linear models. Detailed results are presented in Table A4. Each linear model is compared with its various non-linear counterparts (with at least one of the independent variables in the model non-linear). We first compute the average performance across all corresponding non-linear models and then compare it with the performance of the linear model. We do this for standard models and then also, separately, for the three transformations (WLS; logarithmic and WLS+logarithmic). At the median, introducing non-linearities improves performance in 6 out of 9 cases (three countries $\times$ three criteria) for the standard linear model and for the WLS linear models. Nonlinear models are worst for the US in terms of the RMSE and for Germany in terms of the QLIKE criterion. For logarithmic and WLS/logarithmic models, non-linearities provide only improvement in 3, respectively 1 of the 9 cases at the median. Of course, it is conceivable that just a few of the non-linear models drag down the performance of the average non-linear corresponding model. The maximum changes are, with just a few exceptions, always better for nonlinear models relative to the corresponding linear model, and in the case of QLIKE, the percent improvement is very large (varying between 1.5% and 48%).

Overall, both data transformation and non-linear models have the potential to substantially improve on our linear benchmark models.

4.2 Cross-Validation Results

Our first step in the model selection procedure is to use the standard cross-validation procedure to compare the performance of the various models. Our goal is to find models that are robustly great forecasting models, across models and across performance metrics. We therefore rank the models per country and per performance metric and then also report the average rank, which is our overall ranking criterion. Table 4 produces the top 25 models with their rankings for the various countries and the various performance metrics; the average ranking per country for the three measures, and the overall average ranking. Table A5 in the online appendix reports all models and their respective ranks.

[Insert Table 4]

According to the overall average ranking, 23 out of the top 25 models feature non-linear coefficients and use logarithmic transformation. Of these models, 16, including the top 4, use all four predictive variables, another 5 models use only three predictive variables, leaving out the daily realized variance. Also, "lm4_log" and "lm7_log" are ranked among the top 10 models, which simply use logarithmic transformations of all four predictive variables and of all predictive variables except for the daily realized variance, respectively. These two models are of course quite parsimonious and they are also special in a different way. Table 5 shows the percent improvement of the top 25 models relative to the "lm4" (the full linear model) across all 3 measures and for all 3 countries (hence 9 numbers in a row). The "lm4_log" and "lm7_log" models are among the only 7 models that are uniformly better than the lm4 model. The most discriminating criterion is the RMSE for Germany. Note that the lm4 model is almost uniformly better than the two other benchmark models (lm2 and lm3), as shown in the bottom panel of Table 5, which is why we use lm4 as the benchmark in this table.

[Insert Table 5]

Next, we run the horserace regression of Equation (6), and conduct a t-test of the α coefficient against 0.5. The forecasts used in these regressions are the ones delivered by the cross-validation exercise in each sub-sample. The test verifies whether the model would receive a weight larger than 0.5 when competing with the forecasts of one of the three benchmark models: lm2, lm3, and lm4. A model is considered to beat the benchmark if the t-test yields a t-statistic greater than 1.645 (a 5% one-sided test). In Table 6, we report the number of models that beat each benchmark for each country. The last column indicates how many models beat a particular benchmark model for all countries. The last row reports how many models beat all benchmark models for each country. Table A6 provides a comprehensive list of these models. The number of models beating all three benchmarks per country (last row) is quite large. However, there are much fewer models beating a particular benchmark for all three countries (last column) and there are ultimately 2 models that beat all three benchmarks for all three countries. These models are lm4_log and lm7_log.

[Insert Table 6]

Table 7 reports some properties of these two models. In Panel A we report the t-statistics for the horserace tests relative to the three benchmark models. The t-statistics are invariably very large, being lowest for Germany relative to the lm4 benchmark model, where they are in the 2.5-3.0 range. Panel B reports the correlation of their forecasts with those of the benchmark models, whereas Panel C reports the same correlation statistics during crisis periods. The crisis sample comprises 2.3% of the full sample, and is defined as the union of the periods representing the 1% right tail for any of the four predictive variables. Both winning models generate forecasts that are relatively highly correlated with the benchmark forecasts. Overall, these correlations vary between 0.944 and 0.994. Invariably, these correlations are lower during crisis times, varying between 0.702 and

0.962. This is not surprising as the log transformation has more impact when risk is high.

[Insert Table 7]

One last feature of the winning models we check is their implied incidence to generate negative variance risk premiums. While theoretically the variance risk premium can be negative (see Bekaert and Engstrom, 2017; Bekaert, Engstrom, and Ermolov, 2023, for theoretical explanations based on "good" uncertainty, and Cheng, 2019, for an explanation based on dealer hedging demands), there is a strong prior that the variance risk premium should be predominantly positive. However, according to Panel D of Table 7, the benchmark models generate a large number of negative variance risk premiums, especially the Corsi model ("lm3"), with the problem least severe for the US. The simple lm2 model is best in this regard, generating only 7 negative variance risk premiums for the US during the sample period 1992-2019 while still generating 153 and 256 negative values for Germany and Japan, respectively. It is also clear from the first two rows that the lm4/7_log models are very effective in bringing down the number of negative variance risk premiums, generating fewer negative variance risk premiums than all the benchmark models with one exception.[5] Compared to lm4 – the best benchmark model given our previous evidence – the decrease in negative variance risk premiums is very substantial. This is also mostly true for crisis periods, although the lm2 benchmark model generates the least negative variance premiums in crisis times.

5 Alternative Models

We now consider several extensions of the base models. The first two models go beyond country-by-country estimation by pooling information across countries in a panel model or actually use foreign independent variables in the realized variance projections. We then add alternative independent variables to our projections, including negative returns, to capture leverage effects, as Buncic and Gisler (2017) suggest; jumps, and

[5]The lm4_log model generates 257 negative variance risk premiums for Japan, and the lm2 model 256.

semi-variances. We also consider a model which uses quarticity as an interaction term, and finally, estimate a MIDAS model (Ghysels et al., 2019) which parameterizes a flexible function of the daily variances, generalizing the HAR model. We outline the various models in more detail in Section 5.1 and discuss the results in Section 5.2.

5.1 Extending the Base Models

Panel Model We first estimate a panel model version of our model inspired by Bollerslev, Hood, Huss, and Pedersen (2018). They show that imposing the same coefficients across different asset classes (while accommodating different means) improves out-of-sample forecasting performance for volatility, suggesting that the dynamics of volatility are similar across asset classes. In our international context, it is plausible that the dynamics are similar across countries. We therefore consider a panel model with fixed effects to deal with country-specific means. Specifically, we estimate a panel model with country fixed effects using OLS. The benchmark full linear model (lm4) in a panel setting can be expressed as follows:

$$\mathbf{E}_{t,Panel}\left(RV_{i,t+22}^{(22)}\right) = \alpha_i + \hat{\beta}^m RV_{i,t}^{(22)} + \hat{\beta}^w RV_{i,t}^{(5)} + \hat{\beta}^d RV_{i,t} + \hat{\gamma} IV_{i,t}^2 \tag{7}$$

We perform the standard cross-validation exercise with every subset featuring different fixed effects. We test the panel model versions of lm4, lm4_log, and lm7_log, which we indicate by "panel," e.g. panel_lm4. We then perform the standard horse race test verifying whether country-specific models beat the panel model; for example, the horse race regression for benchmark model lm4 is as follows:

$$RV_{t+22}^{(22)} - \mathbf{E}_{t,lm4}^{Panel}\left[RV_{t+22}^{(22)}\right] = \alpha\left\{\mathbf{E}_{t,lm4}\left[RV_{t+22}^{(22)}\right] - \mathbf{E}_{t,lm4}^{Panel}\left[RV_{t+22}^{(22)}\right]\right\} + \epsilon_{t+22} \tag{8}$$

That is, in testing $\alpha = 0.5$, the alternative panel model serves as the benchmark model.

Global Model Given that there is a large global component in risk variables (see Bekaert, Hoerova, and Xu, 2023), it is conceivable that foreign variables improve forecasting power. Our current forecasts of course likely embed such a global component already. Figure 1 shows rolling correlations of our predicted variances across countries. Note that to interpret these correlations, the country perspective matters, because of the different time zones. Here, the correlations are computed from the US perspective, with the German and Japanese predicted variances taken on the same day (i.e., markets on any particular day open first in Japan, then move to Europe and final market trading occurs in the US). On average, the correlations between the US and Germany are the highest at around 0.86 for both models, Germany and Japan are 0.58 correlated and the US and Japanese forecasted variances show an average correlation of about 0.63.

Figure 1, Panel A, shows that the correlations do vary substantially over time. They were very low in the early part of the sample, but increased in the late nineties, becoming extraordinarily high during and right after the Great Financial Recession. They decrease again to near zero levels around 2015 before increasing back to the 0.6-0.8 range after 2017. In Panel B, we summarize all cross-country correlations in one statistic, namely the ratio of the variance of the average volatility to the average volatility. That is, with $v_{t,j}$ the forecasted variance at time t for country j; the ratio is

$$\frac{\sqrt{Var\left(\frac{\sum v_{t,j}}{N}\right)}}{\sum Vol(v_{t,j})/N},$$

where Vol indicates the standard deviation. This variance ratio statistic is 1 under perfect correlation, and thus is a measure of average correlation.

The graph shows a variance ratio statistic that is invariably above 0.8 and moves close to 1 after the Great Financial Recession. The 1995-1997 and 2017 periods are the only time during which the ratio dips below 0.8. We therefore do not observe trending behavior but low frequency movements around a high-level average correlation.

[Insert Figure 1]

When we consider foreign variables in forecasting, it is important to adjust for time zones. Thus, for the US forecasting equation, German and Japanese variables are from the same day. For Germany, Japanese variables are from the same day, but US variables are from the day before. For Japan, US and German variables are from the day before. Note that this naturally makes the foreign variables slightly more stale than the domestic variables, which may therefore adequately capture the global information. Still, we informally test whether foreign information helps in volatility prediction (at the monthly horizon), by testing whether the other countries' forecasts improve the country specific forecast. That is, for country j, we estimate:

$$RV_{j,t+22}^{(22)} = \omega_{j,j} Prediction_{j,t}^{(22)} + \sum_{i \neq j} \omega_{j,i} Prediction_{i,t}^{(22)} + \varepsilon_{j,t+22} \tag{9}$$

where $Prediction_{i,t}$ is the "best" forecast for country i at time t (from the same model). We minimize the variance of $\varepsilon_{j,t+22}$ with two constraints: (1) the weights adding up to one ($\sum_i \omega_{j,i} = 1$); (2) all weights must be be greater than or equal to zero ($\omega_{j,i} \geq 0$). We estimate the model as a quadratic programming problem. For our three countries case, taking Germany as an example, the model is:

$$\begin{aligned} RV_{DE,t+22}^{(22)} = \omega_{DE,DE}\ Prediction\_DE_t + \omega_{DE,JP}\ Prediction\_JP_t \\ + \omega_{DE,US}\ Prediction\_US_t + \varepsilon_{DE,t+22}, \end{aligned}$$

with $\omega_{DE,DE} + \omega_{DE,JP} + \omega_{DE,US} = 1$ and $\omega_{DE,DE}, \omega_{DE,JP}, \omega_{DE,US} \geq 0$. We minimize $\sum_t (\varepsilon_{j,t+22})^2$ for one country at a time.

The model is estimated over the full sample using the forecasts from our previous standard cross-validation exercises; we consider the benchmark model (lm4) and the two best overall models, lm4_log and lm7_log. Given that we pre-estimate the "best" country specific forecasts and use a full sample estimation, this exercise is slightly less formal than our other alternative models.

We show some key results are in Table 8; the columns indicate the countries and the rows how much weight is assigned to the forecasts of the different countries. If foreign information is not valuable at all, the diagonal elements would all be one. The US forecast has a weight between 9.6% and 11.3% in forecasting Japanese realized variances and a

4.9% weight forecasting the German variance, using the lm4 model. In forecasting US realized variances, the German forecast has a weight of 7.0% using the lm4 model. All other off-diagonal elements are effectively zero. Thus, for the standard cross-validation forecasts, overwhelmingly, foreign information is likely not very helpful.

[Insert Table 8]

"Leverage" Model As a third model, we follow Corsi and Renò (2012) and add negative returns to the standard HAR volatility forecasting model. For example, with $r_{i,t}$ the logged daily return in country i at time t, the variables of interest are negative returns at the monthly, weekly and daily level, defined as

$$r_{i,t}^{(h)-} = Min\left[r_{i,t}^{(h)}, 0\right],$$

where $r_{i,t}^{(h)} = \sum_{t=1}^{t=h} r_{i,t}$ and h takes the values 22, 5, and 1, corresponding to the monthly, weekly and daily frequencies. Specifically, the full linear model with leverage effect (leverage_lm4) is as follows:

$$\begin{aligned}\mathbf{E}_{t,Leverage}\left(RV_{i,t+22}^{(22)}\right) = \hat{\alpha}_i + \hat{\beta}_i^m RV_{i,t}^{(22)} + \hat{\beta}_i^w RV_{i,t}^{(5)} + \hat{\beta}_i^d RV_{i,t} + \hat{\gamma}_i IV_{i,t}^2 \\ + \hat{\delta}_i^m r_{i,t}^{(22)-} + \hat{\delta}_i^w r_{i,t}^{(5)-} + \hat{\delta}_i^d r_{i,t}^{(1)-}\end{aligned} \tag{10}$$

The coefficients on these negative return variables are expected to be negative to capture the well-known asymmetric volatility effect, where conditional volatility and returns are negatively correlated. We create leverage versions of our two preferred models and also of the benchmark lm4 model, which we indicate by "leverage."[6]

"Jump" Model A fourth model splits up quadratic variation into jump and non-jump components, using the concept of bipower variation developed in Barndorff-Nielsen and Shephard (2004). Bipower variation multiplies the absolute values of nearby high frequency returns to lead to an estimate of the "continuous" variation component in

[6] The effect is called "leverage effect" because one purported explanation of asymmetric volatility is that negative returns increase financial leverage and thus volatility. However, Bekaert and Wu (2000) show that asymmetric volatility is more likely driven by time-varying risk premium effects.

realized variances, which we denote as CV. Subtracting that measure from the standard quadratic variation measures delivers the jump component, denoted by J. The Man library also records the bipower variation at the daily level.[7] The resulting "jump" version of the lm4 benchmark model is:

$$\begin{aligned}\mathbf{E}_{t,Jump}\left(RV^{(22)}_{i,t+22}\right) = \hat{\alpha}_i + \hat{\beta}^m_i CV^{(22)}_{i,t} + \hat{\beta}^w_i CV^{(5)}_{i,t} + \hat{\beta}^d_i CV_{i,t} + \hat{\gamma}_i IV^2_{i,t} \\ + \hat{\delta}^m_i J^{(22)}_{i,t} + \hat{\delta}^w_i J^{(5)}_{i,t} + \hat{\delta}^d_i J_{i,t}\end{aligned} \tag{11}$$

"Downside Risk" Model A fifth model tries to embed the notion of 'bad" and "good" volatility, where "bad" volatility is associated with increased downside risk (See e.g. Chen and Ghysels, 2011; Bekaert and Engstrom, 2017). We use the simple implementation of Patton and Sheppard (2015) who create realized semi-variances using positive and negative returns to create "good" and "bad" semi variances. The "bad" semi-variance is denoted as $RS^{(-)}$. The "downside risk" model is as follows:

$$\begin{aligned}\mathbf{E}_{t,Downside}\left(RV^{(22)}_{i,t+22}\right) = \hat{\alpha}_i + \hat{\beta}^m_i RV^{(22)}_{i,t} + \hat{\beta}^w_i RV^{(5)}_{i,t} + \hat{\beta}^d_i RV_{i,t} + \hat{\gamma}_i IV^2_{i,t} \\ + \hat{\delta}^m_i RS^{(22)-}_{i,t} + \hat{\delta}^w_i RS^{(5)-}_{i,t} + \hat{\delta}^d_i RS^{-}_{i,t}\end{aligned} \tag{12}$$

Quarticity Model Our sixth alterative model interacts the realized variances in the HAR part of the model with a measure of quarticity. The noise in measuring realized variances is proportional to quarticity, which depends on returns to the fourth power (See e.g. Barndorff-Nielsen and Shephard, 2002). Bollerslev et al. (2016) focus directly on forecasting future realized variances and suggest estimating an AR(1) model with the AR(1) coefficient interacted with realized quarticity, or an HAR model with some or all of the coefficients interacted with the relevant quarticity measure. Note that the intuition here is quite similar to our interaction with levels of the realized variances in the non-linear base models. Because we do not have high frequency quarticity data, we use the fourth power of daily returns as proxies. Specifically, daily realized quarticity,

[7] Our sample for the Jump model starts from 2000 due to data availability.

RQ, is computed as $RQ_{i,t} = r_{i,t}^4/3$, where $r_{i,t}$ is daily return.[8] The weekly and monthly realized quarticity is the 5- and 22-day rolling average of RQ. Our model is as follows:

$$\begin{aligned}\mathbf{E}_{t,Quarticity}\left(RV_{i,t+22}^{(22)}\right) = {} & \hat{\alpha}_i + \hat{\beta}_i^m RV_{i,t}^{(22)} + \hat{\beta}_i^w RV_{i,t}^{(5)} + \hat{\beta}_i^d RV_{i,t} + \hat{\gamma} IV_{i,t}^2 \\ & + \hat{\delta}_i^m \sqrt{RQ_{i,t}^{(22)}} \cdot RV_{i,t}^{(22)} + \hat{\delta}_i^w \sqrt{RQ_{i,t}^{(5)}} \cdot RV_{i,t}^{(5)} + \hat{\delta}_i^d \sqrt{RQ_{i,t}} \cdot RV_{i,t}\end{aligned} \tag{13}$$

MIDAS Model Our last model is a MIDAS model. The HAR model is a special case of a MIDAS model which puts particular weights on the past daily realized daily variances, whereas the MIDAS model entertains a flexible function of the past 22 realized variances within the month. We parametrize the weights with an (exponential) Almon lag specification as in Ghysels et al. (2019). Specifically, this model can be written as follows:

$$\mathbf{E}_{t,MIDAS}\left(RV_{i,t+22}^{(22)}\right) = \hat{\alpha}_i + \hat{\phi}_i \sum_{j=0}^{K} \left(\hat{w}_i^{(j)} RV_{i,t-j}\right), \tag{14}$$

where $j = 0$ representing the last day in the month, and $j = K$ representing the first day in the month. We set $K = 22$. The weight function depends on two parameters, with the weights adding to one and always positive. The weight for day j is (ignoring country indicators for simplicity):

$$w^{(j)} = \frac{\exp\left(\theta_1 j + \theta_2 j^2\right)}{\sum_{j=0}^{K} \exp\left(\theta_1 j + \theta_2 j^2\right)}$$

Note that this specification has three parameters, two determining the weight function and one "autoregressive" parameter. This weight function can fit almost any pattern, for example, if $\theta_2 < 0$ weights decline to zero eventually. We estimate the model by nonlinear least squares, just as we did for the nonlinear AR model. We use agnostic starting values, setting $\phi = 11$ and $\theta_1 = \theta_2 = 0$, which produces equal $1/(K+1)$ weights. Note that for weights equal to $1/(K+1)$, the usual autoregressive coefficient would correspond to $\phi/(K+1)$.

[8] We subtract the sample mean of the square root of the realized quarticity, that is $\sqrt{RQ_{i,t}^{(n)}} - \overline{\sqrt{RQ_{i,t}^{(n)}}}$. Note that the sample mean will vary through time in our out-of-sample application.

5.2 Empirical Results for the Alternative Models.

To conserve space, we relegate detailed tables with the results for all alternative models to the online Appendix (Table A7 to A12). All these tables have the same format and we reproduce the one for panel models in Table 9 to illustrate. In panel A, we show t-statistics for the null hypothesis $\alpha = 0.5$. When we reject the null with positive numbers, the country specific model dominates the panel forecast, that is, the panel model serves as the benchmark model. On the left, we test the country-specific model against the panel model version of itself (e.g. lm4_log vs panel_lm4_log). On the right, we report the horserace test against the benchmark panel model (e.g. lm4_log vs. panel_lm4 and lm7_log vs. panel_lm4). The panel versions of the three models mostly underperform the corresponding country specific models, with the differences significant in 7 out of 9 cases. The exception is the lm4 model for the US, with the country specific model significantly worse than the panel_lm4 model. The lm4_log and lm7_log models obtain weights significantly higher than 0.5 in all three countries. When compared with the panel counterparts, the lm4_log model and lm7_log model are statistically significantly better for Germany and Japan but worse for the US.

[Insert Table 9]

In Panel B, we show the improvement in performance according to the various model selection criteria, where the benchmark is the lm4 model. Not surprisingly the panel-log models uniformly outperform the lm4 model and also produce less negative variance risk premiums than the panel_lm4 model. That model only improves on the benchmark lm4 model in 3 out of 9 cases. This suggests that the improvements are due to the logarithmic transformation, not the panel estimation. Indeed, the lm4_log and lm7_log models outperform the panel_lm4 model for all criteria. When comparing with the panel version of itself, lm4_log and lm7_log generate rather similar outperformance relative to the lm4 benchmark, confirming that the logarithmic transformation is the source of performance improvement.

In Panel C, we show correlations between forecasts of the lm4, lm4_log and lm7_log models and their panel counterparts. The correlations are generally high, varying between 0.942 and 0.997. Moreover, for the few cases where the panel models win the horserace or improve on a model criterion, their forecasts are more than 99% correlated with those of our preferred models. As a result, we conclude that the overall superior performance of our two selected models, the lm4_log and lm7_log models, remains largely intact.

For the remainder of this section, we characterize the main results across the various models. We start by examining the t-statistics delivered by the horse race tests. In Table 10, we show model comparisons of the alternative models versus the lm4 benchmark in the LHS columns and versus the lm4_log model in the RHS columns. Here we focus on the cross-validation tests reported in Panel A; we discuss the forward-chained tests, reported in Panel B, in Section 7.

[Insert Table 10]

Focusing first on the first three columns, positive values indicate that the alternative model outperforms the lm4 model, which is the benchmark here. The first line simply confirms that the lm4_log model significantly outperforms the lm4 benchmark model for all three countries. Because the model additions are based on the lm4_log model, it is natural to expect that these alternative models also outperform the lm4 model, but of course additional model complexity is often detrimental to out-of-sample forecasting power. We note that 5 of our 7 alternative models indeed continue to outperform the lm4 benchmark model. The exceptions are the jump model which is significantly worse than the lm4 model for all three countries and MIDAS model, which is worse for Germany and Japan.

Turning to the last three columns, we now use the lm4_log model as the benchmark model, so that positive values indicate that the model extension succeeds in beating the lm4_log model.[9] Over all 21 tests (7 alternative models for three countries), this happens 6 times. The global and leverage models outperform for Japan and the US and do so significantly. The downside model outperforms significantly for Germany and the U.S.

[9]Note that in Table 9, we considered the alternative model as the benchmark model.

For the 15 cases where the lm4_log model remains best, its outperformance is statistically significant in 10 cases.

Figure 2 shows the relative performance of these alternative models relative to the lm4_log model for our three performance metrics, BIC, RMSE and QLIKE. The countries are indicated with circles for DE, triangles for JP, and squares for the US. The vertical axis represents the performance statistics for the lm4_log model, and the horizontal axis presents performance statistics for the alternative lm4_log model. The Figure shows 7 x 3 pairs of performance statistics. Because the different performance metrics have very different units, we show the percent improvement over the lm4 benchmark model, rendering all the units similar. Also, when a performance pair is above the 45-degree line, it indicates that the lm4_log model outperforms the alternative model. For the BIC criterion, the overwhelming number of pairs are above the 45-degree line and others are just below it. Most pairs are close to the 45-degree line. For the RMSE, we observe more significant outperformance of the lm4_log model, but we now observe 6 cases where our alternative model performs better, although not dramatically so. For the QLIKE criterion, most pairs line up close to the 45-degree line (with only one exception). More often than not, alternative models outperform, but the performance improvement percentages are mostly quite close. Again, there is no alternative model that consistently outperforms the lm4_log model.

[**Insert Figure 2**]

Finally, Figure 3 focuses on the incidence of negative VRP across all alternative models. We again focus on Panel A, reporting the results for the cross-validation exercise, whereas the forward-chained results are plotted in Panel B of Figure 3. We show the incidence of negative VRPs with bar charts for the three countries for the lm4 model, the lm4_log model and the 7 alternative lm4_log models. We already know that the lm4_log model delivers fewer negative variance risk premiums than the benchmark lm4 model, and the alternative lm4_log models also generally improve upon the lm4 model. Relative to the lm4_log model, some do better, some do worse. In particular, the global, jump, downside risk and MIDAS models deliver fewer negative variance risk premiums. Of these models, the "downside risk" model is the only model that outperforms the lm4_log model in horse race tests for two countries. Clearly, this is a potentially quite successful model.

[**Insert Figure 3**]

To conclude our section on alternative models, we note that only a few models outperform our benchmark lm4_log model in a few cases. The "downside risk" model, splitting up variances in "bad" and "good" semi-variances, seems most promising, in that it also delivers quite low incidence of negative variance risk premiums. Nevertheless, all these not so parsimonious models ultimately generate forecasts that are highly correlated with the forecasts generated by our preferred lm4_log model. For example, the "downside risk" model generates forecasts that are 0.998, 0.993 and 0.997 correlated with the lm4_log forecasts for Germany, Japan and the U.S., respectively.

6 Extending the Sample to Multiple Countries

In this section, we extend our analysis to include more countries, but over a shorter sample period due to data availability. Table 11 summarizes the extended sample: Switzerland (CH), Germany (DE), France (FR), the Euro area (EA), Japan (JP), the Netherlands (NL), the United Kingdom (UK), and the United States (US). For these seven countries and the Euro area, we obtain a balanced panel from January 2000 to December 2019. This gives us about 4500 daily observations for each country.

To have a roughly similar number of observations for each subset (1,000) as in the long sample, three-country tests, we use 4 subsets for the cross-validation and forward-chained tests (instead of 7). We investigate the performance of the three benchmark models (lm4, lm3, lm2) and the two "winning" models (lm4_log and lm7_log).

[**Insert Table 11**]

Table 12 reports the results, with Panel A focusing on horserace tests. As in the previous horse race tests, lm4_log and lm7_log are tested against the three benchmark models (lm4, lm3, and lm2), and we report the results of the t-test for $\alpha = 0.5$. There are only two cases (out of 8x2x3=48 tests) in which a benchmark model beats one of our proposed models. The lm4 model delivers a weight higher than 0.5 relative to the lm7_log model for Japan, but the difference is not statistically significant. Analogously, the lm3 model is slightly but not significantly better than the lm7_log model for the US. In 44 out of 48 cases, the lm4_log and lm7_log models deliver positive and statistically significant t-statistics, in one case at only the 10% significance level, but in most cases at the 1% level. Thus, the superiority of the proposed models, especially the lm4_log model, extends to this larger country sample.

[**Insert Table 12**]

The excellent performance of the lm4_log and lm7_log models in the horserace tests also extends to their relative performance in terms of the BIC, RMSE and QLIKE criteria. In Panel B, we report the percentage improvement of our preferred models relative to the lm4 benchmark. For completeness, the two last lines also report the same statistics for the lm2 and lm3 models showing that for the different sample period and expanded country sample, the lm2 model is actually a competitive model with its performance mostly slightly better than that of the lm4 model. However, our two preferred models continue to be uniformly better than the lm4 model (with the performance differences invariably positive for all criteria and all countries) and also uniformly better than the lm2 model.

Panel C shows that the winning models still generate forecasts that are highly correlated with the forecasts of the benchmark models. These correlations are always larger than 0.9. In fact, the correlations rarely dip below 0.95, but they are substantially lower during crisis periods, especially relative to the lm3 model (see Panel D). The winning models also uniformly generate a lower incidence of negative variance risk premiums, compared to the lm3 or lm4 models (see Panel E). As we indicated before, this is not uniformly true for the lm2 model, with that model generating a lower incidence of negative variance risk premiums for Switzerland, Germany, and Japan and universally so in crisis periods.

We conclude that the lm4_log and lm7_log models not only are easy to estimate but also deliver volatility forecasts that perform well across multiple countries, across different time periods and along several performance criteria.

Finally, it would be of interest to quantify the economic benefits of using a superior volatility model. This is not easy as utility benefits also depend on expected returns. Here, we follow Bollerslev et al. (2018) who maximize a "one risky asset" mean variance utility imposing a constant Sharpe ratio, so that the allocation only varies with variance predictions. They then compare the realized utility based on using a particular volatility model to forecast future volatility to the utility obtained when having the true realized volatility. In essence, they measure the value of market timing under a particular expected return assumption and assume expected is realized future volatility under the perfect model. Instead, we simply compare the utility benefits of our preferred models to using the lm4 model.

Table 13 reports the utility difference between the lm4_log and lm7_log models relative to the lm4 model for our 8 countries/regions under both cross-validation (Panel A) and forward-chained cross-validation (Panel B). The benefits are expressed in annualized percent and can be interpreted as the extra expected return needed under the lm4 model to gain the same utility as under our preferred models (certainty equivalent). For cross-validation in Panel A, the utility benefits of the lm4_log model vary between -13.5 basis points for the Euro area, the only time the lm4_log model is inferior, and +93 basis points for Switzerland. The lm4_log model delivers uniformly better utility benefits than the lm7_log model. For the forward-chained cross-validation, the benefits are similar but larger, varying between -29.6 basis points for the Euro area and 2.41% for Germany for the lm4_log model, which is once again uniformly better than the lm7_log model. The utility benefits exceed 1% also for Switzerland, France, Japan and the Netherlands, for both the lm4_log and lm7_log models. We now further detail the results under the forward-chained validation method.

[Insert Table 13]

7 Robustness: Forward-Chained Validation

We repeat the whole analysis for the forward-chained performance results. To conserve space, we relegate the tables and more detailed discussion to the Online Appendix. When we rank models according to the various model selection criteria, the lm4_log and lm7_log model rank even better than under standard cross-validation, at numbers 4 and 5 respectively (see Table A13).[10] In terms of the other criteria we examine, the models are slightly less dominant than under standard cross-validation (see Table A14). For example, the lm4 model proves to be a very formidable model in terms of the QLIKE criterion for the US, and our two preferred models perform worse on that criterion (while still beating it across all other country/criteria combinations). Only three models uniformly outperform the lm4 model. In terms of the horse race regression, a similar issue arises, with the lm3 model constituting a difficult to beat benchmark model for Japan and US, which only 9 and 6 models can beat (see Table A16). This implies that the set of models beating all three benchmark models for all three countries is empty. However, as shown in Table A17, the lm4_log and lm7_log models are the only two models that beat all three benchmarks for the US and Germany (and they also beat the lm2 and lm4 models for Japan).

Similar to what we discussed under cross-validation, these models generate forecasts that correlate highly with those of the benchmark models (correlations varying between 0.928 and 0.991), with the correlation decreasing substantially during crises (see Table A18). They also generate fewer negative variance risk premiums. While there are now a few models that outperform the lm4_log and lm7_log models, none do so on a consistent basis and the forecast correlations of the best models are invariably high.

In terms of alternative models, Table 10, Panel B, reported the horse race tests relative to the lm4 and, importantly the lm4_log models.[11] We only observe two instances where an alternative model outperforms the lm4_log model, both occurring for Germany, namely the global and downside models. The performance of the global model is not

[10] Table A15 reports ranks for all 320 models.

[11] More results are reported in Online Appendix Table A19 to A25 and Figure A1.

surprising, as foreign information now enters in a more meaningful way, see Table A20 for results on the weights estimation. For Japan, the Japanese forecast has a weight of around 75% for the logarithmic models and about 65% for the lm4 model, with the rest assigned to the US forecast. Forecasting with the logarithmic models, the own country forecast receives weights of 87-89% for Germany and the US; where for Germany the remainder is taken up by the Japanese forecast, whereas for the US it is split between the German and Japanese forecasts (with a bit more weight on Germany). Note that these results suggest that the nearby forecasts in terms of time zone are mostly the more valuable ones (see also Bekaert and Xu, 2024).

Figure 3, Panel B, shows that the panel, global, jump and downside risk models generate mostly fewer negative variance risk premiums than the lm4_log model.

Finally, using forward-chained validation for the extended countries sample, the proposed models deliver statistically significant and positive t-statistics in 40 out of 48 cases, positive and insignificant t-statistics in 3 cases, and negative coefficients in 5 cases (see Appendix Table A26). The latter are only significantly negative for Germany relative to the lm4 model and the US relative to the lm3 model. While not as dominant as for the standard cross-validation exercise, again our proposed models perform overall much better than the benchmark models.

8 Conclusion

In this article, we initially examine 320 different forecasting models for international monthly stock return volatilities, using high frequency realized variances and the implied option variance as the predictor variables. We evaluate models that are easy to estimate, including all possible linear models and all possible non-linear models, where the coefficients depend on the level of the independent variable, so that the dependence on the past independent variables can decrease when volatility is unusually high. The latter model is estimated using non-linear least squares. Importantly, we also consider logarithmically transformed and weighted least squares estimation approaches (and a combination of the two) for all of the possible models. We demonstrate that these transformations improve forecasting accuracy and that for each linear model, a number of corresponding non-linear models outperform.

Our key result is that a logarithmically transformed Corsi (2009) model combined with the option implied variance ("lm4_log") is robustly, across countries and time, among the best forecasting models. A closely related model where the daily realized variance is left out as a predictor ("lm7_log") has almost as good performance. We estimate 7 alternative models, including a panel model as in Bollerslev et al. (2018), a "global" model, a model including negative returns, a model decomposing realized variances into jump and continuous variation components, and one using "bad" and "good" semi-variances, a model incorporating quarticity, and, finally, a MIDAS model over daily realized variances. While some of these models outperform the lm4_log model in a few instances, and also, more frequently, generate fewer negative variance risk premiums, the overall superior performance of the lm4_log model remains impressive. When alternative models have better performance, the forecasts they generate are extremely highly correlated with those of the lm4_log model.

We believe that the models we propose will prove hard to beat convincingly when parsimony, stability and robustness in forecasting are valued. Of course, even more complicated models can be estimated. For example, there is a long literature on model combination forecasts (see e.g. Wang, Ma, Wei, and Wu, 2016 for U.S. volatility), which we have not explored. Alternative non-linear models, for example, regime switching models, are worth exploring. Finally, the original development of the quadratic variation models suggest that the realized variance may follow an ARMA(1,1) process, and this model fares quite well in fitting stock specific idiosyncratic volatilities (see Bekaert et al., 2025). We leave examining such models to future research. However, we should note that our experience in examining a large variety of models for this article strongly suggests that models competitive with our proposed models, end up generating forecasts highly correlated with the "lm4_log" and "lm7_log" forecasts.

References

Andersen, Torben G, Tim Bollerslev, and Francis X Diebold, 2007, Roughing it up: Including jump components in the measurement, modeling, and forecasting of return volatility, *Review of Economics and Statistics* 89, 701–720.

Andersen, Torben G., Tim Bollerslev, Francis X. Diebold, and Paul Labys, 2003, Modeling and Forecasting Realized Volatility, *Econometrica* 71, 579–625.

Barndorff-Nielsen, Ole E, and Neil Shephard, 2002, Econometric analysis of realized volatility and its use in estimating stochastic volatility models, *Journal of the Royal Statistical Society Series B: Statistical Methodology* 64, 253–280.

Barndorff-Nielsen, Ole E, and Neil Shephard, 2004, Power and bipower variation with stochastic volatility and jumps, *Journal of Financial Econometrics* 2, 1–37.

Bekaert, Geert, Mikael Bergbrant, and Haimanot Kassa, 2025, Expected idiosyncratic volatility, *Journal of Financial Economics* 167, 104023.

Bekaert, Geert, and Eric Engstrom, 2017, Asset Return Dynamics under Habits and Bad Environment–Good Environment Fundamentals, *Journal of Political Economy* 125, 713–760.

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

Bekaert, Geert, Marie Hoerova, and Marco Lo Duca, 2013, Risk, uncertainty and monetary policy, *Journal of Monetary Economics* 60, 771–788.

Bekaert, Geert, Marie Hoerova, and Nancy R. Xu, 2023, Risk, Monetary Policy and Asset Prices in a Global World, *Working Paper* .

Bekaert, Geert, and Guojun Wu, 2000, Asymmetric Volatility and Risk in Equity Markets, *The Review of Financial Studies* 13, 1–42.

Bekaert, Geert, Eric Engstrom, and Andrey Ermolov, 2023, The Variance Risk Premium in Equilibrium Models, *Review of Finance* 27, 1977–2014.

Bollerslev, Tim, Benjamin Hood, John Huss, and Lasse Heje Pedersen, 2018, Risk Everywhere: Modeling and Managing Volatility, *The Review of Financial Studies* 31, 2729–2773.

Bollerslev, Tim, Andrew J. Patton, and Rogier Quaedvlieg, 2016, Exploiting the errors: A simple approach for improved volatility forecasting, *Journal of Econometrics* 192, 1–18.

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

Buncic, Daniel, and Katja I. M. Gisler, 2016, Global equity market volatility spillovers: A broader role for the United States, *International Journal of Forecasting* 32, 1317–1339.

Buncic, Daniel, and Katja I. M. Gisler, 2017, The role of jumps and leverage in forecasting volatility in international equity markets, *Journal of International Money and Finance* 79, 1–19.

Chen, Xilong, and Eric Ghysels, 2011, News—good or bad—and its impact on volatility predictions over multiple horizons, *The Review of Financial Studies* 24, 46–81.

Cheng, Ing-Haw, 2019, The VIX Premium, *The Review of Financial Studies* 32, 180–227.

Clements, Adam, and Daniel P. A. Preve, 2021, A Practical Guide to harnessing the HAR volatility model, *Journal of Banking & Finance* 133, 106285.

Corsi, Fulvio, 2009, A Simple Approximate Long-Memory Model of Realized Volatility, *Journal of Financial Econometrics* 7, 174–196.

Corsi, Fulvio, Francesco Audrino, and Roberto Renò, 2012, HAR Modeling for Realized Volatility Forecasting, in *Handbook of Volatility Models and Their Applications*, 363–382.

Corsi, Fulvio, and Roberto Renò, 2012, Discrete-Time Volatility Forecasting With Persistent Leverage Effect and the Link With Continuous-Time Volatility Modeling, *Journal of Business & Economic Statistics* 30, 368–380.

Ghysels, Eric, Alberto Plazzi, Rossen Valkanov, Antonio Rubia, and Asad Dossani, 2019, Direct Versus Iterated Multiperiod Volatility Forecasts, *Annual Review of Financial Economics* 11, 173–195.

Kourtis, Apostolos, Raphael N. Markellos, and Lazaros Symeonidis, 2016, An International Comparison of Implied, Realized, and GARCH Volatility Forecasts, *Journal of Futures Markets* 36, 1164–1193.

Liang, Chao, Yu Wei, and Yaojie Zhang, 2020, Is implied volatility more informative for forecasting realized volatility: An international perspective, *Journal of Forecasting* 39, 1253–1276.

Liang, Chao, Yu Wei, Likun Lei, and Feng Ma, 2020, Global equity market volatility forecasting: New evidence, *International Journal of Finance & Economics* 27, 594–609.

Liu, Lily Y., Andrew J. Patton, and Kevin Sheppard, 2015, Does anything beat 5-minute RV? A comparison of realized measures across multiple asset classes, *Journal of Econometrics* 187, 293–311.

Patton, Andrew J., 2011, Volatility forecast comparison using imperfect volatility proxies, *Journal of Econometrics* 160, 246–256.

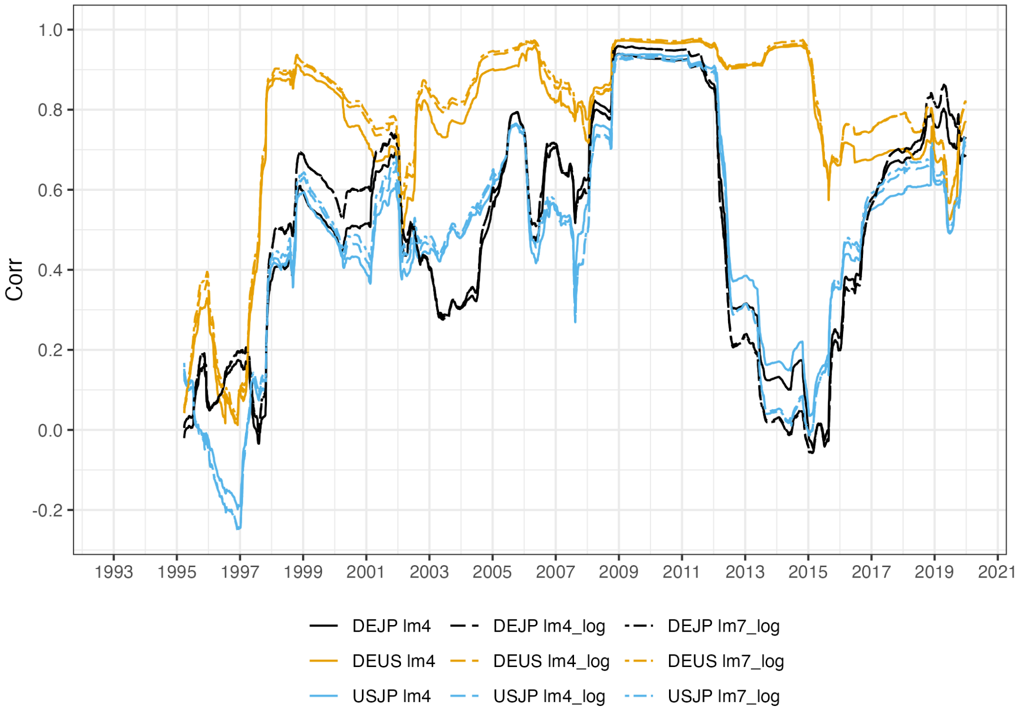
Patton, Andrew J, and Kevin Sheppard, 2015, Good volatility, bad volatility: Signed jumps and the persistence of volatility, *Review of Economics and Statistics* 97, 683–697.

Taylor, Nicholas, 2015, Realized volatility forecasting in an international context, *Applied Economics Letters* 22, 503–509.

Wang, Yudong, Feng Ma, Yu Wei, and Chongfeng Wu, 2016, Forecasting realized volatility in a changing world: A dynamic model averaging approach, *Journal of Banking & Finance* 64, 136–149.

Zhang, Yaojie, Feng Ma, and Yin Liao, 2020, Forecasting global equity market volatilities, *International Journal of Forecasting* 36, 1454–1475.

(a) Panel A: Cross-country correlation



(b) Panel B: Variance Ratio

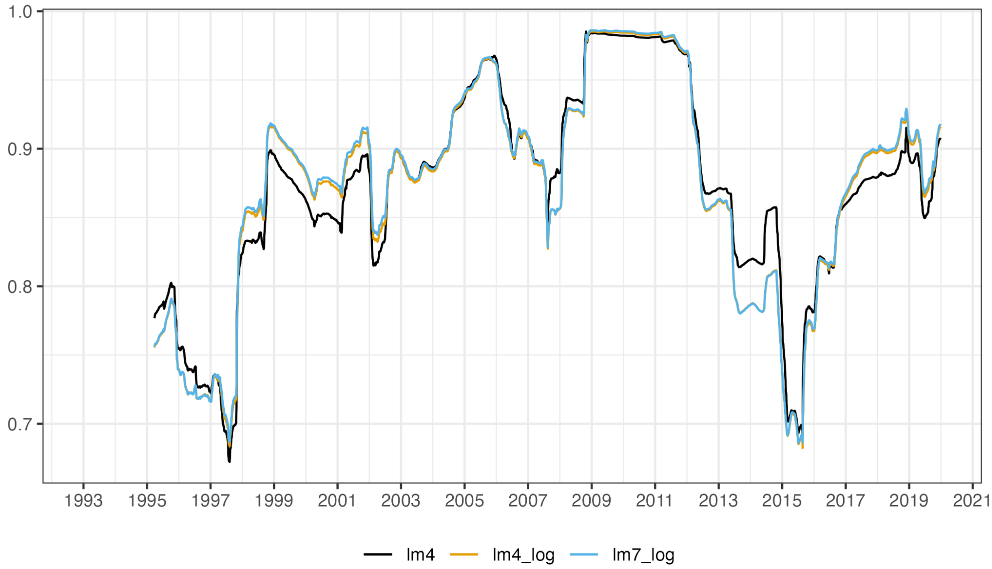


Figure 1: **Time-varying cross-country volatility correlations**

Panel A plots the rolling pairwise correlations of volatility forecasts between two countries, with the country pairs indicated in the legend. Specifically, black lines correspond to the correlations between Germany and Japan; yellow lines to the correlations between Germany and the US; and blue lines to the correlations between Japan and the US. Each pairwise correlation time series uses three models: our benchmark model and our two preferred models (lm4_log and lm7_log). The solid line represents the benchmark model, the dashed line represents lm4_log, and the dotted line represents lm7_log. The model specifications are described in Section 2, with more details in Section A of the appendix. Panel B plots the rolling variance ratio defined as $\sqrt{\mathrm{Var}\left(\frac{\sum v_{t,j}}{N}\right)} / \left(\frac{\sum \mathrm{Vol}(v_{t,j})}{N}\right)$. For more details, refer to Section 5.2 of the paper. In Panel B, the black line corresponds to the benchmark model, while the yellow and blue lines correspond to lm4_log and lm7_log, respectively. The window length for both panels is three years.

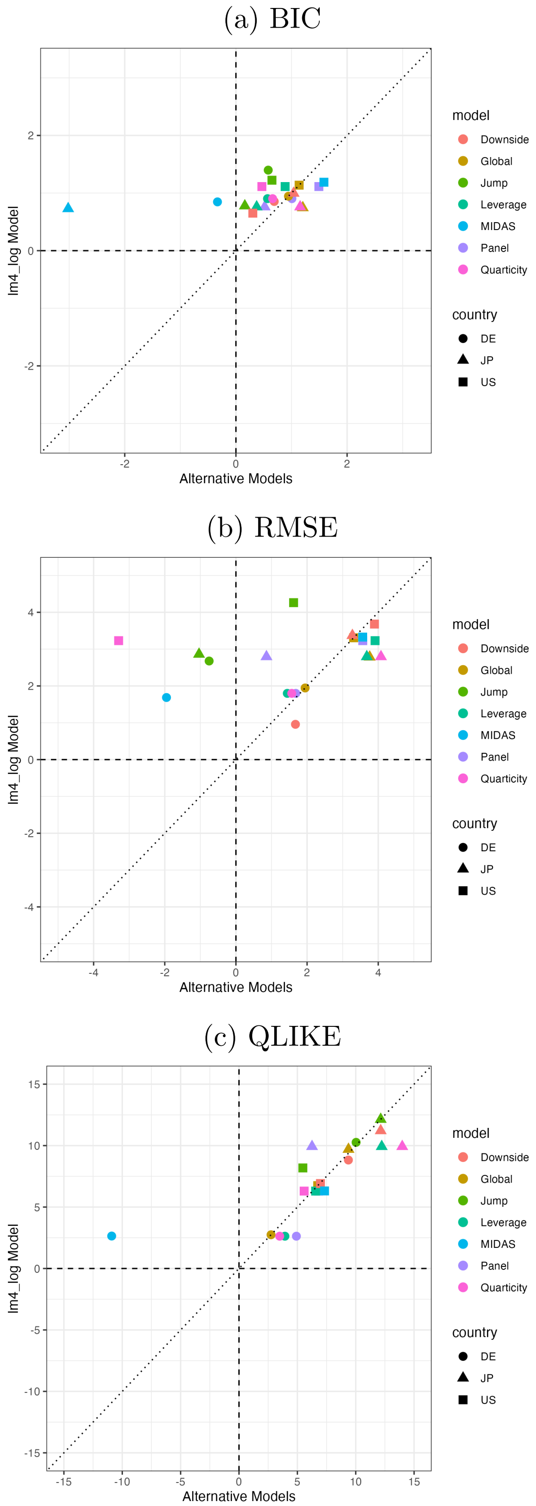
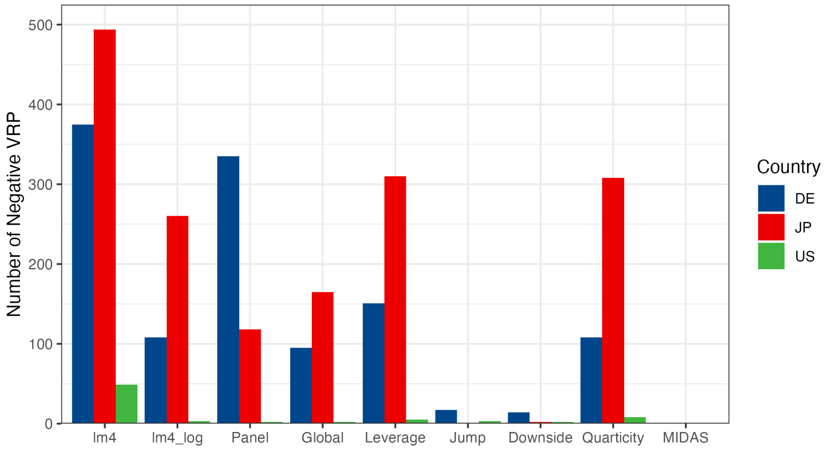


Figure 2: **Alternative Models Performance Comparison: Cross-Validation**

This figure summarizes the performance of different models using cross-validation. The X-axis shows the performance of the alternative models (lm4_log version), while the Y-axis shows the performance of the benchmark lm4_log model. Panels (a), (b), and (c) display results based on the BIC, RMSE, and QLIKE metrics, respectively. Performance is measured as percent improvement relative to the lm4-model.

(a) Cross-Validation



(b) Forward-Chained Cross-Validation

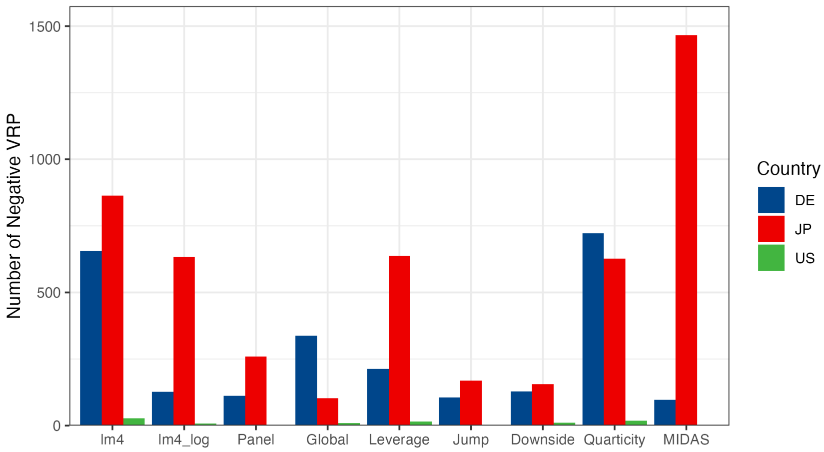


Figure 3: **Incidence of Negative Variance Risk Premiums across Alternative Models**

This figure summarizes the number of negative variance risk premia across different models. Panel (a) presents the results using cross-validation, while Panel (b) shows the results based on forward-chained.

Table 1: Linear Model Specifications

	$RV_t^{(22)}$	$RV_t^{(5)}$	RV_t	IV^2
lm1	Yes	No	No	No
lm2	Yes	No	No	Yes
lm3	Yes	Yes	Yes	No
lm4	Yes	Yes	Yes	Yes
lm5	Yes	Yes	No	No
lm6	Yes	No	Yes	No
lm7	Yes	Yes	No	Yes
lm8	Yes	No	Yes	Yes
lm9	No	Yes	No	No
lm10	No	Yes	Yes	No
lm11	No	Yes	No	Yes
lm12	No	Yes	Yes	Yes
lm13	No	No	Yes	No
lm14	No	No	Yes	Yes
lm15	No	No	No	Yes

Table 2: Model Selection Method

Panel A: Cross-Validation Example		
Iteration	Training Samples	Test Sample
1	[1, 2, 3, 4, 5, 6]	[7]
2	[1, 2, 3, 4, 5, 7]	[6]
3	[1, 2, 3, 4, 6, 7]	[5]
4	[1, 2, 3, 5, 6, 7]	[4]
5	[1, 2, 4, 5, 6, 7]	[3]
6	[1, 3, 4, 5, 6, 7]	[2]
7	[2, 3, 4, 5, 6, 7]	[1]
Panel B: Forward-Chained Example		
1	[1, 2, 3, 4, 5, 6]	[7]
2	[1, 2, 3, 4, 5]	[6]
3	[1, 2, 3, 4]	[5]
4	[1, 2, 3]	[4]
5	[1, 2]	[3]
6	[1]	[2]

Table 3: Cross-Validation: Effect of Transformations for Linear Models

This table reports the distribution of cross-validation performance changes for each transformation method, each model selection criterion, and each country. The three transformation methods are WLS, Log, and Log+WLS. The performance change is calculated as the percentage difference in the performance between the transformed model and the base linear model. The performance measures are BIC, RMSE, and QLIKE. Positive numbers indicate improvement and negative number indicates deterioration. Since there are 15 base linear models, we have 15 pairs of comparison (e.g. lm1_log vs lm1, lm2_log vs lm2, etc). We report the 25th percentile, average, median, 75th percentile, and max of the changes. All numbers are expressed in percent.

	BIC (%)			RMSE (%)			QLIKE (%)		
	DE	JP	US	DE	JP	US	DE	JP	US
P25									
WLS	0.268	0.477	0.161	0.403	-0.219	0.696	0.802	5.522	1.866
Log	0.367	0.543	0.789	1.178	1.610	1.768	0.586	7.685	3.920
Log_WLS	0.493	0.698	1.147	-4.120	1.838	0.411	-9.846	5.884	2.756
Mean									
WLS	0.961	0.997	0.497	1.246	1.110	1.174	3.946	4.685	5.540
Log	0.735	0.607	1.144	1.660	1.696	2.414	-1.680	4.173	9.323
Log_WLS	0.966	1.265	1.517	-3.392	3.185	1.865	-11.093	-3.612	6.385
Median									
WLS	0.491	0.504	0.273	0.655	0.282	1.087	2.209	7.830	5.011
Log	0.913	0.746	1.113	1.799	2.794	2.349	2.977	9.118	9.007
Log_WLS	0.867	0.954	1.308	-2.872	2.677	2.626	-5.385	7.158	5.663
P75									
WLS	1.104	1.253	0.639	1.248	1.750	1.338	9.850	8.068	9.663
Log	1.161	0.913	1.535	3.116	3.075	3.034	7.064	10.131	13.181
Log_WLS	1.027	1.799	1.714	-2.561	5.271	3.175	-2.886	8.318	10.706
Max									
WLS	3.940	2.398	2.060	5.207	5.169	3.694	11.817	8.267	11.538
Log	1.256	1.169	2.086	3.911	4.555	4.867	9.898	12.041	22.782
Log_WLS	3.458	2.467	3.742	-1.590	5.898	3.504	1.224	11.559	11.612

Table 4: Cross-Validation: Top 25 Models

This table reports the cross-validation performance for the top 25 models. Columns (2) to (10) display the ranking for each country and each measure. Column (11) reports the average ranking across all countries and all measures. Columns (12) to (14) display the average ranking across all measures for each country. The table is sorted by column (11). The last three rows report the ranking of three benchmark models (lm2, lm3, and lm4) among all 320 models.

	BIC			RMSE			QLIKE			Ave Rankings			
	DE	JP	US	DE	JP	US	DE	JP	US	Ave	DE	JP	US
nlm4_14_log	1	9	7	80	55	16	1	10	3	20.2	27.3	24.7	8.7
nlm4_11_log	3	4	4	86	39	7	2	48	4	21.9	30.3	30.3	5.0
nlm4_12_log	4	1	4	75	2	23	41	67	17	26.0	40.0	23.3	14.7
nlm4_9_log	15	8	8	131	49	15	3	16	6	27.9	49.7	24.3	9.7
lm4_log	27	76	47	9	18	6	17	62	1	29.2	17.7	52.0	18.0
nlm4_6_log	12	5	15	115	7	44	20	70	12	33.3	49.0	27.3	23.7
nlm4_5_log	13	8	24	130	27	51	9	29	18	34.3	50.7	21.3	31.0
nlm4_13_log	14	10	9	123	20	26	21	79	8	34.4	52.7	36.3	14.3
lm7_log	22	70	56	7	22	18	14	74	42	36.1	14.3	55.3	38.7
nlm7_7_log	5	6	30	73	4	40	38	75	56	36.3	38.7	28.3	42.0
nlm4_8_log	2	35	9	68	82	25	33	83	7	38.2	34.3	66.7	13.7
nlm4_1_log	36	19	17	173	69	22	8	19	21	42.7	72.3	35.7	20.0
nlm4_15_log	24	31	1	148	77	5	23	80	2	43.4	65.0	62.7	2.7
nlm4_4_log	9	39	18	113	84	41	10	88	15	46.3	44.0	70.3	24.7
nlm4_7_log	10	2	73	105	1	142	5	12	72	46.9	40.0	5.0	95.7
nlm7_6_log	17	20	35	124	27	53	16	86	59	48.6	52.3	44.3	49.0
nlm4_10_log	31	34	10	159	81	8	29	82	10	49.3	73.0	65.7	9.3
nlm4_2_log	41	18	39	176	64	76	6	24	34	53.1	74.3	35.3	49.7
nlm4_3_log	29	3	74	152	17	128	4	22	73	55.8	61.7	14.0	91.7
nlm3_5_log	97	15	23	156	42	28	125	4	14	56.0	126.0	20.3	21.7
nlm7_3_log	25	54	27	149	90	17	19	91	43	57.2	64.3	78.3	29.0
nlm7_5_log	40	41	20	160	85	19	32	90	44	59.0	77.3	72.0	27.7
nlm8_4_log	16	68	13	127	138	45	22	60	48	59.7	55.0	88.7	35.3
nlm4_12_log_w	6	36	3	141	60	10	152	132	24	62.7	99.7	76.0	12.3
nlm7_1_log	30	47	42	158	89	38	13	93	69	64.3	67.0	76.3	49.7
Benchmark													
lm2	181	229	102	118	227	123	139	243	211	174.8	146.0	233.0	145.3
lm3	219	222	230	207	150	94	262	206	271	206.8	229.3	192.7	198.3
lm4	168	210	158	95	151	97	68	210	183	148.9	110.3	190.3	146.0

Table 5: Cross-Validation: Top 25 Model Performance Improvements

This table reports the Cross-Validation performance improvements for the top25 models compared to lm4. The table is sorted by the average performance ranking across all countries and all measures. Positive numbers indicate improvement and negative numbers indicate deterioration. All numbers are expressed in percent.

	BIC (%)			RMSE (%)			QLIKE (%)		
	DE	JP	US	DE	JP	US	DE	JP	US
nlm4_14_log	1.209	1.111	1.355	0.432	1.931	2.760	4.475	11.447	6.103
nlm4_11_log	1.200	1.159	1.372	0.266	2.256	3.144	4.185	10.241	6.086
nlm4_12_log	1.170	1.275	1.372	0.485	3.501	2.598	1.560	9.814	5.669
nlm4_9_log	1.058	1.113	1.347	-1.172	2.027	2.768	3.821	11.114	5.950
lm4_log	0.903	0.760	1.113	1.799	2.794	3.230	2.628	9.924	6.299
nlm4_6_log	1.074	1.138	1.306	-0.628	3.049	2.035	2.494	9.776	5.796
nlm4_5_log	1.065	1.113	1.282	-1.160	2.570	1.888	3.266	10.693	5.667
nlm4_13_log	1.061	1.103	1.345	-0.820	2.748	2.568	2.440	9.504	5.929
lm7_log	0.970	0.788	1.041	1.844	2.688	2.735	2.725	9.658	4.736
nlm7_7_log	1.162	1.128	1.247	0.499	3.138	2.184	1.662	9.611	4.345
nlm4_8_log	1.202	0.947	1.345	0.564	1.490	2.570	1.979	9.432	5.937
nlm4_1_log	0.815	1.053	1.300	-4.093	1.676	2.671	3.381	10.939	5.600
nlm4_15_log	0.927	0.964	1.416	-3.075	1.525	3.357	2.323	9.503	6.258
nlm4_4_log	1.111	0.931	1.298	-0.593	1.469	2.079	2.970	9.261	5.754
nlm4_7_log	1.108	1.266	0.886	-0.311	3.620	-3.531	3.544	11.271	3.916
nlm7_6_log	1.049	1.050	1.207	-0.884	2.570	1.874	2.632	9.293	4.285
nlm4_10_log	0.847	0.957	1.341	-3.663	1.506	3.032	2.118	9.458	5.814
nlm4_2_log	0.782	1.059	1.203	-4.207	1.787	0.928	3.447	10.801	4.882
nlm4_3_log	0.875	1.177	0.876	-3.308	2.807	-2.274	3.741	10.894	3.871
nlm3_5_log	0.434	1.079	1.288	-3.565	2.126	2.554	-2.647	12.000	5.758
nlm7_3_log	0.923	0.864	1.273	-3.097	1.426	2.747	2.557	9.238	4.702
nlm7_5_log	0.792	0.924	1.293	-3.680	1.465	2.727	1.988	9.248	4.699
nlm8_4_log	1.053	0.793	1.309	-1.058	0.293	2.034	2.366	9.978	4.593
nlm4_12_log_w	1.148	0.945	1.374	-2.371	1.887	2.928	-3.967	7.626	5.476
nlm7_1_log	0.865	0.894	1.185	-3.639	1.431	2.289	2.762	9.150	4.009
Benchmark									
lm2	-0.070	-0.330	0.455	-0.686	-2.556	-1.206	-3.314	-3.359	-1.186
lm3	-1.038	-0.108	-0.502	-6.443	0.031	0.157	-13.587	0.394	-8.888
lm4	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000

Table 6: Cross-Validation Horserace: Number of Winning Models

This table reports the number of models that beat each benchmark model in the Cross-Validation horserace test for each country. Column (5) lists the number of models that beat each benchmark model in the Cross-Validation horserace test for all countries. The last row reports the number of models that beat all three benchmark models.

Benchmark	DE	JP	US	ALL
lm4	84	133	86	2
lm2	95	219	130	9
lm3	177	101	22	5
ALL	84	99	15	2

Table 7: Properties of Winning Models

Panel A reports the horserace test t-statistics for lm4_log and lm7_log against each benchmark model (lm2, lm3, lm4). Panel B reports the correlation of lm4_log and lm7_log with each benchmark model (lm2, lm3, lm4). Panel C reports the same correlations statistics during the crisis sample, defined as the union of the 1% right tail for any of the four predictive variables. Panel D reports the number of negative variance risk premiums for both the full sample and the crisis periods. The crisis sample comprises 2.3% of the full sample.

Panel A: Horserace t-statistics									
	Benchmark lm4			Benchmark lm3			Benchmark lm2		
	DE	JP	US	DE	JP	US	DE	JP	US
lm4_log	2.658	10.541	16.537	5.058	16.902	26.710	15.240	5.785	14.889
lm7_log	2.969	10.799	12.447	5.341	17.021	26.027	15.485	5.780	12.393
Panel B: Correlation with the benchmark									
	Benchmark lm4			Benchmark lm3			Benchmark lm2		
	DE	JP	US	DE	JP	US	DE	JP	US
lm4_log	0.986	0.969	0.994	0.984	0.948	0.986	0.944	0.948	0.946
lm7_log	0.986	0.972	0.993	0.984	0.949	0.988	0.945	0.950	0.939
Panel C: Correlation with the benchmark during crisis periods									
	Benchmark lm4			Benchmark lm3			Benchmark lm2		
	DE	JP	US	DE	JP	US	DE	JP	US
lm4_log	0.943	0.941	0.845	0.962	0.903	0.799	0.702	0.756	0.834
lm7_log	0.943	0.949	0.820	0.962	0.906	0.798	0.703	0.764	0.807
Panel D: Negative VRP									
	Full Sample			Crisis Periods					
	DE	JP	US	DE	JP	US			
lm4_log	116	257	4	12	45	3			
lm7_log	110	242	4	12	47	3			
lm2	153	256	7	0	8	7			
lm3	1816	654	422	39	73	20			
lm4	375	494	49	22	52	10			

Table 8: Global Model Estimation

This table reports the weights placed on the forecasts from the three countries for three different models (the benchmark lm4 model and the two selected models lm4_log and lm7_log), all considering the standard cross-validation forecasts. The columns indicate the models and the countries for which the forecasts are made, the three rows indicate the actual forecasts from Germany, Japan and the US. Thus, the weights add up to one in each column.

	lm4			lm4_log			lm7_log		
	DE	JP	US	DE	JP	US	DE	JP	US
CV_DE	0.951	0.006	0.070	1.000	0.000	0.000	1.000	0.000	0.000
CV_JP	0.000	0.881	0.000	0.000	0.900	0.000	0.000	0.904	0.000
CV_US	0.049	0.113	0.930	0.000	0.100	1.000	0.000	0.096	1.000

Table 9: Panel Model Results

This table summarize the results for the panel model. Panel A reports the t-statistics of horserace tests (the t-statistics for the test $\alpha = 0.5$) of each model versus the leverage model version of itself (first three columns) or the panel model version of the lm4 model (the last three columns). The sample is based on the cross-validation exercise. Panel B reports performance improvement relative to lm4. Panel C reports the correlation with the lm4, lm4_log, and lm7_log models.

Panel A: Horserace Test

	Test against panel version of itself			Test against lm4_panel		
	DE	JP	US	DE	JP	US
lm4	8.031	19.267	-14.045			
lm4_log	2.821	11.799	4.027	12.208	20.373	-4.518
lm7_log	3.354	11.590	0.679	12.547	20.590	-6.933

Panel B: Performance

	BIC			RMSE			QLIKE			Neg VRP		
	DE	JP	US	DE	JP	US	DE	JP	US	DE	JP	US
panel_lm4	-0.357	-0.027	0.116	-2.136	-3.992	2.417	-5.912	1.647	-3.817	1233	223	11
panel_lm4_log	1.007	0.517	1.492	1.682	0.858	3.565	4.917	6.257	6.818	335	118	2
panel_lm7_log	1.073	0.586	1.560	1.673	0.893	3.600	5.186	6.387	6.944	331	120	2
lm4_log	0.903	0.760	1.113	1.799	2.794	3.230	2.628	9.924	6.299	116	257	4
lm7_log	0.970	0.788	1.041	1.844	2.688	2.735	2.725	9.658	4.736	110	242	4

Panel C: Correlation with the benchmark and winning models

	lm4			lm4_log			lm7_log		
	DE	JP	US	DE	JP	US	DE	JP	US
panel_lm4	0.980	0.987	0.990	0.985	0.942	0.991	0.985	0.945	0.992
panel_lm4_log	0.973	0.979	0.988	0.996	0.984	0.994	0.996	0.985	0.997
panel_lm7_log	0.973	0.978	0.988	0.995	0.985	0.994	0.996	0.986	0.997

Table 10: Summary of Horserace Test Results for Alternative Models

This table summarizes the horserace test results for all alternative models. Panel A reports results based on cross-validation, while Panel B uses forward-chained. Each row corresponds to one model, and t-statistics are reported. All alternative models are based on the lm4_log specification. The first three columns compare each model to the lm4 benchmark, and the next three columns compare each model to lm4_log. Negative values indicate that the benchmark model (lm4 or lm4_log) outperforms the alternative; positive values indicate that the alternative model performs better.

Panel A: Cross-Validation

	Test against lm4			Test against lm4_log		
	DE	JP	US	DE	JP	US
lm4_log	2.658	10.541	16.537			
Panel	0.644	0.667	10.228	-2.821	-11.799	-4.027
Global	3.143	13.845	16.340	-0.258	**5.744**	**5.405**
Leverage	1.059	12.426	13.033	-1.042	**5.960**	**6.277**
Jump	-10.696	-8.433	-4.501	-17.579	-13.021	-16.956
Downside	5.261	1.071	14.939	**7.391**	-10.486	**6.893**
Quarticity	0.784	6.597	-21.432	-1.148	1.122	-25.825
MIDAS	-8.110	-36.664	6.662	-7.678	-36.269	-9.042

Panel B: Forward-Chained Cross-Validation

	Test against lm4			Test against lm4_log		
	DE	JP	US	DE	JP	US
lm4_log	7.837	7.805	5.848			
Panel	1.100	-22.418	-9.472	-5.249	-24.499	-11.645
Global	7.492	7.462	4.093	**2.399**	1.464	-1.932
Leverage	-5.680	-1.393	0.587	-9.461	-5.217	-2.376
Jump	-4.958	1.416	-26.430	-4.888	-2.432	-29.034
Downside	0.944	1.562	4.188	**7.359**	-5.788	0.010
Quarticity	-21.970	-2.472	-18.310	-25.225	-7.913	-18.252
MIDAS	-4.483	-23.646	-3.203	-13.911	-23.703	-15.088

Table 11: Extended Sample Summary

Country	Sample Size	Starting Date	Ending Date
CH	5008	2000-01-04	2019-12-30
DE	5070	2000-01-03	2019-12-30
EA	5098	2000-01-03	2019-12-31
FR	5098	2000-01-03	2019-12-31
JP	4886	2000-01-04	2019-12-30
NL	5098	2000-01-03	2019-12-31
UK	5043	2000-01-04	2019-12-31
US	5017	2000-01-03	2019-12-31

Table 12: Extended sample

The table summarizes the results for the extended sample. Panel A reports the horserace t-statistics for each country's lm4_log and lm7_7 log models against each benchmark model. Panel B reports the performance improvement for each country in terms of each criterion. Panels C and D report the correlation with each benchmark model for the full sample and during crisis periods. Panel E reports the number of negative variance risk premiums for the full sample and crisis periods.

Panel A: Horserace t-statistics

	Benchmark lm4								Benchmark lm3								Benchmark lm2							
	CH	DE	EA	FR	JP	NL	UK	US	CH	DE	EA	FR	JP	NL	UK	US	CH	DE	EA	FR	JP	NL	UK	US
lm4_log	15.254	4.318	4.260	7.818	2.540	7.342	15.445	16.678	12.447	3.911	10.632	8.776	6.084	7.695	8.231	0.230	14.575	8.253	5.706	10.109	8.618	11.150	14.152	19.156
lm7_log	13.598	1.656	3.991	5.780	-1.272	5.471	13.763	14.804	10.508	2.236	10.552	7.497	3.834	6.489	7.847	-0.368	13.858	6.082	5.486	9.003	6.232	10.069	13.260	18.499

Panel B: Performance improvement

	BIC								RMSE								QLIKE							
	CH	DE	EA	FR	JP	NL	UK	US	CH	DE	EA	FR	JP	NL	UK	US	CH	DE	EA	FR	JP	NL	UK	US
lm4_log	1.514	1.714	0.461	1.078	0.657	0.964	0.775	1.399	6.858	4.437	1.304	3.204	2.151	3.272	2.909	4.589	16.254	12.971	2.333	13.422	10.706	13.858	11.033	8.660
lm7_log	1.501	1.666	0.535	1.077	0.599	0.991	0.847	1.447	6.385	3.830	1.292	2.887	1.293	2.977	2.801	4.425	15.401	12.102	2.298	12.022	9.744	13.133	10.697	7.583
lm2	0.404	-0.060	0.690	0.462	0.421	0.764	0.937	1.256	0.335	-1.417	0.858	0.150	-0.898	0.247	1.731	1.359	-1.622	-0.007	1.472	1.662	-2.581	-0.043	4.055	2.385
lm3	-0.790	-0.575	-2.913	-1.398	-0.768	-1.708	-2.360	-0.510	-1.747	-1.524	-10.204	-4.186	-3.115	-4.973	-6.415	0.696	-11.322	-17.718	-39.461	-17.475	-4.908	-12.187	-26.135	-9.528

Panel C: Correlation with the benchmark

	Benchmark lm4								Benchmark lm3								Benchmark lm2							
	CH	DE	EA	FR	JP	NL	UK	US	CH	DE	EA	FR	JP	NL	UK	US	CH	DE	EA	FR	JP	NL	UK	US
lm4_log	0.979	0.991	0.995	0.987	0.990	0.989	0.988	0.995	0.960	0.976	0.921	0.972	0.966	0.972	0.919	0.966	0.960	0.982	0.992	0.981	0.967	0.980	0.988	0.991
lm7_log	0.978	0.988	0.995	0.983	0.988	0.986	0.987	0.995	0.956	0.976	0.921	0.968	0.963	0.968	0.918	0.965	0.963	0.980	0.991	0.981	0.970	0.980	0.988	0.992

Panel D: Correlation with the benchmark during crisis periods

	Benchmark lm4								Benchmark lm3								Benchmark lm2							
	CH	DE	EA	FR	JP	NL	UK	US	CH	DE	EA	FR	JP	NL	UK	US	CH	DE	EA	FR	JP	NL	UK	US
lm4_log	0.956	0.973	0.976	0.954	0.988	0.962	0.951	0.977	0.768	0.909	0.834	0.920	0.912	0.888	0.773	0.877	0.902	0.949	0.969	0.944	0.938	0.963	0.966	0.968
lm7_log	0.943	0.957	0.974	0.922	0.978	0.942	0.944	0.974	0.732	0.906	0.834	0.897	0.897	0.867	0.775	0.873	0.913	0.934	0.967	0.934	0.950	0.956	0.964	0.974

Panel E: Negative VRP

	Full Sample								Crisis Periods							
	CH	DE	EA	FR	JP	NL	UK	US	CH	DE	EA	FR	JP	NL	UK	US
lm4_log	5	10	0	11	1	1	32	3	4	6	0	3	0	0	4	2
lm7_log	8	12	0	13	4	1	24	3	4	7	0	3	0	0	5	3
lm2	1	0	0	27	0	6	336	7	0	0	0	0	0	0	1	0
lm3	193	577	1078	626	37	421	1746	396	6	20	17	11	5	8	16	11
lm4	13	10	8	117	9	68	699	48	4	8	7	5	3	5	11	6

Table 13: Economic Benefits

This table reports the economic benefits of using superior volatility models. We compute economic benefits of different volatility models follow Bollerslev et al. (2018) who maximize a "one risky asset" mean variance utility imposing a constant Sharpe ratio, so that the allocation only varies with variance predictions. Specifically, the realized utility based on using a particular volatility model is compared with the utility obtained using the benchmark lm4 model. Positive (negative) number means utility improvement (deterioration). The benefits are expressed in annualized percent and can be interpreted as the extra expected return needed under the lm4 model to gain the same utility as under our preferred models (certainty equivalent). Panel A reports the results using cross-validation and Panel B shows forward-chained results.

Panel A: Cross-Validation								
	CH	DE	EA	FR	JP	NL	UK	US
lm2	-0.177	-0.055	0.004	0.038	-0.199	-0.108	0.233	-0.340
lm3	-0.827	-1.297	-2.576	-1.206	-0.728	-0.857	-1.833	-0.876
lm4	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
lm4_log	0.933	0.518	-0.135	0.604	0.338	0.707	0.620	0.203
lm7_log	0.852	0.447	-0.136	0.452	0.278	0.634	0.580	0.034
Panel B: Forward-Chained Cross-Validation								
	CH	DE	EA	FR	JP	NL	UK	US
lm2	-0.073	-0.015	0.034	0.006	-0.354	0.058	0.158	-0.278
lm3	-1.532	-1.671	-3.330	-1.542	-0.492	-1.414	-2.099	-1.681
lm4	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
lm4_log	1.171	2.431	-0.296	1.745	1.089	1.663	0.773	0.400
lm7_log	1.101	2.346	-0.298	1.593	1.016	1.585	0.715	0.287

A Online Appendix

Table A1: Full Non-Linear Model Specification

	$RV_t^{(22)}$	$RV_t^{(5)}$	RV_t	IV^2
nlm4-1	NL	NL	NL	NL
nlm4-2	L	NL	NL	NL
nlm4-3	NL	L	NL	NL
nlm4-4	NL	NL	L	NL
nlm4-5	NL	NL	NL	L
nlm4-6	NL	NL	L	L
nlm4-7	NL	L	NL	L
nlm4-8	NL	L	L	NL
nlm4-9	L	NL	NL	L
nlm4-10	L	NL	L	NL
nlm4-11	L	L	NL	NL
nlm4-12	NL	L	L	L
nlm4-13	L	NL	L	L
nlm4-14	L	L	NL	L
nlm4-15	L	L	L	NL

Table A2: Rest of Non-Linear Model Specification

	$RV_t^{(22)}$	$RV_t^{(5)}$	RV_t	IV^2
nlm1-1	NL	No	No	No
nlm9-1	No	NL	No	No
nlm13-1	No	No	NL	No
nlm15-1	No	No	No	NL
nlm2-1	NL	No	No	NL
nlm2-2	L	No	No	NL
nlm2-3	NL	No	No	L
nlm5-1	NL	NL	No	No
nlm5-2	L	NL	No	No
nlm5-3	NL	L	No	No
nlm6-1	NL	No	NL	No
nlm6-2	L	No	NL	No
nlm6-3	NL	No	L	No
nlm10-1	No	NL	NL	No
nlm10-2	No	L	NL	No
nlm10-3	No	NL	L	No
nlm11-1	No	NL	No	NL
nlm11-2	No	L	No	NL
nlm11-3	No	NL	No	L
nlm14-1	No	No	NL	NL
nlm14-2	No	No	L	NL
nlm14-3	No	No	NL	L
nlm3-1	NL	NL	NL	No
nlm3-2	L	NL	NL	No
nlm3-3	NL	L	NL	No
nlm3-4	NL	NL	L	No
nlm3-5	L	L	NL	No
nlm3-6	L	NL	L	No
nlm3-7	NL	L	L	No
nlm7-1	NL	NL	No	NL
nlm7-2	L	NL	No	NL
nlm7-3	NL	L	No	NL
nlm7-4	NL	NL	No	L
nlm7-5	L	L	No	NL
nlm7-6	L	NL	No	L
nlm7-7	NL	L	No	L
nlm8-1	NL	No	NL	NL
nlm8-2	L	No	NL	NL
nlm8-3	NL	No	L	NL
nlm8-4	NL	No	NL	L
nlm8-5	L	No	L	NL
nlm8-6	L	No	NL	L
nlm8-7	NL	No	L	L
nlm12-1	No	NL	NL	NL
nlm12-2	No	L	NL	NL
nlm12-3	No	NL	L	NL
nlm12-4	No	NL	NL	L
nlm12-5	No	L	L	NL
nlm12-6	No	L	NL	L
nlm12-7	No	NL	L	L

A.1 Additional Cross-Validation Results

Table A3: Cross-Validation: Effect of Transformations for Non-Linear Models

This table reports the distribution of cross-validation performance changes for each transformation method, each model selection criterion, and each country. Three transformation methods are WLS, Log, and Log+WLS. The performance change is calculated as a percentage change in the performance measures between the transformed model and the baseline non-linear model. The performance measures are BIC, RMSE, and QLIKE. Positive numbers indicate improvement and negative numbers indicate deterioration. Since there are 65 base non-linear models, we have 65 pair of comparison (e.g. nlm4_1_log vs nlm4_1, nlm4_2_log vs nlm4_2, etc). We report the 25th percentile, the average, the median, the 75th percentile, and the maximum changes. All numbers are expressed in percent.

	BIC (%)			RMSE (%)			QLIKE (%)		
	DE	JP	US	DE	JP	US	DE	JP	US
P25									
WLS	0.019	0.081	0.006	-0.133	0.212	0.567	1.266	-0.765	-0.610
Log	0.286	0.112	1.090	-4.693	-0.565	7.785	0.937	-1.256	3.164
Log_WLS	-7.255	-1.838	-0.151	-172.052	-20.250	-9.950	-4.866	-4.319	2.654
Mean									
WLS	0.326	0.422	0.237	0.170	0.966	2.485	5.392	1.239	0.395
Log	0.461	0.143	1.359	-3.735	-0.873	9.993	7.067	2.318	8.414
Log_WLS	-4.664	-1.059	0.759	-97.612	-14.873	2.294	0.680	-1.019	7.598
Median									
WLS	0.297	0.253	0.195	0.330	0.707	2.694	3.833	0.565	0.246
Log	0.553	0.272	1.399	-2.653	0.348	9.886	7.043	1.449	4.699
Log_WLS	-5.105	-0.553	1.139	-79.738	-5.688	7.674	1.321	-1.016	3.913
P75									
WLS	0.501	0.596	0.422	0.786	1.208	4.411	9.506	3.707	2.437
Log	0.860	0.410	1.570	-1.465	0.941	11.896	11.340	5.821	7.980
Log_WLS	-2.861	-0.021	1.435	-43.704	-2.344	11.251	4.281	2.269	7.339
Max									
WLS	2.466	2.182	2.105	1.715	5.135	7.154	23.143	7.021	7.273
Log	1.709	0.761	2.178	2.574	1.748	17.925	25.435	10.325	46.955
Log_WLS	1.551	0.697	1.822	1.309	1.259	13.438	17.610	8.415	47.316

Table A4: Cross-Validation: Effect of Nonlinearity

This table reports the distribution of cross-validation performance changes of using nonlinearity for each linear model category, each model selection criteria, and each country. There are multiple non-linear counterparts for each linear model. For example, lm4 is compared to nlm4-1, nlm4-2, etc and lm3 is compared to nlm3-1, nlm3-2, etc. We first compute the average performance across all corresponding non-linear models and then compare it with the linear model. We also compare the transformed non-linear model with the transformed linear model, e.g. lm4_log vs nlm4-1_log, nlm4-2_log, etc. We report the 25th percentile, the average, the median, the 75th percentile, and the maximum changes. The change is expressed as the percentage difference between the transformed and the base model. Positive numbers indicate improvement and negative numbers indicate deterioration. All numbers are expressed in percent.

	BIC (%)			RMSE (%)			QLIKE (%)		
	DE	JP	US	DE	JP	US	DE	JP	US
P25									
NLM	0.005	0.648	-0.342	0.372	1.290	-11.386	-10.726	2.612	-2.420
NLM_w	-0.057	0.315	-0.405	0.405	1.941	-10.232	-2.491	0.542	-4.742
NLM_log	-0.103	0.106	0.038	-5.585	-1.342	-1.236	-0.210	-0.467	-0.704
NLM_log_w	-7.808	-1.929	-1.598	-162.777	-21.550	-21.604	-0.454	-1.014	-0.447
Mean									
NLM	0.343	0.736	-0.158	1.379	1.945	-9.341	-4.013	7.019	-2.738
NLM_w	0.049	0.418	-0.232	0.802	2.313	-7.760	-1.509	1.713	-7.513
NLM_log	-0.020	0.197	0.082	-4.179	-1.138	-0.984	1.587	1.280	-0.519
NLM_log_w	-5.070	-1.315	-0.760	-89.019	-15.860	-9.327	3.255	0.773	0.470
Median									
NLM	0.218	0.718	-0.170	1.030	1.658	-8.985	-4.117	7.926	0.191
NLM_w	0.058	0.410	-0.243	0.990	2.163	-7.357	-1.953	2.305	-2.706
NLM_log	0.024	0.229	0.141	-3.618	-0.790	-0.441	0.571	-0.119	-0.395
NLM_log_w	-5.360	-0.793	-0.552	-70.245	-7.508	-2.106	1.807	-0.230	0.079
P75									
NLM	0.633	0.824	0.019	1.906	2.418	-5.850	3.432	10.601	2.116
NLM_w	0.155	0.503	-0.001	1.484	2.610	-3.272	-0.439	2.803	-1.122
NLM_log	0.161	0.333	0.186	-2.213	-0.072	-0.010	1.896	1.609	-0.054
NLM_log_w	-3.602	-0.211	-0.120	-38.194	-2.016	-0.283	5.039	0.789	0.853
Max									
NLM	1.499	1.225	0.452	7.019	4.342	-3.560	14.841	12.951	10.244
NLM_w	0.422	1.053	0.221	3.573	4.601	-0.476	1.408	3.933	1.213
NLM_log	0.305	0.532	0.300	-0.028	0.849	1.014	47.131	31.919	1.511
NLM_log_w	0.440	0.255	0.233	3.530	-0.130	1.396	41.334	47.770	8.411

Table A5: Cross-Validation: All 320 Model Ranking

	BIC			RMSE			QLIKE			Rankings			
	DE	JP	US	DE	JP	US	DE	JP	US	Ave	DE	JP	US
nlm4_14_log	1	9	7	80	55	16	1	10	3	20.2	27.3	24.7	8.7
nlm4_11_log	3	4	4	86	39	7	2	48	4	21.9	30.3	30.3	5.0
nlm4_12_log	4	1	4	75	2	23	41	67	17	26.0	40.0	23.3	14.7
nlm4_9_log	15	8	8	131	49	15	3	16	6	27.9	49.7	24.3	9.7
lm4_log	27	76	47	9	18	6	17	62	1	29.2	17.7	52.0	18.0
nlm4_6_log	12	5	15	115	7	44	20	70	12	33.3	49.0	27.3	23.7
nlm4_5_log	13	8	24	130	27	51	9	29	18	34.3	50.7	21.3	31.0
nlm4_13_log	14	10	9	123	20	26	21	79	8	34.4	52.7	36.3	14.3
lm7_log	22	70	56	7	22	18	14	74	42	36.1	14.3	55.3	38.7
nlm7_7_log	5	6	30	73	4	40	38	75	56	36.3	38.7	28.3	42.0
nlm4_8_log	2	35	9	68	82	25	33	83	7	38.2	34.3	66.7	13.7
nlm4_1_log	36	19	17	173	69	22	8	19	21	42.7	72.3	35.7	20.0
nlm4_15_log	24	31	1	148	77	5	23	80	2	43.4	65.0	62.7	2.7
nlm4_4_log	9	39	18	113	84	41	10	88	15	46.3	44.0	70.3	24.7
nlm4_7_log	10	2	73	105	1	142	5	12	72	46.9	40.0	5.0	95.7
nlm7_6_log	17	20	35	124	27	53	16	86	59	48.6	52.3	44.3	49.0
nlm4_10_log	31	34	10	159	81	8	29	82	10	49.3	73.0	65.7	9.3
nlm4_2_log	41	18	39	176	64	76	6	24	34	53.1	74.3	35.3	49.7
nlm4_3_log	29	3	74	152	17	128	4	22	73	55.8	61.7	14.0	91.7
nlm3_5_log	97	15	23	156	42	28	125	4	14	56.0	126.0	20.3	21.7
nlm7_3_log	25	54	27	149	90	17	19	91	43	57.2	64.3	78.3	29.0
nlm7_5_log	40	41	20	160	85	19	32	90	44	59.0	77.3	72.0	27.7
nlm8_4_log	16	68	13	127	138	45	22	60	48	59.7	55.0	88.7	35.3
nlm4_12_log_w	6	36	3	141	60	10	152	132	24	62.7	99.7	76.0	12.3
nlm7_1_log	30	47	42	158	89	38	13	93	69	64.3	67.0	76.3	49.7
nlm8_6_log	11	111	12	120	187	36	12	71	40	66.7	47.7	123.0	29.3
nlm3_2_log	137	11	28	201	36	27	137	9	20	67.3	158.3	18.7	25.0
nlm7_4_log	21	25	77	126	25	159	11	59	126	69.9	52.7	36.3	120.7
nlm3_3_log	129	6	40	196	30	55	155	2	29	71.3	160.0	12.7	41.3
nlm3_1_log	143	7	36	214	26	43	143	7	26	71.7	166.7	13.3	35.0
lm4_log_w	51	93	14	143	40	4	163	125	19	72.4	119.0	86.0	12.3
nlm4_12_w	53	29	140	28	13	155	57	64	118	73.0	46.0	35.3	137.7
nlm3_6_log	136	23	25	202	24	29	161	47	22	74.3	166.3	31.3	25.3
nlm7_7_w	49	33	145	27	8	152	59	39	159	74.6	45.0	26.7	152.0
lm3_log	141	69	45	144	9	52	177	30	25	76.9	154.0	36.0	40.7
nlm4_14_log_w	8	154	5	122	234	2	71	108	5	78.8	67.0	165.3	4.0
lm7_log_w	39	86	32	142	46	9	161	135	63	79.2	114.0	89.0	34.7
nlm4_12	45	32	116	18	16	143	229	3	127	81.0	97.3	17.0	128.7
nlm7_2_log	33	52	55	163	105	106	24	96	105	82.1	73.3	84.3	88.7
lm8_log	34	170	44	62	160	21	77	148	33	83.2	57.7	159.3	32.7
nlm3_7_log	112	24	37	185	19	90	190	44	54	83.9	162.3	29.0	60.3
nlm8_5_log	7	174	2	65	230	20	72	158	35	84.8	48.0	187.3	19.0
nlm4_7_w	61	16	167	23	3	214	65	68	147	84.9	49.7	29.0	176.0
nlm8_2_log	47	123	16	183	203	32	34	78	50	85.1	88.0	134.7	32.7
nlm8_1_log	46	118	26	180	197	39	35	73	53	85.2	87.0	129.3	39.3
nlm12_5_log	23	150	66	21	142	60	37	230	39	85.3	27.0	174.0	55.0
nlm4_14_w	76	46	107	59	117	114	62	120	68	85.4	65.7	94.3	96.3
nlm4_8_w	62	61	144	48	48	154	64	51	142	86.0	58.0	53.3	146.7
nlm4_13_w	99	42	138	39	37	158	74	76	111	86.0	70.7	51.7	135.7
nlm4_6_w	44	80	155	1	114	165	46	58	115	86.4	30.3	84.0	145.0
nlm7_7	42	38	129	16	21	151	244	5	137	87.0	100.7	21.3	139.0
nlm4_11_w	104	13	121	91	35	122	93	119	85	87.0	96.0	55.7	109.3
lm5_log	133	57	65	146	11	83	160	25	104	87.1	146.3	31.0	84.0
nlm4_4_w	92	63	148	32	121	161	26	13	141	88.6	50.0	65.7	150.0
nlm4_15_w	112	14	109	129	51	100	81	129	74	88.8	107.3	64.7	94.3
nlm4_3_w	71	26	169	33	29	208	69	38	166	89.9	57.7	31.0	181.0
nlm4_1_w	91	37	166	25	31	194	36	69	161	90.0	50.7	45.7	173.7

Table A5: Cross-Validation: All 320 Model Ranking *(continued)*

	BIC			RMSE			QLIKE			Rankings			
	DE	JP	US	DE	JP	US	DE	JP	US	Ave	DE	JP	US
nlm7_3_w	55	64	149	44	51	150	66	66	168	90.3	55.0	60.3	155.7
nlm5_2_log	142	26	59	212	23	71	140	45	98	90.7	164.7	31.3	76.0
nlm7_5_w	86	22	111	98	59	89	92	144	116	90.8	92.0	75.0	105.3
nlm4_5_w	69	62	167	6	65	210	51	56	150	92.9	42.0	61.0	175.7
nlm5_1_log	148	21	65	218	14	86	148	35	108	93.7	171.3	23.3	86.3
nlm5_3_log	117	12	64	190	6	113	176	32	139	94.3	161.0	16.7	105.3
nlm4_10_w	105	48	142	52	45	164	84	99	134	97.0	80.3	64.0	146.7
lm12_log	54	199	93	17	165	56	42	226	37	98.8	37.7	196.7	62.0
nlm3_4_log	140	17	87	213	5	149	171	28	79	98.8	174.7	16.7	105.0
nlm7_4	96	53	217	20	32	251	30	6	191	99.6	48.7	30.3	219.7
lm4_w	114	144	132	83	156	69	25	124	60	100.8	74.0	141.3	87.0
nlm4_9_w	103	30	147	78	34	185	94	114	123	100.9	91.7	59.3	151.7
nlm7_6_w	96	49	162	34	41	193	78	89	170	101.3	69.3	59.7	175.0
nlm8_3_log	35	166	6	172	228	24	87	157	41	101.8	98.0	183.7	23.7
nlm7_4_w	67	40	201	10	10	258	75	34	226	102.3	50.7	28.0	228.3
nlm4_2_w	106	45	156	46	38	192	83	104	154	102.7	78.3	62.3	167.3
lm2_log	26	155	83	60	154	81	58	127	202	105.1	48.0	145.3	122.0
lm7_w	102	140	131	82	166	72	31	133	91	105.3	71.7	146.3	98.0
nlm2_3_log	18	115	63	110	129	95	86	128	214	106.4	71.3	124.0	124.0
lm11_log	52	213	92	8	181	62	44	244	81	108.6	34.7	212.7	78.3
nlm7_1_w	61	75	202	13	63	255	27	54	237	109.7	33.7	64.0	231.3
nlm8_6_log_w	19	242	18	132	263	13	100	155	46	109.8	83.7	220.0	25.7
nlm7_2_w	100	56	168	43	53	196	88	109	181	110.4	77.0	72.7	181.7
nlm12_6_log	38	207	76	111	245	64	7	203	47	110.9	52.0	218.3	62.3
nlm12_7_log	56	157	84	137	155	85	43	234	58	112.1	78.7	182.0	75.7
lm12_log_w	75	58	53	164	70	57	210	239	101	114.1	149.7	122.3	70.3
nlm8_7_w	74	79	154	51	67	212	112	84	196	114.3	79.0	76.7	187.3
nlm4_8_log_w	252	55	7	274	83	14	198	137	16	115.1	241.3	91.7	12.3
nlm8_7_log	20	134	62	112	157	166	101	146	145	115.9	77.7	145.7	124.3
nlm4_4	43	99	184	3	108	213	227	8	164	116.6	91.0	71.7	187.0
nlm4_6	57	47	241	4	28	241	209	1	224	116.9	90.0	25.3	235.3
nlm4_13_log_w	243	135	11	265	169	1	89	130	9	116.9	199.0	144.7	7.0
nlm8_5_w	109	60	108	119	126	117	126	169	129	118.1	118.0	118.3	118.0
lm8_log_w	50	190	19	153	188	12	203	188	61	118.2	135.3	188.7	30.7
nlm6_2_log	135	124	39	194	186	54	191	63	82	118.7	173.3	124.3	58.3
lm3_log_w	189	65	41	234	15	82	275	77	93	119.0	232.7	52.3	72.0
nlm4_5	59	76	238	2	33	262	204	11	197	120.2	88.3	40.0	232.3
nlm8_7	58	87	136	30	78	202	258	49	189	120.8	115.3	71.3	175.7
nlm8_6_w	93	93	114	87	163	131	106	164	136	120.8	95.3	140.0	127.0
lm11_log_w	64	59	54	166	88	61	221	262	130	122.8	150.3	136.3	81.7
nlm8_2_w	136	50	120	133	96	129	133	163	153	123.7	134.0	103.0	134.0
nlm4_8	94	89	119	54	106	188	296	37	135	124.2	148.0	77.3	147.3
nlm12_5_w	80	28	151	90	131	134	98	254	157	124.8	89.3	137.7	147.3
nlm8_4_w	89	67	182	45	44	256	123	87	234	125.2	85.7	66.0	224.0
lm5_log_w	174	51	57	233	12	103	270	57	175	125.8	225.7	40.0	111.7
nlm12_1_log	121	162	82	208	191	68	28	205	67	125.8	119.0	186.0	72.3
nlm3_5_log_w	122	170	51	199	244	42	228	42	34	125.8	183.0	152.0	42.3
nlm12_4_log	68	194	91	145	226	115	15	204	76	126.0	76.0	208.0	94.0
nlm7_7_log_w	218	102	33	257	127	31	146	152	71	126.3	207.0	127.0	45.0
nlm4_3	144	75	141	69	73	207	298	26	106	126.6	170.3	58.0	151.3
lm8_w	118	163	122	99	201	87	70	167	113	126.7	95.7	177.0	107.3
nlm7_6_log_w	232	147	43	256	179	11	67	140	66	126.8	185.0	155.3	40.0
nlm4_4_log_w	261	92	29	277	125	35	165	136	23	127.0	234.3	117.7	29.0
nlm4_9_log_w	242	186	21	264	243	3	63	112	11	127.2	189.7	180.3	11.7
nlm3_5_w	171	66	130	178	107	124	141	113	119	127.7	163.3	95.3	124.3
nlm12_6_w	72	91	139	56	182	130	82	266	132	127.8	70.0	179.7	133.7
nlm6_1_log	172	113	49	225	161	74	211	52	103	128.9	202.7	108.7	75.3

Table A5: Cross-Validation: All 320 Model Ranking *(continued)*

	BIC			RMSE			QLIKE			Rankings			
	DE	JP	US	DE	JP	US	DE	JP	US	Ave	DE	JP	US
nlm11_3_log	65	196	89	139	185	93	47	255	96	129.4	83.7	212.0	92.7
nlm2_2_log	37	169	60	174	231	80	61	148	206	129.6	90.7	182.7	115.3
nlm8_3_w	70	122	160	55	130	211	104	106	210	129.8	76.3	119.3	193.7
nlm12_2_log	108	180	88	197	216	91	18	202	70	130.0	107.7	199.3	83.0
nlm3_6_log_w	234	103	71	258	119	75	186	92	32	130.0	226.0	104.7	59.3
nlm2_1_log	32	180	58	169	236	84	60	150	205	130.4	87.0	188.7	115.7
nlm11_2_log	90	198	75	179	200	58	48	248	83	131.0	105.7	215.3	72.0
nlm4_6_log_w	262	120	31	278	134	33	164	134	27	131.4	234.7	129.3	30.3
nlm12_3_log	107	195	81	192	202	65	52	236	65	132.8	117.0	211.0	70.3
nlm4_13	152	83	175	84	47	220	136	159	143	133.2	124.0	96.3	179.3
nlm12_2_w	124	44	152	125	123	132	110	253	144	134.1	119.7	140.0	142.7
nlm12_7_w	84	71	200	22	93	231	79	215	212	134.1	61.7	126.3	214.3
nlm3_6_w	195	78	177	167	66	184	129	46	167	134.3	163.7	63.3	176.0
lm11_w	82	152	146	63	220	98	40	260	152	134.8	61.7	210.7	132.0
lm12_w	101	160	157	66	218	96	39	259	122	135.3	68.7	212.3	125.0
nlm11_2_w	111	43	150	128	137	133	97	251	172	135.8	112.0	143.7	151.7
nlm12_4_w	83	70	207	14	87	243	80	220	220	136.0	59.0	125.7	223.3
nlm4_7_log_w	255	125	34	272	183	34	184	121	28	137.3	237.0	143.0	32.0
nlm8_1_w	81	114	185	41	118	252	99	107	240	137.4	73.7	113.0	225.7
nlm11_3_w	77	77	209	12	98	240	76	216	235	137.8	55.0	130.3	228.0
nlm4_9	176	100	183	61	54	182	175	193	117	137.9	137.3	115.7	160.7
nlm7_6	139	94	193	50	57	238	131	168	171	137.9	106.7	106.3	200.7
nlm8_4	94	73	204	36	62	253	265	50	209	138.4	131.7	61.7	222.0
nlm3_7_w	175	90	192	155	71	219	113	15	216	138.4	147.7	58.7	209.0
nlm12_6_log_w	28	251	61	136	268	63	119	218	104	138.7	94.3	245.7	76.0
nlm4_1	150	88	194	58	92	246	231	23	169	139.0	146.3	67.7	203.0
lm2_log_w	39	165	70	150	184	79	188	174	208	139.7	125.7	174.3	119.0
nlm4_7	63	139	182	19	52	218	243	191	151	139.8	108.3	127.3	183.7
nlm7_3	153	100	125	92	113	181	303	41	160	140.9	182.7	84.7	155.3
nlm8_5	157	133	105	121	148	191	49	208	158	141.1	109.0	163.0	151.3
nlm3_2_w	193	72	189	161	61	222	130	61	182	141.2	161.3	64.7	197.7
nlm2_3_w	60	130	155	42	122	229	156	123	263	142.2	86.0	125.0	215.7
nlm4_15	169	98	103	85	124	172	247	198	86	142.4	167.0	140.0	120.3
nlm2_1_w	48	149	153	35	147	227	124	138	264	142.8	69.0	144.7	214.7
nlm7_1	145	105	187	57	112	244	235	27	179	143.4	145.7	81.3	203.3
nlm11_1_log	115	208	86	195	207	70	54	258	102	143.9	121.3	224.3	86.0
nlm5_2_log_w	254	107	97	268	116	102	178	72	110	144.9	233.3	98.3	103.0
lm6_log	182	172	50	191	146	78	253	141	94	145.2	208.7	153.0	74.0
nlm3_3_w	183	72	222	147	43	264	120	18	241	145.6	150.0	44.3	242.3
nlm7_4_log_w	264	138	52	280	149	47	157	143	84	146.0	233.7	143.3	61.0
nlm3_2_log_w	238	179	72	259	239	73	169	55	31	146.1	222.0	157.7	58.7
lm7	156	206	138	96	162	99	73	211	177	146.4	108.3	193.0	138.0
lm3_w	180	153	159	189	173	92	121	111	146	147.1	163.3	145.7	132.3
nlm4_14	179	126	161	103	121	160	154	199	121	147.1	145.3	148.7	147.3
nlm2_2_w	116	112	101	135	152	125	158	186	239	147.1	136.3	150.0	155.0
nlm4_11	186	82	133	100	100	209	248	195	75	147.6	178.0	125.7	139.0
nlm4_5_log_w	265	181	38	279	241	37	147	110	30	147.6	230.3	177.3	35.0
nlm7_5	184	81	110	107	97	177	251	200	124	147.9	180.7	126.0	137.0
nlm11_3	66	187	236	5	143	254	45	213	185	148.2	38.7	181.0	225.0
nlm3_7	165	96	188	165	75	230	138	31	251	148.8	156.0	67.3	223.0
nlm5_3_w	177	104	214	162	76	224	118	21	243	148.8	152.3	67.0	227.0
lm4	168	210	158	95	151	97	68	210	183	148.9	110.3	190.3	146.0
nlm6_3_log	155	142	48	220	136	104	259	145	131	148.9	211.3	141.0	94.3
lm2_w	103	190	112	97	223	101	109	185	229	149.9	103.0	199.3	147.3
nlm2_3	73	136	117	37	128	215	292	103	248	149.9	134.0	122.3	193.3
nlm12_7	78	183	247	15	140	249	55	212	174	150.3	49.3	178.3	223.3
nlm4_10	149	114	163	53	109	248	226	178	133	152.6	142.7	133.7	181.3

Table A5: Cross-Validation: All 320 Model Ranking *(continued)*

	BIC			RMSE			QLIKE			Rankings			
	DE	JP	US	DE	JP	US	DE	JP	US	Ave	DE	JP	US
nlm12_4	79	185	254	11	139	261	50	219	178	152.9	46.7	181.0	231.0
nlm12_3_w	88	125	205	29	170	234	91	225	221	154.2	69.3	173.3	220.0
nlm3_4_log_w	274	86	94	289	99	148	242	94	62	154.2	268.3	93.0	101.3
nlm3_7_log_w	266	82	81	281	102	137	260	102	77	154.2	269.0	95.3	98.3
lm5_w	166	151	170	193	177	107	127	116	187	154.9	162.0	148.0	154.7
nlm12_1_w	87	121	216	26	153	247	90	231	230	155.7	67.7	168.3	231.0
nlm4_2	162	108	176	67	101	257	220	180	140	156.8	149.7	129.7	191.0
nlm5_2_w	197	95	224	171	74	237	134	59	225	157.3	167.3	76.0	228.7
nlm11_1_w	82	131	215	24	171	242	85	227	245	158.0	63.7	176.3	234.0
nlm5_3	167	109	211	170	86	239	144	40	259	158.3	160.3	78.3	236.3
nlm3_3	206	74	235	168	50	268	167	33	255	161.8	180.3	52.3	252.7
nlm3_4_w	193	101	253	154	72	278	114	14	278	161.9	153.7	62.3	269.7
lm11	113	246	164	71	247	120	56	282	162	162.3	80.0	258.3	148.7
nlm3_1_w	192	85	260	151	58	283	116	17	299	162.3	153.0	53.3	280.7
nlm12_7_log_w	229	188	78	254	233	48	103	242	89	162.7	195.3	221.0	71.7
nlm8_7_log_w	250	193	22	273	219	30	225	189	78	164.3	249.3	200.3	43.3
nlm8_1	132	137	178	64	141	245	301	95	190	164.8	165.7	124.3	204.3
lm8	178	227	137	109	206	112	102	229	192	165.8	129.7	220.7	147.0
nlm7_2	160	119	179	72	120	259	222	182	180	165.9	151.3	140.3	206.0
lm12	127	250	180	76	246	119	53	281	163	166.1	85.3	259.0	154.0
nlm8_3	126	146	135	74	172	228	304	100	217	166.9	168.0	139.3	193.3
nlm5_3_log_w	267	87	99	282	103	175	257	85	156	167.9	268.7	91.7	143.3
nlm6_2_log_w	159	241	67	215	262	77	252	139	109	169.0	208.7	214.0	84.3
nlm12_6	130	228	196	79	225	180	107	279	107	170.1	105.3	244.0	161.0
nlm8_4_log_w	259	233	46	276	255	49	195	165	64	171.3	243.3	217.7	53.0
nlm8_6	188	161	172	116	176	190	153	214	173	171.4	152.3	183.7	178.3
nlm5_1_log_w	275	110	120	291	111	187	236	81	138	172.1	267.3	100.7	148.3
nlm4_11_log_w	260	260	80	294	288	66	168	122	13	172.3	240.7	223.3	53.0
nlm5_1_w	195	113	271	157	79	291	122	20	304	172.4	158.0	70.7	288.7
nlm6_2_w	205	129	143	209	168	153	207	154	199	174.1	207.0	150.3	165.0
nlm8_2	196	128	150	106	145	233	264	201	149	174.7	188.7	158.0	177.3
nlm12_4_log_w	251	245	79	269	261	50	105	222	90	174.7	208.3	242.7	73.0
lm2	181	229	102	118	227	123	139	243	211	174.8	146.0	233.0	145.3
nlm3_4	209	106	266	177	80	280	135	36	288	175.2	173.7	74.0	278.0
nlm3_3_log_w	271	168	96	285	238	168	246	53	57	175.8	267.3	153.0	107.0
nlm3_1_log_w	276	173	98	290	232	173	232	65	52	176.8	266.0	156.7	107.7
lm6_log_w	201	177	49	240	164	105	297	183	176	176.9	246.0	174.7	110.0
nlm3_1	212	97	270	175	68	286	150	40	301	177.7	179.0	68.3	285.7
lm1_log	190	159	128	206	144	167	239	101	267	177.9	211.7	134.7	187.3
nlm2_1	128	167	130	70	193	226	308	126	262	178.9	168.7	162.0	206.0
nlm12_5	125	200	181	81	205	217	149	277	186	180.1	118.3	227.3	194.7
nlm3_6	214	132	239	182	94	250	162	151	219	182.6	186.0	125.7	236.0
nlm11_1	85	219	249	31	215	269	111	233	232	182.7	75.7	222.3	250.0
nlm11_3_log_w	256	221	78	270	248	59	132	263	120	183.0	219.3	244.0	85.7
nlm7_5_log_w	283	27	231	307	56	313	205	142	86	183.3	265.0	75.0	210.0
nlm5_1	211	116	272	178	91	292	142	43	308	183.7	177.0	83.3	290.7
nlm1_1_log	187	141	134	232	135	195	250	105	276	183.9	223.0	127.0	201.7
nlm12_1	98	216	256	38	210	271	117	235	218	184.3	84.3	220.3	248.3
nlm12_3	95	218	252	40	214	267	128	232	213	184.3	87.7	221.3	244.0
nlm12_2	123	217	213	49	221	236	183	273	165	186.7	118.3	237.0	204.7
lm6_w	200	197	165	216	218	118	202	162	203	186.8	206.0	192.3	162.0
nlm11_2	110	225	174	47	237	216	194	278	201	186.9	117.0	246.7	197.0
nlm3_5	220	145	198	198	110	171	240	197	204	187.0	219.3	150.7	191.0
nlm2_3_log_w	253	182	68	275	212	88	224	179	215	188.4	250.7	191.0	123.7
nlm3_2	216	127	253	181	87	266	170	166	242	189.8	189.0	126.7	253.7
nlm14_3_log	146	284	115	177	301	127	96	287	193	191.8	139.7	290.7	145.0
nlm8_2_log_w	291	272	69	308	295	67	199	177	51	192.1	266.0	248.0	62.3

Table A5: Cross-Validation: All 320 Model Ranking *(continued)*

	BIC			RMSE			QLIKE			Rankings			
	DE	JP	US	DE	JP	US	DE	JP	US	Ave	DE	JP	US
nlm2_2	198	164	104	114	196	198	287	224	250	192.8	199.7	194.7	184.0
lm1_log_w	204	156	124	242	146	179	294	115	277	193.0	246.7	139.0	193.3
lm14_w	134	231	191	94	258	135	159	292	253	194.1	129.0	260.3	193.0
nlm5_2	215	140	264	184	104	272	166	161	256	195.8	188.3	135.0	264.0
lm14_log	171	281	126	117	290	109	189	301	184	196.4	159.0	290.7	139.7
nlm6_3_w	202	158	243	188	175	273	187	97	280	200.3	192.3	143.3	265.3
nlm12_5_log_w	279	268	19	304	302	46	219	268	99	200.4	267.3	279.3	54.7
lm14	151	274	223	102	274	156	108	295	231	201.6	120.3	281.0	203.3
nlm6_1_w	210	142	262	186	132	290	197	98	303	202.2	197.7	124.0	285.0
nlm14_3_w	120	204	199	77	253	223	193	296	260	202.8	130.0	251.0	227.3
nlm6_3_log_w	269	201	90	283	213	141	288	187	155	203.0	280.0	200.3	128.7
nlm14_1_log	207	278	113	237	293	116	115	286	198	204.8	186.3	285.7	142.3
nlm6_3	203	157	245	200	170	275	200	118	289	206.3	201.0	148.3	269.7
lm9_log_w	231	84	141	249	95	206	309	269	275	206.6	263.0	149.3	207.3
lm3	219	222	230	207	150	94	262	206	271	206.8	229.3	192.7	198.3
lm10_log_w	233	117	123	250	115	186	310	267	265	207.3	264.3	166.3	191.3
nlm14_3_log_w	119	290	95	187	307	121	213	293	247	208.0	173.0	296.7	154.3
nlm14_1_w	170	168	190	134	249	200	216	291	258	208.4	173.3	236.0	216.0
nlm6_1_log_w	273	238	100	286	257	176	268	153	125	208.4	275.7	216.0	133.7
nlm14_2_w	161	176	195	138	252	203	208	290	257	208.9	169.0	239.3	218.3
lm15_w	131	248	171	89	273	163	214	305	287	209.0	144.7	275.3	207.0
nlm6_1	213	143	268	203	133	289	223	117	302	210.1	213.0	131.0	286.3
lm15_log	147	280	186	110	291	140	173	298	270	210.6	143.3	289.7	198.7
lm14_log_w	154	243	85	211	267	108	282	307	246	211.4	215.7	272.3	146.3
nlm14_2_log	199	285	106	226	299	110	196	302	188	212.3	207.0	295.3	134.7
lm5	217	220	246	210	158	111	266	207	281	212.9	231.0	195.0	212.7
nlm4_2_log_w	268	259	210	287	280	316	95	160	55	214.4	216.7	233.0	193.7
nlm10_1_log	246	205	229	252	194	139	279	221	195	217.8	259.0	206.7	187.7
lm15	158	283	208	108	283	189	145	303	292	218.8	137.0	289.7	229.7
lm1_w	191	223	219	229	240	169	238	184	279	219.1	219.3	215.7	222.3
lm10_log	241	224	212	238	167	145	299	240	207	219.2	259.3	210.3	188.0
lm15_log_w	138	244	118	204	270	136	278	306	285	219.9	206.7	273.3	179.7
nlm10_2_log	240	211	228	245	211	147	280	217	200	219.9	255.0	213.0	191.7
nlm14_3	167	266	261	104	266	239	151	294	233	220.1	140.7	275.3	244.3
nlm10_3_log	245	212	227	251	180	138	289	256	194	221.3	261.7	216.0	186.3
nlm15_1_w	163	226	171	140	259	201	245	304	286	221.7	182.7	263.0	219.3
nlm4_3_log_w	287	247	205	295	271	312	192	156	36	222.3	258.0	224.7	184.3
nlm6_2	230	189	251	230	174	221	272	209	254	225.6	244.0	190.7	242.0
nlm4_1_log_w	290	263	218	301	287	314	172	147	49	226.8	254.3	232.3	193.7
nlm1_1_log_w	270	191	197	284	198	232	284	131	272	228.8	279.3	173.3	233.7
nlm4_15_log_w	278	252	233	299	285	318	185	172	38	228.9	254.0	236.3	196.3
lm9_log	239	219	246	239	159	174	295	252	244	229.7	257.7	210.0	221.3
nlm15_1_log	194	286	173	221	300	146	174	299	274	229.7	196.3	295.0	197.7
lm6	224	234	250	231	208	126	285	228	283	229.9	246.7	223.3	219.7
nlm10_2_w	221	148	257	235	190	199	286	280	261	230.8	247.3	206.0	239.0
nlm7_3_log_w	284	239	219	298	265	305	215	171	87	231.4	265.7	225.0	203.7
nlm4_10_log_w	294	254	232	302	286	317	180	175	45	231.7	258.7	238.3	198.0
nlm1_1_w	208	203	277	205	199	303	230	149	314	232.0	214.3	183.7	298.0
nlm8_5_log_w	280	267	127	305	298	307	233	196	80	232.6	272.7	253.7	171.3
nlm9_1_log	247	215	258	253	178	170	281	264	238	233.8	260.3	219.0	222.0
nlm10_2_log_w	228	232	202	248	256	157	305	246	249	235.9	260.3	244.7	202.7
nlm1_1	209	202	278	217	195	302	237	170	316	236.2	221.0	189.0	298.7
nlm7_2_log_w	299	256	226	310	284	311	181	173	95	237.2	263.3	237.7	210.7
nlm7_1_log_w	296	255	221	303	281	306	201	176	97	237.3	266.7	237.3	208.0
nlm14_2	164	265	259	93	264	263	263	289	282	238.0	173.3	272.7	268.0
nlm14_1	173	264	273	88	260	276	254	288	268	238.2	171.7	270.7	272.3
nlm8_3_log_w	281	261	203	296	278	308	241	194	88	238.9	272.7	244.3	199.7

Table A5: Cross-Validation: All 320 Model Ranking *(continued)*

	BIC			RMSE			QLIKE			Rankings			
	DE	JP	US	DE	JP	US	DE	JP	US	Ave	DE	JP	US
lm10_w	223	214	267	243	242	144	277	276	266	239.1	247.7	244.0	225.7
nlm8_1_log_w	292	271	206	300	294	309	217	181	92	240.2	269.7	248.7	202.3
lm9_w	222	209	269	244	245	162	276	275	269	241.2	247.3	243.0	233.3
nlm12_2_log_w	295	253	241	311	277	294	179	223	100	241.4	261.7	251.0	211.7
nlm10_3_log_w	289	171	265	292	192	205	293	271	228	245.1	291.3	211.3	232.7
nlm10_3_w	226	184	279	222	204	284	271	249	290	245.4	239.7	212.3	284.3
lm1	225	240	274	236	229	178	300	245	295	246.9	253.7	238.0	249.0
nlm15_1	185	275	255	101	275	265	283	300	296	248.3	189.7	283.3	272.0
nlm10_1_w	223	178	280	219	197	296	274	265	309	249.0	238.7	213.3	295.0
nlm9_1_log_w	293	175	276	293	189	225	291	270	252	251.6	292.3	211.3	251.0
nlm9_1_w	227	192	283	224	209	300	269	250	313	251.9	240.0	217.0	298.7
nlm12_1_log_w	302	279	244	312	305	293	182	241	112	252.2	265.3	275.0	216.3
nlm10_3	237	236	286	228	222	295	256	237	291	254.2	240.3	231.7	290.7
nlm12_3_log_w	301	249	281	309	269	310	202	261	114	255.1	270.7	259.7	235.0
nlm10_1_log_w	286	230	263	288	254	204	290	257	227	255.4	288.0	247.0	231.3
nlm2_1_log_w	285	269	220	297	296	287	234	192	223	255.9	272.0	252.3	243.3
nlm10_1	235	235	287	223	217	301	261	247	307	257.0	239.7	233.0	298.3
nlm9_1	236	237	288	227	224	304	255	238	311	257.8	239.3	233.0	301.0
nlm2_2_log_w	282	266	242	306	297	298	218	190	222	257.9	268.7	251.0	254.0
nlm11_2_log_w	288	270	248	313	304	288	212	274	128	258.3	271.0	282.7	221.3
nlm11_1_log_w	305	277	240	314	303	285	206	272	148	261.1	275.0	284.0	224.3
lm10	249	262	284	246	250	183	306	284	293	261.9	267.0	265.3	253.3
nlm10_2	244	241	282	241	235	260	302	283	273	262.3	262.3	253.0	271.7
lm9	248	258	285	247	251	197	307	285	297	263.9	267.3	264.7	259.7
nlm14_1_log_w	304	293	225	317	309	270	249	297	242	278.4	290.0	299.7	245.7
nlm14_2_log_w	297	291	234	315	310	274	273	312	236	282.4	295.0	304.3	248.0
nlm15_1_log_w	298	292	237	316	311	235	267	308	284	283.1	293.7	303.7	252.0
lm13_w	258	273	291	261	279	279	314	313	306	286.0	277.7	288.3	292.0
lm13_log_w	263	276	275	266	282	281	318	317	305	287.0	282.3	291.7	287.0
nlm13_1_w	257	257	289	255	272	315	316	315	315	287.9	276.0	281.3	306.3
nlm13_1	277	282	294	260	276	319	312	310	312	293.6	283.0	289.3	308.3
lm13_log	303	288	290	267	292	277	317	316	298	294.2	295.7	298.7	288.3
nlm13_1_log_w	272	294	292	263	308	299	315	314	300	295.2	283.3	305.3	297.0
lm13	300	287	295	262	289	297	313	309	310	295.8	291.7	295.0	300.7
nlm13_1_log	306	289	293	271	306	282	311	311	294	295.9	296.0	302.0	289.7

Table A6: Cross-Validation Horserace: Winning Models

DE	JP	US	ALL
lm11	lm3_log	lm3_w	lm4_log
lm12	lm4_log	lm3_log	lm7_log
lm4_log	lm5_log	lm4_log	
lm7_log	lm7_log	lm7_log	
lm11_log	lm3_log_w	lm8_log	
lm12_log	lm4_log_w	lm11_log	
nlm4_1	lm5_log_w	lm12_log	
nlm4_2	lm7_log_w	nlm3_2_log	
nlm4_3	lm9_log_w	nlm3_5_log	
nlm4_4	nlm4_1	nlm3_6_log	
nlm4_5	nlm4_2	nlm4_9_log_w	
nlm4_6	nlm4_3	nlm4_13_log_w	
nlm4_7	nlm4_5	nlm4_14_log_w	
nlm4_8	nlm4_6	nlm3_5_log_w	
nlm4_9	nlm4_7	nlm7_6_log_w	
nlm4_10	nlm4_9		
nlm4_11	nlm4_10		
nlm4_12	nlm4_11		
nlm4_13	nlm4_12		
nlm4_14	nlm4_13		
nlm4_15	nlm3_1		
nlm7_1	nlm3_2		
nlm7_2	nlm3_3		
nlm7_3	nlm3_4		
nlm7_4	nlm3_5		
nlm7_5	nlm3_6		
nlm7_6	nlm3_7		
nlm7_7	nlm7_4		
nlm8_1	nlm7_5		
nlm8_2	nlm7_6		
nlm8_3	nlm7_7		
nlm8_4	nlm8_4		
nlm8_7	nlm8_7		
nlm12_1	nlm5_1		
nlm12_2	nlm5_2		
nlm12_3	nlm5_3		
nlm12_4	nlm4_1_w		
nlm12_5	nlm4_2_w		
nlm12_6	nlm4_3_w		
nlm12_7	nlm4_5_w		
nlm2_1	nlm4_6_w		
nlm2_2	nlm4_7_w		
nlm2_3	nlm4_8_w		
nlm11_1	nlm4_9_w		
nlm11_2	nlm4_10_w		
nlm11_3	nlm4_11_w		
nlm14_1	nlm4_12_w		
nlm14_2	nlm4_13_w		
nlm14_3	nlm4_15_w		
nlm15_1	nlm3_1_w		
nlm4_1_w	nlm3_2_w		
nlm4_2_w	nlm3_3_w		
nlm4_3_w	nlm3_4_w		
nlm4_4_w	nlm3_6_w		
nlm4_5_w	nlm3_7_w		
nlm4_6_w	nlm7_1_w		
nlm4_7_w	nlm7_2_w		
nlm4_8_w	nlm7_3_w		

Table A6: Cross-Validation Horserace: Winning Models *(continued)*

DE	JP	US	ALL
nlm4_9_w	nlm7_4_w		
nlm4_10_w	nlm7_5_w		
nlm4_12_w	nlm7_6_w		
nlm4_13_w	nlm7_7_w		
nlm4_14_w	nlm8_4_w		
nlm7_1_w	nlm8_7_w		
nlm7_2_w	nlm12_4_w		
nlm7_3_w	nlm5_1_w		
nlm7_4_w	nlm5_2_w		
nlm7_6_w	nlm5_3_w		
nlm7_7_w	nlm4_3_log		
nlm8_1_w	nlm4_5_log		
nlm8_3_w	nlm4_6_log		
nlm8_4_w	nlm4_7_log		
nlm8_7_w	nlm4_9_log		
nlm12_1_w	nlm4_11_log		
nlm12_3_w	nlm4_12_log		
nlm12_4_w	nlm4_13_log		
nlm12_6_w	nlm4_14_log		
nlm12_7_w	nlm3_1_log		
nlm2_1_w	nlm3_2_log		
nlm2_3_w	nlm3_3_log		
nlm11_1_w	nlm3_4_log		
nlm11_3_w	nlm3_5_log		
nlm14_3_w	nlm3_6_log		
nlm12_5_log	nlm3_7_log		
	nlm7_4_log		
	nlm7_6_log		
	nlm7_7_log		
	nlm5_1_log		
	nlm5_2_log		
	nlm5_3_log		
	nlm4_8_log_w		
	nlm4_12_log_w		
	nlm3_4_log_w		
	nlm3_6_log_w		
	nlm3_7_log_w		
	nlm7_5_log_w		
	nlm5_1_log_w		
	nlm5_2_log_w		
	nlm5_3_log_w		

Table A7: Global Model Summary

Panel A reports performance improvement relative to the lm4 benchmark model. Panel B reports correlations of the global volatility forecasts with the lm4, lm4_log, and lm7_log volatility forecasts.

Panel A: Performance

	BIC			RMSE			QLIKE			Neg VRP		
	DE	JP	US	DE	JP	US	DE	JP	US	DE	JP	US
global_lm4	0.235	0.495	-0.000	0.438	1.116	-0.001	0.770	0.743	-0.002	77	258	26
global_lm4_log	0.945	1.203	1.137	1.945	3.760	3.308	2.745	9.379	6.744	95	165	2
global_lm7_log	1.015	1.225	1.048	2.007	3.615	2.698	2.866	9.153	4.976	90	162	3
lm4_log	0.945	0.745	1.137	1.945	2.784	3.308	2.745	9.696	6.744	95	245	2
lm7_log	1.015	0.777	1.048	2.007	2.680	2.698	2.866	9.422	4.976	90	244	3

Panel B: Correlation with the benchmark and winning models

	Benchmark lm4			lm4_log			lm7_log		
	DE	JP	US	DE	JP	US	DE	JP	US
global_lm4	1.000	0.995	1.000	0.986	0.962	0.994	0.986	0.964	0.993
global_lm4_log	0.986	0.973	0.994	1.000	0.996	1.000	1.000	0.996	0.998
global_lm7_log	0.986	0.974	0.993	1.000	0.996	0.998	1.000	0.997	1.000

Table A8: Leverage Model Summary

This table summarize the results for the leverage model. Panel A reports the t-statistics of horserace tests (the t-statistics for the test $\alpha = 0.5$) of each model versus the leverage model version of itself (first three columns) or the leverage model version of the lm4 model (the last three columns). The sample is based on the cross-validation exercise. Panel B reports performance improvement relative to lm4 (expressed in %). Panel C reports the correlation with the lm4, lm4_log, and lm7_log models.

Panel A: Horserace Test

	Test against leverage version of itself			Test against leverage_lm4		
	DE	JP	US	DE	JP	US
lm4	0.279	-11.323	4.588			
lm4_log	1.042	-5.960	-6.277	2.365	-2.523	12.648
lm7_log	0.772	-8.220	-7.908	2.619	-2.694	10.469

Panel B: Performance

	BIC			RMSE			QLIKE			Neg VRP		
	DE	JP	US	DE	JP	US	DE	JP	US	DE	JP	US
leverage_lm4	-0.787	-0.915	-1.065	-1.100	0.779	-2.227	-0.444	4.721	-5.137	321	507	93
leverage_lm4_log	0.568	0.373	0.885	1.444	3.678	3.911	3.934	12.228	6.565	151	310	5
leverage_lm7_log	0.714	0.537	0.865	1.560	3.911	3.502	3.683	11.629	5.026	128	295	4
lm4_log	0.903	0.760	1.113	1.799	2.794	3.230	2.628	9.924	6.299	108	260	3
lm7_log	0.970	0.788	1.041	1.844	2.688	2.735	2.725	9.658	4.736	100	258	3

Panel C: Correlation with the benchmark and winning models

	Benchmark lm4			lm4_log			lm7_log		
	DE	JP	US	DE	JP	US	DE	JP	US
leverage_lm4	0.993	0.950	0.984	0.980	0.939	0.977	0.980	0.941	0.974
leverage_lm4_log	0.980	0.945	0.981	0.982	0.959	0.985	0.982	0.959	0.983
leverage_lm7_log	0.983	0.952	0.986	0.987	0.965	0.988	0.987	0.967	0.988

Table A9: Jump Model Summary

This table summarize the results for the jump model. Panel A reports the t-statistics of horserace tests (the t-statistics for the test $\alpha = 0.5$) of each model versus the jump model version of itself (first three columns) or the jump model version of the lm4 model (the last three columns). The sample is based on the cross-validation exercise. Panel B reports performance improvement relative to lm4 (expressed in %). Panel C reports the correlation with the lm4, lm4_log, and lm7_log models.

Panel A: Horserace Test

	Test against Jump version of itself			Test against Jump_lm4		
	DE	JP	US	DE	JP	US
lm4	-1.126	-8.535	6.495			
lm4_log	17.579	13.021	16.956	4.437	-0.513	15.141
lm7_log	18.035	15.009	19.288	2.675	-3.909	14.063

Panel B: Performance

	BIC			RMSE			QLIKE			Neg VRP		
	DE	JP	US	DE	JP	US	DE	JP	US	DE	JP	US
jump_lm4	-0.394	0.245	-0.812	-0.345	2.119	-2.666	-1.574	3.589	-2.460	33	4	168
jump_lm4_log	0.581	0.160	0.648	-0.753	-1.037	1.621	10.032	12.141	5.471	17	1	3
jump_lm7_log	0.729	0.159	0.770	-0.991	-3.006	0.857	9.181	10.911	4.242	18	0	3
lm4_log	1.398	0.777	1.223	2.677	2.861	4.262	10.265	12.136	8.183	11	1	3
lm7_log	1.374	0.666	1.267	2.289	1.869	4.086	9.258	10.768	7.129	12	4	3

Panel C: Correlation with the benchmark and winning models

	Benchmark lm4			lm4_log			lm7_log		
	DE	JP	US	DE	JP	US	DE	JP	US
jump_lm4	0.997	0.986	0.983	0.988	0.979	0.975	0.986	0.977	0.975
jump_lm4_log	0.984	0.962	0.980	0.991	0.966	0.986	0.991	0.965	0.985
jump_lm7_log	0.981	0.954	0.979	0.989	0.954	0.983	0.992	0.961	0.985

Table A10: Downside Risk Model Summary

This table summarize the results for the downside risk model. Panel A reports the t-statistics of horserace tests (the t-statistics for the test $\alpha = 0.5$) of each model versus the downside risk model version of itself (first three columns) or the downside risk model version of the lm4 model (the last three columns). Panel B reports performance improvement relative to lm4 (expressed in %). Panel C reports the correlation with the lm4, lm4_log, and lm7_log models.

Panel A: Horserace Test

	Test against downside version of itself			Test against Downside_lm4		
	DE	JP	US	DE	JP	US
lm4	-0.672	17.729	-3.878			
lm4_log	-5.236	9.173	-10.629	3.077	14.487	11.255
lm7_log	-5.522	8.978	-10.188	1.872	10.609	9.785

Panel B: Performance

	BIC			RMSE			QLIKE			Neg VRP		
	DE	JP	US	DE	JP	US	DE	JP	US	DE	JP	US
downside_lm4	-0.271	-0.862	-0.201	-0.073	-3.037	0.091	2.264	-4.721	1.188	39	13	84
downside_lm4_log	1.128	0.703	1.181	2.915	2.418	5.259	10.419	13.201	8.913	16	2	2
downside_lm7_log	1.202	0.657	1.284	2.593	1.451	5.029	9.610	11.863	8.053	23	6	2
lm4_log	1.398	0.777	1.223	2.677	2.861	4.262	10.265	12.136	8.183	11	1	3
lm7_log	1.374	0.666	1.267	2.289	1.869	4.086	9.258	10.768	7.129	12	4	3

Panel C: Correlation with the benchmark and winning models

	Benchmark lm4			lm4_log			lm7_log		
	DE	JP	US	DE	JP	US	DE	JP	US
downside_lm4	0.988	0.994	0.995	0.977	0.979	0.987	0.974	0.981	0.986
downside_lm4_log	0.989	0.975	0.989	0.998	0.993	0.997	0.996	0.988	0.996
downside_lm7_log	0.988	0.975	0.989	0.997	0.987	0.995	0.998	0.993	0.997

Table A11: Quarticity Model Summary

This table summarize the results for the quarticity model. Panel A reports the t-statistics of horserace tests (the t-statistics for the test $\alpha = 0.5$) of each model versus the quarticity model version of itself (first three columns) or the jump model version of the lm4 model (the last three columns). The sample is based on the cross-validation exercise. Panel B reports performance improvement relative to lm4 (expressed in %). Panel C reports the correlation with the lm4, lm4_log, and lm7_log models.

Panel A: Horserace Test

	Test against Quarticity version of itself			Test against quarticity_lm4		
	DE	JP	US	DE	JP	US
lm4	6.506	-0.364	49.757			
lm4_log	1.148	-1.122	25.825	7.444	7.071	50.115
lm7_log	-6.053	-0.267	26.605	7.594	7.066	48.923

Panel B: Performance

	BIC			RMSE			QLIKE			Neg VRP		
	DE	JP	US	DE	JP	US	DE	JP	US	DE	JP	US
quarticity_lm4	-0.190	0.296	-1.444	-1.719	0.961	-21.902	-7.045	6.789	-9.457	94	275	52
quarticity_lm4_log	0.660	1.155	0.467	1.569	4.077	-3.297	3.488	13.985	5.581	108	308	8
quarticity_lm7_log	0.923	1.246	0.456	2.668	3.765	-4.178	3.618	13.660	3.961	100	289	13
lm4_log	0.903	0.760	1.113	1.799	2.794	3.230	2.628	9.924	6.299	108	260	3
lm7_log	0.970	0.788	1.041	1.844	2.688	2.735	2.725	9.658	4.736	100	258	3

Panel C: Correlation with the benchmark and winning models

	Benchmark lm4			lm4_log			lm7_log		
	DE	JP	US	DE	JP	US	DE	JP	US
quarticity_lm4	0.926	0.934	0.601	0.917	0.939	0.594	0.918	0.941	0.587
quarticity_lm4_log	0.968	0.882	0.921	0.976	0.930	0.926	0.976	0.929	0.923
quarticity_lm7_log	0.979	0.881	0.915	0.990	0.931	0.920	0.990	0.931	0.921

Table A12: MIDAS Model Summary

This table summarize the results for the jump model. Panel A reports the t-statistics of horserace tests (the t-statistics for the test $\alpha = 0.5$) of each model versus the jump model version of itself (first three columns) or the jump model version of the lm4 model (the last three columns). The sample is based on the cross-validation exercise. Panel B reports performance improvement relative to lm4 (expressed in %). Panel C reports the correlation with the lm4, lm4_log, and lm7_log models.

Panel A: Horserace Test

	Test against MIDAS version of itself			Test against MIDAS		
	DE	JP	US	DE	JP	US
lm4	-0.899	-4.414	-8.370			
lm4_log	7.678	36.269	9.042	2.762	8.019	11.656
lm7_log				3.139	8.068	8.609

Panel B: Performance

	BIC			RMSE			QLIKE			Neg VRP		
	DE	JP	US	DE	JP	US	DE	JP	US	DE	JP	US
MIDAS	0.305	0.109	0.089	0.430	0.397	0.386	1.615	0.302	0.078	116	382	36
MIDAS_log	-0.332	-3.018	1.583	-1.952	-21.216	3.563	-10.907	-44.427	7.328	0	0	0
lm4_log	0.847	0.729	1.188	1.686	2.801	3.327	2.641	9.962	6.310	100	251	3
lm7_log	0.916	0.759	1.108	1.734	2.697	2.812	2.732	9.694	4.750	94	249	3

Panel C: Correlation with the benchmark and winning models

	Benchmark lm4			lm4_log			lm7_log		
	DE	JP	US	DE	JP	US	DE	JP	US
MIDAS	0.998	0.991	0.998	0.990	0.967	0.993	0.990	0.968	0.992
MIDAS_log	0.988	0.815	0.988	0.972	0.751	0.996	0.972	0.754	0.995

A.2 Additional Forward-Chained Results

Table A13: Forward-Chained Validation: Top 25 Model Ranking

This table reports the Forward-Chained performance for the top 25 models. Columns (2) to (10) display the ranking for each country and each measure. Column (11) reports the average ranking across all countries and all measures. Columns (12) to (14) display the average ranking across all measures for each country. The table is sorted by column (11). The last three rows report the ranking of three benchmark models (lm2, lm3, and lm4) among all 320 models.

	BIC			RMSE			QLIKE			Rankings			
	DE	JP	US	DE	JP	US	DE	JP	US	Ave	DE	JP	US
nlm4_14_log	4	86	26	22	87	22	27	16	108	44.2	17.7	63.0	52.0
nlm4_13_log	6	95	23	31	102	20	34	40	84	48.3	23.7	79.0	42.3
nlm7_6_log	5	94	29	30	98	31	32	39	105	51.4	22.3	77.0	55.0
lm4_log	49	116	42	13	99	18	47	23	80	54.1	36.3	79.3	46.7
lm7_log	44	106	41	14	91	23	45	22	102	54.2	34.3	73.0	55.3
nlm8_6_log	10	105	24	38	113	32	51	43	114	58.9	33.0	87.0	56.7
nlm7_5_log	1	202	8	10	240	4	22	48	8	60.3	11.0	163.3	6.7
nlm4_1_log	7	171	16	46	211	36	26	45	9	63.0	26.3	142.3	20.3
nlm3_2_log	2	1	60	158	2	46	171	3	139	64.7	110.3	2.0	81.7
nlm8_4_log	8	114	25	51	109	35	50	102	104	66.4	36.3	108.3	54.7
nlm4_2_log	26	172	9	103	212	6	20	46	7	66.8	49.7	143.3	7.3
nlm4_11_log	32	182	6	112	230	2	15	30	3	68.0	53.0	147.3	3.7
nlm3_6_log	13	7	63	167	8	53	177	8	132	69.8	119.0	7.7	82.7
nlm3_5_log	22	2	65	170	3	45	190	1	156	72.7	127.3	2.0	88.7
nlm4_10_log	30	180	15	99	216	10	33	55	24	73.6	54.0	150.3	16.3
nlm4_9_log	3	91	52	40	93	122	24	24	215	73.8	22.3	69.3	129.7
nlm4_15_log	35	203	2	109	242	1	29	50	2	74.8	57.7	165.0	1.7
lm8_log	59	143	36	27	125	28	78	88	97	75.7	54.7	118.7	53.7
nlm3_5_w	39	17	105	160	15	61	201	63	23	76.0	133.3	31.7	63.0
lm2_log	47	128	44	28	116	68	75	78	133	79.7	50.0	107.3	81.7
nlm4_4_log	31	100	38	91	95	62	36	68	210	81.2	52.7	87.7	103.3
nlm5_2_log	11	4	117	169	5	85	176	7	171	82.8	118.7	5.3	124.3
nlm6_2_log	41	10	76	179	13	64	202	11	164	84.4	140.7	11.3	101.3
nlm12_5_log	29	209	20	19	184	15	35	222	46	86.6	27.7	205.0	27.0
lm3_log	90	16	109	180	10	41	199	4	134	87.0	156.3	10.0	94.7
Benchmark													
lm2	211	265	39	139	253	89	186	285	65	170.2	178.7	267.7	64.3
lm3	210	85	110	184	60	33	236	270	57	138.3	210.0	138.3	66.7
lm4	202	258	66	111	220	42	175	277	25	152.9	162.7	251.7	44.3

Table A14: Forward-Chained Validation: Top 25 Model Performance Improvements

This table reports the Forward-Chained performance improvements for the top25 models compared to lm4. The table is sorted by the average performance ranking across all countries and all measures. Positive numbers indicate improvement and negative numbers indicate deterioration. All numbers are expressed in percent.

	BIC (%)			RMSE (%)			QLIKE (%)		
	DE	JP	US	DE	JP	US	DE	JP	US
nlm4_14_log	1.872	2.334	0.539	2.809	7.057	0.866	14.085	18.872	-10.874
nlm4_13_log	1.859	2.141	0.584	2.285	6.553	0.887	13.599	17.171	-8.853
nlm7_6_log	1.866	2.174	0.505	2.287	6.728	0.341	13.740	17.267	-10.490
lm4_log	1.408	1.816	0.269	3.315	6.724	1.232	12.964	18.033	-8.427
lm7_log	1.483	1.920	0.269	3.304	6.918	0.863	13.069	18.113	-9.904
nlm8_6_log	1.791	1.993	0.569	2.065	5.359	0.317	12.656	17.064	-11.337
nlm7_5_log	1.916	0.795	0.940	3.465	-1.557	2.397	14.354	16.922	1.960
nlm4_1_log	1.841	1.090	0.674	1.889	0.523	0.170	14.185	16.997	1.949
nlm3_2_log	1.887	4.198	0.027	-3.355	14.116	-0.076	0.854	21.124	-15.518
nlm8_4_log	1.813	1.849	0.562	1.826	5.619	0.192	12.677	14.802	-10.268
nlm4_2_log	1.654	1.088	0.932	0.224	0.488	2.223	14.398	16.959	2.523
nlm4_11_log	1.612	1.004	0.955	-0.013	-0.716	2.500	14.876	17.795	2.816
nlm3_6_log	1.745	4.039	0.013	-3.911	13.567	-0.496	-0.143	20.316	-14.647
nlm3_5_log	1.684	4.188	0.002	-3.997	14.008	-0.038	-2.965	21.898	-16.632
nlm4_10_log	1.635	1.019	0.839	0.362	0.312	1.917	13.705	16.447	0.051
nlm4_9_log	1.881	2.254	0.083	2.035	6.813	-3.857	14.322	18.011	-25.556
nlm4_15_log	1.570	0.787	0.985	0.034	-1.738	2.684	14.068	16.862	2.840
lm8_log	1.315	1.391	0.348	2.383	4.443	0.528	10.606	15.388	-9.658
nlm3_5_w	1.529	3.612	-0.190	-3.618	12.531	-0.811	-5.397	16.179	0.084
lm2_log	1.429	1.574	0.248	2.378	4.926	-1.096	10.853	15.692	-14.825
nlm4_4_log	1.620	2.067	0.289	0.675	6.753	-0.817	13.409	16.046	-24.919
nlm5_2_log	1.785	4.103	-0.268	-3.951	13.750	-1.826	-0.014	20.500	-18.568
nlm6_2_log	1.492	3.886	-0.075	-5.168	12.684	-0.999	-5.976	19.769	-17.285
nlm12_5_log	1.644	0.708	0.616	2.908	1.530	1.346	13.494	8.680	-2.651
lm3_log	1.034	3.659	-0.221	-5.242	13.172	0.005	-4.946	21.071	-14.838
Benchmark									
lm2	-0.159	-0.406	0.273	-1.089	-3.373	-2.018	-1.695	-2.656	-6.251
lm3	-0.151	2.356	-0.231	-5.845	9.964	0.300	-19.050	2.626	-4.469
lm4	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000

Table A15: Forward-Chained Validation: All 320 Model Ranking

	BIC			RMSE			QLIKE			Rankings			
	DE	JP	US	DE	JP	US	DE	JP	US	Ave	DE	JP	US
nlm4_14_log	4	86	26	22	87	22	27	16	108	44.2	17.7	63.0	52.0
nlm4_13_log	6	95	23	31	102	20	34	40	84	48.3	23.7	79.0	42.3
nlm7_6_log	5	94	29	30	98	31	32	39	105	51.4	22.3	77.0	55.0
lm4_log	49	116	42	13	99	18	47	23	80	54.1	36.3	79.3	46.7
lm7_log	44	106	41	14	91	23	45	22	102	54.2	34.3	73.0	55.3
nlm8_6_log	10	105	24	38	113	32	51	43	114	58.9	33.0	87.0	56.7
nlm7_5_log	1	202	8	10	240	4	22	48	8	60.3	11.0	163.3	6.7
nlm4_1_log	7	171	16	46	211	36	26	45	9	63.0	26.3	142.3	20.3
nlm3_2_log	2	1	60	158	2	46	171	3	139	64.7	110.3	2.0	81.7
nlm8_4_log	8	114	25	51	109	35	50	102	104	66.4	36.3	108.3	54.7
nlm4_2_log	26	172	9	103	212	6	20	46	7	66.8	49.7	143.3	7.3
nlm4_11_log	32	182	6	112	230	2	15	30	3	68.0	53.0	147.3	3.7
nlm3_6_log	13	7	63	167	8	53	177	8	132	69.8	119.0	7.7	82.7
nlm3_5_log	22	2	65	170	3	45	190	1	156	72.7	127.3	2.0	88.7
nlm4_10_log	30	180	15	99	216	10	33	55	24	73.6	54.0	150.3	16.3
nlm4_9_log	3	91	52	40	93	122	24	24	215	73.8	22.3	69.3	129.7
nlm4_15_log	35	203	2	109	242	1	29	50	2	74.8	57.7	165.0	1.7
lm8_log	59	143	36	27	125	28	78	88	97	75.7	54.7	118.7	53.7
nlm3_5_w	39	17	105	160	15	61	201	63	23	76.0	133.3	31.7	63.0
lm2_log	47	128	44	28	116	68	75	78	133	79.7	50.0	107.3	81.7
nlm4_4_log	31	100	38	91	95	62	36	68	210	81.2	52.7	87.7	103.3
nlm5_2_log	11	4	117	169	5	85	176	7	171	82.8	118.7	5.3	124.3
nlm6_2_log	41	10	76	179	13	64	202	11	164	84.4	140.7	11.3	101.3
nlm12_5_log	29	209	20	19	184	15	35	222	46	86.6	27.7	205.0	27.0
lm3_log	90	16	109	180	10	41	199	4	134	87.0	156.3	10.0	94.7
nlm4_12	12	119	137	11	126	141	10	94	146	88.4	11.0	113.0	141.3
nlm8_2_log	50	223	4	142	250	9	37	71	13	88.8	76.3	181.3	8.7
nlm7_7	14	121	131	15	127	139	13	96	154	90.0	14.0	114.7	141.3
lm3_w	55	44	151	166	18	84	203	105	1	91.9	141.3	55.7	78.7
nlm7_2_log	28	199	21	98	231	98	31	100	21	91.9	52.3	176.7	46.7
nlm4_7	9	118	170	8	124	172	3	77	162	93.7	6.7	106.3	168.0
lm5_log	78	12	134	181	9	66	198	2	168	94.2	152.3	7.7	122.7
nlm4_14_log_w	38	96	67	104	108	107	152	69	109	94.4	98.0	91.0	94.3
lm4_log_w	101	110	55	135	89	26	168	107	79	96.7	134.7	102.0	53.3
lm7_log_w	95	102	59	137	81	38	170	99	95	97.3	134.0	94.0	64.0
lm5_w	56	40	160	174	22	100	205	115	16	98.7	145.0	59.0	92.0
nlm3_1_log	209	3	73	233	1	55	160	5	151	98.9	200.7	3.0	93.0
nlm8_5_log	62	235	1	146	251	8	68	119	12	100.2	92.0	201.7	7.0
nlm4_13_w	117	148	62	83	141	39	122	133	64	101.0	107.3	140.7	55.0
nlm3_6_w	48	20	182	157	20	154	192	31	123	103.0	132.3	23.7	153.0
nlm3_4_log	216	8	87	234	7	71	166	10	145	104.9	205.3	8.3	101.0
nlm8_6_log_w	40	103	91	110	112	140	158	80	110	104.9	102.7	98.3	113.7
nlm6_2_w	61	43	129	182	50	118	207	113	49	105.8	150.0	68.7	98.7
nlm4_14_w	120	169	40	101	176	34	129	174	22	107.2	116.7	173.0	32.0
nlm2_2_log	58	253	12	145	261	25	62	114	40	107.8	88.3	209.3	25.7
lm12_log	60	244	56	18	197	37	39	215	117	109.2	39.0	218.7	70.0
lm8_log_w	109	117	53	143	106	47	182	137	90	109.3	144.7	120.0	63.3
nlm12_6_log	19	215	49	29	214	94	19	192	153	109.3	22.3	207.0	98.7
nlm3_2_w	51	21	191	159	21	186	191	29	136	109.4	133.7	23.7	171.0
lm11_log	52	243	50	20	199	40	41	223	124	110.2	37.7	221.7	71.3
nlm4_5_log	256	93	27	270	90	30	95	47	89	110.8	207.0	76.7	48.7
nlm6_1_log	219	13	78	237	14	67	185	19	173	111.7	213.7	15.3	106.0
nlm4_13	112	155	135	71	145	96	66	165	61	111.8	83.0	155.0	97.3
nlm12_7_log	16	239	43	53	227	59	30	225	119	112.3	33.0	230.3	73.7
nlm4_10_w	270	178	79	1	175	58	57	147	59	113.8	109.3	166.7	65.3
nlm5_1_log	214	5	120	236	4	93	165	9	178	113.8	205.0	6.0	130.3
nlm12_2_log	36	267	28	107	270	49	11	208	50	114.0	51.3	248.3	42.3

Table A15: Forward Chaining: All 320 Model Ranking *(continued)*

	BIC			RMSE			QLIKE			Rankings			
	DE	JP	US	DE	JP	US	DE	JP	US	Ave	DE	JP	US
lm6_w	83	59	157	183	53	109	211	154	19	114.2	159.0	88.7	95.0
nlm3_3_log	208	6	94	232	6	56	163	6	258	114.3	201.0	6.0	136.0
lm6_log	154	47	115	196	46	76	210	26	166	115.1	186.7	39.7	119.0
nlm12_1_log	43	268	22	126	271	24	14	221	47	115.1	61.0	253.3	31.0
nlm4_5	33	125	245	12	129	240	23	52	190	116.6	22.7	102.0	225.0
nlm12_3_log	45	275	17	117	278	14	25	237	45	117.0	62.3	263.3	25.3
nlm11_2_log	53	271	19	130	273	16	17	231	48	117.6	66.7	258.3	27.7
lm2_log_w	98	108	88	144	101	88	184	127	122	117.8	142.0	112.0	99.3
nlm11_3_log	20	245	45	55	235	70	28	229	135	118.0	34.3	236.3	83.3
nlm11_1_log	46	272	18	119	274	21	21	245	51	118.6	62.0	263.7	30.0
nlm7_4_log	253	101	30	268	97	52	101	65	107	119.3	207.3	87.7	63.0
lm4_w	173	218	57	105	181	17	135	191	6	120.3	137.7	196.7	26.7
lm7_w	168	212	51	108	183	19	137	198	18	121.6	137.7	197.7	29.3
nlm4_11_w	171	167	111	66	166	103	65	176	70	121.7	100.7	169.7	94.7
nlm8_7	37	160	166	23	178	188	18	146	186	122.4	26.0	161.3	180.0
nlm4_9_w	116	151	132	82	142	162	119	131	71	122.9	105.7	141.3	121.7
nlm7_1_log	248	195	13	256	221	11	77	93	14	125.3	193.7	169.7	12.7
nlm4_4	27	153	229	21	179	233	5	79	205	125.7	17.7	137.0	222.3
nlm5_2_w	54	24	237	162	24	201	195	36	198	125.7	137.0	28.0	212.0
lm3_log_w	213	33	123	227	25	57	303	35	125	126.8	247.7	31.0	101.7
nlm4_6	91	133	230	34	135	220	52	67	184	127.3	59.0	111.7	211.3
nlm4_3	99	122	185	43	133	189	64	87	226	127.6	68.7	114.0	200.0
nlm7_5_w	165	176	101	92	171	79	76	197	98	128.3	111.0	181.3	92.7
nlm12_6_w	81	208	83	85	201	83	111	228	81	129.0	92.3	212.3	82.3
nlm4_1	94	127	224	33	136	242	55	59	195	129.4	60.7	107.3	220.3
nlm4_12_log	119	99	124	168	83	195	71	84	233	130.7	119.3	88.7	184.0
lm5_log_w	207	18	138	228	19	81	304	28	158	131.2	246.3	21.7	125.7
nlm4_9	104	158	200	47	146	219	56	168	83	131.2	69.0	157.3	167.3
nlm6_3_log	232	39	90	238	49	99	197	62	175	131.2	222.3	50.0	121.3
nlm8_4	34	156	219	17	173	237	9	129	209	131.4	20.0	152.7	221.7
nlm7_3	130	126	147	72	140	163	63	108	235	131.6	88.3	124.7	181.7
nlm8_6_w	140	201	46	116	217	86	139	201	38	131.6	131.7	206.3	56.7
nlm12_4_log	15	232	104	67	232	168	16	209	143	131.8	32.7	224.3	138.3
nlm3_5	160	64	126	171	51	112	204	244	56	132.0	178.3	119.7	98.0
nlm4_6_w	63	132	197	24	117	210	81	116	249	132.1	56.0	121.7	218.7
nlm7_6_w	125	154	144	88	143	167	124	135	111	132.3	112.3	144.0	140.7
nlm4_3_log	261	213	11	262	241	5	115	85	5	133.1	212.7	179.7	7.0
nlm7_4	92	140	236	44	139	235	46	73	203	134.2	60.7	117.3	224.7
nlm7_2_w	106	181	159	69	180	185	94	151	103	136.4	89.7	170.7	149.0
lm8_w	176	241	48	125	224	27	148	220	20	136.6	149.7	228.3	31.7
nlm5_3_log	215	11	181	235	11	207	169	12	188	136.6	206.3	11.3	192.0
nlm4_5_w	76	131	193	35	118	216	96	110	256	136.8	69.0	119.7	221.7
nlm3_6	105	45	231	165	34	212	180	103	159	137.1	150.0	60.7	200.7
nlm4_7_log	254	87	77	266	75	183	110	54	128	137.1	210.0	72.0	129.3
nlm4_8_log	250	240	5	257	247	3	108	120	4	137.1	205.0	202.3	4.0
nlm7_3_log	251	236	10	259	244	7	106	111	10	137.1	205.3	197.0	9.0
nlm7_6	102	161	202	50	149	221	60	175	115	137.2	70.7	161.7	179.3
lm11_log_w	96	145	98	150	128	77	188	195	165	138.0	144.7	156.0	113.3
nlm3_5_log_w	156	9	189	222	12	193	300	14	147	138.0	226.0	11.7	176.3
nlm12_6_log_w	42	197	64	140	225	60	173	162	181	138.2	118.3	194.7	101.7
lm3	210	85	110	184	60	33	236	270	57	138.3	210.0	138.3	66.7
nlm3_7	79	22	187	188	37	182	213	60	281	138.8	160.0	39.7	216.7
nlm4_8	122	124	175	59	138	173	80	106	277	139.3	87.0	122.7	208.3
nlm4_6_log	255	89	106	267	71	177	104	70	116	139.4	208.7	76.7	133.0
lm12_log_w	100	162	107	149	130	72	187	194	157	139.8	145.3	162.0	112.0
lm11_w	114	242	97	93	219	50	123	252	69	139.9	110.0	237.7	72.0
lm1_log	145	31	214	199	40	153	212	25	240	139.9	185.3	32.0	202.3

Table A15: Forward Chaining: All 320 Model Ranking *(continued)*

	BIC			RMSE			QLIKE			Rankings			
	DE	JP	US	DE	JP	US	DE	JP	US	Ave	DE	JP	US
nlm12_7	17	227	228	6	188	215	6	240	137	140.4	9.7	218.3	193.3
nlm3_7_log	193	14	167	226	16	205	172	15	268	141.8	197.0	15.0	213.3
lm12_w	127	246	113	90	218	48	121	249	66	142.0	112.7	237.7	75.7
nlm8_1_log	249	249	7	260	257	12	103	130	15	142.4	204.0	212.0	11.3
nlm3_3	84	19	227	186	30	213	208	32	286	142.8	159.3	27.0	242.0
nlm2_3	57	184	150	41	205	208	44	166	234	143.2	47.3	185.0	197.3
nlm4_12_w	124	134	152	127	119	150	150	140	199	143.9	133.7	131.0	167.0
nlm7_7_log	252	97	102	265	79	190	113	81	118	144.1	210.0	85.7	136.7
nlm2_3_log	259	123	32	269	115	82	133	143	141	144.1	220.3	127.0	85.0
nlm3_2	110	41	249	164	31	249	174	98	182	144.2	149.3	56.7	226.7
nlm3_7_w	66	25	239	175	28	192	232	53	289	144.3	157.7	35.3	240.0
lm5	212	83	121	189	63	54	240	272	67	144.6	213.7	139.3	80.7
nlm4_2_w	195	177	155	153	177	181	48	145	72	144.8	132.0	166.3	136.0
lm1_w	134	69	188	198	69	164	222	187	78	145.4	184.7	108.3	143.3
nlm5_3	88	27	206	193	43	191	214	64	284	145.6	165.0	44.7	227.0
nlm12_6	69	266	130	26	249	126	90	292	63	145.7	61.7	269.0	106.3
nlm7_7_w	129	137	148	129	121	149	155	142	204	146.0	137.7	133.3	167.0
lm6_log_w	221	50	136	229	48	91	307	74	163	146.6	252.3	57.3	130.0
nlm4_2	131	144	226	52	151	252	49	159	155	146.6	77.3	151.3	211.0
nlm8_3_log	257	254	3	261	265	13	126	124	17	146.7	214.7	214.3	11.0
nlm4_14	166	222	93	89	204	114	130	268	35	146.8	128.3	231.3	80.7
nlm4_10	152	147	207	60	153	232	86	161	126	147.1	99.3	153.7	188.3
nlm12_4	18	229	246	7	190	238	7	238	152	147.2	10.7	219.0	212.0
nlm7_1	123	142	242	65	156	248	40	82	228	147.3	76.0	126.7	239.3
nlm4_15_w	192	170	99	151	168	65	206	189	86	147.3	183.0	175.7	83.3
nlm12_7_w	70	183	192	57	154	206	83	185	196	147.3	70.0	174.0	198.0
lm2_w	186	250	31	134	238	44	159	233	58	148.1	159.7	240.3	44.3
nlm3_3_w	72	28	251	176	27	218	227	42	295	148.4	158.3	32.3	254.7
nlm12_5	115	90	180	70	72	174	93	275	272	149.0	92.7	145.7	208.7
nlm4_4_w	149	159	154	120	150	106	142	139	223	149.1	137.0	149.3	161.0
nlm2_2_w	147	230	47	96	229	102	136	235	120	149.1	126.3	231.3	89.7
nlm5_3_w	77	32	248	178	32	194	234	57	291	149.2	163.0	40.3	244.3
nlm4_7_w	126	138	168	124	122	171	149	132	214	149.3	133.0	130.7	184.3
nlm11_3	21	231	235	9	193	234	12	246	169	150.0	14.0	223.3	212.7
nlm11_2_w	75	207	161	36	192	133	89	242	221	150.7	66.7	213.7	171.7
nlm7_2	136	150	217	56	157	253	54	172	167	151.3	82.0	159.7	212.3
nlm12_4_w	71	186	204	58	158	222	82	184	201	151.8	70.3	176.0	209.0
nlm6_2_log_w	170	15	244	223	17	236	301	18	144	152.0	231.3	16.7	208.0
nlm8_3	86	174	184	49	198	209	38	158	273	152.1	57.7	176.7	222.0
lm4	202	258	66	111	220	42	175	277	25	152.9	162.7	251.7	44.3
nlm4_13_log_w	240	194	68	251	248	119	97	86	73	152.9	196.0	176.0	86.7
nlm8_2_w	158	200	119	115	206	147	140	206	88	153.2	137.7	204.0	118.0
lm6	218	98	127	195	73	75	254	276	68	153.8	222.3	149.0	90.0
lm7	201	257	54	114	223	51	179	280	30	154.3	164.7	253.3	45.0
nlm4_9_log_w	241	198	86	252	255	131	92	61	74	154.4	195.0	171.3	97.0
nlm12_5_w	73	219	163	39	186	132	91	265	225	154.8	67.7	223.3	173.3
nlm4_8_w	236	185	158	2	191	148	59	203	212	154.9	99.0	193.0	172.7
nlm5_2	121	46	257	172	38	256	183	118	207	155.3	158.7	67.3	240.0
nlm7_6_log_w	242	187	85	254	245	121	98	75	92	155.4	198.0	169.0	99.3
nlm6_2	194	73	164	185	67	169	215	259	77	155.9	198.0	133.0	136.7
lm11	148	277	74	68	259	74	154	299	55	156.4	123.3	278.3	67.7
nlm11_3_w	74	188	211	64	159	225	87	186	219	157.0	75.0	177.7	218.3
nlm4_11	169	163	172	63	155	184	84	250	177	157.4	105.3	189.3	177.7
nlm12_2_w	93	205	171	78	185	155	112	226	194	157.7	94.3	205.3	173.3
lm12	159	279	95	62	258	69	151	298	52	158.1	124.0	278.3	72.0
nlm12_3	25	226	253	4	210	254	2	236	217	158.6	10.3	224.0	241.3
nlm4_6_log_w	289	120	72	294	144	80	249	97	82	158.6	277.3	120.3	78.0

Table A15: Forward Chaining: All 320 Model Ranking *(continued)*

	BIC			RMSE			QLIKE			Rankings			
	DE	JP	US	DE	JP	US	DE	JP	US	Ave	DE	JP	US
nlm1_1_log	234	26	201	242	45	170	200	56	257	159.0	225.3	42.3	209.3
nlm2_1_log	260	260	14	264	268	29	131	164	43	159.2	218.3	230.7	28.7
nlm4_3_w	233	164	179	100	165	176	53	148	216	159.3	128.7	159.0	190.3
nlm3_4_w	80	37	265	177	41	268	226	41	305	160.0	161.0	39.7	279.3
nlm3_1	111	23	266	191	35	274	216	21	306	160.3	172.7	26.3	282.0
nlm3_4	108	29	262	192	39	269	218	33	298	160.9	172.7	33.7	276.3
nlm12_3_w	87	216	223	75	203	230	8	204	202	160.9	56.7	207.7	218.3
nlm12_1	24	224	256	3	208	267	1	232	237	161.3	9.3	221.3	253.3
nlm7_3_w	146	189	145	123	194	129	141	190	197	161.6	136.7	191.0	157.0
lm8	205	262	58	128	243	63	181	283	34	161.9	171.3	262.7	51.7
nlm3_1_w	97	36	267	187	36	271	217	37	309	161.9	167.0	36.3	282.3
nlm11_1	23	228	254	5	213	263	4	239	230	162.1	10.7	226.7	249.0
nlm14_1_log	144	291	34	155	293	78	61	296	113	162.8	120.0	293.3	75.0
nlm2_1	89	206	165	61	228	224	43	181	274	163.4	64.3	205.0	221.0
nlm12_1_w	64	214	233	37	200	246	70	202	208	163.8	57.0	205.3	229.0
nlm8_5_w	196	204	92	156	207	87	209	218	106	163.9	187.0	209.7	95.0
nlm10_2_w	190	38	209	202	23	143	281	169	220	163.9	224.3	76.7	190.7
nlm4_15	179	175	142	74	169	152	117	264	211	164.8	123.3	202.7	168.3
nlm14_3_log	82	284	61	97	286	108	69	295	206	165.3	82.7	288.3	125.0
nlm8_7_log	258	135	96	263	123	200	134	149	131	165.4	218.3	135.7	142.3
nlm3_2_log_w	217	57	259	241	94	262	220	13	130	165.9	226.0	54.7	217.0
nlm7_4_w	138	139	218	122	120	229	144	122	271	167.0	134.7	127.0	239.3
nlm11_1_w	67	220	232	42	209	245	72	207	218	168.0	60.3	212.0	231.7
lm1_log_w	220	35	225	230	42	157	309	66	229	168.1	253.0	47.7	203.7
nlm8_1	139	168	216	77	196	247	73	141	261	168.7	96.3	168.3	241.3
nlm5_1	128	34	271	194	44	279	219	38	311	168.7	180.3	38.7	287.0
nlm7_5	185	179	143	80	172	161	120	266	213	168.8	128.3	205.7	172.3
nlm8_6	182	247	122	102	234	159	147	273	53	168.8	143.7	251.3	111.3
nlm3_6_log_w	223	63	258	246	103	255	223	20	129	168.9	230.7	62.0	214.0
nlm12_2	68	248	198	16	226	202	42	279	242	169.0	42.0	251.0	214.0
nlm5_1_w	113	42	274	190	47	276	221	49	316	169.8	174.7	46.0	288.7
nlm12_5_log_w	277	173	37	311	189	43	225	200	75	170.0	271.0	187.3	51.7
lm2	211	265	39	139	253	89	186	285	65	170.2	178.7	267.7	64.3
nlm14_2_log	177	292	33	161	294	73	88	312	112	171.3	142.0	299.3	72.7
nlm5_2_log_w	238	51	263	255	82	260	224	17	160	172.2	239.0	50.0	227.7
nlm10_1_log	224	76	205	217	54	138	233	170	241	173.1	224.7	100.0	194.7
nlm8_7_w	153	165	176	133	160	196	161	178	244	174.0	149.0	167.7	205.3
lm10_w	198	55	247	206	29	151	293	210	183	174.7	232.3	98.0	193.7
nlm4_7_log_w	271	92	133	276	86	124	196	112	282	174.7	247.7	96.7	179.7
nlm11_2	85	255	173	25	236	175	74	291	270	176.0	61.3	260.7	206.0
nlm7_7_log_w	280	130	112	293	152	111	253	155	100	176.2	275.3	145.7	107.7
lm9_w	197	52	250	207	33	160	294	212	187	176.9	232.7	99.0	199.0
nlm10_2_log	228	81	196	216	62	123	272	157	259	177.1	238.7	100.0	192.7
lm14_log	163	288	75	84	290	97	107	308	185	177.4	118.0	295.3	119.0
nlm4_4_log_w	279	193	103	286	246	113	237	104	44	178.3	267.3	181.0	86.7
nlm8_2	183	196	186	79	202	214	100	256	191	178.6	120.7	218.0	197.0
nlm12_7_log_w	235	259	81	249	275	92	116	163	138	178.7	200.0	232.3	103.7
nlm8_4_log_w	286	129	139	301	161	130	258	123	87	179.3	281.7	137.7	118.7
nlm10_3_log	230	88	203	218	68	135	238	196	239	179.4	228.7	117.3	192.3
nlm8_7_log_w	283	136	100	299	131	110	277	179	101	179.6	286.3	148.7	103.7
nlm6_1	141	48	264	200	57	277	239	91	310	180.8	193.3	65.3	283.7
nlm6_1_w	133	54	270	197	59	281	228	90	315	180.8	186.0	67.7	288.7
lm14	178	289	116	81	287	117	157	313	91	181.0	138.7	296.3	108.0
nlm8_5	189	210	149	87	215	166	125	271	222	181.6	133.7	232.0	179.0
nlm12_4_log_w	237	274	89	250	289	95	109	152	140	181.7	198.7	238.3	108.0
nlm8_4_w	155	166	194	132	163	239	156	167	265	181.9	147.7	165.3	232.7
nlm6_3	137	53	252	201	66	251	248	136	296	182.2	195.3	85.0	266.3

Table A15: Forward Chaining: All 320 Model Ranking *(continued)*

	BIC			RMSE			QLIKE			Rankings			
	DE	JP	US	DE	JP	US	DE	JP	US	Ave	DE	JP	US
nlm14_3	103	285	177	48	283	198	114	311	121	182.2	88.3	293.0	165.3
nlm15_1_log	175	293	35	163	295	105	85	310	179	182.2	141.0	299.3	106.3
nlm10_2_log_w	222	30	260	231	26	211	310	83	267	182.2	254.3	46.3	246.0
nlm11_3_log_w	239	269	80	253	284	90	118	171	148	183.6	203.3	241.3	106.0
nlm7_1_w	143	192	222	121	195	228	138	153	262	183.8	134.0	180.0	237.3
lm14_w	164	270	140	113	267	116	132	284	170	184.0	136.3	273.7	142.0
lm10_log_w	246	60	195	240	56	158	312	138	263	185.3	266.0	84.7	205.3
lm9_log_w	244	49	221	239	52	165	311	126	269	186.2	264.7	75.7	218.3
lm15_log	151	286	82	86	288	128	105	307	246	186.6	114.0	293.7	152.0
nlm9_1_log	227	84	243	219	65	156	235	193	260	186.9	227.0	114.0	219.7
lm10_log	247	111	215	220	74	115	289	182	231	187.1	252.0	122.3	187.0
lm1	231	115	174	204	92	142	284	282	161	187.2	239.7	163.0	159.0
nlm4_5_log_w	291	238	71	298	269	104	244	95	76	187.3	277.7	200.7	83.7
nlm14_3_log_w	65	276	108	147	282	120	178	267	243	187.3	130.0	275.0	157.0
nlm14_3_w	132	263	156	106	263	187	127	278	193	189.4	121.7	268.0	178.7
nlm7_4_log_w	290	225	84	295	264	101	243	109	99	190.0	276.0	199.3	94.7
nlm2_3_w	174	191	162	141	182	226	167	199	276	190.9	160.7	190.7	221.3
nlm2_3_log_w	281	146	128	300	174	125	265	177	127	191.4	282.0	165.7	126.7
nlm2_2	200	237	114	118	233	180	143	274	236	192.8	153.7	248.0	176.7
nlm3_3_log_w	282	58	283	278	78	280	283	51	142	192.8	281.0	62.3	235.0
lm9_log	243	104	240	221	70	134	288	183	253	192.9	250.7	119.0	209.0
nlm6_3_w	172	56	261	205	64	257	308	117	304	193.8	228.3	79.0	274.0
lm15	188	294	69	95	292	146	164	317	200	196.1	149.0	301.0	138.3
nlm4_1_w	199	190	220	173	187	223	189	150	245	197.3	187.0	175.7	229.3
nlm8_3_w	162	233	183	138	239	199	153	224	247	197.6	151.0	232.0	209.7
lm15_w	184	278	70	131	280	144	145	297	255	198.2	153.3	285.0	156.3
nlm8_1_w	161	211	212	136	222	243	146	188	266	198.3	147.7	207.0	240.3
lm14_log_w	150	256	125	152	256	137	193	290	232	199.0	165.0	267.3	164.7
nlm14_2_w	191	264	190	54	262	197	67	288	280	199.2	104.0	271.3	222.3
nlm10_3_w	204	62	279	210	58	275	291	125	302	200.7	235.0	81.7	285.3
nlm5_3_log_w	284	66	273	279	100	266	285	89	172	201.6	282.7	85.0	237.0
nlm14_1_w	118	261	213	73	260	231	102	281	279	202.0	97.7	267.3	241.0
nlm10_1_w	203	61	289	208	55	286	286	121	313	202.4	232.3	79.0	296.0
nlm3_1_log_w	285	78	295	280	162	284	274	27	149	203.8	279.7	89.0	242.7
nlm1_1_w	181	71	293	203	76	292	246	156	319	204.1	210.0	101.0	301.3
nlm2_1_w	180	234	169	148	237	227	162	213	275	205.0	163.3	228.0	223.7
nlm14_2	135	282	210	45	279	217	79	309	290	205.1	86.3	290.0	239.0
nlm3_4_log_w	288	80	281	281	167	278	282	44	150	205.7	283.7	97.0	236.3
nlm14_1	107	280	255	32	276	261	58	302	285	206.2	65.7	286.0	267.0
nlm6_1_log_w	292	67	299	284	88	289	290	76	174	206.6	288.7	77.0	254.0
nlm5_1_log_w	287	74	302	282	148	282	279	34	176	207.1	282.7	85.3	253.3
nlm15_1_w	157	273	146	94	277	204	128	303	288	207.8	126.3	284.3	212.7
nlm9_1_w	206	65	300	214	61	288	292	128	317	207.9	237.3	84.7	301.7
lm15_log_w	142	251	141	154	252	178	194	286	278	208.4	163.3	263.0	199.0
nlm1_1	187	68	276	211	77	290	276	173	318	208.4	224.7	106.0	294.7
nlm10_1_log_w	267	77	315	273	105	296	262	58	252	211.7	267.3	80.0	287.7
nlm6_3_log_w	293	82	269	283	114	258	295	144	180	213.1	290.3	113.3	235.7
nlm3_7_log_w	278	70	272	277	104	259	271	101	287	213.2	275.3	91.7	272.7
nlm9_1_log_w	268	72	317	275	96	297	270	72	264	214.6	271.0	80.0	292.7
nlm10_2	245	152	241	212	111	179	287	294	224	216.1	248.0	185.7	214.7
nlm15_1	167	287	178	76	285	241	99	316	300	216.6	114.0	296.0	239.7
nlm10_3_log_w	269	79	314	274	107	295	269	92	251	216.7	270.7	92.7	286.7
nlm4_12_log_w	320	141	153	320	164	136	306	160	283	220.3	315.3	155.0	190.7
nlm10_3	226	109	278	213	84	273	299	251	294	225.2	246.0	148.0	281.7
nlm8_2_log_w	311	300	118	310	303	203	257	219	11	225.8	292.7	274.0	110.7
nlm1_1_log_w	294	75	307	285	110	285	296	134	248	226.0	291.7	106.3	280.0
lm10	265	221	234	224	132	127	302	305	227	226.3	263.7	219.3	196.0

Table A15: Forward Chaining: All 320 Model Ranking *(continued)*

	BIC			RMSE			QLIKE			Rankings			
	DE	JP	US	DE	JP	US	DE	JP	US	Ave	DE	JP	US
nlm10_1	225	107	286	209	80	283	297	243	308	226.4	243.7	143.3	292.3
lm9	266	217	238	225	137	145	305	306	238	230.8	265.3	220.0	207.0
nlm9_1	229	112	298	215	85	287	298	253	314	232.3	247.3	150.0	299.7
nlm4_1_log_w	295	295	208	287	296	291	247	180	26	236.1	276.3	257.0	175.0
nlm13_1_w	263	113	304	243	134	293	315	293	312	252.2	273.7	180.0	303.0
nlm4_2_log_w	305	307	288	304	300	304	229	205	32	252.7	279.3	270.7	208.0
nlm7_2_log_w	304	302	284	302	297	308	230	211	37	252.8	278.7	270.0	209.7
nlm7_1_log_w	301	303	280	290	298	300	255	217	36	253.3	282.0	272.7	205.3
nlm4_3_log_w	298	304	277	291	306	298	256	234	31	255.0	281.7	281.3	202.0
nlm4_11_log_w	308	297	292	307	301	307	241	214	29	255.1	285.3	270.7	209.3
lm13_w	264	149	303	245	170	264	316	304	299	257.1	275.0	207.7	288.7
nlm4_10_log_w	303	306	296	303	299	306	231	216	54	257.1	279.0	273.7	218.7
nlm4_15_log_w	306	299	291	306	304	309	252	230	28	258.3	288.0	277.7	209.3
nlm4_8_log_w	296	308	282	288	310	301	260	257	27	258.8	281.3	291.7	203.3
nlm7_5_log_w	307	298	297	305	302	311	251	227	33	259.0	287.7	275.7	213.7
nlm13_1_log_w	262	157	320	248	147	313	317	287	301	261.3	275.7	197.0	311.3
nlm8_1_log_w	302	312	285	297	308	302	266	247	41	262.2	288.3	289.0	209.3
nlm8_3_log_w	299	310	275	292	313	299	275	261	42	262.9	288.7	294.7	205.3
nlm8_5_log_w	310	305	290	308	309	310	267	248	39	265.1	295.0	287.3	213.0
nlm2_2_log_w	309	301	301	309	305	312	264	241	60	266.9	294.0	282.3	224.3
nlm2_1_log_w	300	313	287	296	314	303	273	262	62	267.8	289.7	296.3	217.3
nlm12_1_log_w	313	318	309	313	315	314	245	254	94	275.0	290.3	295.7	239.0
nlm12_3_log_w	312	316	311	314	312	316	242	260	93	275.1	289.3	296.0	240.0
nlm11_1_log_w	314	315	310	312	311	315	250	258	96	275.7	292.0	294.7	240.3
nlm11_2_log_w	316	311	313	315	316	317	263	269	85	278.3	298.0	298.7	238.3
lm13_log_w	272	252	268	271	254	272	320	314	303	280.7	287.7	273.3	281.0
nlm13_1_log	274	281	305	247	266	250	313	315	292	282.6	278.0	287.3	282.3
nlm12_2_log_w	315	314	199	317	317	270	259	263	320	286.0	297.0	298.0	263.0
nlm7_3_log_w	297	309	306	289	307	305	261	255	254	287.0	282.3	290.3	288.3
lm13_log	276	296	294	272	291	244	319	320	293	289.4	289.0	302.3	277.0
lm13	275	290	308	258	281	265	318	319	297	290.1	283.7	296.7	290.0
nlm13_1	273	283	312	244	272	294	314	318	307	290.8	277.0	291.0	304.3
nlm14_1_log_w	317	320	316	319	319	318	268	289	192	295.3	301.3	309.3	275.3
nlm14_2_log_w	319	319	318	318	318	319	280	301	189	297.9	305.7	312.7	275.3
nlm15_1_log_w	318	317	319	316	320	320	278	300	250	304.2	304.0	312.3	296.3

Table A16: Forward-Chained Horserace: Number of Winning Models

This table reports the number of models that beat each benchmark model in the Forward-Chained Horserace test for each country. Column (5) lists the number of models that beat each benchmark model in the Cross-Validation Horserace test for all countries. The last row reports the number of models that beat all three benchmark models.

Benchmark	DE	JP	US	ALL
lm4	65	111	38	6
lm2	88	191	111	21
lm3	168	9	6	0
ALL	65	9	6	0

Table A17: Forward Chain Horserace: Winning Models

DE	JP	US	ALL
lm11	nlm3_5	lm4_w	
lm12	nlm12_5	lm4_log	
lm14	nlm3_1_log	lm7_log	
lm4_log	nlm3_2_log	nlm4_11_log	
lm7_log	nlm3_3_log	nlm4_15_log	
lm8_log	nlm3_4_log	nlm12_3_log	
lm11_log	nlm3_5_log		
lm12_log	nlm5_1_log		
nlm4_1	nlm10_1_log		
nlm4_2			
nlm4_3			
nlm4_4			
nlm4_5			
nlm4_6			
nlm4_7			
nlm4_8			
nlm4_9			
nlm4_10			
nlm4_11			
nlm4_12			
nlm4_13			
nlm4_14			
nlm4_15			
nlm7_1			
nlm7_2			
nlm7_3			
nlm7_4			
nlm7_5			
nlm7_6			
nlm7_7			
nlm8_1			
nlm8_2			
nlm8_3			
nlm8_4			
nlm8_5			
nlm8_6			
nlm8_7			
nlm12_1			
nlm12_2			
nlm12_3			
nlm12_4			
nlm12_5			
nlm12_6			
nlm12_7			
nlm2_2			
nlm11_1			
nlm11_2			
nlm11_3			
nlm14_1			
nlm14_2			
nlm14_3			
nlm15_1			
nlm4_5_w			
nlm4_6_w			
nlm4_8_w			
nlm4_10_w			
nlm7_2_w			
nlm12_1_w			

Table A17: Forward-Chained Horserace: Winning Models *(continued)*

DE	JP	US	ALL
nlm12_5_w			
nlm11_1_w			
nlm11_2_w			
nlm14_1_w			
nlm4_14_log			
nlm7_5_log			
nlm12_5_log			

Table A18: Properties of of Winning Model – Forward-Chained Validation

Panel A reports the horserace test t-statistics for lm4_log and lm7_log again each benchmark model (lm2, lm3, lm4). Panel B reports the correlation of lm4_log and lm7_log with each benchmark model (lm2, lm3, lm4). Panel C reports the same correlations statistics during the crisis sample, defined as the union of the 1% right tail for any of the four predictive variables. The crisis sample comprises 2.3% of the full sample. Panel D reports the number of negative variance risk premiums for both the full sample and the crisis periods. The crisis sample comprises 2.3% of the full sample.

Panel A: Horserace t-statistics									
	Benchmark lm4			Benchmark lm3			Benchmark lm2		
	DE	JP	US	DE	JP	US	DE	JP	US
lm4_log	7.837	7.805	5.848	17.973	-4.871	9.118	11.741	18.948	17.339
lm7_log	7.585	8.355	3.697	17.995	-4.587	8.169	11.520	19.468	16.221
Panel B: Correlation with the benchmark									
	Benchmark lm4			Benchmark lm3			Benchmark lm2		
	DE	JP	US	DE	JP	US	DE	JP	US
lm4_log	0.991	0.976	0.993	0.926	0.939	0.981	0.991	0.970	0.985
lm7_log	0.991	0.975	0.991	0.926	0.940	0.980	0.991	0.970	0.985
Panel C: Correlation with the benchmark during crisis									
	Benchmark lm4			Benchmark lm3			Benchmark lm2		
	DE	JP	US	DE	JP	US	DE	JP	US
lm4_log	0.973	0.951	0.962	0.974	0.956	0.915	0.711	0.831	0.962
lm7_log	0.973	0.951	0.953	0.974	0.957	0.921	0.714	0.831	0.955
Panel D: Negative VRP									
	Full Sample			Crisis Periods					
	DE	JP	US	DE	JP	US			
lm4_log	127	633	8	12	54	2			
lm7_log	94	627	10	12	58	3			
lm2	673	819	8	0	20	0			
lm3	1206	1129	128	42	80	20			
lm4	655	863	27	20	63	11			

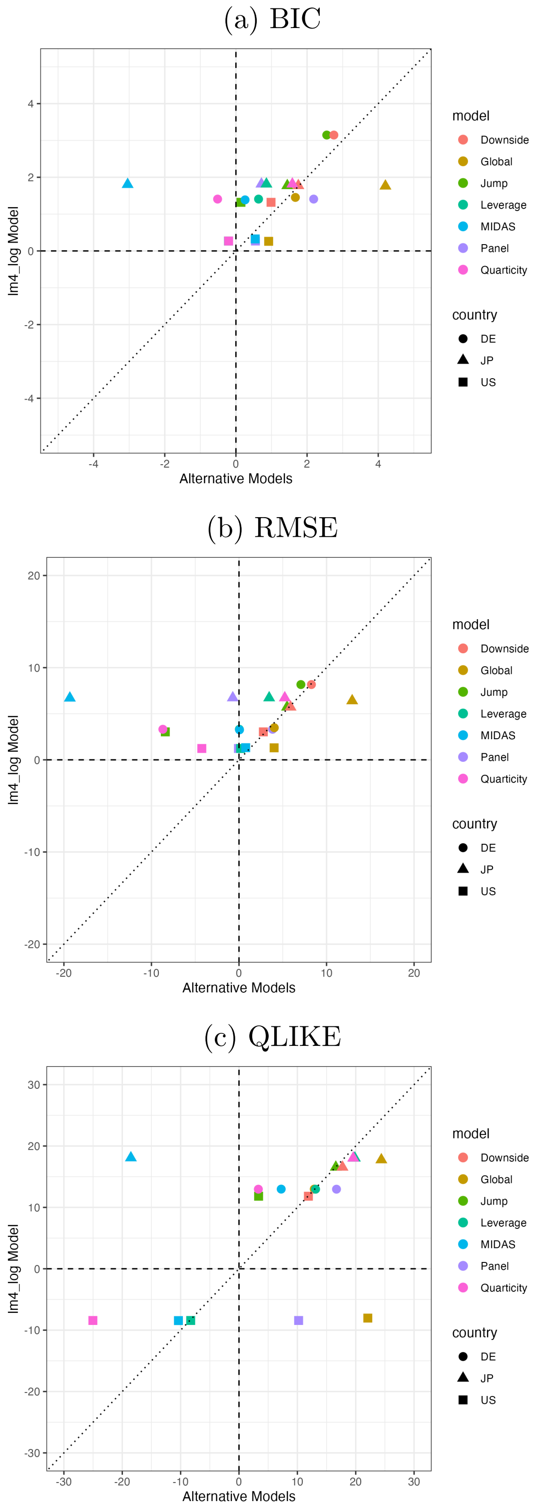


Figure A1: **Alternative Models Performance Comparison: Forward-Chained Validation**

This figure summarizes the performance of different models using forward-chained validation. The X-axis shows the performance of the alternative models (lm4_log version), while the Y-axis shows the performance of the benchmark lm4_log model. Panels (a), (b), and (c) display results based on the BIC, RMSE, and QLIKE metrics, respectively.

Table A19: Panel Model Results – Forward-Chained Validation

This table summarize the results for the panel model using forward-chained validation. Panel A reports the t-statistics of horserace tests (the t-statistics for the test $\alpha = 0.5$) of each model versus the leverage model version of itself (first three columns) or the panel model version of the lm4 model (the last three columns). The sample is based on the cross-validation exercise. Panel B reports performance improvement relative to lm4. Panel C reports the correlation with the lm4, lm4_log, and lm7_log models.

Panel A: Horserace Test

	Test against panel version of itself			Test against lm4_panel		
	DE	JP	US	DE	JP	US
lm4	-7.590	19.157	-0.500			
lm4_log	5.249	24.499	11.645	-2.716	19.617	2.503
lm7_log	5.212	24.340	10.634	-2.946	19.966	1.535

Panel B: Performance

	BIC			RMSE			QLIKE			Neg VRP		
	DE	JP	US	DE	JP	US	DE	JP	US	DE	JP	US
panel_lm4	0.829	-0.176	1.112	2.884	-4.865	4.083	-0.040	7.578	5.955	946	625	7
panel_lm4_log	2.185	0.717	0.548	3.832	-0.709	-0.065	16.705	19.607	10.217	112	259	3
panel_lm7_log	2.250	0.844	0.613	3.811	-0.415	-0.063	16.678	19.794	10.187	114	251	3
lm4_log	1.408	1.816	0.269	3.315	6.724	1.232	12.964	18.033	-8.427	127	633	8
lm7_log	1.483	1.920	0.269	3.304	6.918	0.863	13.069	18.113	-9.904	94	627	10

Panel C: Correlation with the benchmark and winning models

	lm4			lm4_log			lm7_log		
	DE	JP	US	DE	JP	US	DE	JP	US
panel_lm4	0.985	0.969	0.964	0.990	0.940	0.959	0.990	0.941	0.958
panel_lm4_log	0.970	0.986	0.984	0.990	0.974	0.991	0.989	0.974	0.993
panel_lm7_log	0.970	0.986	0.984	0.990	0.973	0.992	0.990	0.974	0.993

Table A20: Global Model Estimation – Forward-Chained Validation

This table reports the weights placed on the forecasts from the three countries for three different models (the benchmark lm4 model and the two selected models lm4_log and lm7_log), all considering the forward-chained forecasts. The columns indicate the models and the countries for which the forecasts are made, the three rows indicate the actual forecasts from Germany, Japan and the US. Thus, the weights add up to one in each column.

	lm4			lm4_log			lm7_log		
	DE	JP	US	DE	JP	US	DE	JP	US
FC_DE	0.788	0.000	0.107	0.882	0.000	0.090	0.892	0.000	0.096
FC_JP	0.188	0.640	0.023	0.118	0.741	0.038	0.108	0.747	0.036
FC_US	0.024	0.360	0.870	0.000	0.259	0.872	0.000	0.253	0.868

Table A21: Global Model Summary – Forward-Chained Validation

Panel A reports performance improvement relative to the lm4 benchmark model. Panel B reports correlations of the global volatility forecasts with the lm4, lm4_log, and lm7_log volatility forecasts.

Panel A: Performance

	BIC			RMSE			QLIKE			Neg VRP		
	DE	JP	US	DE	JP	US	DE	JP	US	DE	JP	US
global_lm4	0.524	3.728	0.618	1.671	11.014	2.815	0.699	16.027	19.376	874	109	41
global_lm4_log	1.671	4.202	0.920	4.031	12.928	4.012	12.895	24.389	22.071	338	103	10
global_lm7_log	1.737	4.267	0.946	3.973	12.955	3.695	12.979	24.354	21.613	302	106	15
lm4_log	1.453	1.761	0.261	3.482	6.405	1.304	12.981	17.751	-8.036	121	579	7
lm7_log	1.530	1.870	0.250	3.460	6.607	0.844	13.081	17.845	-9.657	90	573	9

Panel B: Correlation with the benchmark and winning models

	Benchmark lm4			lm4_log			lm7_log		
	DE	JP	US	DE	JP	US	DE	JP	US
global_lm4	0.996	0.959	0.994	0.992	0.918	0.988	0.992	0.918	0.986
global_lm4_log	0.988	0.972	0.988	0.999	0.972	0.996	0.998	0.973	0.995
global_lm7_log	0.988	0.972	0.985	0.999	0.973	0.994	0.999	0.974	0.995

Table A22: Jump Model Summary – Forward-Chained Validation

This table summarize the results for the jump model using the forward-chained validation. Panel A reports the t-statistics of horserace tests (the t-statistics for the test $\alpha = 0.5$) of each model versus the jump model version of itself (first three columns) or the jump model version of the lm4 model (the last three columns). Panel B reports performance improvement relative to lm4 (expressed in %). Panel C reports the correlation with the lm4, lm4_log, and lm7_log models.

Panel A: Horserace Test

	Test against Jump version of itself			Test against Jump_lm4		
	DE	JP	US	DE	JP	US
lm4	0.768	-4.402	12.384			
lm4_log	4.888	2.432	29.034	-1.958	0.552	12.194
lm7_log	2.722	3.408	29.763	-3.211	-0.881	11.137

Panel B: Performance

	BIC			RMSE			QLIKE			Neg VRP		
	DE	JP	US	DE	JP	US	DE	JP	US	DE	JP	US
jump_lm4	0.030	-0.051	-0.749	0.798	0.831	-3.398	2.169	0.738	-0.190	885	448	271
jump_lm4_log	2.553	1.444	0.141	7.069	5.490	-8.417	30.200	16.586	3.359	106	169	4
jump_lm7_log	2.697	1.531	0.285	6.948	4.647	-8.925	29.528	15.705	2.302	106	176	4
lm4_log	3.147	1.773	1.320	8.167	5.722	3.033	30.808	16.550	11.820	101	159	6
lm7_log	3.086	1.680	1.361	7.586	4.998	2.814	29.954	15.567	11.011	105	172	6

Panel C: Correlation with the benchmark and winning models

	Benchmark lm4			lm4_log			lm7_log		
	DE	JP	US	DE	JP	US	DE	JP	US
jump_lm4	0.998	0.985	0.985	0.985	0.977	0.977	0.983	0.971	0.977
jump_lm4_log	0.980	0.983	0.927	0.991	0.990	0.943	0.990	0.988	0.943
jump_lm7_log	0.978	0.977	0.924	0.990	0.984	0.940	0.991	0.988	0.941

Table A23: Downside Risk Model Summary – Forward-Chained Validation

This table summarize the results for the downside risk model using the forward-chained validation. Panel A reports the t-statistics of horserace tests (the t-statistics for the test $\alpha = 0.5$) of each model versus the downside risk model version of itself (first three columns) or the downside risk model version of the lm4 model (the last three columns). Panel B reports performance improvement relative to lm4 (expressed in %). Panel C reports the correlation with the lm4, lm4_log, and lm7_log models.

Panel A: Horserace Test

	Test against downside version of itself			Test against Downside_lm4		
	DE	JP	US	DE	JP	US
lm4	1.878	9.633	0.493			
lm4_log	-7.359	5.788	-0.010	-0.725	8.944	3.874
lm7_log	-7.140	5.243	1.482	-1.797	6.681	2.248

Panel B: Performance

	BIC			RMSE			QLIKE			Neg VRP		
	DE	JP	US	DE	JP	US	DE	JP	US	DE	JP	US
downside_lm4	-0.322	-0.755	-0.209	-0.692	-2.115	-0.093	-0.474	-2.323	1.970	920	394	243
downside_lm4_log	2.753	1.754	0.987	8.273	5.888	2.782	30.531	17.731	11.890	129	156	11
downside_lm7_log	2.800	1.714	1.090	7.753	5.179	2.505	29.809	16.764	11.181	135	171	11
lm4_log	3.147	1.773	1.320	8.167	5.722	3.033	30.808	16.550	11.820	101	159	6
lm7_log	3.086	1.680	1.361	7.586	4.998	2.814	29.954	15.567	11.011	105	172	6

Panel C: Correlation with the benchmark and winning models

	Benchmark lm4			lm4_log			lm7_log		
	DE	JP	US	DE	JP	US	DE	JP	US
downside_lm4	0.990	0.992	0.998	0.974	0.980	0.991	0.971	0.976	0.990
downside_lm4_log	0.985	0.983	0.994	0.997	0.996	1.000	0.995	0.992	0.999
downside_lm7_log	0.984	0.979	0.993	0.997	0.993	0.999	0.997	0.996	1.000

Table A24: Quarticity Model Summary – Forward-Chained Validation

This table summarize the results for the quarticity model using the forward-chained validation. Panel A reports the t-statistics of horserace tests (the t-statistics for the test $\alpha = 0.5$) of each model versus the jump model version of itself (first three columns) or the jump model version of the lm4 model (the last three columns). The sample is based on the cross-validation exercise. Panel B reports performance improvement relative to lm4 (expressed in %). Panel C reports the correlation with the lm4, lm4_log, and lm7_log models.

Panel A: Horserace Test

	Test against Jump version of itself			Test against Jump_lm4		
	DE	JP	US	DE	JP	US
lm4	6.297	11.535	53.422			
lm4_log	25.225	7.913	18.252	8.663	14.750	51.035
lm7_log	21.412	7.764	18.970	8.566	15.048	50.362

Panel B: Performance

	BIC			RMSE			QLIKE			Neg VRP		
	DE	JP	US	DE	JP	US	DE	JP	US	DE	JP	US
quarticity_lm4	0.165	0.313	-2.262	-0.966	-0.332	-34.011	5.167	5.420	-82.718	145	740	50
quarticity_lm4_log	-0.515	1.586	-0.206	-8.697	5.230	-4.248	3.316	19.578	-25.035	722	627	18
quarticity_lm7_log	0.054	1.810	-0.158	-4.721	5.625	-4.981	3.540	19.569	-26.748	697	615	18
lm4_log	1.408	1.816	0.269	3.315	6.724	1.232	12.964	18.033	-8.427	127	633	8
lm7_log	1.483	1.920	0.269	3.304	6.918	0.863	13.069	18.113	-9.904	94	627	10

Panel C: Correlation with the benchmark and winning models

	Benchmark lm4			lm4_log			lm7_log		
	DE	JP	US	DE	JP	US	DE	JP	US
quarticity_lm4	0.937	0.906	0.192	0.923	0.891	0.183	0.923	0.890	0.173
quarticity_lm4_log	0.921	0.910	0.906	0.927	0.937	0.907	0.927	0.935	0.905
quarticity_lm7_log	0.953	0.912	0.900	0.961	0.943	0.902	0.961	0.941	0.903

Table A25: MIDAS Model Summary – Forward-Chained Validation

This table summarize the results for the jump model using the forward-chained validation. Panel A reports the t-statistics of horserace tests (the t-statistics for the test $\alpha = 0.5$) of each model versus the jump model version of itself (first three columns) or the jump model version of the lm4 model (the last three columns). The sample is based on the cross-validation exercise. Panel B reports performance improvement relative to lm4 (expressed in %). Panel C reports the correlation with the lm4, lm4_log, and lm7_log models.

Panel A: Horserace Test

	Test against MIDAS version of itself			Test against MIDAS		
	DE	JP	US	DE	JP	US
lm4	7.849	-20.671	-5.412			
lm4_log	13.911	23.703	15.088	10.358	2.008	3.569
lm7_log				10.149	2.656	1.635

Panel B: Performance

	BIC			RMSE			QLIKE			Neg VRP		
	DE	JP	US	DE	JP	US	DE	JP	US	DE	JP	US
MIDAS	0.244	0.314	-0.045	-0.202	1.474	-0.107	-0.822	0.731	-4.149	710	839	32
MIDAS_log	0.257	-3.044	0.542	0.059	-19.321	0.764	7.231	-18.524	-10.377	96	1466	2
lm4_log	1.387	1.804	0.326	3.254	6.699	1.318	12.977	18.071	-8.444	127	620	8
lm7_log	1.459	1.909	0.319	3.240	6.896	0.931	13.082	18.154	-9.924	94	613	10

Panel C: Correlation with the benchmark and winning models

	Benchmark lm4			lm4_log			lm7_log		
	DE	JP	US	DE	JP	US	DE	JP	US
MIDAS	0.998	0.998	0.998	0.990	0.978	0.992	0.990	0.978	0.990
MIDAS_log	0.991	0.939	0.981	0.992	0.905	0.995	0.993	0.904	0.994

Table A26: Extended sample – Forward-Chained Validation

The table summarizes the results for the extended sample using forward-chained validation. Panel A reports the horserace t-statistics for each country's lm4_log and lm7_7 log models against each benchmark model. Panel B reports the performance improvement for each country in terms of each criterion. Panels C and D report the correlation with each benchmark model for the full sample and during crisis periods. Panel E reports the number of negative variance risk premiums for the full sample and crisis periods.

Panel A: Horserace t-statistics

	Benchmark lm4								Benchmark lm3								Benchmark lm2							
	CH	DE	EA	FR	JP	NL	UK	US	CH	DE	EA	FR	JP	NL	UK	US	CH	DE	EA	FR	JP	NL	UK	US
lm4_log	15.847	-2.471	5.945	3.507	3.847	4.569	17.774	4.336	9.239	0.107	11.151	5.397	8.978	6.722	9.479	-4.111	14.543	1.697	6.739	5.278	9.767	7.476	16.990	16.194
lm7_log	13.429	-3.808	5.800	2.500	1.450	3.261	16.701	2.327	7.406	-0.801	11.113	4.472	7.167	5.476	9.030	-4.667	13.979	0.441	6.712	4.534	8.266	6.587	16.926	15.384

Panel B: Performance improvement

	BIC								RMSE								QLIKE							
	CH	DE	EA	FR	JP	NL	UK	US	CH	DE	EA	FR	JP	NL	UK	US	CH	DE	EA	FR	JP	NL	UK	US
lm4_log	1.383	3.147	0.383	1.565	1.662	1.328	0.943	1.325	6.276	8.167	1.284	4.653	5.402	4.716	3.562	3.043	18.914	30.808	2.184	27.168	15.962	25.412	14.316	11.847
lm7_log	1.426	3.086	0.466	1.539	1.566	1.349	1.017	1.363	5.941	7.586	1.307	4.229	4.649	4.357	3.447	2.814	18.196	29.954	2.159	25.969	14.954	24.790	13.911	11.032
lm2	0.430	-0.016	0.751	0.434	0.316	0.942	0.897	1.167	-0.086	-1.138	0.897	0.110	-1.642	0.722	1.143	0.231	0.361	0.343	2.405	1.058	-3.926	2.589	3.915	3.218
lm3	-1.444	-1.157	-4.264	-1.860	-0.697	-2.247	-2.899	-1.488	-4.178	-3.550	-14.775	-5.963	-2.882	-7.328	-7.806	-1.931	-23.679	-19.021	-54.451	-20.267	-2.571	-19.201	-32.661	-18.102

Panel C: Correlation with the benchmark

	Benchmark lm4								Benchmark lm3								Benchmark lm2							
	CH	DE	EA	FR	JP	NL	UK	US	CH	DE	EA	FR	JP	NL	UK	US	CH	DE	EA	FR	JP	NL	UK	US
lm4_log	0.982	0.988	0.995	0.982	0.989	0.981	0.992	0.994	0.960	0.965	0.911	0.967	0.964	0.975	0.928	0.958	0.955	0.976	0.989	0.972	0.962	0.962	0.989	0.986
lm7_log	0.981	0.986	0.995	0.976	0.984	0.978	0.991	0.993	0.952	0.964	0.911	0.963	0.958	0.970	0.927	0.957	0.959	0.974	0.989	0.971	0.961	0.961	0.989	0.986

Panel D: Correlation with the benchmark during crisis periods

	Benchmark lm4								Benchmark lm3								Benchmark lm2							
	CH	DE	EA	FR	JP	NL	UK	US	CH	DE	EA	FR	JP	NL	UK	US	CH	DE	EA	FR	JP	NL	UK	US
lm4_log	0.952	0.978	0.981	0.946	0.983	0.949	0.969	0.974	0.788	0.904	0.828	0.912	0.921	0.930	0.794	0.877	0.869	0.938	0.957	0.922	0.907	0.901	0.966	0.944
lm7_log	0.940	0.963	0.979	0.907	0.960	0.917	0.963	0.968	0.731	0.902	0.827	0.882	0.894	0.893	0.791	0.872	0.887	0.924	0.955	0.910	0.915	0.888	0.966	0.946

Panel E: Negative VRP

	Full Sample								Crisis Periods							
	CH	DE	EA	FR	JP	NL	UK	US	CH	DE	EA	FR	JP	NL	UK	US
lm4_log	17	101	3	22	192	21	34	6	5	17	3	7	1	6	4	4
lm7_log	17	105	5	28	199	20	26	6	6	15	5	7	0	6	5	3
lm2	4	850	0	521	417	198	516	10	0	0	0	0	0	0	1	0
lm3	551	1447	1506	1234	455	1003	2107	975	10	20	20	12	5	9	19	17
lm4	23	860	11	618	413	353	996	198	4	10	7	1	3	2	6	7