ESSAYS ON UNCERTAINTY MEASURES AND FORECASTING

by

CHUN YAO

A dissertation submitted to the

School of Graduate Studies

Rutgers, The State University of New Jersey

In partial fulfillment of the requirements

For the degree of

Doctor of Philosophy

Graduate Program in Economics

Written under the direction of

Norman Rasmus Swanson

And approved by

____

____

____

____

New Brunswick, New Jersey

May, 2020

ABSTRACT OF THE DISSERTATION

Essays on Uncertainty Measures and Forecasting

By CHUN YAO

Dissertation Director:

Norman Rasmus Swanson

The recent decade saw the rapid increase of data size and frequency available for economic and financial analysis. This also brings the opportunity to gain new insights into the interplay between uncertainty, financial markets, and the macro economy, utilizing recent advances in high-frequency financial econometrics, as well as in macroeconometrics.

In Chapter 2, we introduce a class of multi-frequency macroeconomic and financial volatility risk factors. The factors are designed to measure uncertainty, and are latent variables extracted from a state space model that includes multiple different frequencies of non-parametrically estimated components of quadratic variation. When forecasting growth rates of monthly frequency macroeconomic variables, including housing starts, industrial production and nonfarm payroll employment, use of the new risk factors results in significant improvements in predictive performance.

Additionally, when used to forecast corporate yields, the risk factors result in monotonically increasing predictive accuracy gains, as one moves from predicting bonds with higher ratings to predicting bonds with lower ratings. This is consistent with the existence of a natural pricing channel wherein financial risk is more important, predictively, for lower grade bonds. Although the above results are promising, it should be noted that there are

exceptions. In particular, we find that when forecasting personal consumption, consumer sentiment and price growth rates, the use of simple daily volatility measures often yield superior predictions. Nevertheless, the preponderance of evidence presented in this paper points to impressive predictive gains associated with the use of the new volatility risk factors. Finally, it is worth noting that a variety of other risk factors, including the Aruoba et al. (2009b) business conditions index as well as a new financial-macroeconomic risk factor based on our multi-frequency approach often also contain marginal predictive content for the variables that we examine, although their inclusion does not reduce the importance of our multi-frequency volatility risk factor.

In Chapter 3, we examine the usefulness of a large variety of machine learning methods for forecasting daily and monthly sector level equity returns. We also examine the usefulness of three new latent risk factors that are designed to capture key forecasting information associated with financial market stress, market uncertainty, and macroeconomic fundamentals. The factors are variously based on the decomposition (using high frequency financial data) of the quadratic covariation between two assets into continuous and jump components, and the extraction of latent factors from mixed frequency state space models populated with nonparametrically estimated components of quadratic variation and/or low frequency macroeconomic data. In addition to constructing predictions using standard machine learning methods such as random forest, gradient boosting, support vector machine learning, penalized regression, and neural networks, among others, we also investigate the predictive performance of a group of hybrid machine learning methods that combine least absolute shrinkage operator and neural network specification methods. Overall, at the monthly frequency, we find that machine learning methods significantly improve forecasting performance, as measured using mean square forecast error (MSFE) and directional predictive accuracy rate (DPAR), relative to the random walk and linear benchmark alternatives. The "best" method is clearly the random forest method, which "wins" in almost all permutations at the monthly frequency, across all of the "target" variables that we predict. It is also worth noting that our hybrid machine learning methods often outperform individual methods, when forecasting daily data, although predictive gains associated with the use of any machine learning method are substantially reduced when forecasting at a daily versus monthly frequency. Finally, the novel uncertainty factors that we build are present in almost all of our "MSFE-best" and directional "accuracy-best" models, suggesting that the risk factors constructed using both high frequency financial data (e.g., 5-minute frequency S&P500 and sector ETF data) and aggregate low frequency macroeconomic data, are useful for predicting returns.

In Chapter 4[1], we evaluate the development of new tests and methods used in the evaluation of time series forecasts and forecasting models remains as important today as it has been for the last 50 years. Paraphrasing what Sir Clive W.J. Granger (arguably the father of modern day time series forecasting) said in the 1990s at a conference in Svinkloev, Denmark, 'OK, the model looks like an interesting extension, but can it forecast better than existing models.' Indeed, the forecast evaluation literature continues to expand, with interesting new tests and methods being developed at a rapid pace. In this chapter, we discuss a selected group of predictive accuracy tests and model selection methods that have been developed in recent years, and that are now widely used in the forecasting literature. We begin by reviewing several tests for comparing the relative forecast accuracy of different models, in the case of point forecasts. We then broaden the scope of our discussion by introducing density-based predictive accuracy tests. We conclude by noting that predictive accuracy is typically assessed in terms of a given loss function, such as mean squared forecast error or mean absolute forecast error. Most tests, including those discussed here, are consequently loss function dependent, and the relative forecast superiority of predictive models is therefore also dependent on specification of a loss function. In light of this fact, we conclude this chapter by discussing loss function robust predictive density accuracy tests that have recently been developed using principles of stochastic dominance.

[1]This chapter is published in "Macroeconomic Forecasting in the Era of Big Data" (Fuleky (2020)) as a chapter entitled "Forecasting Evaluation" (Cheng et al. (2020)).

Acknowledgements

I first and foremost would like to express my deep gratitude to my advisor, Norman R. Swanson. Without his guidance, support, enlightenment, and encouragement, this work would not be possible. He gives me the best supervision one can have. His acumen and rigorousness would always be the inspiration for my future pursuit. I am also very grateful to Xiye Yang, for his kind support and valuable advice for my research and career. For so many times, he sheds light on my questions and motivates my work with his knowledge and enthusiasm to the field.

I would like to thank my committee members, John Landon-Lane, and also outside member, John Chao, for taking the time and providing incisive suggestions to help me improve the quality of this dissertation.

I feel very fortunate to have the opportunity to interact with many faculty members at the Rutgers economics department, who also provide me rigorous training and extraordinary education. In particular, I would like to thank Roger Klein, for his generous help and careful guidance during my study. I also greatly appreciate administrative staff in my department for their heartening support, especially Linda Zullinger. Their excellent work and care are crucial in keeping my graduate study running smoothly.

I would also like to thank my colleagues and classmates for their accompany and help over this journey. A lot of memorable and meaningful moments in the past five years are bonded with them.

Lastly, I cannot thank enough to my parents. Their love and support keep me going forward. My last acknowledgment goes to my partner Weijia Peng, for her love and sharing of every ups and downs with me.

Dedication

To my parents

Table of Contents

List of Tables

List of Figures

Chapter 1

Introduction

In the second chapter, entitled "*Measuring Uncertainty Using Mixed Frequency Macroeconomic and Financial Volatility Risk Factors*", we introduce a new class of latent macroeconomic and financial risk (or volatility) factors, based on the use of high dimensional, high frequency, and multi-frequency data. My objective is to add to the large and relatively nascent literature exploring the construction of and uses for measures of uncertainty in the context of financial and macroeconomic forecasting (see e.g. Jurado et al. (2015a) and Baker et al. (2016)). The latent factors that we propose are extracted from a state space model that includes multiple different frequencies of non-parametrically estimated components of quadratic variation, all of which are extracted from high frequency financial data. The state space model may (or may not) also contain (multi-frequency) macroeconomic variables such as initial job claims and income growth rates, for example. By carrying out a series of ex-ante forecasting experiments, we demonstrate that when forecasting growth rates of monthly frequency macroeconomic variables, including housing starts, industrial production and nonfarm payroll employment, use of the new uncertainty measures results in significant improvement in predictive accuracy. Additionally, when used to forecast corporate yields, the uncertainty measures result in monotonically increasing predictive accuracy gains, as one moves from predicting bonds with higher ratings to predicting bonds with lower ratings. This is consistent with the existence of a natural pricing channel wherein financial risk is more important, predictively, for lower grade bonds.

The third chapter, entitled "*Forecasting Sector Level Equity Returns with Latent Risk Factors and Machine Learning*"(joint with Weijia Peng), develops additional new uncertainty measures based on the use of multi-frequency macroeconomic data and high frequency financial data. In particular, in this paper, we explore the usefulness of various new uncertainty measures constructed using methodology related to that discussed above, in the context of forecasting sector level asset returns. In this context, we explore the usefulness of a large number of machine learning, shrinkage and dimension reduction methods when specifying prediction models, with the objective of not only ascertaining the marginal predictive content of our new uncertainty measures, but also providing new evidence of the usefulness of said methods. In this sense, we add not only to the recent literature on both uncertainty measures (and their uses in finance), but also to the nascent literature on empirical asset pricing using machine learning (see e.g. Gu et al. (2018)). More specifically, we use big data to construct three different categories of latent uncertainty or risk factors, including ones based on macroeconomic fundamentals, market (financial) volatility, and financial market sector contagion. Using these new measures, we perform an extensive set of prediction experiments, in which we predict both returns and directional movements of daily and monthly returns for the S&P500 and various SPDR sector ETFs, including finance (XLF), technology (XLK), health care (XLV), and consumer discretionary (XLY). Relative performance of our various machine learning methods is then compared against random walk and linear benchmark models using conditional Diebold-Mariano and Pesaran-Timmermann directional accuracy tests. The machine learning methods that we utilize include random forest, gradient boosting, support vector machine learning, penalized regression (shrinkage) and neural networks (deep learning). We also evaluate related machine learning classifier methods including latent discriminant analysis, nave Bayes, support vector classifier, k-nearest-neighbors, and gradient boosting. Additionally, we propose and evaluate a group of hybrid two-step machine learning methods that combine least absolute shrinkage (lasso) and neural network methods. Interestingly, we find evidence of substantial forecasting improvement when our latent uncertainty factors are included in our different forecasting models, indicating the usefulness of mapping the information in big datasets into a small number of key latent factors, when forecasting at market and sector levels. Our results also indicate that models constructed using machine learning methods yield significantly lower mean square forecast errors and higher directional accuracy ratesthan various benchmark linear models. Moreover, the random forest method dominates all other machine learning methods. Finally, our hybrid machine learning methods often outperform non-hybrid methods.

In the fourth chapter, entitled "*Forecasting Evaluation*", coauthored with Mingmian Chen and Norman R. Swanson and published in *Macroeconomic Forecasting in the Era of Big Data*, Springer, Cham, 2020, 495-537, we discuss a select group of recently developed predictive accuracy tests and model selection methods, focusing primarily on those tests and methods that are becoming widely used in the forecasting literature. We begin by reviewing several tests for comparing the relative forecast accuracy of different models, in the case of point forecasts. We then broaden the scope of our discussion by introducing density-based predictive accuracy tests. We conclude by noting that predictive accuracy is typically assessed in terms of a given loss function, such as mean squared forecast error or mean absolute forecast error. Most tests, including those discussed in the paper, are consequently loss function dependent, and the relative forecast superiority of predictive models istherefore also dependent on specification of a loss function. In light of this fact, we conclude by discussing loss function robust predictive density accuracy tests that have recently been developed using principles of stochastic dominance. These tests are potentially important, as they do not require the specification of a loss function when comparing models. Thus, a model selected as best remains so, irrespective of the loss function an investigator is interested in using.

Chapter 2

Measuring Uncertainty Using Mixed Frequency Macroeconomic and Financial Volatility Risk Factors

2.1 Introduction

The impact of uncertainty on macroeconomic activity has drawn considerable attention from researchers, practitioners, and policy makers in recent years. This is in no small part due to the severity of the recent Great Recession of 2008. Other factors, such as recent slowdowns in business spending and industrial activity associated with uncertainty due to trade disputes and monetary policy outlooks have also led to interest in refining extant measures of uncertainty used to gauge economic conditions.[1] In this vein, several models have been proposed to explain the relationship between uncertainty and economic activity. For example, in a key recent paper in this literature, Bloom (2009) argues that higher uncertainty expectations cause firms to shrink their production and then freeze reallocations across industrial networks, ultimately resulting in productivity growth decreases. In a related paper, Basu and Bundick (2017) introduce a theoretical model that demonstrates how uncertainty shocks can affect aggregate demand. Jurado et al. (2015b) takes a different approach, and proposes a class of uncertainty measures that exploit data rich environments.

In this paper, we build on the above literature, and in particular on the work of Jurado et al. (2015b), by introducing a class of multi-frequency macroeconomic and financial volatility risk factors that are aimed at measuring market uncertainty. The new risk factors are latent variables extracted from state space models that include multiple different frequencies of macroeconomic and financial variables, as well as multiple different frequencies of non-parametrically estimated components of quadratic variation. For this reason,

[1]For example, in his September 2019 FOMC meeting press conference opening remarks, Federal Reserve Chair Powell said "... elevated uncertainty is weighing on U.S. investment and exports. Our business contacts around the country have been telling us that uncertainty about trade policy has discouraged them from investing in their businesses."(FOMC (2019))

our models include data frequencies ranging from 5-minutes to quarterly. The state space models are specified in one of two ways. First, they may be specified solely using latent components of quadratic variation, including continuous and jump component variation measures extracted from high frequency S&P500 data. Alternatively, they may be specified using quadratic variation components as well as additional observed variables, including macroeconomic indicators such as interest rates, employment, and production. Related papers that utilize mixed-frequency state space models include Mariano and Murasawa (2003), Frale et al. (2008), Aruoba et al. (2009b) and Marcellino et al. (2016). None of these papers, however, include multiple frequencies of the same latent variable, as is done in this paper.[2]

Our objective is thus to add to the large nascent literature exploring measures of uncertainty using financial data. For instance, we build on the work of Bloom (2009) and Basu and Bundick (2017), who examine uncertainty measures based on the VIX and the VXO, respectively. In other recent work, Gilchrist et al. (2014) examines the importance of realized volatility from a micro-level firm-specific asset return dataset. Carriero et al. (2016) incorporates both volatility uncertainty measures and "target" forecasting variables in a VAR setting. Other papers employ factor and state space models to build measures of uncertainty. For example, Jo and Sekkel (2017) use forecasting errors from macroeconomic indicators in the Survey of Professional Forecasters and extract common factors using a stochastic volatility model, and Carriero et al. (2015) builds a Bayesian model to extract latent factors.[3] Another important paper in this area which is closely related to ours is Chauvet et al. (2015). These authors implement a dynamic factor model to extract common components from realized volatilities of stocks and bonds. The main difference between our approach and that of Chauvet et al. (2015) is that while they include high frequency

[2]An interesting alternative method for handling mixed-frequency data is the mixed data sampling (MIDAS) approach proposed by Ghysels et al. (2007). The idea underlying this method is to establish a regression relation between a low-frequency variable and a set of higher-frequency variables that are aggregated by dynamic weighting functions. Following this idea, Andreou et al. (2013) demonstrate how daily financial data can be incorporated into a forecasting model for quarterly GDP.

[3]Other authors investigate uncertainty measures constructed using more exotic data. For example, Bachmann et al. (2013) use survey data regarding firms' business conditions and equate forecast "disagreement" with uncertainty. Baker et al. (2016) develop a policy uncertainty index based on the frequency of news coverage in leading newspapers.

based measures of uncertainty in their analysis, all factors are estimated using data of the same frequency. Our multi-frequency approach instead builds on the work of Corsi (2009), where the use of heterogeneous autoregressive realized volatility models is motivated by arguing that agents with different decision horizons react to, and cause, different volatility dynamics. In particular, it is argued that there are short-term traders with daily (or higher) trading frequencies, medium-term investors who typically rebalance their positions weekly, and long-term investors who induce low frequency volatility dynamics. Our approach is to consider frequencies of daily, bi-daily, tri-daily, and weekly, in order to capture effects associated with short-term and medium-term agents.

Finally, we would be remiss if we did not cite the key paper by Aruoba et al. (2009b), in which a business conditions index is constructed by extracting a latent factor from macroeconomic variables of different observational frequencies. Both of these last two papers utilize a state space modeling framework in order to estimate their latent factors. We do likewise. The main difference between our approach and that of Aruoba et al. (2009b) is that they do not include nonparametric measures of uncertainty constructed using high frequency data. Instead, they analyze a model that includes macroeconomic indicators. In order to compare our results with theirs, we include a variant of our model which nests their model.

Our key findings can be summarized as follows. First, when our multi-frequency volatility risk factors are incorporated into the forecasting models for macroeconomic variables including housing starts, industrial production and nonfarm payroll employment, we observe substantial and significant improvements in predictive performance, relative to benchmark models including simple autoregressive models, as well as mixed frequency models driven solely by macroeconomic indicators. Moreover, the largest predictive gains are associated with models that include factors constructed using both macroeconomic and high frequency financial data, as opposed to purely financial or purely macroeconomic factors. These results confirm the existence of financial-macroeconomic linkages discussed in Bloom (2009), where uncertainty influences credit conditions as well as a firm's expansion decisions. They also highlight the potential predictive usefulness of the new uncertainty measures introduced in this paper.

Second, an interesting pattern emerges when using our multi-frequency volatility risk factors to forecast corporate bond yields. In particular, the risk factors result in monotonically increasing predictive accuracy gains (as measured by mean square forecast error (MSFE)), as one moves from predicting bonds with higher ratings to predicting bonds with lower ratings. This is consistent with the existence of a natural pricing channel wherein financial risk is more important, predictively, for lower grade bonds. For example, models that include risk factors are generally associated with 30% to 40% MSFE drops for bonds with ratings lower than BB; are associated with 10% to 20% drops for A and BBB rated bonds; and are associated with no drops for AAA and AA rated bonds. Summarizing, the highest rated investment grade bonds seem to act as 'safe haven', in the sense that they show little dependence on volatility.

Third, four different measure of volatility are used in our analysis, including realized volatility (RV_t), truncated realized volatility (TRV_t), bi-power variation (BPV_t), and jump variation ($JV_t = RV_t - BPV_t$). In our forecasting experiments, TRV_t is clearly the most effective measure to use when constructing multi-frequency volatility risk factors. This is perhaps not surprising, given the difficulties noted in the financial econometrics literature associated with extracting useful predictive content from jump variation measures. It is also not surprising, thus, that we find little predictive content in factors constructed using JV_t.

Fourth, we provide empirical evidence that multi-frequency volatility risk factors react more quickly to changing economic conditions than macroeconomic factors. This is particularly apparent when observing the behavior of these risk factors after the recession of 2008.

2.2 Volatility, Macroeconomic, and Macroeconomic-volatility Risk Factors

In this section, we outline the methodology used in the construction of the risk factors analyzed in the sequel. We begin by summarizing the method that we use to address temporal aggregation as well as missing observations. We then briefly review the high frequency measures of volatility used, followed by a detailed explanation of the state space modeling framework implemented in order to estimate multi-frequency volatility risk factors, macroeconomic risk factors, and "mixed" volatility-macroeconomic risk factors.

2.2.1 Inter-temporal Aggregation

As discussed in Aruoba et al. (2009b) and Aruoba et al. (2009a) various issues regarding the temporal aggregation of variables of different frequencies, as well as stock and flow features of the variables that we examine are worth mentioning.

First, note that observed values of the flow variables are accumulated over past periods, while stock variables reflect quantities at a particular point in time. This is important, because in our state space modeling framework, we incorporate latent flow variable factors. For example, if the state space system is evolving at a daily frequency, and observed values of a monthly flow variable are accumulated value over past 30 days or so. As the state variable or risk factor that we interested in is evolving on daily basis, each time this monthly flow variable updates it introduces a 'shock' to the factor, unless flow variable observations are properly introduced into the system. This is done by introducing a new latent variable that sums past flow values across the period at which the flow variable is observed.

For a stock variable, the above complication does not arise. In particular, the observed value for a stock variable can simply be expressed as a function of the current state factor, possibly lags, and a stochastic disturbance term. As an example, let F_t denote the state factor at time t, and define

$$y_t^s = \beta_1 F_t + u_t,$$

where F_t is a latent factor, and u_t is a stochastic disturbance term. In this equation, the stock variable, y_t^s.

On the other hand, and as discussed above, a flow variable instead reflects a quantity that is aggregated throughout the observational time interval. Thus, a flow variable can be defined as follows:

$$y_t^f = \sum_{i=0}^{K_j-1} y_{t-i}^f,$$

where indices i and j denote the i^{th} time point within the j^{th} observational interval, and K_j is the length of the interval between two observational time points (i.e., time points for which observations are available - namely, between the $(j-1)^{th}$ and j^{th} time points).

Since the value of flow variable is inter-temporally accumulated over a given period of time, we can thus form a state vector that sums all lags of states during the corresponding period, hence addressing inter-temporal aggregation. However, given that our highest sampling frequency is daily, when the flow variable is monthly or quarterly, this approach will result in the specification of a very large state vector containing all lags across the flow variable sampling frequency. For example, to represent all lags of quarterly real GDP in a daily updating system, roughly 120 lags of the state variable need to be included in our system. This would lead to a system with so many parameters that convergence under maximum likelihood estimation would be difficult to achieve. For this reason, we instead implement the aggregated states approach of Aruoba et al. (2009a) in order to account for flow variables in our system. Namely, we define

$$y_t^f = \beta C_t + \gamma y_{t-M}^f + w_t,$$

where C_t is a latent state variable defined specifically for flow variables, M is the observational lag length, and the w_t are serially uncorrelated error terms. Here, C_t sums over its past values within the observational period of the flow variables. Namely,

$$C_{t+1} = \psi_{t+1} C_t + \rho F_t,$$

where

$$\psi_t = \begin{cases} 0, & \text{if } t \text{ is the first observation of the period} \\ 1, & \text{otherwise,} \end{cases}$$

and where ψ_t is an indicator that controls for the observational frequency of the flow variable. Hence, if a flow variable is updated at time t, then the value of C_{t+1} will be refreshed to be $0 + \rho F_t$, while if a flow variable is not updated at time t, then $C_{t+1} = C_t + \rho F_t$, and thus includes its past value in the sum.

2.2.2 High frequency measures of volatility and jump risk

Let X_t be the log-price of an asset at time t. Assume that the log-price process follows a jump-diffusion model (hence, almost surely, its paths are right continuous with left limits). Namely,

$$X_t = X_0 + \int_0^t b_s ds + \int_0^t \sigma_s dB_s + \sum_{s \leq t} \Delta X_s.$$

In the above expression, B is a standard Brownian motion and $\Delta X_s := X_s - X_{s-}$, where $X_{s-} := \lim_{u \uparrow s} X_u$, represents the possible jump of the process X, at time s.

Consider a finite time horizon, $[0, t]$ that contains n high-frequency observations of the log-price process. A typical time horizon is one day. Let $\Delta_n = t/n$ be the sampling frequency. Then intra-daily returns can be expressed as $r_{i,n} = X_{i\Delta_n} - X_{(i-1)\Delta_n}$.

A well-established result in the high frequency econometrics literature is that realized volatility is a consistent estimator of the total quadratic variation. Namely,

$$RV_t = \sum_{i=1}^{n} r_{i,n}^2 \xrightarrow{\text{u.c.p.}} \int_0^t \sigma_s^2 ds + \sum_{s \leq T} (\Delta X_s)^2 = QV_t = IV_t + JV_t,$$

where $\xrightarrow{\text{u.c.p.}}$ denotes convergence in probability, uniformly in time. There are many estimators of integrated volatility (IV_t), which is the variation due to the continuous component of quadratic variation (QV_t). For example, multipower variations are defined as follows:

$$V_t = \sum_{i=j+1}^{n} |r_{i,n}|^{r_1} |r_{i-1,n}|^{r_2} ... |r_{i-j,n}|^{r_j},$$

where $r_1, r_2, ..., r_j$ are positive, such that $\sum_{i=1}^{j} r_i = k$, say. An important special case of this estimator is bipower variation (BPV_t), which was introduced by Barndorff-Nielsen and Shephard (2004). Namely,

$$BPV_t = (\mu_1)^{-2} \sum_{i=2}^{n} |r_{i,n}||r_{i-1,n}|$$

where $\mu_1 = E(|Z|) = 2^{1/2}\Gamma(1)/\Gamma(1/2) = \sqrt{2/\pi}$, with Z a standard normal random variable, and where $\Gamma(\cdot)$ denotes the gamma function. Another useful estimator is truncated bipower variation ($TBPV_t$), which combines the truncation method proposed by Mancini (2009) and the bipower variation (BPV_t) estimator discussed above. Namely,

$$TBPV_t = (\mu_1)^{-2} \sum_{i=2}^{n} |\overline{r}_{i,n}||\overline{r}_{i-1,n}|, \quad \overline{r}_{i,n} = r_{i,n} 1_{\{|r_{i,n}| < \alpha_n\}},$$

where $\alpha_n = \alpha \Delta_n^{\varpi}, \varpi \in (0, \frac{1}{2})$. Similarly, truncated realized variance (TRV_t) is defined as

$$TRV_t = \sum_{i=1}^{n} \overline{r}_{i,n}^2.$$

Finally, jump variation (JV_t) can be estimated as $JV_t = RV_t - BPV_t$ or $JV_t = RV_t - TBPV_t$, for example. In the sequel, we shall utilize RV_t, TRV_t, BPV_t and $JV_t = RV_t - BPV_t$.

Under certain regularity conditions (refer to the above cited papers and Jacod and Protter (2011) and Aït-Sahalia and Jacod (2014) for details), all BPV_t, $TBPV_t$ and TRV_t are consistent estimators of the integrated volatility $\mathrm{IV}_t := \int_0^t \sigma_s^2 ds$. Hence, the corresponding JV_t estimators are also consistent. Moreover, it is also well-established that these estimators converge stably in law at the rate $\sqrt{1/\Delta_n}$, or equivalently, $\sqrt{n}$. Let T be the total number of such representative finite time horizon $[0, t]$ (e.g., day, week, month or quarter). If $\Delta_n T \to 0$, then the impact of estimating the latent volatility and jump risk factors are asymptotically negligible, since the parameters in our state space model converge at rate $\sqrt{T}$.

2.2.3 Volatility risk factors

In order to extract pure volatility risk factors we utilize a standard state space model. The model that we implement is closest to that used in Chauvet et al. (2015), and also follows

Aruoba et al. (2009b), although the latter authors do not consider volatility measures in their analysis. While Chauvet et al. (2015) implement a very interesting strategy for extracting a latent volatility factor from various different realized stock and bond volatility measures, we instead focus solely on SP500 returns in our analysis and consider a model incorporating different frequencies of volatility. In this sense, the structure of our model resembles that of a heterogeneous autoregressive realized volatility type model of the variety introduced in Corsi (2009) and Corsi and Renò (2012). The approach, thus, is meant to capture the heavy persistence in volatility. Moreover, we consider various different volatility estimators, including RV_t, TRV_t, BPV_t, and JV_t.

Summarizing, the variable $y_t = (y_t^1, y_t^2, y_t^3, y_t^4)$ corresponds to data measured at 4 different time horizons, including daily (denoted by d), bi-daily (denoted by $2d$), tri-daily (denoted by $3d$), and weekly (denoted by w). In our setup, y_t is alternately set equal to RV_t, TRV_t, BPV_t, or JV_t. The latent risk factor that we are interested in extracting is called MF_t^{vol}. Finally, the elements of y_t, which are aggregated, are flow variables. Therefore, we include three aggregated state variables, i.e., C_t^1, C_t^2 and C_t^3, to address the aggregation issues discussed above. The state space model is:

Observation Equation:

$$\begin{pmatrix} y_t^d \\ y_t^{2d} \\ y_t^{3d} \\ y_t^w \end{pmatrix} = \begin{pmatrix} \beta_1 & 0 & 0 & 0 & 1 & 0 & 0 & 0 \\ 0 & \beta_2 & 0 & 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & \beta_3 & 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & \beta_4 & 0 & 0 & 0 & 1 \end{pmatrix} \begin{pmatrix} MF_t^{vol} \\ C_t^1 \\ C_t^2 \\ C_t^3 \\ u_t^1 \\ u_t^2 \\ u_t^3 \\ u_t^4 \end{pmatrix}$$

State Equation:

$$\begin{pmatrix} MF^{vol}_{t+1} \\ C^1_{t+1} \\ C^2_{t+1} \\ C^3_{t+1} \\ u^1_{t+1} \\ u^2_{t+1} \\ u^3_{t+1} \\ u^4_{t+1} \end{pmatrix} = \begin{pmatrix} \rho & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ \rho & \psi^1_{t+1} & 0 & 0 & 0 & 0 & 0 & 0 \\ \rho & 0 & \psi^2_{t+1} & 0 & 0 & 0 & 0 & 0 \\ \rho & 0 & 0 & \psi^3_{t+1} & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & \eta_1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & \eta_2 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & \eta_3 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & \eta_4 \end{pmatrix} \begin{pmatrix} MF^{vol}_{t} \\ C^1_{t} \\ C^2_{t} \\ C^3_{t} \\ u^1_{t} \\ u^2_{t} \\ u^3_{t} \\ u^4_{t} \end{pmatrix}$$

$$+ \begin{pmatrix} 1 & 0 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 1 \end{pmatrix} \begin{pmatrix} e^1_t \\ e^2_t \\ e^3_t \\ e^4_t \\ e^5_t \end{pmatrix},$$

where the error terms $e^i_t \overset{i.i.d}{\sim} N(0, \sigma^2_i)$, with $i = 1, ..., 5$.

As mentioned above, the three aggregated variables in the state vector, C^1_t, C^2_t and C^3_t, are designed to handle bi-daily, tri-daily and weekly updating of our volatility series, respectively. Also, ψ_1, ψ_2 and ψ_3 are binary-valued parameters for the aggregated state variables, and are defined as follows:

$$\psi^1_t = \begin{cases} 0, & \text{if t is an odd number} \\ 1, & \text{otherwise,} \end{cases},$$

for the bi-daily updating series;

$$\psi^2_t = \begin{cases} 0, & \text{if t is the first day of every three days} \\ 1, & \text{otherwise,} \end{cases},$$

for the tri-daily updating series; and

$$\psi^3_t = \begin{cases} 0, & \text{if t is the first day of every week} \\ 1, & \text{otherwise,} \end{cases}$$

for the weekly series.

In the above observation equation, only the highest frequency variable, y_t^d, is directly connected with the factor, MF_t^{vol}, via β_1. The three other volatility variables are connected with MF_t^{vol} via the aggregated state variables (i.e, C_t^1, C_t^2 and C_t^3) and via β_2, β_3 and β_4. Coupled with the setup of the binary-valued parameters (i.e., ψ_1, ψ_2 and ψ_3) in the state equation, this ensures the proper inter-temporal aggregation of the flow variables in the system. and refreshes the quantity at the beginning of each period. Finally, the u_t are stochastic disturbance terms, and are assumed to follow autoregressive processes, as in Aruoba et al. (2009b). In the state equation, the first four state variables are connected with MF_t^{vol} via ρ. Of these four state variables, the last three (i.e., C_t^1, C_t^2 and C_t^3) are defined such that their previous values are added to ρMF_t^{vol} whenever flow aggregation is required.

In the state space model given above, as well as the state space models discussed in subsequent sections, we make straightforward modifications to account for missing values. Namely, the state vector is assumed to remain constant when missing values as a result of our mixed-frequency setup are encountered. In particular, we denote a mixed-frequency dataset as y_t^*. Thus, the observed vector of data in the above model is actually $y_t^* = [y_t^d, y_t^{2d}, y_t^{3d}, y_t^w]'$. Under a standard state space model setup, the observation vector $y_t^* = y_t$, as there are no missing values. More generally, let y_t be the same dimension as y_t^*, and connect these variables using a mapping matrix, M_t, such that $y_t^* = M_t y_t$, and the elements of M_t equal 1 if the corresponding value in y_t^* is non-missing and equal 0 for missing values. Then, we can use a standard estimation algorithm for the mixed-frequency dataset y_t^*. Namely, let S_t be a $m \times q$ vector of state variables. For $t = 1, \cdots, T$, the compact form of the above state space model can be written as:

$$y_t^* = H_t^* S_t$$

$$S_{t+1} = AS_t + B\eta_t,$$

where $\eta_t \overset{i.i.d.}{\sim} N(0, Q)$. According to Anderson and Moore (2012), when the observation y_t becomes available in the standard state space model, the joint distribution between $y_t - E(y_t|y_{t-1})$ and S_t would update $S_{t|t}$ and the state equation would then yield S_{t+1}. By incorporating the mapping from y_t to y_t^*, we can apply the estimation procedure used in the standard state space model to our mixed-frequency dataset. More specifically, we have

$$y_t^* = M_t y_t, \quad H_t^* = M_t H_t, \quad \beta_t^* = M_t \beta_t.$$

Then similar to the standard state space model, we obtain the following joint distribution:

$$\begin{pmatrix} S_t \\ y_t^* - E(y_t^*|y_{t-1}^*) \end{pmatrix} \sim N\left[\begin{pmatrix} S_{t|t-1} \\ 0 \end{pmatrix}, \begin{pmatrix} P_t H_t^{*'} & P_t H_t^{*'} \\ P_t H_t^{*'} & V_t \end{pmatrix}\right],$$

where P_t denotes the variance of S_t given y_{t-1}^*, and V_t denotes the variance of $y_t^* - E(y_t^*|y_{t-1}^*)$. (Refer to Anderson and Moore (2012) for a detailed derivation of mean, covariance and variance in the above discussion.)

From the joint distribution of these variables, we obtain:

$$\alpha_{t|t} = S_t + P_t H_t^{*'} F_t^{-1}[y_t^* - E(y_t^*|y_{t-1}^*)]$$
$$P_{t|t} = P_t + P_t H_t^{*'} F_t^{-1} H_t P_t'.$$

According to the state equation, we can then estimate the next step state vector as:

$$\alpha_{t+1} = A\alpha_{t|t}$$
$$P_{t+1} = AP_tA' + BQB'$$

Alternatively, when y_t has missing values due to different updating frequencies, the state vector will not update and instead adheres to the following law of motion:

$$\alpha_{t+1} = A\alpha_t$$
$$P_{t+1} = AP_tA' + BQB'.$$

2.2.4 Macroeconomic risk factors

We again begin with $y_t = (y_t^1, y_t^2, y_t^3, y_t^4)$. In this section, the data are measured at daily (denoted by d), weekly (denoted by w), monthly (denoted by m), and quarterly (denoted by w) frequencies. This allows us to construct a "benchmark" risk factor corresponding to the business conditions index analyzed by Aruoba et al. (2009b).[4] In particular, following Aruoba et al. (2009b), we use four macroeconomic variables with different sampling frequencies, including: (1) the daily yield curve spread (y_t^1), defined to be the difference between the 10-year U.S. Treasury bond yield and the 3-month Treasury bill yield; (2) weekly initial claims for unemployment insurance (y_t^2); (3) nonfarm payroll employment (y_t^3); and (4) quarterly gross domestic product (y_t^4). The corresponding state space model used to extract our risk factor, called MF_t^{mac} is:

Observation equation:

$$\begin{pmatrix} y_t^1 \\ y_t^2 \\ y_t^3 \\ y_t^4 \end{pmatrix} = \begin{pmatrix} \beta_1 & 0 & 0 & 1 \\ 0 & \beta_2 & 0 & 0 \\ \beta_3 & 0 & 0 & 0 \\ 0 & 0 & \beta_4 & 0 \end{pmatrix} \begin{pmatrix} MF_t^{mac} \\ C_t^1 \\ C_t^2 \\ u_t^1 \end{pmatrix} + \begin{pmatrix} 0 & 0 & 0 \\ \gamma_2 & 0 & 0 \\ 0 & \gamma_3 & 0 \\ 0 & 0 & \gamma_4 \end{pmatrix} \begin{pmatrix} y_{t-W}^2 \\ y_{t-M}^3 \\ y_{t-Q}^4 \end{pmatrix} + \begin{pmatrix} 0 \\ w_t^2 \\ w_t^3 \\ w_t^4 \end{pmatrix}.$$

State equation:

$$\begin{pmatrix} MF_{t+1}^{mac} \\ C_{t+1}^1 \\ C_{t+1}^2 \\ u_{t+1}^1 \end{pmatrix} = \begin{pmatrix} \rho & 0 & 0 & 0 \\ \rho & \psi_{t+1}^1 & 0 & 0 \\ \rho & 0 & \psi_{t+1}^2 & 0 \\ 0 & 0 & 0 & \gamma_1 \end{pmatrix} \begin{pmatrix} MF_t^{mac} \\ C_t^1 \\ C_t^2 \\ u_t^1 \end{pmatrix} + \begin{pmatrix} 1 & 0 \\ 1 & 0 \\ 1 & 0 \\ 0 & 1 \end{pmatrix} \begin{pmatrix} e_t^1 \\ e_t^2 \end{pmatrix},$$

where the error terms $e_t^i \overset{i.i.d}{\sim} N(0, \sigma_i^2)$, with $i = 1, 2$.

The variables in this model include observed variables, the y_t; our latent risk factor, MF_t^{mac}; aggregate state variables, C_t^1 and C_t^2; and stochastic disturbance terms, u_t^1, w_t^2, w_t^3, and w_t^4. Note that in this model, only y_t^2 and y_t^4 are flow variables in this model, and hence there are only two aggregate state variables. Accordingly, we also define two binary-valued variables ψ_1 and ψ_2 for these aggregated state variables. Namely,

$$\psi_t^1 = \begin{cases} 0, & \text{if t is the first day of the week} \\ 1, & \text{otherwise,} \end{cases}$$

[4] A 6-variable variant of this index is updated regularly on the Philadelphia Federal Reserve Bank website.

and

$$\psi_t^2 = \begin{cases} 0, & \text{if t is the first day of the quarter} \\ 1, & \text{otherwise.} \end{cases}$$

2.2.5 Macroeconomic-volatility and macroeconomic-volatility square roots risk factors

In order to construct our third variety of risk factors, we combine macroeconomic and volatility variables. The basic notion behind this risk factor is that convoluting both types of data may yield a more complete picture of the interaction between risks directly affecting macroeconomic variables, and risks that are transmitted through financial market volatility. Namely, we are interested in ascertaining the usefulness of combining uncertainty measures of the variety analyzed by Bloom (2009) with those analyzed by Chauvet et al. (2015), as well as Aruoba et al. (2009b).

We begin with $y_t = (y_t^1, y_t^2, y_t^3, y_t^4, y_t^5)$. Here, y_t^1 is alternatively set equal to daily RV_t, TRV_t, BPV_t or JV_t. The rest of the observed variables in our model are the same as those use when constructing MF_t^{mac}. Namely, they are: (1) the daily yield curve spread (y_t^2), defined to be the difference between the 10-year U.S. treasury bond yield and the 3-month treasury bond yield; (2) weekly initial claims for unemployment insurance (y_t^3); (3) nonfarm payroll employment (y_t^4); and (4) quarterly gross domestic product (y_t^5). The risk factor extracted in this setup depends on the definition of y_t^1. Namely, we extract, first, macroeconomic-volatility risk factors $MF_t^{conv} = MF_t^{mac-RV}$, $MF_t^{mac-TRV}$, $MF_t^{mac-BPV}$, or MF_t^{mac-JV}, for each of y_t^1 equal to RV_t, TRV_t, BPV_t or JV_t, respectively; and second, macroeconomic-volatility square roots risk factors $MF_t^{conv-sq} = MF_t^{mac-RV-sq}$, $MF_t^{mac-TRV-sq}$, $MF_t^{mac-BPV-sq}$, or $MF_t^{mac-JV-sq}$, for each of y_t^1 equal to the square roots of RV_t, TRV_t, BPV_t or JV_t, respectively. The state space model is:

Observation equation:

$$\begin{pmatrix} y_t^1 \\ y_t^2 \\ y_t^3 \\ y_t^4 \\ y_t^5 \end{pmatrix} = \begin{pmatrix} \beta_0 & 0 & 0 & 0 & 1 \\ \beta_1 & 0 & 0 & 1 & 0 \\ 0 & \beta_2 & 0 & 0 & 0 \\ \beta_3 & 0 & 0 & 0 & 0 \\ 0 & 0 & \beta_4 & 0 & 0 \end{pmatrix} \begin{pmatrix} MF_t^{\text{conv}} \\ C_t^1 \\ C_t^2 \\ u_t^1 \\ u_t^0 \end{pmatrix} + \begin{pmatrix} 0 & 0 & 0 \\ 0 & 0 & 0 \\ \gamma_2 & 0 & 0 \\ 0 & \gamma_3 & 0 \\ 0 & 0 & \gamma_4 \end{pmatrix} \begin{pmatrix} y_{t-W}^3 \\ y_{t-M}^4 \\ y_{t-Q}^5 \end{pmatrix} + \begin{pmatrix} 0 \\ 0 \\ w_t^2 \\ w_t^3 \\ w_t^4 \end{pmatrix}.$$

State equation:

$$\begin{pmatrix} MF_{t+1}^{\text{conv}} \\ C_{t+1}^1 \\ C_{t+1}^2 \\ u_{t+1}^1 \\ u_{t+1}^0 \end{pmatrix} = \begin{pmatrix} \rho & 0 & 0 & 0 & 0 \\ \rho & \psi_{t+1}^1 & 0 & 0 & 0 \\ \rho & 0 & \psi_{t+1}^2 & 0 & 0 \\ 0 & 0 & 0 & \gamma_1 & 0 \\ 0 & 0 & 0 & 0 & \gamma_0 \end{pmatrix} \begin{pmatrix} MF_t^{\text{conv}} \\ C_t^1 \\ C_t^2 \\ u_t^1 \\ u_t^0 \end{pmatrix} + \begin{pmatrix} 1 & 0 & 0 \\ 1 & 0 & 0 \\ 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix} \begin{pmatrix} e_t^1 \\ e_t^2 \\ e_t^3 \end{pmatrix},$$

where the error terms $e_t^i \overset{i.i.d}{\sim} N(0, \sigma_i^2)$, with $i = 1, 2, 3$.

The variables in this model include observed variables, the y_t; our latent risk factor, MF_t^{conv}; aggregate state variables, C_t^1 and C_t^2; and stochastic disturbance terms, u_t^1, u_t^0, w_t^2, w_t^3, and w_t^4. As above, only y_t^2 and y_t^4 are flow variables in this model, and hence there are only two aggregate state variables. Accordingly, we also define two binary-valued variables ψ_1 and ψ_2 for these aggregated state variables. Namely,

$$\psi_t^1 = \begin{cases} 0, & \text{if t is the first day of the week} \\ 1, & \text{otherwise} \end{cases}$$

and

$$\psi_t^2 = \begin{cases} 0, & \text{if t is the first day of the quarter} \\ 1, & \text{otherwise} \end{cases}.$$

2.3 Experimental Setup

In our forecasting experiments we evaluate the predictive content of the risk factors discussed in Section 2. This is done by forecasting the following monthly macroeconomic variables: industrial production (IP), nonfarm payroll employees (PAYEMS), housing starts (HOUST), personal consumption expenditures (PCE), the University of Michigan consumer sentiment index (SENTI), and core consumption price index (CPI) which excludes food and energy. Additionally, we predict a number of financial variables, including monthly effective yields for Fitch-rated AAA, AA, A, BBB, BB, B, and CCC corporate bonds. Prior to carrying out experiments, all macroeconomic data are log-differenced, and all financial data are differenced.

A summary of forecasting models used in our experiments is given below, and further details are contained in Table 3.2.

Autoregressive (AR) Models:

An AR model is used as a benchmark, and is specified as follows:

$$y_{t+h} = c + \alpha' W_t + \epsilon_{t+h}, \tag{2.1}$$

where y_t is the "target" forecast variable of interest, h denotes the forecast horizon, W_t contains autoregressive lags, and α a conformably defined coefficient vector. Lags are chosen anew, prior to the construction of each monthly forecast, using both the Akaike Information Criterion (AIC) and the Schwarz Information Criterion (SIC). Tabulated mean square forecast errors (MSFEs) reported in the sequel are for the case with lags selected using the AIC. Results for the SIC case were qualitatively the same, and are available upon request.

Autoregressive Models with Risk Factors:

Let F_t denote latent risk factors, which belong to the union of MF_t^{vol}, MF_t^{mac}, and MF_t^{conv} (see Section 2 for complete details). We estimate the following factor augmented forecasting model:

$$y_{t+h} = c + \alpha' W_t + \rho' F_t + \epsilon_{t+h},$$

where F_t includes either 1 or 2 risk factors, and may include lags of F_t. The rest of the terms in this model are defined as described above. Lags (of F_t and y_t) are again chosen using both the AIC.

Autoregressive Models with Daily Volatility Measures:

In addition to estimating models with multi-frequency volatility risk factors, we also estiamted models using standard daily quadratic variation component measures, including

RV_t, TRV_t, BPV_t and JV_t. Let X_t denote any one of these volatility measures, as well as lags thereof. We estimate the following volatility augmented forecasting model:

$$y_{t+h} = c + \alpha' W_t + \gamma' X_t + \epsilon_{t+h}.$$

All terms in this model are as defined above, and lags are again selected using the AIC.

Autoregressive Models with Risk Factors and Daily Volatility Measures:

Combining the two previous specifications, we also estimate the following forecasting model:

$$y_{t+h} = c + \alpha' W_t + \rho' F_t^x + \gamma' X_t + \epsilon_{t+h},$$

Again, all terms in this model are as defined above, and lags are again selected using the AIC.

As mentioned above, although some components in the vector process y are estimates obtained from high frequency data, as long as the sampling interval length Δ_n shrinks to zero fast enough, their associated estimation errors have asymptotically negligible effects on all the parameters of interest. An alternative and possibly more convincing argument goes as follows. We can view those high frequency estimates as observed quantities associated with the latent factors in the state equation. Then the high frequency estimation errors are naturally embedded in the residuals of the observation equation. Since those high frequency estimation errors converge to zero, they are definitely bounded in probability, hence perfectly in line with our assumption on the residuals.

All forecasts are constructed using rolling windows of data consisting of either 36 or 72 months of observations, and all coefficients and lag specifications are re-estimated prior to the construction of each new prediction. Hence, our forecasting period is July 2009 - December 2018 (i.e. 67 monthly forecasts). For an analysis of the use of rolling versus recursive and alternative windowing techniques in the context of forecasting, see Clark and McCracken (2009) and Hansen and Timmermann (2012) and Rossi and Inoue (2012). Estimation is carried out using maximum likelihood, and forecasts are generated for horizons $h = 1$, 2, 3, and 6.

As mentioned above, results are evaluated by comparing mean square forecast error (MSFEs). These statistics are additionally analyzed using conditional Diebold and Mariano (1995b) predictive accuracy tests of the variety developed by Giacomini and White (2006). The null hypothesis of the test is: $H_0 : \mathrm{E}[L(\epsilon_{t+h}^{(1)})] - \mathrm{E}[L(\epsilon_{t+h}^{(2)})] = 0$, where the $\epsilon_{t+h}^{(i)}$ is the prediction error of model i, for $i = 1, 2$. In our analysis, $L(\cdot)$ is a quadratic loss function. The test statistic that we utilize is:

$$\mathrm{DM}_P = P^{-1} \sum_{t=1}^{P} \frac{d_{t+h}}{\hat{\sigma}_{\bar{d}}},$$

where $d_{t+h} = [\hat{\epsilon}_{t+h}^{(1)}]^2 - [\hat{\epsilon}_{t+h}^{(2)}]^2$, $\bar{d}$ denotes the mean of d_{t+h}, $\hat{\sigma}_{\bar{d}}$ is a heteroskedasticity and autocorrelation consistent estimate of the standard deviation of $\bar{d}$, and P denotes the number of ex ante predictions used to construct the test statistic.[5] If the DM_P statistic has a negative value, Model 1 is preferred to Model 2. If the test rejects the null hypothesis, the difference between Model 1 and Model 2 is statistically significant. In the sequel, we assume that the test statistic is asymptotically N(0,1), following Giacomini and White (2006). It should be noted, though, that the unconditional variant of the test proposed by Diebold and Mariano (1995b) requires modified critical values in cases where models being compared are nested (see McCracken (2000b) and Corradi and Swanson (2006c) for complete details). For an interesting discussion of alternative approaches to assessing forecasting performance, see Rossi and Sekhposyan (2011).

In addition to comparing MSFEs, we also compare the directional predictive accuracy of our different models. This is done by examining contingency tables, as in Swanson and White (1995). To this end, we first code the directional movement assocaited with a prediction as 1 (for positive incremental changes) and -1 (for negative incremental changes). Namely, define:

---

[5]Giacomini and White (2006) also discuss a Wald version of this test statistic, which we do not utilize in this paper.

$$D_{t+h}^{actual} = \begin{cases} 1 & y_{t+h}^{actual} - y_t^{actual} \geq 0 \\ -1 & y_{t+h}^{actual} - y_t^{actual} < 0 \end{cases} \qquad D_{t+h}^{pred} = \begin{cases} 1 & y_{t+h}^{pred} - y_t^{actual} \geq 0 \\ -1 & y_{t+h}^{pred} - y_t^{actuall} < 0 \end{cases}$$

where h denotes the forecast horizon. The classical contingency table associated with these directional prediction signals is:

		actual	
		down	up
predicted	down	d_1	d_2
	up	d_3	d_4

Here d_1 (d_4) is the number of correct forecasts of downward (upward) movements and d_2 (d_3) is the number of incorrect forecasts of downward (upward) movements. Define $N_1 = d_1 + d_3$, $N_2 = d_2 + d_4$, and $N = N_1 + N_2$. The null hypothesis is that there is no predictive information in our forecasts concerning the direction of change in y_{t+h}. One appropriate test for this hypothesis is thus the classical chi-square test of independence (see e.g., Pesaran and Timmermann (1994)). When reporting the results of this test, we also report the so-called directional accuracy rate, defined as $(d_1 + d_4)/N$. A higher rate, thus, indicates a higher probability of successfully predicting the direction of change.

2.4 Data

Various data are utilized in this paper, at frequencies ranging from 5-minutes to quarterly. However, our state space models, from which latent factors are extracted uses only daily, monthly and quarterly data. For daily data, the sample period is January 03, 2006 to December 31, 2018. For weekly data, the sample period is the first week of January 2006 through the last week of December 2018. For monthly and quarterly data, we collect observations for the period January 2009 - December 2018.

A number of our models utilize daily nonparametric volatility estimators of components of the quadratic variation of the S&P500, in addition to various daily macroeconomic and financial variables. These volatility estimators are constructed using 5-minute SPY (SPDR S&P 500 ETF Trust) transaction prices, which were downloaded from the TAQ database.

Our macroeconomic variables and bond yields are obtained from the FRED-MD database at the St. Louis Federal Reserve Bank. More specifically, the following macroeconomic variables were collected for use in our state space models: (1) the daily yield curve spread, defined as the difference between the 10-year U.S. Treasury bond yield and the 3-month Treasury bill yield; (2) weekly initial claims for unemployment insurance; (3) the monthly number of nonfarm payroll employees; and (4) quarterly gross domestic product. All of these variables are log differenced in all calculations in order to induce stationarity, and then standardized, with the exception of yield spreads, which are standardized.

Additionally, the following monthly macroeconomic variables were used as “target” variables in our forecasting experiments: industrial production (IP), the monthly number of total nonfarm payroll employees (PAY), housing starts (HS), personal consumption expenditures (PCE), the University of Michigan consumer sentiment index (SI), and core consumption price index (CPI), which excludes food and energy. The first three targets are key measurements related to business spending and residential investment activities, while the last three targets are direct reflections of consumption sector as well as price level. All of these variables are also log differenced in all calculations in order to induce stationarity, except for the housing starts and sentiment index as suggested by the FRED-MD database appendix to respectively take log and perform first order difference.

Finally, we also collected various corporate bond yield series for use as additional “target” variables in our forecasting experiments. These data consist of effective monthly bond yields for Fitch-rated AAA to CCC bonds (see Table 2 for details). The source of these data is the ICE Benchmark Administration Limited (IBA), and they are available at from the FRED database of the Federal Reserve Bank of St. Louis. All data are seasonally adjusted, and are differenced prior to our analysis. Details and transformations of macroeconomic and financial variables used in our research is listed in Table 2.1.

2.5 Empirical Findings

Tables 2.3 and 2.4 list relative MSFE (relative to the AR benchmark) for the case where the target variable is IP in our prediction experiments. Directional predictive accuracy results for IP are given in Tables 2.5 and 2.6. In these four tables, there are two distinct sample periods reported on. The first sample period is 2006:1-2018:12. For this period, $h = 1$ to $h = 6$-month ahead predictions are calculated for the period 2012:1-2018:12, and rolling windows of either 36 or 72 observations are utilized in model estimation. The second sample period is 2009:1-2018:12. For this period, 1 to 6-month ahead predictions are calculated for the period 2015:1-2018:12, and rolling windows of either 36 or 72 observations are utilized in model estimation. In Tables 2.3 and 2.4, the "MSFE-best" or"directional accuracy rate-best" models are denoted in bold, and starred entries indicate rejection of the Giacomini-White null hypothesis of equal predictive accuracy (in Tables 2.3 and 2.4, where all comparisons use our AR model as the benchmark, against which the models list in the first column of entries in the tables are compared). In Tables 2.5 and 2.6, the "directional accuracy rate-best" models are denoted in bold, and starred entries indicate rejection of the null hypothesis of statistical independence.[6] In addition, MSFE and directional accuracy rate results for all 13 target variables (see Table 3.2) are summarized in Tables 2.7 - 2.10. In these tables, only the "MSFE-best" models are listed, for a forecast horizon, window length, and sample period. Finally, Table 3.5 contains relative MSFEs analogous to those reported in Tables 2.3 and 2.4, except that a small subset of the most widely successful models listed in Table 2 are reported on, for all of the corporate bond yields analyzed in our experiments. The rest of this section summarizes our findings based on the results presented in these tables.

---

[6] Analogous detailed findings for all of our other 12 target variables are reported separately in an appendix, for the sake of brevity.

2.5.1 Variables related to business spending and investment

Our macroeconomic indicators related to business spending and residential investment include housing starts (HS), industrial production (IP) and nonfarm payroll employment (PAY). Our experimental findings for these variables are summarized in Tables 2.3 - 2.6, Tables 2.7 - 2.10, and in the appendix. Our conclusions based on analysis of these results are as follows.

First, for HS and IP, out-of-sample MSFEs significantly decrease in various factor augmented models, including RV, TRV, BPV, and JV, for example. Consider HS (see Tables 2.3 to 2.6). For our longer sample beginning in January 2006 (called Sample 1), use of the BPV model results in MSFE decreases (relative to the AR benchmark) of 10.3% when h=2 and w=36, and drops by 22.1%, when h=2 and w=72. Notice that the longer rolling window yields substantially lower MSFEs for the BPV model, which is our MSFE-best model. This finding characterizes many of our target variables, as evidenced upon inspection of 2.7 and 2.8, in which MSFE-best models are summarized, across all forecast horizons, windows, and sample periods. Interestingly, the maximum MSFE reduction of 53.7%, and is achieved for the TRV model, when h=6. In our shorter sample beginning in January 2009 (called Sample 2), RV, TRV, and BPV volatility factor augmented models again appreciably reduce MSFEs, for h=1 or 2, and w=36 or 72. Here, RV volatility factor augmented models reduce MFSE the most (15.8% when h=2 and w=36). Interestingly, our macro-volatility "convolution factor" augmented models also yield forecasting improvements, when w=72. For instance, the use of CMTRV1 decreases the MSFE by 7.3%, relative to the AR benchmark. Turning to IP, we note that our findings are qualitatively the same as those reported for HS, although MSFE reductions are at most 7%, across all horizons, windows, and sample periods.

As an aid to understanding the difference between the different risk factors in our analysis, we also provide a series of figures plotting the factors. In particular, refer to Figures 2.1 - 2.4 where all of the factors utilized in our experiments are plotted across our longer sample period (i.e. Sample 1). Notice, for example, that the macroeconomic factor plotted in Figure 2.1 is very different from the volatility risk factors plotted in Figures 2.2 - 2.4.

Second, for PAY, macro-volatility "square root convolution factors" (i.e., see models CMJV2, CMTRV2, CMBPV2, and CMJV2), lead to substantial reductions in MSFEs for both our shorter and longer sample periods. For example, for Sample 1, CMRV2, CMTRV2, and CMBPV2 augmented models result in MSFE decreases of 8.8%, 8%, and 3.2%, when h=1 and w=36; and CMTRV2 and CMBPV2 models result in MSFE decreases of 9.2% and 8.4%, when h=1 and w=72. For Sample 2, the strong results for our CMRV2, CMTRV2, and CMBPV2 models are replicated, as MSFE reductions are all greater than 10% (i.e., MSFE decreases are 17.2%, 15.6%, and 9.8%, for w=36, and MSFE decreases are 7.8%, 13.6%, and 11.6% for w=72, respectively).

Third, directional prediction results are also promising, indicating significant predictive accuracy gains, particularly for models with macro-volatility factors and "square root" factors (i.e., models CMJV1, CMTRV1, CMBPV1, CMJV1, CMJV2, CMTRV2, CMBPV2, and CMJV2). Still, it is worth noting that models associated with the largest directional accuracy rates are not always the same as those associated with the smallest relative MSFEs. For example, for HS, the directional accuracy "best"model is CMTRV2, in which the directional forecasting accuracy rate is 76.7%, for Sample 2 and h=1, which occurs when w=72. On the other hand, the analogous "MSFE-best" model is CMJV1. Nevertheless, the overall conclusion from inspection of this group of target variables is that volatility type factors are very useful for reducing MSFE and for increasing directional predictive accuracy, when considering business spending and residential investment related variables. Moreover, our "convolution" type factors appear to perform the best.

2.5.2 Variables related to consumer spending

Our macroeconomic indicators related to consumer spending include the consumer sentiment index (SI), the consumer price index (CPI), and personal consumption expenditures (PCE). Results from our prediction experiments using these variables are gathered in Tables 2.7 - 2.10, as well as in the appendix. Our conclusions based on analysis of these results are as follows.

First, the “MSFE-best” models often include factor augmented models. However, unlike the case of HS, IP, and PAY, where a large number of augmented models yield lower MSFEs than our benchmark AR model, we only observe occasional MSFE improvements for CPI, PCE, and SI. Still, even in the worst performing scenarios, such as in the case of CPI, there is some indication that risk factors may be useful. For example, for CPI at the h=5 horizon, the MRV model results in MSFE reductions of 9.6% (for w=36) and 5.1% (for w=72), under Sample 1 and MSFE reductions of 4.4% (w=36) and 5.4% (w=72), under Sample 2. Additionally, for SI at the h=1 horizon, the MBPV model results in MSFE reductions of 6.6% and 14.7%, for Sample 1 and Sample 2, respectively, when w=72.

Second, directional forecast accuracy rates are comparable to rates achieved for our business spending and residential investment variables, when factor augmented models are utilized for directional prediction. Still, as evidenced in Tables 2.7 and 2.8, AR models do sometimes yield the highest directional forecast accuracy rates. Given that this also occurs when predicting the direction of change for HS, IP, and PAY, we have evidence that risk factors are more useful for predicting absolute magnitudes of our variables than turning points. Drilling down into our findings more deeply, note that for CPI, our RV, TRV, and BPV models, as well as our factors or MRV, MTRV, and MBPV models generally result in around 5% to 6% increases in directional predictive accuracy, relative to the AR benchmark, when h=5 and w=72, for both Sample 1 and Sample 2. For PCE, the MRV, MTRV, and MBPV models yield increases in directional accuracy of comparable (and greater) magnitudes, when h=5 and w=72, under both sample periods.

2.5.3 Corporate bond yields

In our experiments, we also investigate the importance of market uncertainty on the corporate bond market, using yield series for AAA, AA, A, BBB, BB, and B rated bonds. Results for these experiments are gathered in Tables 2.7 - 2.10 and Table 3.5. Our conclusions based on analysis of these results are as follows.

First, it is very clear upon inspection of the MSFEs in Table 3.5 that predictive accuracy associated with the use of our risk factors increases as the quality of the bond deteriorates. The greatest gains associated with the use of our risk factors are associated with predicting junk bonds, while the is little to gain by using risk factors at all when predicting AAA rated bonds. Take the case of $w = 36$, which is reported in Table 3.5. Here, bonds with B and CCC ratings show the largest MSFE reductions from amongst all bonds, when the RV, TRV, and BPV factor augmented models are utilized. For example, for CCC-rated bond yield forecasting, the TRV model results in MSFE reductions of 9.3%, 22.4%, and 13.9%, for h=4 to 6, respectively, under Sample 1; and results in MSFE reductions of 12.9%, 35.6%, and 35.8% under Sample 2. For B-rated bonds, the TRV model results in MSFE reductions of 22.5%, 25.3%, and 12.4%, for h=4 to 6, respectively, under Sample 2.

However, predictive gains deteriorate as the investment quality of the bond improves. For example, for BB rated bonds under Sample 2, the TRV model results in MSFE reductions of 16.8%, 11.1%, and 3.3%, for h=4 to 6, respectively. All of these percentages are lower than the analogous percentages for CCC and B-rated bonds. The same result holds when comparing BBB versus BB-rated bonds, and A versus BBB-rated bonds etc. Thus, we have strong evidence of the usefulness of our risk factors for predicting corporate bond yields that involve financial risk, as might be expected.

Second, notice that the fourth row of entries in each panel of Table 3.5 are for the JV model. In this model, the only risk factor is our jump variation type factor. Needless to say, results are less than starling in these case. Apparently, jump-based risk factors are of little use when predicting corporate bond yields, and instead our risk factors that capture the continuous components of quadratic variation yield the most promising results.

2.6 Concluding Remarks

In this paper three macroeconomic and financial volatility risk factors are developed. The new risk factors are latent variables extracted from state space models that include multiple different frequencies of macroeconomic and financial variables, as well as multiple different frequencies of non-parametrically estimated components of quadratic variation. The state space models are specified in one of two ways. First, they may be specified solely using latent components of quadratic variation, including continuous and jump component variation measures extracted from high frequency S&P500 data. Alternatively, they may be specified using quadratic variation components as well as additional observed variables, including macroeconomic indicators such as interest rates, employment, and production. Finally, three types of risk factors are constructed. One type involves extracting volatility factors using only high frequency financial data, one type involves using only low frequency macroeconomic data, and onte type involves using both high frequency and low frequency data.

Our key findings can be summarized as follows. First, our multi-frequency financial and financial-macroeconomic volatility risk factors yield significantly improved predictions for a number of variables including housing starts, industrial production and nonfarm payroll, relative to benchmark models including simple autoregressive models, as well as mixed frequency models driven solely by macroeconomic indicators. Second, the same risk factors are useful for predicting low-grade corporate bond yields; but not high-grade corporate bond yields, which underscores the importance of the investment grade of bonds in withstanding turbulent market conditions, as might be expected. Third, four different measures of volatility are used in our analysis, including realized volatility (RV_t), truncated realized volatility (TRV_t), bi-power variation (BPV_t), and jump variation ($JV_t = RV_t - BPV_t$). In our forecasting experiments, TRV_t is clearly the most effective measure to use when constructing volatility risk factors. Moreover, factors constructed using JV_t fare quite poorly in our prediction experiments.

Table 2.1: Macroeconomic and Financial Variables

Name	Frequency	Description	Treatment
SPY	Daily	SPDR S&P 500 ETF Trust Price	$\Delta log(x_t)$
SPR	Daily	Yield Curve Spread	no transformation
		(10-year Treasury Bond Yield Minus 3-month Yield)	
IC	Weekly	Initial Claims for Unemployment Insurance	$\Delta log(x_t)$
PAY	Monthly	Number of Employees on Non-agricultural Payrolls	$\Delta log(x_t)$
GDP	Quarterly	Real Gross Domestic Product	$\Delta log(x_t)$
IP	Monthly	Industrial Production Index	$\Delta log(x_t)$
HS	Monthly	Housing Starts	$log(x_t)$
PCE	Monthly	Personal Consumption Expenditures	$\Delta log(x_t)$
SI	Monthly	University of Michigan Consumer Sentiment Index	Δx_t
CPI	Monthly	Consumer Price Index Less Food and Energy	$\Delta log(x_t)$
AAA	Monthly	US Corporate AAA Effective Yield	Δx_t
AA	Monthly	US Corporate AA Effective Yield	Δx_t
A	Monthly	US Corporate A Effective Yield	Δx_t
BBB	Monthly	US Corporate BBB Effective Yield	Δx_t
BB	Monthly	US High Yield BB Effective Yield	Δx_t
B	Monthly	US High Yield B Effective Yield	Δx_t
CCC	Monthly	US High Yield CCC or Below Effective Yield	Δx_t

*Notes: SPY data is downloaded from the WRDS Trade and Quotes (TAQ) database. All remaining series are obtained from the FRED-MD database of the St. Louis Federal Reserve Bank and are seasonally adjusted. Bond classifications from AAA to CCC are based on S&P500 and Fitch standards.

Table 2.2: Forecasting Models

Model	Description
Benchmark Model:	
Autoregression (AR)	$y_{t+h} = c + \sum_{i=1}^{r} \alpha_i y_{t-i} + \epsilon_t$
Macroeconomic Factor Augmented Model:	
MAC: $AR + MF^{mac}$	$y_{t+h} = c + \sum_{i=1}^{r} \alpha_i y_{t-i} + \rho_1 MF_t^{mac} + \epsilon_t$
Volatility Factor Augmented Models:	
RV: $AR + MF^{RV}$	$y_{t+h} = c + \sum_{i=1}^{r} \alpha_i y_{t-i} + \rho_3 MF_t^{RV} + \epsilon_t$
TRV: $AR + MF^{TRV}$	$y_{t+h} = c + \sum_{i=1}^{r} \alpha_i y_{t-i} + \rho_3 MF_t^{TRV} + \epsilon_t$
BPV: $AR + MF^{BPV}$	$y_{t+h} = c + \sum_{i=1}^{r} \alpha_i y_{t-i} + \rho_3 MF_t^{BPV} + \epsilon_t$
JV: $AR + MF^{JV}$	$y_{t+h} = c + \sum_{i=1}^{r} \alpha_i y_{t-i} + \rho_3 MF_t^{JV} + \epsilon_t$
Macro-Volatility Convolution Factor Augmented Models:	
CMRV1: $AR + MF^{mac-RV}$	$y_{t+h} = c + \sum_{i=1}^{r} \alpha_i y_{t-i} + \rho_2 MF_t^{mac-RV} + \epsilon_t$
CMTRV1: $AR + MF^{mac-TRV}$	$y_{t+h} = c + \sum_{i=1}^{r} \alpha_i y_{t-i} + \rho_2 MF_t^{\text{mac-TRV}} + \epsilon_t$
CMBPV1: $AR + MF^{mac-BPV}$	$y_{t+h} = c + \sum_{i=1}^{r} \alpha_i y_{t-i} + \rho_2 MF_t^{mac-BPV} + \epsilon_t$
CMJV1: $AR + MF^{mac-JV}$	$y_{t+h} = c + \sum_{i=1}^{r} \alpha_i y_{t-i} + \rho_2 MF_t^{mac-JV} + \epsilon_t$
CMRV2: $AR + MF^{mac-RVsqrt}$	$y_{t+h} = c + \sum_{i=1}^{r} \alpha_i y_{t-i} + \rho_2 MF_t^{mac-RVsqrt} + \epsilon_t$
CMTRV2: $AR + MF^{mac-TRVsqrt}$	$y_{t+h} = c + \sum_{i=1}^{r} \alpha_i y_{t-i} + \rho_2 MF_t^{\text{mac-TRVsqrt}} + \epsilon_t$
CMBPV2: $AR + MF^{mac-BPVsqrt}$	$y_{t+h} = c + \sum_{i=1}^{r} \alpha_i y_{t-i} + \rho_2 MF_t^{mac-BPVsqrt} + \epsilon_t$
CMJV2: $AR + MF^{mac-JVsqrt}$	$y_{t+h} = c + \sum_{i=1}^{r} \alpha_i y_{t-i} + \rho_2 MF_t^{mac-JVsqrt} + \epsilon_t$
Volatility Augmented Models:	
VRV: $AR + RV$	$y_{t+h} = c + \sum_{i=1}^{r} \alpha_i y_{t-i} + \gamma RV_t + \epsilon_t$
VTRV: $AR + TRV$	$y_{t+h} = c + \sum_{i=1}^{r} \alpha_i y_{t-i} + \gamma TRV_t + \epsilon_t$
VBPV: $AR + BPV$	$y_{t+h} = c + \sum_{i=1}^{r} \alpha_i y_{t-i} + \gamma BPV_t + \epsilon_t$
VJV: $AR + JV$	$y_{t+h} = c + \sum_{i=1}^{r} \alpha_i y_{t-i} + \gamma JV_t + \epsilon_t$
Macroeconomic Factor and Volatility Factor Augmented Models:	
MRV: $AR + MF^{mac} + MF^{RV}$	$y_{t+h} = c + \sum_{i=1}^{r} \alpha_i y_{t-i} + \rho_1 MF_t^{mac} + \rho_2 MF_t^{RV} + \epsilon_t$
MTRV: $AR + MF^{mac} + MF^{TRV}$	$y_{t+h} = c + \sum_{i=1}^{r} \alpha_i y_{t-i} + \rho_1 MF_t^{mac} + \rho_2 MF_t^{TRV} + \epsilon_t$
MBPV: $AR + MF^{mac} + MF^{BPV}$	$y_{t+h} = c + \sum_{i=1}^{r} \alpha_i y_{t-i} + \rho_1 MF_t^{mac} + \rho_2 MF_t^{BPV} + \epsilon_t$
MJV: $AR + MF^{mac} + MF^{JV}$	$y_{t+h} = c + \sum_{i=1}^{r} \alpha_i y_{t-i} + \rho_1 MF_t^{mac} + \rho_2 MF_t^{JV} + \epsilon_t$
Macroeconomic Factor and Volatility Augmented Models:	
MVRV: $AR + MF^{mac} + RV$	$y_{t+h} = c + \sum_{i=1}^{r} \alpha_i y_{t-i} + \rho_1 MF_t^{mac} + \gamma RV_t + \epsilon_t$
MVTRV: $AR + MF^{mac} + TRV$	$y_{t+h} = c + \sum_{i=1}^{r} \alpha_i y_{t-i} + \rho_1 MF_t^{mac} + \gamma TRV_t + \epsilon_t$
MVBPV: $AR + MF^{mac} + BPV$	$y_{t+h} = c + \sum_{i=1}^{r} \alpha_i y_{t-i} + \rho_1 MF_t^{mac} + \gamma BPV_t + \epsilon_t$
MVJV: $AR + MF^{mac} + JV$	$y_{t+h} = c + \sum_{i=1}^{r} \alpha_i y_{t-i} + \rho_1 MF_t^{mac} + \gamma JV_t + \epsilon_t$

*Notes: For the AR model, the number of lags, r, is determined using the AIC (results for the case where the SIC is instead used are qualitatively the same). All MF volatility risk factors are constructed as detailed in Section 2. In this table, MF^{mac} denote purely macroeconomic risk factors, while MF^{RV}, MF^{TRV}, MF^{BPV}, and MF^{JV} denote volatility risk factors based on the use of RV, TRV, BPV, and JV estimators, respectively. Here, RV, TRV, BPV, and JV correspond to realized volatility, truncated realized volatility, bi-power variation, and the jump component of quadratic variation, respectively. Additionally, MF^{mac-RV}, $MF^{mac-TRV}$, $MF^{mac-BPV}$, and MF^{mac-JV} are convolution factors that combine macroeconomic variables and one of either RV, TRV, BPV, and JV in formulating the risk factor. Finally, $MF^{mac-RVsqrt}$, $MF^{mac-TRVsqrt}$, $MF^{mac-BPVsqrt}$, $MF^{mac-JVsqrt}$ are of the same, but use square roots of RV, TRV, BPV, and JV estimators in their construction.

Table 2.3: Ex-Ante Relative MSFEs for Housing Starts (Sample 1: 2006:1 - 2018:12)

Model	Forecast horizon					
	1-month	2-month	3-month	4-month	5-month	6-month
			rolling window size = 36			
AR	1.000	1.000	1.000	1.000	1.000	1.000
$AR + MF^{mac}$	1.173***	1.092***	1.104***	1.027	0.981***	**0.938*****
$AR + MF^{RV}$	**0.93*****	**0.897*****	0.999***	1.006	1.037	1.038***
$AR + MF^{TRV}$	0.93***	0.902***	1.015	1.013	1.043	1.04***
$AR + MF^{BPV}$	0.931***	0.897***	1.012	1.002**	1.036	1.035***
$AR + MF^{JV}$	2.277***	2.178***	2.01***	1.726***	1.747***	1.602***
$AR + MF^{mac-RV}$	0.999***	1.096***	1.161***	1.028***	1.004*	0.968***
$AR + MF^{mac-TRV}$	1.019	1.003*	**0.998*****	1.004*	1.058***	1.047
$AR + MF^{mac-BPV}$	0.997***	1.096**	1.158***	1.033***	1.032***	0.967***
$AR + MF^{mac-JV}$	1.011	1.133***	1.078***	1.022	0.979***	0.952***
$AR + MF^{mac-RV\,sqrt}$	0.992***	0.974***	1.011	**0.982*****	1.079***	1.031
$AR + MF^{mac-TRV\,sqrt}$	0.991***	0.989***	1.055***	1.004*	1.062***	1.044
$AR + MF^{mac-BPV\,sqrt}$	1.007	1.02	1.016	0.992***	1.076***	1.059
$AR + MF^{mac-JV\,sqrt}$	1.014	0.991***	1.048***	1.031***	**0.962*****	0.963***
$AR + RV$	1.04***	1.022	1.052	1.041***	1.037***	0.983***
$AR + TRV$	1.037***	1.021	1.053	1.041***	1.038***	0.985***
$AR + BPV$	1.043***	1.023	1.058	1.049***	1.043***	0.982***
$AR + JV$	1.044*	0.952***	0.999***	0.998***	0.962***	0.964***
$AR + MF^{mac} + MF^{RV}$	1.152***	1.09***	1.165***	1.321***	1.194***	1.208***
$AR + MF^{mac} + MF^{TRV}$	1.151***	1.091***	1.181***	1.336***	1.199***	1.218***
$AR + MF^{mac} + MF^{BPV}$	1.146***	1.075***	1.16***	1.326***	1.194***	1.193***
$AR + MF^{mac} + MF^{JV}$	1.253***	1.254***	1.096	1.174***	1.252***	1.286***
$AR + MF^{mac} + RV$	1.272***	1.197***	1.116***	1.129***	1.005*	1.037
$AR + MF^{mac} + TRV$	1.272***	1.2***	1.115***	1.132***	1.016	1.04
$AR + MF^{mac} + BPV$	1.272***	1.196***	1.119***	1.129***	1.011	1.032
$AR + MF^{mac} + JV$	1.253***	1.156***	1.103***	1.016	1.029	1.085
			rolling window size = 72			
AR	1.000	1.000	1.000	1.000	1.000	1.000
$AR + MF^{mac}$	0.925***	0.851***	0.841***	0.7***	0.623***	0.591***
$AR + MF^{RV}$	**0.849*****	0.78***	0.711***	0.569***	0.484***	0.479***
$AR + MF^{TRV}$	0.852***	0.798***	0.724***	**0.561*****	**0.477*****	**0.463*****
$AR + MF^{BPV}$	0.855***	**0.779*****	**0.71*****	0.567***	0.482***	0.477***
$AR + MF^{JV}$	1.081	1.061	1.034	0.796***	0.63***	0.559***
$AR + MF^{mac-RV}$	0.92***	0.882***	0.841***	0.76***	0.709***	0.741***
$AR + MF^{mac-TRV}$	0.889***	0.883***	0.884***	0.832***	0.872***	0.881***
$AR + MF^{mac-BPV}$	0.916***	0.877***	0.833***	0.754***	0.704***	0.749***
$AR + MF^{mac-JV}$	0.906***	0.892***	0.88***	0.812***	0.756***	0.792***
$AR + MF^{mac-RV\,sqrt}$	0.91***	0.922***	0.919***	0.853***	0.875***	0.857***
$AR + MF^{mac-TRV\,sqrt}$	0.873***	0.868***	0.856***	0.834***	0.875***	0.812***
$AR + MF^{mac-BPV\,sqrt}$	0.866***	0.877***	0.899***	0.866***	0.872***	0.832***
$AR + MF^{mac-JV\,sqrt}$	0.922***	0.915***	0.911***	0.805***	0.721***	0.77***
$AR + RV$	0.922***	0.836***	0.766***	0.753***	0.642***	0.664***
$AR + TRV$	0.917***	0.837***	0.764***	0.753***	0.629***	0.66***
$AR + BPV$	0.923***	0.837***	0.764***	0.754***	0.637***	0.667***
$AR + JV$	0.938***	0.863***	0.883***	0.794***	0.698***	0.723***
$AR + MF^{mac} + MF^{RV}$	0.886***	0.806***	0.785***	0.66***	0.607***	0.615***
$AR + MF^{mac} + MF^{TRV}$	0.887***	0.813***	0.789***	0.668***	0.607***	0.616***
$AR + MF^{mac} + MF^{BPV}$	0.884***	0.804***	0.782***	0.655***	0.605***	0.61***
$AR + MF^{mac} + MF^{JV}$	1.011	0.894***	0.801***	0.712***	0.631***	0.658***
$AR + MF^{mac} + RV$	0.93***	0.823***	0.838***	0.706***	0.627***	0.639***
$AR + MF^{mac} + TRV$	0.928***	0.825***	0.842***	0.706***	0.627***	0.633***
$AR + MF^{mac} + BPV$	0.93***	0.82***	0.834***	0.699***	0.626***	0.64***
$AR + MF^{mac} + JV$	0.92***	0.88***	0.839***	0.727***	0.629***	0.632***

*Notes: This table reports mean square forecast errors (MSFEs) relative to the AR benchmark model. The forecasting model is given in the first column (see Table 2 for a description of the models). Starred entries indicate rejections of the Giacomini and White (2006) test of conditional predictive accuracy. In particular, ***, **, and * indicate rejection at the 1%, 5%, and 10% levels, respectively. The entire sample period used in the forecasting experiment is 2006:1-2018:12, and ex-ante rolling window MSFEs correspond to predictions made for the period 2012:1 to 2018:12.

Table 2.4: Ex-Ante Relative MSFEs for Housing Starts (Sample 2: 2009:1 - 2018:12)

Model	Forecast horizon					
	1-month	2-month	3-month	4-month	5-month	6-month
			rolling window size = 36			
AR	1.000	1.000	1.000	1.000	1.000	1.000
$AR + MF^{mac}$	1.231***	1.114**	1.166	0.97***	0.967***	0.969***
$AR + MF^{RV}$	0.954***	0.842***	0.944***	0.924***	1.193***	1.064
$AR + MF^{TRV}$	**0.95*****	0.846***	0.981***	0.938***	1.205***	1.061
$AR + MF^{BPV}$	0.958***	0.847***	0.952***	**0.91*****	1.187***	1.05
$AR + MF^{JV}$	2.977***	3.567***	3.028***	1.923***	2.313***	2.173***
$AR + MF^{mac-RV}$	1.121***	1.191***	1.304***	1.056***	1.147***	0.924***
$AR + MF^{mac-TRV}$	1.03	0.993***	**0.927*****	1.074***	1.05***	0.99***
$AR + MF^{mac-BPV}$	1.122***	1.181***	1.296***	1.074***	1.149***	**0.92*****
$AR + MF^{mac-JV}$	1.121***	1.267***	1.148***	0.988***	1.083***	0.936***
$AR + MF^{mac-RV sqrt}$	0.958***	0.995***	0.944***	1.023	1.114***	0.986***
$AR + MF^{mac-TRV sqrt}$	0.994***	0.984***	1.081	1.052***	1.061***	1.009
$AR + MF^{mac-BPV sqrt}$	1.003*	0.987***	0.97***	1.03	1.098***	0.988***
$AR + MF^{mac-JV sqrt}$	1.058**	0.969***	1.189***	1.008	0.997***	0.94***
$AR + RV$	1.117***	1.145***	1.262***	1.023	1.047	1.007
$AR + TRV$	1.107***	1.136***	1.259***	1.021	1.045	1.008
$AR + BPV$	1.125***	1.153***	1.279***	1.032	1.061	1.012
$AR + JV$	1.137***	**0.818*****	0.989***	1.058***	**0.966*****	0.972***
$AR + MF^{mac} + MF^{RV}$	1.231***	1.123	1.259***	1.479***	1.474***	1.637***
$AR + MF^{mac} + MF^{TRV}$	1.232***	1.121	1.316***	1.512***	1.476***	1.652***
$AR + MF^{mac} + MF^{BPV}$	1.231***	1.099	1.243***	1.489***	1.473***	1.588***
$AR + MF^{mac} + MF^{JV}$	1.127	0.993***	1.048	1.024	1.212	1.513***
$AR + MF^{mac} + RV$	1.383***	1.393***	1.226***	1.219**	1.105	1.155
$AR + MF^{mac} + TRV$	1.379***	1.399***	1.224***	1.224**	1.107	1.156
$AR + MF^{mac} + BPV$	1.389***	1.4***	1.225***	1.217*	1.102	1.148
$AR + MF^{mac} + JV$	1.235***	1.245***	1.296***	1.028	1.095	1.345***
			rolling window size = 72			
AR	1.000	1.000	1.000	1.000	1.000	1.000
$AR + MF^{mac}$	1.064	1.043	1.03	1.02	1.05	**0.949*****
$AR + MF^{RV}$	0.989***	0.963***	0.998***	1.001**	0.994***	0.971***
$AR + MF^{TRV}$	0.985***	0.97***	1.001**	0.974***	0.997***	0.972***
$AR + MF^{BPV}$	1.014	0.958***	0.994***	0.996***	0.992***	0.971***
$AR + MF^{JV}$	1.42***	1.447***	1.356***	1.239***	1.192***	1.117***
$AR + MF^{mac-RV}$	1.037	0.95***	0.978***	0.965***	0.958***	0.98***
$AR + MF^{mac-TRV}$	1.033	**0.927*****	1.022***	**0.951*****	0.975***	1.033
$AR + MF^{mac-BPV}$	1.043	0.952***	0.98***	0.963***	0.956***	0.962***
$AR + MF^{mac-JV}$	**0.938*****	0.958***	1.009	1.049	0.991***	1.079***
$AR + MF^{mac-RV sqrt}$	0.989***	0.994***	1.028***	0.999***	1.004	1.048
$AR + MF^{mac-TRV sqrt}$	0.971***	0.974***	1.036***	0.979***	1.004	0.979***
$AR + MF^{mac-BPV sqrt}$	0.988***	0.947***	1.038***	0.963***	0.984***	1.028
$AR + MF^{mac-JV sqrt}$	0.955***	0.943***	1.024***	1.036*	0.983***	1.083***
$AR + RV$	0.999***	1.008	0.956***	1.017	0.949***	0.984***
$AR + TRV$	0.997***	1.002**	0.958***	1.021	0.952***	1.001**
$AR + BPV$	1.002*	1.014	**0.955*****	1.016	**0.947*****	0.993***
$AR + JV$	1.019	0.961***	1.026	0.998***	1.006	1.032*
$AR + MF^{mac} + MF^{RV}$	1.106*	1.027	1.049	1.053	1.087	0.982***
$AR + MF^{mac} + MF^{TRV}$	1.105*	1.032	1.053	1.056	1.091	0.987***
$AR + MF^{mac} + MF^{BPV}$	1.106*	1.022	1.046	1.049	1.085	0.981***
$AR + MF^{mac} + MF^{JV}$	1.267***	1.332***	1.188***	1.222***	1.141*	1.144
$AR + MF^{mac} + RV$	1.147***	0.991***	1.04	0.984***	1.06	0.986***
$AR + MF^{mac} + TRV$	1.141***	0.99***	1.042	0.985***	1.058	1.009*
$AR + MF^{mac} + BPV$	1.153***	0.993***	1.04	0.984***	1.057	0.988***
$AR + MF^{mac} + JV$	1.008*	1.087	0.976***	1.026	1.008*	0.983***

*Notes: See notes to Table 2.3. The entire sample period used in the forecasting experiment is 2006:1-2018:12, and ex-ante rolling window MSFEs correspond to predictions made for the period 2015:1 to 2018:12.

Table 2.5: Ex-ante Directional Accuracy Rates for Housing Starts (Sample 1: 2006:1 - 2018:12)

Model	Forecast horizon					
	1-month	2-month	3-month	4-month	5-month	6-month
			rolling window size $= 36$			
AR	65.8%***	70.9%***	67.1%***	73.4%***	65.8%***	59.5%*
$AR + MF^{mac}$	62%***	65.8%***	65.8%***	69.6%***	67.1%***	64.6%***
$AR + MF^{RV}$	65.8%***	**73.4%*****	60.8%*	68.4%***	69.6%***	55.7%
$AR + MF^{TRV}$	65.8%***	73.4%***	62%**	69.6%***	**70.9%*****	54.4%
$AR + MF^{BPV}$	65.8%***	73.4%***	60.8%*	69.6%***	70.9%***	55.7%
$AR + MF^{JV}$	48.1%	49.4%	55.7%	57%	55.7%	55.7%
$AR + MF^{mac-RV}$	64.6%***	69.6%***	59.5%**	70.9%***	63.3%***	64.6%***
$AR + MF^{mac-TRV}$	69.6%***	70.9%***	65.8%***	**75.9%*****	67.1%***	58.2%
$AR + MF^{mac-BPV}$	64.6%***	69.6%***	59.5%**	70.9%***	63.3%***	64.6%***
$AR + MF^{mac-JV}$	67.1%***	68.4%***	67.1%***	70.9%***	63.3%***	62%**
$AR + MF^{mac-RVsqrt}$	68.4%***	69.6%***	**68.4%*****	75.9%***	67.1%***	60.8%**
$AR + MF^{mac-TRVsqrt}$	**70.9%*****	69.6%***	67.1%***	75.9%***	67.1%***	57%
$AR + MF^{mac-BPVsqrt}$	69.6%***	69.6%***	67.1%***	75.9%***	67.1%***	58.2%
$AR + MF^{mac-JVsqrt}$	68.4%***	67.1%***	62%**	67.1%***	65.8%***	64.6%***
$AR + RV$	65.8%***	67.1%***	60.8%**	67.1%***	69.6%***	57%
$AR + TRV$	63.3%***	67.1%***	60.8%**	67.1%***	69.6%***	55.7%
$AR + BPV$	64.6%***	67.1%***	62%**	67.1%***	69.6%***	57%
$AR + JV$	68.4%***	68.4%***	63.3%**	74.7%***	68.4%***	60.8%**
$AR + MF^{mac} + MF^{RV}$	59.5%**	63.3%***	62%**	68.4%***	67.1%***	62%**
$AR + MF^{mac} + MF^{TRV}$	58.2%*	63.3%***	62%**	67.1%***	67.1%***	62%**
$AR + MF^{mac} + MF^{BPV}$	60.8%**	63.3%***	63.3%**	67.1%***	67.1%***	62%**
$AR + MF^{mac} + MF^{JV}$	60.8%**	55.7%	62%**	64.6%***	64.6%***	**68.4%*****
$AR + MF^{mac} + RV$	57%*	65.8%***	65.8%***	67.1%***	67.1%***	67.1%***
$AR + MF^{mac} + TRV$	57%*	65.8%***	65.8%***	65.8%***	67.1%***	67.1%***
$AR + MF^{mac} + BPV$	57%*	65.8%***	65.8%***	67.1%***	65.8%***	67.1%***
$AR + MF^{mac} + JV$	62%**	65.8%***	65.8%***	74.7%***	68.4%***	63.3%**
			rolling window size $= 72$			
AR	72.2%***	69.6%***	63.3%***	58.2%**	48.1%	48.1%
$AR + MF^{mac}$	70.9%***	73.4%***	70.9%***	70.9%***	63.3%***	64.6%***
$AR + MF^{RV}$	70.9%***	74.7%***	67.1%***	**79.7%*****	**73.4%*****	68.4%***
$AR + MF^{TRV}$	72.2%***	74.7%***	68.4%***	79.7%***	73.4%***	**70.9%*****
$AR + MF^{BPV}$	70.9%***	74.7%***	67.1%***	79.7%***	73.4%***	67.1%***
$AR + MF^{JV}$	64.6%***	60.8%**	65.8%***	74.7%***	63.3%	68.4%***
$AR + MF^{mac-RV}$	72.2%***	73.4%***	**75.9%*****	64.6%***	57%*	53.2%
$AR + MF^{mac-TRV}$	**74.7%*****	72.2%***	69.6%***	63.3%***	53.2%	57%*
$AR + MF^{mac-BPV}$	72.2%***	73.4%***	73.4%***	64.6%***	57%*	50.6%
$AR + MF^{mac-JV}$	74.7%***	72.2%***	69.6%***	58.2%**	53.2%	51.9%
$AR + MF^{mac-RVsqrt}$	73.4%***	69.6%***	65.8%***	65.8%***	54.4%	58.2%*
$AR + MF^{mac-TRVsqrt}$	73.4%***	72.2%***	69.6%***	64.6%***	50.6%	55.7%
$AR + MF^{mac-BPVsqrt}$	73.4%***	72.2%***	69.6%***	65.8%***	51.9%	57%*
$AR + MF^{mac-JVsqrt}$	72.2%***	70.9%***	67.1%***	59.5%***	58.2%**	55.7%
$AR + RV$	74.7%***	**78.5%*****	70.9%***	60.8%***	62%**	60.8%**
$AR + TRV$	74.7%***	78.5%***	70.9%***	60.8%***	62%**	60.8%**
$AR + BPV$	74.7%***	77.2%***	70.9%***	60.8%***	62%**	58.2%*
$AR + JV$	73.4%***	75.9%***	68.4%***	60.8%***	60.8%**	60.8%**
$AR + MF^{mac} + MF^{RV}$	68.4%***	72.2%***	70.9%***	73.4%***	62%**	62%***
$AR + MF^{mac} + MF^{TRV}$	68.4%***	74.7%***	70.9%***	73.4%***	63.3%***	60.8%**
$AR + MF^{mac} + MF^{BPV}$	68.4%***	73.4%***	70.9%***	73.4%***	64.6%***	62%***
$AR + MF^{mac} + MF^{JV}$	63.3%***	72.2%***	67.1%***	70.9%***	62%***	58.2%**
$AR + MF^{mac} + RV$	65.8%***	74.7%***	70.9%***	70.9%***	62%**	62%***
$AR + MF^{mac} + TRV$	65.8%***	74.7%***	69.6%***	70.9%***	62%***	60.8%**
$AR + MF^{mac} + BPV$	65.8%***	74.7%***	70.9%***	72.2%***	62%**	62%***
$AR + MF^{mac} + JV$	67.1%***	75.9%***	70.9%***	72.2%***	62%**	62%***

*Notes: See notes to Table 2.3. Entries in this table are direction accuracy rates, and starred entries denote rejection of the directional accuracy test based on the contingency tables discussed in Section 3 and Pesaran and Timmermann (1994).

Table 2.6: Ex-ante Directional Accuracy Rates for Housing Starts (Sample 2: 2009:1 - 2018:12)

Model	Forecast horizon					
	1-month	2-month	3-month	4-month	5-month	6-month
			rolling window size = 36			
AR	62.8%	76.7%***	79.1%***	72.1%***	65.1%**	65.1%**
$AR + MF^{mac}$	65.1%	69.8%***	72.1%***	76.7%***	72.1%***	69.8%***
$AR + MF^{RV}$	62.8%	**79.1%*****	69.8%***	69.8%***	74.4%***	62.8%**
$AR + MF^{TRV}$	62.8%	79.1%***	72.1%***	69.8%***	**76.7%*****	60.5%*
$AR + MF^{BPV}$	62.8%**	79.1%***	69.8%***	69.8%***	76.7%***	62.8%**
$AR + MF^{JV}$	41.9%	46.5%	53.5%	58.1%	62.8%*	55.8%
$AR + MF^{mac-RV}$	62.8%**	69.8%***	67.4%***	74.4%***	60.5%*	74.4%***
$AR + MF^{mac-TRV}$	69.8%***	72.1%***	76.7%***	76.7%***	67.4%***	67.4%***
$AR + MF^{mac-BPV}$	62.8%**	69.8%***	67.4%***	74.4%***	60.5%*	74.4%***
$AR + MF^{mac-JV}$	65.1%**	67.4%***	76.7%***	72.1%***	60.5%*	69.8%***
$AR + MF^{mac-RVsqrt}$	69.8%***	69.8%***	79.1%***	76.7%***	67.4%***	67.4%***
$AR + MF^{mac-TRVsqrt}$	**72.1%*****	69.8%***	76.7%***	76.7%***	67.4%***	65.1%**
$AR + MF^{mac-BPVsqrt}$	69.8%***	72.1%***	**81.4%*****	76.7%***	67.4%***	67.4%***
$AR + MF^{mac-JVsqrt}$	67.4%**	72.1%***	69.8%***	72.1%***	65.1%**	69.8%***
$AR + RV$	65.1%**	69.8%***	69.8%***	72.1%***	72.1%***	62.8%**
$AR + TRV$	65.1%**	69.8%***	69.8%***	72.1%***	72.1%***	60.5%*
$AR + BPV$	62.8%*	69.8%***	69.8%***	72.1%***	72.1%***	62.8%**
$AR + JV$	69.8%***	72.1%***	72.1%***	74.4%***	69.8%***	67.4%***
$AR + MF^{mac} + MF^{RV}$	58.1%	62.8%**	67.4%***	72.1%***	67.4%***	69.8%***
$AR + MF^{mac} + MF^{TRV}$	58.1%	62.8%**	67.4%***	69.8%***	67.4%***	69.8%***
$AR + MF^{mac} + MF^{BPV}$	60.5%*	62.8%**	69.8%***	69.8%***	67.4%***	69.8%***
$AR + MF^{mac} + MF^{JV}$	60.5%*	60.5%*	67.4%***	67.4%***	69.8%***	72.1%***
$AR + MF^{mac} + RV$	55.8%	67.4%***	74.4%***	67.4%**	69.8%***	**76.7%*****
$AR + MF^{mac} + TRV$	55.8%	67.4%***	74.4%***	65.1%**	69.8%***	76.7%***
$AR + MF^{mac} + BPV$	55.8%	67.4%***	74.4%***	67.4%***	67.4%***	76.7%***
$AR + MF^{mac} + JV$	62.8%**	67.4%***	72.1%***	**79.1%*****	72.1%***	72.1%***
			rolling window size = 72			
AR	74.4%***	74.4%***	79.1%***	74.4%***	62.8%*	62.8%**
$AR + MF^{mac}$	67.4%***	69.8%***	76.7%***	74.4%***	62.8%*	**69.8%*****
$AR + MF^{RV}$	72.1%***	69.8%***	74.4%***	76.7%***	**67.4%*****	67.4%***
$AR + MF^{TRV}$	72.1%***	69.8%***	76.7%***	76.7%***	67.4%***	67.4%***
$AR + MF^{BPV}$	72.1%***	69.8%***	74.4%***	76.7%***	67.4%***	65.1%***
$AR + MF^{JV}$	58.1%**	60.5%*	72.1%***	**79.1%*****	58.1%	65.1%**
$AR + MF^{mac-RV}$	**76.7%*****	74.4%***	**81.4%*****	79.1%***	62.8%*	65.1%**
$AR + MF^{mac-TRV}$	74.4%***	72.1%***	81.4%***	76.7%***	60.5%	65.1%***
$AR + MF^{mac-BPV}$	76.7%***	74.4%***	76.7%***	79.1%***	62.8%*	65.1%**
$AR + MF^{mac-JV}$	76.7%***	69.8%***	72.1%***	72.1%***	65.1%**	65.1%***
$AR + MF^{mac-RVsqrt}$	76.7%***	69.8%***	81.4%***	79.1%***	62.8%*	65.1%***
$AR + MF^{mac-TRVsqrt}$	76.7%***	69.8%***	81.4%***	76.7%***	58.1%	65.1%***
$AR + MF^{mac-BPVsqrt}$	76.7%***	72.1%***	81.4%***	79.1%***	60.5%	65.1%***
$AR + MF^{mac-JVsqrt}$	74.4%***	67.4%***	72.1%***	74.4%***	65.1%**	65.1%***
$AR + RV$	76.7%***	**76.7%*****	79.1%***	74.4%***	67.4%***	65.1%**
$AR + TRV$	76.7%***	76.7%***	79.1%***	74.4%***	67.4%***	65.1%**
$AR + BPV$	76.7%***	74.4%***	79.1%***	74.4%***	67.4%***	60.5%*
$AR + JV$	74.4%***	74.4%***	76.7%***	76.7%***	67.4%***	69.8%***
$AR + MF^{mac} + MF^{RV}$	60.5%*	72.1%***	74.4%***	76.7%***	62.8%*	69.8%***
$AR + MF^{mac} + MF^{TRV}$	60.5%*	72.1%***	74.4%***	76.7%***	62.8%*	69.8%***
$AR + MF^{mac} + MF^{BPV}$	60.5%*	72.1%***	74.4%***	76.7%***	65.1%**	69.8%***
$AR + MF^{mac} + MF^{JV}$	60.5%**	69.8%***	74.4%***	76.7%***	67.4%***	65.1%**
$AR + MF^{mac} + RV$	60.5%	72.1%***	76.7%***	76.7%***	62.8%*	69.8%***
$AR + MF^{mac} + TRV$	60.5%	72.1%***	74.4%***	76.7%***	62.8%*	67.4%***
$AR + MF^{mac} + BPV$	60.5%	72.1%***	76.7%***	79.1%***	62.8%*	69.8%***
$AR + MF^{mac} + JV$	62.8%**	72.1%***	76.7%***	79.1%***	62.8%*	67.4%***

*Notes: See notes to Table 2.5.

Table 2.7: Best Models in Ex-Ante Relative MSFEs (Sample 1: 2006:1 - 2018:12)

Targets	Forecast horizon 1-month	2-month	3-month	4-month	5-month	6-month
	rolling window size = 36					
HS	RV	RV	CMTRV1	CMRV2	CMJV2	MAC
	0.93^{***}	0.897^{***}	0.998^{***}	0.982^{***}	0.962^{***}	0.938^{***}
IP	CMTRV1	MVJV	MVRV	MAC	CMJV2	TRV
	0.98667***	0.93895***	0.95182***	1.0098	0.97259***	0.99748***
PAY	CMRV2	VBPV	MAC	CMTRV1	MAC	VJV
	0.912^{***}	1.024	0.978^{***}	0.926^{***}	0.956^{***}	1.01^{*}
CPI	VJV	VJV	CMJV2	CMJV2	CMRV2	VBPV
	1.025^{***}	0.998^{***}	0.995^{***}	0.972^{***}	0.904^{***}	1.024
PCE	CMRV2	CMBPV1	VJV	CMJV2	VJV	CMJV2
	1.008	1^{***}	1.02	0.986^{***}	0.943^{***}	0.973^{***}
SI	MVRV	CMJV2	VRV	MAC	CMTRV2	CMJV2
	1.006	0.998^{***}	1.028^{***}	1.022	1.012	0.965^{***}
AAA	VJV	CMBPV2	TRV	BPV	TRV	MAC
	1.0065	0.98697	0.87274	0.8946	0.99271	0.97587
AA	VJV	CMTRV2	BPV	VJV	VJV	CMTRV1
	0.90654	0.99637	0.91598	0.89562	0.97354	1.0031
A	VJV	CMTRV1	BPV	VJV	VJV	CMTRV1
	0.98338	0.97003	0.91351	0.85667	0.99297	1.0158
BBB	VJV	VJV	CMJV2	VJV	BPV	CMTRV1
	0.99678	1.0118	0.94862	0.93972	0.98152	0.93416
BB	CMTRV1	MAC	CMJV2	CMJV2	VJV	CMJV1
	1.0498	1.0286	0.94474	0.98845	0.94701	0.88478
B	VJV	VJV	CMJV1	VJV	CMJV2	CMJV2
	1.0539	0.99583	0.9106	0.82958	0.96374	0.94034
CCC	VBPV	CMJV2	CMJV1	TRV	BPV	CMJV2
	1.0079	0.99705	0.95569	0.90124	0.77627	0.8299
	rolling window size = 72					
HS	RV	BPV	BPV	TRV	TRV	TRV
	0.849^{***}	0.779^{***}	0.71^{***}	0.561^{***}	0.477^{***}	0.463^{***}
IP	MAC	CMRV2	MVBPV	MVRV	MVRV	MVRV
	0.95362***	0.9852***	0.88961***	0.9239***	0.95624***	0.9122***
PAY	MAC	VTRV	VJV	VTRV	VBPV	RV
	0.832^{***}	0.931^{***}	0.837^{***}	0.795^{***}	0.788^{***}	0.678^{***}
CPI	CMRV1	CMTRV2	CMBPV2	CMRV1	CMRV2	CMJV2
	0.98^{***}	1^{***}	1.007	0.993^{***}	0.949^{***}	1.001^{**}
PCE	CMBPV2	CMTRV2	MVRV	MVTRV	MVBPV	MAC
	0.977^{***}	1.008	0.991^{***}	1.025	0.948^{***}	0.952^{***}
SI	CMBPV1	CMJV2	CMBPV2	CMJV1	CMTRV2	CMTRV1
	0.946^{***}	1.001	0.991^{***}	1.008	0.984^{***}	0.985^{***}
AAA	MRV	MAC	CMJV1	MBPV	MVBPV	MVBPV
	0.86201	0.94152	0.94201	0.90781	0.88774	0.86193
AA	MTRV	TRV	MVJV	VJV	MVRV	VTRV
	0.91194	0.85887	0.87634	0.8722	0.89221	0.90176
A	MTRV	TRV	MVJV	VJV	VTRV	VBPV
	0.97151	0.91153	0.92619	0.93121	0.92312	0.90901
BBB	MAC	MTRV	MVJV	VJV	CMJV2	CMTRV2
	1.0113	0.92004	0.89056	0.88122	0.969	0.90507
BB	MVJV	MVJV	MVJV	VJV	CMJV1	CMJV1
	1.0122	1.0111	0.88293	0.97694	0.94214	0.88274
B	MAC	CMJV2	CMJV2	CMBPV2	CMJV1	CMJV2
	0.97675	0.98626	0.91961	0.98139	0.94611	0.90126
CCC	MAC	MVRV	TRV	CMTRV1	CMJV1	CMTRV1
	0.93422	0.90557	0.9944	0.96925	0.93447	0.96305

Table 2.8: MSFE-Best Models (Sample 2: 2009:1 - 2018:12)

Targets	Forecast horizon 1-month	2-month	3-month	4-month	5-month	6-month
			rolling window size = 36			
HS	TRV	VJV	CMTRV1	BPV	VJV	CMBPV1
	0.95***	0.818***	0.927***	0.91***	0.966***	0.92***
IP	CMTRV1	MVJV	MVRV	CMRV2	CMJV2	TRV
	0.93117***	0.87597***	0.95576***	1.0175	0.9749***	0.97453***
PAY	CMRV2	VJV	CMRV1	CMTRV1	MAC	CMJV2
	0.828***	1.04	0.996***	0.91***	0.993***	0.994***
CPI	CMJV2	VJV	CMTRV2	CMJV1	CMRV1	VBPV
	1.015***	0.988***	0.986***	0.946***	0.954***	1.043
PCE	CMRV2	CMRV2	CMJV2	CMJV2	VRV	CMTRV1
	0.992***	0.977***	1.011	1.016	0.992***	0.962***
SI	CMJV2	CMJV2	VJV	CMJV2	CMTRV2	CMJV2
	1.036	1.013***	1.012	1.028***	0.993***	0.976***
AAA	CMJV2	CMTRV1	VJV	VJV	CMJV2	CMJV2
	0.99522	0.93567	0.90953	0.95673	0.9595	0.98716
AA	BPV	TRV	BPV	VJV	VJV	CMTRV2
	0.99941	0.95162	0.85747	0.8146	0.84053	0.93124
A	CMJV2	TRV	BPV	VJV	VJV	TRV
	1.009	0.93303	0.8267	0.708	0.8238	0.92889
BBB	CMTRV1	VJV	BPV	BPV	BPV	CMTRV1
	1.0232	1.0067	0.97528	0.82123	0.85947	0.84668
BB	VBPV	TRV	CMJV2	MTRV	BPV	CMJV2
	0.85991	1.0043	0.9211	0.88011	0.87015	0.77122
B	VRV	CMJV2	MTRV	MTRV	MTRV	CMJV2
	0.94503	0.93527	0.85066	0.72681	0.66881	0.82391
CCC	VBPV	CMJV2	TRV	TRV	BPV	BPV
	0.99349	0.96359	0.88827	0.84801	0.64426	0.64237
			rolling window size = 72			
HS	CMJV1	CMTRV1	VBPV	CMTRV1	VBPV	MAC
	0.938***	0.927***	0.955***	0.951***	0.947***	0.949***
IP	MAC	TRV	MVBPV	VRV	CMJV1	MVJV
	0.95002***	0.96174***	0.97213***	0.96263***	0.96238***	0.93786***
PAY	CMTRV2	CMJV2	CMRV2	CMJV2	MRV	VTRV
	0.864***	1.048*	0.975***	0.957***	0.902***	0.953***
CPI	CMRV1	CMTRV2	CMBPV2	CMRV2	CMRV2	TRV
	0.97***	0.993***	0.985***	0.983***	0.946***	0.995***
PCE	MVJV	CMTRV2	MVRV	MVBPV	MVBPV	MAC
	0.979***	0.994***	0.934***	0.962***	0.894***	0.858***
SI	CMBPV1	CMJV2	CMBPV2	MAC	CMRV2	CMTRV1
	0.853***	1***	0.982***	0.948***	0.987***	0.943***
AAA	MTRV	MTRV	CMBPV2	MBPV	MVBPV	MJV
	0.88002	0.94108	0.94533	0.89043	0.86167	0.89758
AA	VJV	MTRV	MTRV	BPV	BPV	VRV
	0.87022	0.84778	0.88517	0.78887	0.85026	0.91993
A	VJV	MJV	TRV	TRV	BPV	MTRV
	0.89387	0.85091	0.80556	0.72695	0.81355	0.88302
BBB	VJV	MJV	TRV	TRV	TRV	BPV
	0.96114	0.82558	0.84309	0.78041	0.81628	0.86507
BB	VBPV	CMJV2	VTRV	VJV	CMJV1	CMJV2
	0.84221	0.91287	0.89463	0.87601	0.91901	0.83331
B	VBPV	CMJV2	MVJV	VJV	CMJV2	CMJV2
	0.83521	0.93668	0.92042	0.86939	0.89025	0.84088
CCC	MAC	MVTRV	MTRV	VTRV	BPV	CMJV2
	0.94713	0.89172	0.97729	0.92905	0.9243	0.96774

*Notes: See notes to Table 2.3. Results are analogous to those depicted in Table 2.7, except that Sample 2 is used instead of Sample 1 in all prediction experiments.

Table 2.9: Best Models in Ex-ante Directional Accuracy Rate (Sample 1: 2006:1 - 2018:12)

Targets	Forecast horizon					
	1-month	2-month	3-month	4-month	5-month	6-month
			rolling window size $= 36$			
HS	CMTRV2	RV	CMRV2	CMTRV1	TRV	MJV
	70.9%***	73.4%***	68.4%***	75.9%***	70.9%***	68.4%***
IP	CMJV1	CMTRV1	CMBPV2	MJV	MVRV	MRV
	72.2% ***	73.4% ***	65.8% ***	72.2% ***	74.7% ***	64.6% **
PAY	MVJV	CMTRV2	MVJV	MAC	MAC	CMRV2
	78.5%***	81%***	72.2%***	72.2%***	75.9%***	75.9%***
CPI	BPV	MAC	BPV	CMBPV1	RV	BPV
	72.2%***	81%***	81%***	83.5%***	82.3%***	77.2%***
PCE	JV	MTRV	CMTRV1	JV	BPV	MAC
	75.9%***	74.7%***	74.7%***	81%***	77.2%***	72.2%***
SI	CMJV2	MAC	CMBPV1	MAC	CMRV2	CMRV2
	75.9%***	72.2%***	75.9%***	75.9%***	73.4%***	74.7%***
			rolling window size $= 72$			
HS	CMTRV1	VRV	CMRV1	RV	RV	TRV
	74.7%***	78.5%***	75.9%***	79.7%***	73.4%***	70.9%***
IP	CMTRV2	BPV	CMBPV2	MVRV	MJV	JV
	75.9%	70.9% ***	69.6% ***	78.5% ***	75.9% ***	65.8% *
PAY	VTRV	VRV	MJV	MVTRV	RV	MAC
	79.7%***	77.2%***	77.2%***	77.2%***	73.4%***	79.7%***
CPI	CMJV1	VRV	CMJV2	CMRV2	MRV	MJV
	69.6%***	81%***	79.7%***	78.5%***	81%***	78.5%***
PCE	CMTRV1	CMBPV1	VRV	VRV	MTRV	VRV
	77.2%***	70.9%***	74.7%***	78.5%***	75.9%***	72.2%***
SI	CMJV2	MRV	VBPV	MAC	CMRV2	CMJV1
	75.9%***	68.4%***	79.7%***	73.4%***	73.4%***	70.9%***

*Notes: See notes to Table 2.8. This table is analogous to Table 2.8, except that directional accuracy rates are tabulated, with starred entries denoting rejection of the independence null (see footnote to Table 2.5 for further details).

Table 2.10: Best Models in Ex-ante Directional Accuracy Rate (Sample 2: 2009:1 - 2018:12)

Targets	Forecast horizon					
	1-month	2-month	3-month	4-month	5-month	6-month
			rolling window size = 36			
HS	CMTRV2	RV	CMBPV2	MVJV	TRV	MVRV
	72.1%***	79.1%***	81.4%***	79.1%***	76.7%***	76.7%***
IP	CMTRV1	CMJV2	CMRV1	MVRV	MVRV	MRV
	72.1%***	76.7%***	67.4%*	76.7%***	86%***	65.1%*
PAY	TRV	TRV	TRV	MAC	MVJV	CMJV2
	83.7%***	86%***	72.1%***	72.1%***	81.4%***	81.4%***
CPI	BPV	MAC	BPV	BPV	RV	MAC
	72.1%***	81.4%***	81.4%***	79.1%***	81.4%***	79.1%***
PCE	JV	MTRV	CMTRV1	JV	CMRV2	MAC
	81.4%***	72.1%***	72.1%***	79.1%***	79.1%***	72.1%***
SI	VJV	JV	CMRV2	MAC	MVRV	CMRV1
	79.1%***	74.4%***	74.4%***	79.1%***	76.7%***	72.1%***
			rolling window size = 72			
HS	CMRV1	VRV	CMRV1	JV	RV	MAC
	76.7%***	76.7%***	81.4%***	79.1%***	67.4%***	69.8%***
IP	CMTRV2	MAC	RV	MRV	JV	JV
	74.4%	74.4%***	67.4%***	76.7%***	81.4%***	67.4%
PAY	CMBPV2	CMJV1	CMRV1	CMJV2	MAC	RV
	86%***	86%***	81.4%***	74.4%***	79.1%***	83.7%***
CPI	VRV	CMJV2	CMJV2	MAC	JV	MJV
	74.4%***	81.4%***	83.7%***	74.4%***	81.4%***	81.4%***
PCE	CMTRV1	MAC	VRV	MAC	MRV	MAC
	81.4%***	69.8%***	69.8%***	79.1%***	76.7%***	72.1%***
SI	CMTRV1	MRV	VBPV	MAC	CMRV2	VRV
	76.7%***	74.4%***	79.1%***	79.1%***	74.4%***	72.1%***

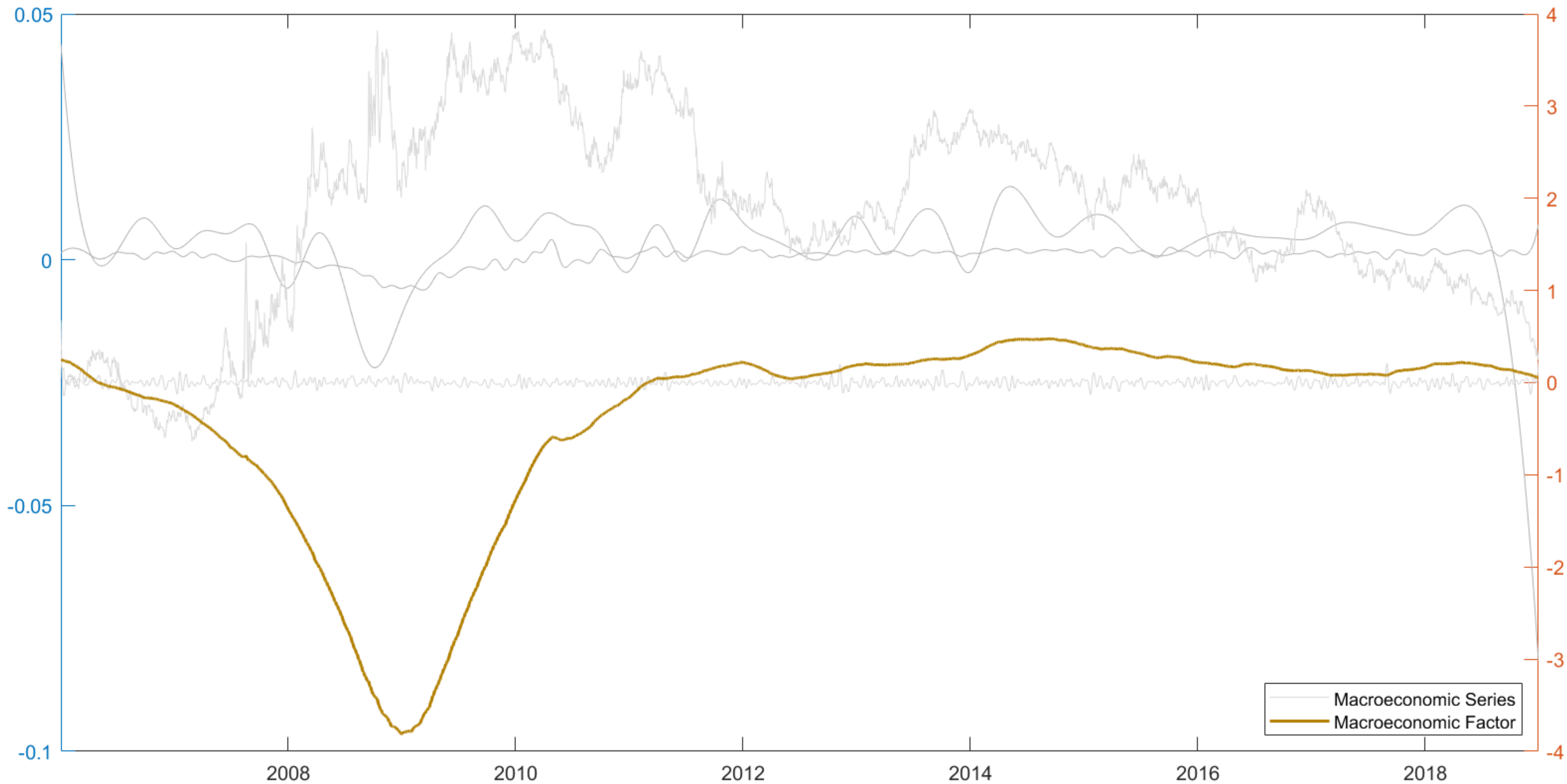
*Notes: See notes to Table 2.8. Results are analogous to those depicted in Table 2.8, except that Sample 2 is used instead of Sample 1 in all prediction experiments.

Table 2.11: Ex-Ante Relative MSFEs for Corporate Bond Yields

Model	Forecast horizon					
	1-month	2-month	3-month	4-month	5-month	6-month
	Sample 1: 2006:1 - 2018:12					
AR	1.000	1.000	1.000	1.000	1.000	1.000
AAA	1.029	1.028	0.873***	0.896***	0.999***	1.084***
	1.031	1.024	0.873***	0.897***	0.993***	1.085***
	1.03	1.029	0.876***	0.895***	1***	1.087***
	1.921***	1.922***	1.706***	1.533***	1.429***	1.601***
AA	1.039	1.001***	0.916***	0.962***	1.03	1.079
	1.039	1***	0.921***	0.968***	1.036	1.086
	1.039	1***	0.916***	0.959***	1.026	1.077
	1.796***	1.817***	1.922***	2.094***	1.892***	2.09***
A	1.08***	1.007*	0.935***	1.007*	1.054	1.093
	1.081***	1.006*	0.938***	1.01*	1.06	1.098
	1.082***	1.008*	0.914***	1.004**	1.051	1.102
	1.902***	1.877***	2.045***	2.26***	2.301***	2.217***
BBB	1.136***	1.053	0.966***	0.951***	1.001***	1.052
	1.138***	1.055	0.974***	0.959***	1.01*	1.062
	1.137***	1.077***	0.958***	0.943***	0.982***	1.044
	3.204***	2.601***	2.156***	2.441***	3.105***	3.087***
BB	1.584***	1.237***	1.051	1.102	1.154*	1.137*
	1.594***	1.244***	1.083	1.119	1.165**	1.137*
	1.602***	1.238***	1.046	1.096	1.148	1.141*
	3.016***	2.633***	2.647***	3.26***	3.587***	3.16***
B	1.794***	1.191***	0.983***	0.967***	1.064	1.17***
	1.816***	1.188***	0.991***	0.978***	0.985***	1.18***
	1.805***	1.197***	0.982***	0.955***	1.056	1.168***
	4.167***	2.991***	2.461***	3.115***	3.966***	3.654***
CCC or below	1.569***	1.327***	1.059	0.909***	0.783***	0.89***
	1.591***	1.331***	0.973***	0.901***	0.794***	0.879***
	1.568***	1.336***	1.041	0.917***	0.776***	0.861***
	3.939***	4.188***	4.223***	2.619***	2.781***	3.27***
	Sample 2: 2009:1 - 2018:12					
AR	1.000	1.000	1.000	1.000	1.000	1.000
AAA	1.128	1.097	0.94***	1.016	1.049	1.065
	1.132	1.09	0.942***	1.017	1.028	1.057
	1.125	1.1	0.945***	1.019	1.058	1.076
	2.264***	2.55***	2.235***	2.013***	2.027***	2.164***
AA	1.001***	0.954***	0.861***	0.864***	0.943***	0.941***
	1***	0.952***	0.864***	0.868***	0.945***	0.946***
	0.999***	0.954***	0.857***	0.858***	0.939***	0.938***
	1.967***	1.972***	2.006***	2.194***	1.738***	1.988***
A	1.056	0.937***	0.867***	0.875***	0.897***	0.93***
	1.053	0.933***	0.868***	0.875***	0.897***	0.929***
	1.057	0.939***	0.827***	0.874***	0.893***	0.948***
	2.325***	2.018***	2.226***	2.153***	2.232***	2.023***
BBB	1.154***	1.031	0.981***	0.831***	0.891***	0.957***
	1.156***	1.033	0.99***	0.838***	0.899***	0.967***
	1.149***	1.059	0.975***	0.821***	0.859***	0.942***
	5.195***	3.54***	2.623***	2.685***	3.748***	3.86***
BB	1.217***	1.005*	0.93***	0.94***	0.877***	0.811***
	1.215***	1.004*	0.971***	0.961***	0.888***	0.812***
	1.227***	1.006	0.923***	0.927***	0.87***	0.814***
	2.943***	2.236***	2.518***	3.642***	3.968***	2.558***
B	1.519***	0.989***	0.88***	0.765***	0.874***	0.87***
	1.53***	0.983***	0.882***	0.775***	0.747***	0.876***
	1.526***	0.997***	0.878***	0.751***	0.865***	0.872***
	4.508***	2.832***	2.324***	3.361***	4.406***	3.452***
CCC or below	1.278***	1.127***	0.997***	0.861***	0.649***	0.672***
	1.29***	1.12*	0.888***	0.848***	0.659***	0.654***
	1.294***	1.139***	0.977***	0.871***	0.644***	0.642***
	3.667***	3.555***	4.292***	2.561***	2.617***	2.916***

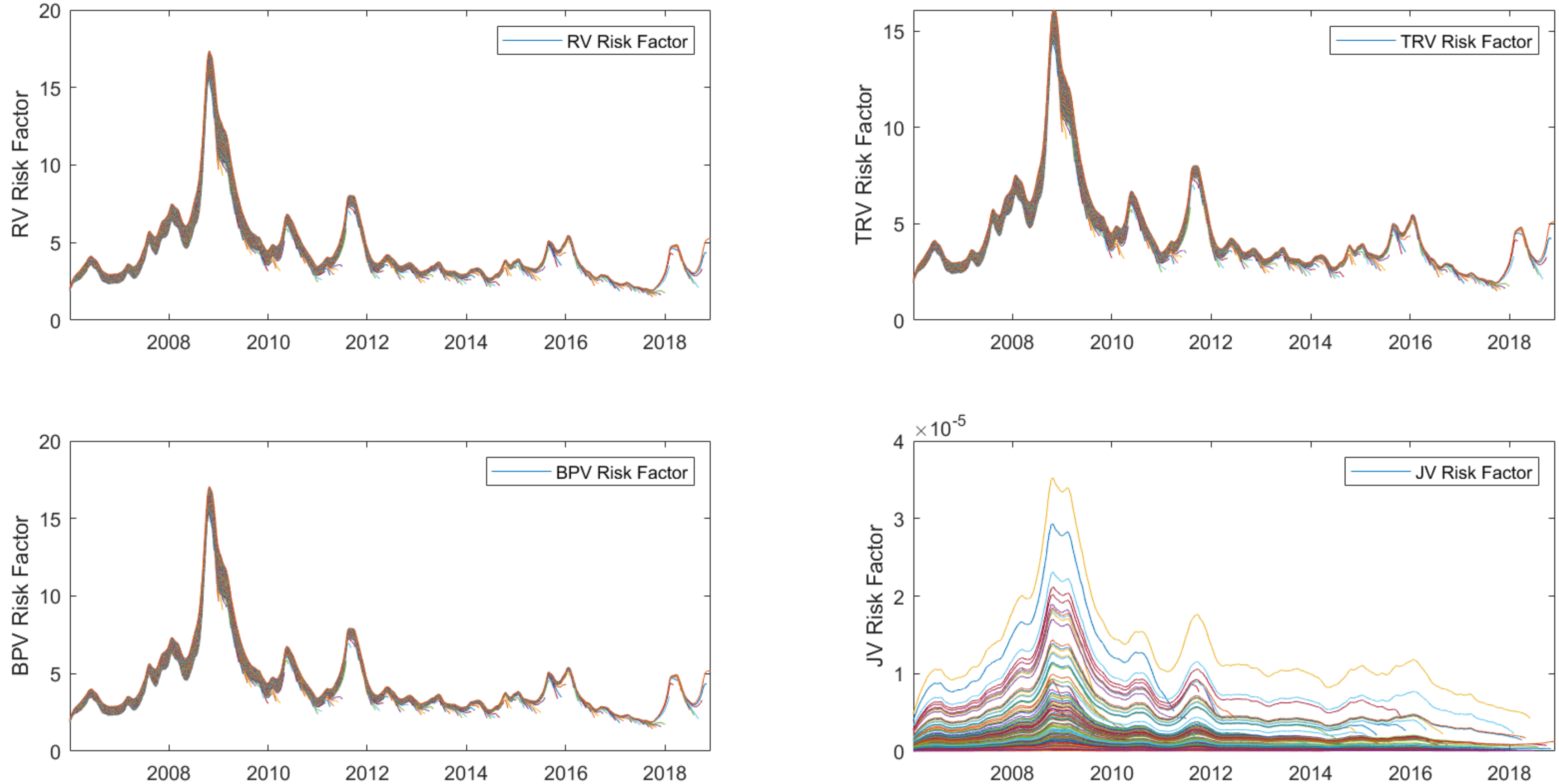
*Notes: See notes to Table 2.3. This table reports results for a select set of models that include those that are "MSFE-best", relative to the AR benchmark, for the corporate bond yield target variables examined in our prediction experiments. See Section 3 for a discussion of these variables, and Section 4 for a summary of these empirical results. Entries in the table are in blocks of 4 rows for each variable. The four rows contain MSFE for the following models, in this order: RV, TRV, BPV, and JV, as depicted in Table 2.2.

Figure 2.1: Macroeconomic Factor and Macroeconomic Series



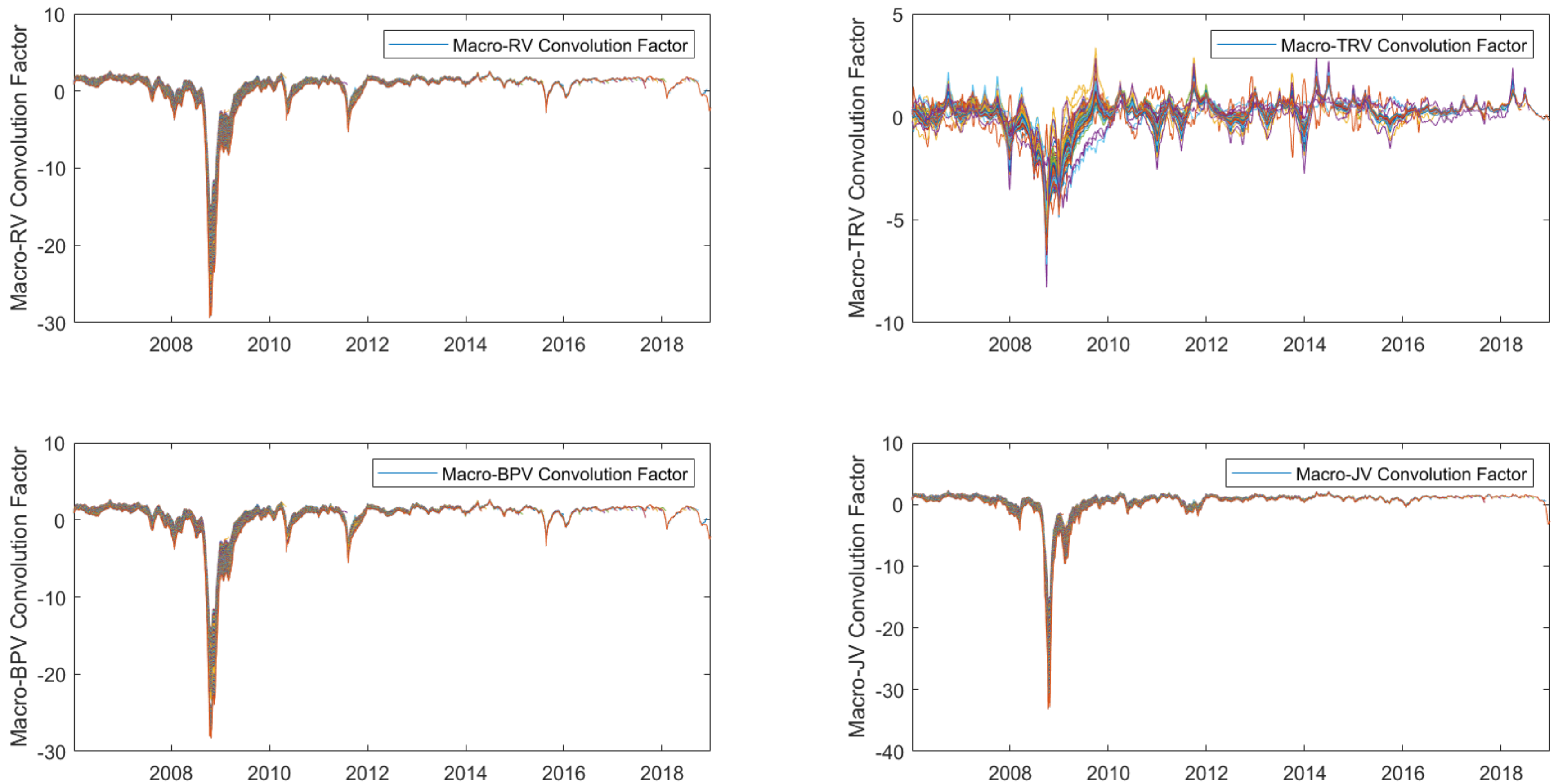
*Notes: In this figure, in order to plot all series in a readily interpretable manner, the ICSA variable is multiplied by minus one. See Section 4 for a description of the 4 macroeconomic variables plotted in gray in this figure, and see Section 3 for a discussion of the methodology used to construct the factor.

Figure 2.2: Volatility Factors



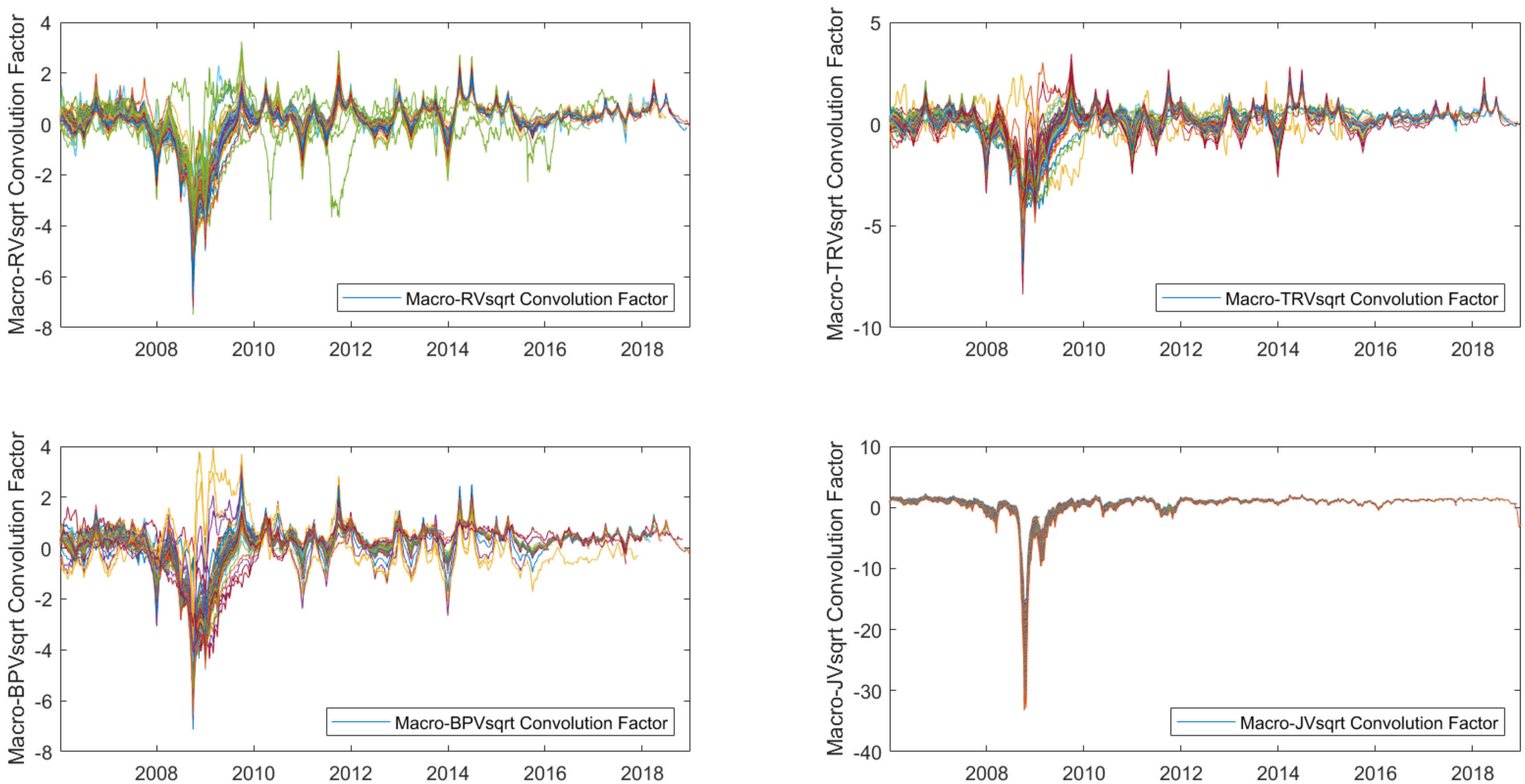
*Notes: In this figure, in order to plot all series in a readily interpretable manner, the TRV and BPV factors are multiplied by minus one. See Section 3 for a discussion of the methodology used to construct the factors in this figure.

Figure 2.3: Volatility-macroeconomic Convolution Factors



*Notes: See Section 3 for a discussion of the methodology used to construct the factors in this figure.

Figure 2.4: Volatility-macroeconomic Convolution Factors



*Notes: See Section 3 for a discussion of the methodology used to construct the factors in this figure.

Chapter 3

Forecasting Sector Level Equity Returns Using Big Data Factors and Machine Learning Models

Note: This chapter is coauthored with Weijia Peng.

3.1 Introduction

The equity risk premium is one of the most widely studied topics in finance, and is crucial to both the understanding of the financial market and portfolio management. A small group of early key works in this area include Fama and French (1992), Fama and French (2015), Welch and Goyal (2007) and Rapach and Zhou (2013), who identify characteristics that have correlation with the equity returns and develop time series models useful for forecasting equity risk premia.

Since the advent of the "big data" era, research into this field of empirical finance has grown ever more rapidly, and numerous researchers have developed and championed the use of ever more sophisticated models for understanding the equity risk premium. In this paper, we add to this nascent literature by examining the marginal predictive content of a large number of machine learning methods for daily and monthly market and sector level equity returns. The novel feature of the modeling approach that we take in this paper is that we not only utilize multi-frequency and multi-dimensional datasets, but we also create a group of latent economic factors including market correlation indices, volatility risk measures, and macro risk factors. This paper, thus, adds to the literature on equity returns forecasting in two ways. First, we build on the work of Aruoba et al. (2009b), Bloom (2009), Jurado et al. (2015b), Aït-Sahalia and Xiu (2016) and others by introducing a class of multi-frequency macroeconomic/financial volatility risk factors and market correlation risk indices that are aimed at measuring market uncertainty. Our state space models are specified in one of two ways, referring to Yao (2019). First, volatility risk factors are specified and estimated using a state space model that includes latent components of quadratic variation, including realized variance (RV_t), truncated realized variance (TRV_t), bi-power variation (BPV_t), and jump variation (JV_t); and also mixed frequency macroeconomic indicators.[1] Alternatively, macroeconomic risk factors are specified and estimated using a state space model that only includes mixed frequency macroeconomic indicators such as interest rates, employment, and production. Finally, we also construct and evaluate market correlation risk indices (or "correlation risk" factors), which are based on estimates of quadratic covariation constructed using high frequency market and sector level returns data. The construction of these indices follows Aït-Sahalia and Xiu (2016), who decompose the quadratic covariation between two assets into continuous and jumps components using high-frequency asset price data, and construct the continuous correlation indices by measuring the correlation between continuous returns, and jump correlation indices using the correlation between jump returns. [2]

Second, we utilize a large number of potentially interesting machine learning methods to allow for a rich variety of model specifications, when forecasting returns. We thus build on previous literature that discusses the difficulties in predicting equity returns, particularly at higher frequencies, such as daily returns (see e.g. Christoffersen and Diebold (2006)). In general, a large machine learning related literature has developed in recent years in the field of financial econometrics. For instance, Hutchinson et al. (1994) develop a nonparametric method for estimating the pricing formula of a derivative asset using neural network models. Rapach et al. (2013) applies adaptive elastic net estimation to predict monthly stock returns in industrialized countries. Other related work includes, but not limit to, Harvey and Liu (2018), Kim and Swanson (2016) and Swanson and Xiong (2017). More recently, in an interesting paper, Gu et al. (2018) conduct a comprehensive study using machine learning methods to predict individual stock risk premia and construct investment portfolios.

---

[1] All of our measures of integrated volatility are extracted from high frequency S&P500 data.

[2] Related papers that utilize mixed-frequency state space models include Mariano and Murasawa (2003), Frale et al. (2008), Aruoba et al. (2009b) and Marcellino et al. (2016). None of these papers, however, include multiple frequencies of the same latent variable, as is done in this paper. Additionally, see Ghysels et al. (2007) for an introduction to the alternative approach of using MIDAS for mixed frequency modeling.

More specifically, we evaluate machine learning methods including random forest, gradient boosting, support vector machine, penalized regressions and neural network (deep learning). Additionally, we evaluate machine learning classifier models including latent discriminant analysis, naive Bayes, support vector classifier, k-nearest-neighbors, random forest, and deep learning. Finally, we propose a group of hybrid machine learning models based on a two-step method that combines the least absolute shrinkage operator (lasso) and neural network methods. As discussed above, our objective is to forecast returns and indicators of directional change. Specifically, we predict both level and directional changes of daily and monthly returns for a variety of target variables, including the S&P500 (SPY) and four SPDR sector ETFs: financials (XLF), technology (XLK), health care (XLV), and consumer discretionary (XLY).[3] The predictors that we use in our analysis include both a small set of mixed frequency variables (for use in our state space models) as well as a variety of other predictors that have been examined previously by Neely et al. (2014), Fama and French (2015) and Welch and Goyal (2007).

Our experimental findings are based on the construction of 1-day and 1-month ahead predictions, formed using rolling and recursive estimation window strategies, for the sample period from 2009 - 2017. Our one-month-ahead forecasts are calculated by aggregating daily forecasts for each month. We also construct two types of directional forecasts. The first type is derived from our returns forecasts, in the sense that returns forecasts are classified as "upward signals" if forecasts are positive, and are otherwise classified as "downward signals". The second type is constructed by utilizing machine learning classifiers to directly generate directional predictions. Our main findings are summarized as follows.

First, based on mean square forecasting error (MSFE) and directional predictive accuracy rates (DPAR), machine learning models yield forecasts that are significantly superior to the random walk and linear regression benchmark forecasts, when predicting monthly

[3]The SPY is the largest exchange-traded fund in the world which is designed to track the S&P 500 stock market index. The XLF, XLK, XLV, and XLY are designed to represent the financial sector, technology sector, healthcare sector, and consumer discretionary sector of the S&P 500 index. The four selected sectors are the largest four S&P 500 sectors, by market cap, as of April 2019, according to Fidelity Research.

returns. Not surprisingly, though, daily results indicate little to choose between our alternative models. Indeed, it is only when we aggregate daily predictions to form monthly predictions, that machine learning methods dominate, for all target assets (i.e. different sector returns), regardless of estimation window strategy.

Second, the random forest method is clearly the preferred machine learning approach. These results are statistically significant (when forecasting monthly returns), and prevail for all of our target variables and estimation strategies. Moreover, these results continue to hold regardless of the set of predictor variables utilized in our different models (we evaluate predictor sets both with and without the latent uncertainty factors discussed above).

Third, "deep" learning models outperform "shallow" learning models. For instance, deep learning models with two to four hidden layers have statistically smaller MSFEs and higher DPARs than shallow learning models with only one hidden layer. Again, this result holds across all windowing strategies used to estimate our models, and for all targets and predictor sets.

Fourth, hybrid machine learning models, which combine lasso and neural network models, often outperform individual models based on both the MSFE and the DPAR. For example, these models usually yield smaller MSFEs and higher DPARs than models based on solely the lasso or neural networks.

Fifth, all three novel risk factors, including market correlation indices, volatility risk factors, and macro risk factors, are shown to contain significant marginal predictive content. In particular, "MSFE-best" and "DPAR-best" forecasting models yield significantly smaller MSFEs and higher DPARs than models without risk factors. Moreover, for the majority of our machine learning models that are not "MSFE-best" or "DPAR-best" best, these three types of factors also prove to be useful.

Sixth, the market as well as all of the sectors that we analyze have different levels of sensitivity to input information, in the form of the predictor set used when constructing the "MSFE-best" or "DPAR-best" model. For the S&P500, a broad range of predictor sets, including sets consisting of (i) all variables, (ii) all variables except macro variables, and (iii) all variables except one of our uncertainty factors have marginal predictive content, when used as machine learning inputs. The exception is our set of "technical variables", which includes trading volume and price trend indicators. When these variables are excluded from the set of predictor variables, MSFEs and DPARs generally improve, for all of our target variables. This may be because useful predictive information contained in our technical variables is also included in our latent uncertainty factors. However, even if this is the case, it is clear from our findings that our latent uncertainty factors have further information embedded in them that is also useful for predicting returns.

Finally, it is worth noting that our correlation indices based on jump variation surges during 2008 and 2011, and drops when market volatility is stable, while our correlation indices based on continuous variation moves in the opposite direction, except for the energy sector. This suggests that the rise in correlation across markets and sectors is largely driven by co-jump behavior. Moreover, When exiting the financial crisis period around 2008, volatility risk factors generally move together across all of the sectors that we analyze in our experiments, including the market (i.e., the S&P500).

The rest of this paper is organized as follows. Section 3.2 summarizes our setup, including a discussion of the latent uncertainty factors that we examine. Section 3.3 discusses our experiment setup and briefly outlines all of the machine learning methods used in the sequel. Finally, Section 3.4 contains a description of the data used in our experiments, and summarizes our empirical findings, and Section 3.5 concludes.

3.2 Market Correlation Indices, Volatility Risk Factors, and Macroeconomic Risk Factors

In this section, we outline the methodology used in the construction of risk factors[4] and market correlation indices analyzed in the sequel. We first introduce the measurements of

[4]We use the notation and setup as in Yao (2019)

high-frequency volatility and continuous and jump volatility parts, and layout the construction of correlation indices. We then turn to the state space framework used to estimate our volatility risk and macroeconomic risk factor, and address temporal aggregation and missing observations problems while working with mixed-frequency series.

3.2.1 High frequency measures of volatility and jump risk

Let X_t be the log-price of an asset at time t. Assume that the log-price process follows a jump-diffusion model (hence, almost surely, its paths are right continuous with left limits). Namely,

$$X_t = X_0 + \int_0^t b_s ds + \int_0^t \sigma_s dB_s + \sum_{s \le t} \Delta X_s. \tag{3.1}$$

In the above expression, B is a standard Brownian motion and $\Delta X_s := X_s - X_{s-}$, where $X_{s-} := \lim_{u \uparrow s} X_u$, represents the possible jump of the process X, at time s.

Consider a finite time horizon, $[0, t]$ that contains n high-frequency observations of the log-price process. A typical time horizon is one day. Let $\Delta_n = t/n$ be the sampling frequency. Then intra-daily returns can be expressed as $r_{i,n} = X_{i\Delta_n} - X_{(i-1)\Delta_n}$.

A well-established result in the high frequency econometrics literature is that realized volatility is a consistent estimator of the total quadratic variation. Namely,

$$RV_t = \sum_{i=1}^{n} r_{i,n}^2 \xrightarrow{\text{u.c.p.}} \int_0^t \sigma_s^2 ds + \sum_{s \le T} (\Delta X_s)^2 = QV_t = IV_t + JV_t, \tag{3.2}$$

where $\xrightarrow{\text{u.c.p.}}$ denotes convergence in probability, uniformly in time. There are many estimators of integrated volatility (IV_t), which is the variation due to the continuous component of quadratic variation (QV_t). For example, multipower variations are defined as follows:

$$V_t = \sum_{i=j+1}^{n} |r_{i,n}|^{r_1} |r_{i-1,n}|^{r_2} ... |r_{i-j,n}|^{r_j}, \tag{3.3}$$

where $r_1, r_2, ..., r_j$ are positive, such that $\sum_{i=1}^{j} r_i = k$, say. An important special case of this estimator is bipower variation (BPV_t), which was introduced by Barndorff-Nielsen and

Shephard (2004). Namely,

$$BPV_t = (\mu_1)^{-2} \sum_{i=2}^{n} |r_{i,n}||r_{i-1,n}| \tag{3.4}$$

where $\mu_1 = E(|Z|) = 2^{1/2}\Gamma(1)/\Gamma(1/2) = \sqrt{2/\pi}$, with Z a standard normal random variable, and where $\Gamma(\cdot)$ denotes the gamma function. Another useful estimator is truncated bipower variation ($TBPV_t$), which combines the truncation method proposed by Mancini (2009) and the bipower variation (BPV_t) estimator discussed above. Namely,

$$TBPV_t = (\mu_1)^{-2} \sum_{i=2}^{n} |\overline{r}_{i,n}||\overline{r}_{i-1,n}|, \quad \overline{r}_{i,n} = r_{i,n}1_{\{|r_{i,n}|<\alpha_n\}}, \tag{3.5}$$

where $\alpha_n = \alpha\Delta_n^{\varpi}, \varpi \in (0, \frac{1}{2})$. Similarly, truncated realized variance (TRV_t) is defined as

$$TRV_t = \sum_{i=1}^{n} \overline{r}_{i,n}^2. \tag{3.6}$$

Finally, jump variation (JV_t) can be estimated as $JV_t = RV_t - BPV_t$ or $JV_t = RV_t - TBPV_t$, for example. In the sequel, we shall utilize RV_t, TRV_t, BPV_t and $JV_t = RV_t - BPV_t$.

Under certain regularity conditions [5], BPV_t, $TBPV_t$ and TRV_t are consistent estimators of unobserved integrated volatility $\mathrm{IV}_t := \int_0^t \sigma_s^2 ds$, and JV_t is the consistent estimator of jump volatility. Moreover, it is also well-established that these estimators converge stably in law at the rate $\sqrt{1/\Delta_n}$, or equivalently, $\sqrt{n}$. Let T be the total number of such representative finite time horizon $[0, t]$ (e.g., day, week, month or quarter). If $\Delta_n T \to 0$, then the impact of estimating the latent volatility and jump risk factors are asymptotically negligible, since the parameters in our state space model converge at rate $\sqrt{T}$.

3.2.2 Market correlation indices

In the high frequency literature, previous research focusing on covariation under multivariate settings has focused mainly on solving three challenges: i) High frequency data tends to be

[5] See papers cited above and Jacod and Protter (2011) and Aït-Sahalia and Jacod (2014) for details about regularity conditions.

severely contaminated with the microstructure noise; ii) non-synchronous high frequency data lead to estimation bias when constructing covariation measures; and iii) covariation matrices must be positive semi-definite in order to guarantee the existence of stable inverses thereof. Notably, Aït-Sahalia and Xiu (2016) address these issues, and develop estimators to decompose quadratic covariation between two assets into continuous and jump components.

Following Aït-Sahalia and Xiu (2016), the quadratic covariation between X_i and X_j is equal to the sum of continuous component quadratic covariation and jump component covariation:

$$[X_i, X_j]_t = [X_i, X_j]_t^c + [X_i, X_j]_t^d, \tag{3.7}$$

where $[X_i, X_j]_t$ is the quadratic covariation between X_i and X_j, $[X_i, X_j]_t^c$ is the continuous quadratic covariation component, and $[X_i, X_j]_t^d$ is the quadratic covariation associated with the discontinuous (jump) component of a process. Andersen et al. (2003) propose realized measures of quadratic covariation, named realized covariance, that are based on the sum of the product of intra-day returns between two assets:

$$cov_{i,j}(t; n) = \sum_{k=1}^{n} r_{i,k,t} \times r_{j,k,t} \tag{3.8}$$

where $cov_{i,j}(t; n)$ denotes the realized covariance between asset i and asset j, at day t. Here, $r_{i,k,t}$ is the intra-daily return of asset i at time interval k, during day t. Realized covariance is an error free estimator of quadratic covariation $[X_i, X_j]_t$, when the length of each intra-daily interval approaches 0 (i.e. the number of intra-daily intervals $n \to \infty$). Namely,

$$\lim_{n \to \infty} cov_{i,j}(t; n) = [X_i, X_j]_t \tag{3.9}$$

In Aït-Sahalia and Xiu (2016), the correlation, $\rho_{i,j}^c$, between asset X_i and asset X_j, which is derived from the continuous component is:

$$\rho_{i,j}^c = \frac{[X_i, X_j]^c}{\sqrt{[X_i, X_i]}\sqrt{[X_j, X_j]}} \tag{3.10}$$

The correlation, $\rho_{i,j}^d$, between asset X_i and asset X_j, which is derived from the jump component is:

$$\rho_{i,j}^d = \frac{[X_i, X_j]^d}{\sqrt{[X_i, X_i]}\sqrt{[X_j, X_j]}} \tag{3.11}$$

where $[X_i, X_i]$ and $[X_j, X_j]$ denote the quadratic variations of assets X_i and X_j, respectively. The quadratic variation in the above formulae is estimated using realized volatility, as in equation (2). The jump component of quadratic covariation, $[X_i, X_j]^d$, is equal to $\sum_{s \leq t} \Delta X_{i,s} \Delta X_{j,s}^T$, where $\Delta X_{i,s}$ represents the jump in X_i, at time s.

In our empirical experiments, we estimate $[X_i, X_j]^d$ using a "jump-test" approach. Namely, the realized covariance associated with jumps is:

$$cov_{i,j}(t; n)^d = \sum_{k=1}^{n} (r_{i,k,t} * I_{jump,i,k,t}) \times (r_{j,k,t} * I_{jump,j,k,t}). \tag{3.12}$$

The jump indicators, $I_{jump,i,k,t}$ and $I_{jump,j,k,t}$, are identified using the Lee and Mykland (2007) jump test. More specifically, Lee and Mykland (2007) use the ratio of realized returns to estimated instantaneous volatility, and construct a nonparametric jump test to identify the exaxt timing of jumps at the intra-day level. The test statisic which identifies whether there is a jump during the interval $(t + l/n, t + (l + 1)/n)$ is:

$$L_{(t+(l+1)/n)} = \frac{X_{t+(l+1)/n} - X_{t+l/n}}{\sigma_{\widehat{t+(l+1)/n}}}, \tag{3.13}$$

where

$$\sigma_{\widehat{t+(l+1)/n}}^2 \equiv \frac{1}{K-2} \sum_{i=l-K+1}^{l-2} |X_{t+(i+1)/n} - X_{t+i/n} || X_{t+i/n} - X_{t+(i-1)/n}|. \tag{3.14}$$

Here K is the window size of a local movement of the process. We choose K = 10 and use the 5-minute sampling frequency (i.e., the number of intra-day observations, n, equals 78). These authors show that

$$\frac{max_{l \in \bar{A}_n} |L_{(t+(l+1)/n)}| - C_n}{S_n} \to \varepsilon, \quad as \;\; \Delta t \to 0, \tag{3.15}$$

where ε has a cumulative distribution function $P(\varepsilon \leq x) = exp(-e^{-x})$,

$$C_n = \frac{(2logn)^{1/2}}{c} - \frac{log\pi + log(logn)}{2c(2logn)^{1/2}} \quad and \quad S_n = \frac{1}{c(2logn)^{1/2}} \tag{3.16}$$

c ≈ 0.7979 and $\bar{A}_n$ is the set of $l \in \{0, 1, ..., n\}$, so that there are no jumps in $(t + l/n, t + (l+1)/n]$. We choose a 10% significance level when applying this test. If the test statistic, $L_{(t+(l+1)/n)}$, lies in the critical region of the null distribution at 10% significance level, then we reject the null hypothesis that there is no jump during $(t + l/n, t + (l+1)/n]$, and the jump indicator, I_{jump}, is set equal to 1. Otherwise, the jump indicator is set equal to 0.

Finally, the continuous component of quadratic covariation is estimated as the difference between realized covariance and discontinuous (jump) realized covariance. Namely,

$$cov_{i,j}(t;n)^c = cov_{i,j}(t;n) - cov_{i,j}(t;n)^d \tag{3.17}$$

3.2.3 Volatility risk factors

Using the state space model setup in Yao (2019), the variable $y_t = (y_t^1, y_t^2, y_t^3, y_t^4)$ corresponding to data measured at 4 different time horizons, including daily (denoted by d), bi-daily (denoted by $2d$), tri-daily (denoted by $3d$), and weekly (denoted by w). In our setup, y_t is alternately set equal to TRV_t. The latent risk factor that we are interested in extracting is called MF_t^{vol}. Finally, the elements of y_t, which are aggregated, are flow variables. Therefore, we include three aggregated state variables, i.e., C_t^1, C_t^2 and C_t^3, to address the aggregation issues discussed above. The state space model is:

Observation Equation:

$$\begin{pmatrix} y_t^d \\ y_t^{2d} \\ y_t^{3d} \\ y_t^w \end{pmatrix} = \begin{pmatrix} \beta_1 & 0 & 0 & 0 & 1 & 0 & 0 & 0 \\ 0 & \beta_2 & 0 & 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & \beta_3 & 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & \beta_4 & 0 & 0 & 0 & 1 \end{pmatrix} \begin{pmatrix} MF_t^{vol} \\ C_t^1 \\ C_t^2 \\ C_t^3 \\ u_t^1 \\ u_t^2 \\ u_t^3 \\ u_t^4 \end{pmatrix}$$

State Equation:

$$\begin{pmatrix} MF^{vol}_{t+1} \\ C^1_{t+1} \\ C^2_{t+1} \\ C^3_{t+1} \\ u^1_{t+1} \\ u^2_{t+1} \\ u^3_{t+1} \\ u^4_{t+1} \end{pmatrix} = \begin{pmatrix} \rho & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ \rho & \psi^1_{t+1} & 0 & 0 & 0 & 0 & 0 & 0 \\ \rho & 0 & \psi^2_{t+1} & 0 & 0 & 0 & 0 & 0 \\ \rho & 0 & 0 & \psi^3_{t+1} & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & \eta_1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & \eta_2 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & \eta_3 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & \eta_4 \end{pmatrix} \begin{pmatrix} MF^{vol}_{t} \\ C^1_{t} \\ C^2_{t} \\ C^3_{t} \\ u^1_{t} \\ u^2_{t} \\ u^3_{t} \\ u^4_{t} \end{pmatrix}$$

$$+ \begin{pmatrix} 1 & 0 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 1 \end{pmatrix} \begin{pmatrix} e^1_t \\ e^2_t \\ e^3_t \\ e^4_t \\ e^5_t \end{pmatrix},$$

where the error terms $e^i_t \overset{i.i.d}{\sim} N(0, \sigma^2_i)$, with $i = 1, ..., 5$.

As mentioned above, the three aggregated variables in the state vector, C^1_t, C^2_t and C^3_t, are designed to handle bi-daily, tri-daily and weekly updating of our volatility series, respectively. Also, ψ_1, ψ_2 and ψ_3 are binary-valued parameters for the aggregated state variables, and are defined as follows:

$$\psi^1_t = \begin{cases} 0, & \text{if t is an odd number} \\ 1, & \text{otherwise,} \end{cases},$$

for the bi-daily updating series;

$$\psi^2_t = \begin{cases} 0, & \text{if t is the first day of every three days} \\ 1, & \text{otherwise,} \end{cases},$$

for the tri-daily updating series; and

$$\psi^3_t = \begin{cases} 0, & \text{if t is the first day of every week} \\ 1, & \text{otherwise,} \end{cases}$$

for the weekly series.

In the above observation equation, only the highest frequency variable, y_t^d, is directly connected with the factor, MF_t^{vol}, via β_1. The three other volatility variables are connected with MF_t^{vol} via the aggregated state variables (i.e, C_t^1, C_t^2 and C_t^3) and via β_2, β_3 and β_4. Coupled with the setup of the binary-valued parameters (i.e., ψ_1, ψ_2 and ψ_3) in the state equation, this ensures the proper inter-temporal aggregation of the flow variables in the system. and refreshes the quantity at the beginning of each period. Finally, the u_t are stochastic disturbance terms, and are assumed to follow autoregressive processes, as in Aruoba et al. (2009b). In the state equation, the first four state variables are connected with MF_t^{vol} via ρ. Of these four state variables, the last three (i.e., C_t^1, C_t^2 and C_t^3) are defined such that their previous values are added to ρMF_t^{vol} whenever flow aggregation is required.

3.2.4 Macroeconomic risk factors

We again begin with $y_t = (y_t^1, y_t^2, y_t^3, y_t^4)$. In this section, the data are measured at daily (denoted by d), weekly (denoted by w), monthly (denoted by m), and quarterly (denoted by w) frequencies. This allows us to construct a "benchmark" risk factor corresponding to the business conditions index analyzed by Aruoba et al. (2009b). In particular, following Aruoba et al. (2009b), we use four macroeconomic variables with different sampling frequencies, including: (1) the daily yield curve spread (y_t^1), defined to be the difference between the 10-year U.S. Treasury bond yield and the 3-month Treasury bill yield; (2) weekly initial claims for unemployment insurance (y_t^2); (3) nonfarm payroll employment (y_t^3); and (4) quarterly gross domestic product (y_t^4). The corresponding state-space model used to extract our risk factor, called MF_t^{mac} is: *Observation equation:*

$$\begin{pmatrix} y_t^1 \\ y_t^2 \\ y_t^3 \\ y_t^4 \end{pmatrix} = \begin{pmatrix} \beta_1 & 0 & 0 & 1 \\ 0 & \beta_2 & 0 & 0 \\ \beta_3 & 0 & 0 & 0 \\ 0 & 0 & \beta_4 & 0 \end{pmatrix} \begin{pmatrix} MF_t^{mac} \\ C_t^1 \\ C_t^2 \\ u_t^1 \end{pmatrix} + \begin{pmatrix} 0 & 0 & 0 \\ \gamma_2 & 0 & 0 \\ 0 & \gamma_3 & 0 \\ 0 & 0 & \gamma_4 \end{pmatrix} \begin{pmatrix} y_{t-W}^2 \\ y_{t-M}^3 \\ y_{t-Q}^4 \end{pmatrix} + \begin{pmatrix} 0 \\ w_t^2 \\ w_t^3 \\ w_t^4 \end{pmatrix}.$$

State equation:

$$\begin{pmatrix} MF_{t+1}^{mac} \\ C_{t+1}^{1} \\ C_{t+1}^{2} \\ u_{t+1}^{1} \end{pmatrix} = \begin{pmatrix} \rho & 0 & 0 & 0 \\ \rho & \psi_{t+1}^{1} & 0 & 0 \\ \rho & 0 & \psi_{t+1}^{2} & 0 \\ 0 & 0 & 0 & \gamma_1 \end{pmatrix} \begin{pmatrix} MF_{t}^{mac} \\ C_{t}^{1} \\ C_{t}^{2} \\ u_{t}^{1} \end{pmatrix} + \begin{pmatrix} 1 & 0 \\ 1 & 0 \\ 1 & 0 \\ 0 & 1 \end{pmatrix} \begin{pmatrix} e_t^1 \\ e_t^2 \end{pmatrix},$$

where the error terms $e_t^i \overset{i.i.d}{\sim} N(0, \sigma_i^2)$, with $i = 1, 2$.

The variables in this model include observed variables, the y_t; our latent risk factor, MF_t^{mac}; aggregate state variables, C_t^1 and C_t^2; and stochastic disturbance terms, u_t^1, w_t^2, w_t^3, and w_t^4. Note that in this model, only y_t^2 and y_t^4 are flow variables in this model, and hence there are only two aggregate state variables. Accordingly, we also define two binary-valued variables ψ_1 and ψ_2 for these aggregated state variables. Namely,

$$\psi_t^1 = \begin{cases} 0, & \text{if t is the first day of the week} \\ 1, & \text{otherwise,} \end{cases}$$

and

$$\psi_t^2 = \begin{cases} 0, & \text{if t is the first day of the quarter} \\ 1, & \text{otherwise.} \end{cases}$$

3.2.5 Technical indicators

Technical indicators have been widely used by practitioners in asset pricing applications. Two key papers discussing different technical indicators include Fama and Blume (1966) and Brock et al. (1992). These papers explore the usefulness of various technical indicators, including filter rules, moving averages, and momentum, when designing trading strategies. Neely et al. (2014) shows the usefulness of technical indicators for predicting the equity risk premium. These authors analyze 14 common technical indicators.

In our experiments, we use two types of technical indicators, including moving average indicators and volume-based trend indicators, following Neely et al. (2014). The moving average technical indicators are derived by using moving-average (MA) rules to generate long or short signals at the end of each trading day, t. Namely, define:

$$D_t = \begin{cases} 1 & \text{if } P_t^{MA(k)} \geqslant P_t^{MA(s)} \\ 0 & \text{otherwise} \end{cases} \tag{3.18}$$

where

$$P_t^{MA(q)} = \frac{1}{q}\sum_{i=0}^{q-1} P_{t-i} \quad \text{for} \quad q = k, s \tag{3.19}$$

Here $D_t = 1$ represents the long signal and $D_t = 0$ represents the short signal at day t. We use 30-, 90-, and 120-day moving-averages of asset prices, P_t, in our experiments.[6] This allows us to obtain potentially useful price trend indicators. The values for q are thus set equal to $30, 90, 120$, representing monthly, quarterly, and semiannually time periods, and yielding three price trend indicators. Table 3.1A lists the technical indicators in detail.

Our volume-based trend indicators are constructed by combining trading volume and prices in order to identify volume trends in the market. The daily "net "" volume is defined as:

$$V_{net,t} = V_t \times S_t \tag{3.20}$$

where V_t is the trading volume at day, t. The dummy variable, S_t, is:

$$S_t = \begin{cases} 1 & \text{if } P_t \geqslant P_{t-1} \\ -1 & \text{otherwise.} \end{cases} \tag{3.21}$$

We use the "net" volume, $V_{net,t}$, to generate the trading signals D_t, where:

$$D_t = \begin{cases} 1 & \text{if } V_{net,t}^{MA(k)} \geqslant V_{net,t}^{MA(s)} \\ 0 & \text{otherwise} \end{cases} \tag{3.22}$$

with

$$V_{net,t}^{MA(q)} = \frac{1}{q}\sum_{i=0}^{q-1} V_{net,t-i} \quad \text{for} \quad q = k, s \tag{3.23}$$

[6]Here P_t is the asset price, measured at the end of each trading day, t

Here $D_t = 1$ represents the "long" signal and $D_t = 0$ represents the "short" signal, on day t. We utilize 30-, 90-, and 120-day moving averages of the "net" volume in order to construct our volume-based trend indicators. The parameters q is thus set as $30, 90, 120$ to represent monthly, quarterly, and semiannually time periods.

3.3 Experiment Setup

In this section, we introduce our experimental setups and models used to predict asset returns. First, we detail the splitting of sample data into validation, training and test parts. The validation dataset is established to estimate hyperparameters in machine learning models and avoid potential overfitting problem. Second, we show specific setups of forecasting models we use in the experiment. For example, in machine learning models, we discuss the tuning process of hyperparameter, the objective function of the model, and parameter estimation algorithm or solver.

In our empirical experiment, forecast targets are the return of the following financial assets: SPY (SPDR S&P 500 ETF Trust), XLF (Financial Select Sector SPDR Fund), XLK (Technology Select Sector SPDR Fund), XLV (Health Care SPDR), and XLY (Consumer Discretionary SPDR). The SPY is the largest exchange-traded fund in the world which is designed to track the S&P 500 stock market index. The XLF, XLK, XLV and XLY are designed to represent the financial sector, technology sector, healthcare sector, and consumer discretionary sector of the S&P 500 index. We forecast one-day-ahead daily returns and directional changes for each target asset using both rolling and recursive estimation windows. The rolling window size is $T = 500$. We denote daily returns of targeted assets at day t as $r_{i,t}$, where i corresponds to one of the five sectors mentioned earlier. We also calcualte one-month-ahead forecasts by aggregating daily forecasts. Finally, as as mentioned above, we also construct two types of directional forecasts. The first type is derived from our returns forecasts, in the sense that returns forecasts are classified as "upward signals" if forecasts are positive, and are otherwise classified as "downward signals". The second type is constructed by utilizing machine learning classifiers to directly generate directional predictions.

3.3.1 Linear models

We use a random walk model as our main benchmark. Namely, forecasts are constructed using:

$$r_{t+1} = a + \epsilon_{t+1} \tag{3.24}$$

where ϵ_{t+1} is a stochastic disturbance term, and a is constant. In our experiments, a is estimated under both rolling and recursive data windows. For the rolling scheme, the window size $T = 500$. For the recursive scheme, a is constructed using asset returns from $t = 251^{th} - t^*$, where t^* denotes the last trading day prior to the period being forecasted. We also estimate linear models with the following specification:

$$r_{t+1} = c + \alpha' W_t + \epsilon_{t+1}, \tag{3.25}$$

where $r_{i,t+1}$ is the "target" forecast variable of interest (i.e. daily returns for SPY, XLF, XLK, XLV, and XLY), and the forecast horizon is one-day-ahead. W_t contains explanatory variables at time t, and α is a conformably defined coefficient vector. W_t consists one-day-lagged returns, $r_{i,t}$, and exogenous variables including macroeconomic and financial volatility (risk) factors, and market correlation indeices; as well as the macroeconomic and technical indicators outlined in Table 3.1A. For details regarding the variables in W_t, refer to Table 3.2. Models are estimated using least squares.

3.3.2 Penalized linear models

We utilize two varieties of penalized regression - ridge regression and Least absolute shrinkage operator (lasso)type regression.

Ridge regression

Ridge regression is introduced by Hoerl and Kennard (1970). Estiamtion involves solving the following problem:

$$minL(\lambda, \alpha) = \sum_{t=1}^{T} [r_{t+1} - c - \alpha' W_t]^2 + \lambda |\alpha'|^2 + \lambda c^2 \tag{3.26}$$

where $\alpha = (\alpha_1, ..., \alpha_p)$ and $\lambda \geq 0$. Here $|\alpha'|^2 = \sum_{j=1}^{p} \alpha_j^2$. The tuning parameter, λ, controls the amount of shrinkage.

Lasso regression

Lasso regression is introduced by Tibshirani (1996). Estimation involves solving the following problem:

$$minL(\lambda, \alpha) = \sum_{t=1}^{T} [r_{t+1} - c - \alpha' W_t]^2 + \lambda|\alpha'| + \lambda|c| \tag{3.27}$$

where $\alpha = (\alpha_1, ..., \alpha_p)$ and $\lambda \geq 0$. Here $|\alpha'| = \sum_{j=1}^{p} |\alpha_j|$. We optimize the tuning parameter, λ, using the training' sample, as discussed above.

3.3.3 Logistic regression

Logistic regression is used in several areas including, for example, the bioassay, epidemiology, and machine learning fields. In a key paper, Cox (1966) introduces the multinomial logistic regression model. The purpose of these models is to estimate the probability that categorical response variables, say r_{t+1}, belong to a particular category via use of a linear probability model. In particular, probabilities based on logistic regression are calculated using the logistic function:

$$\begin{aligned} P(r_{t+1} = m|W_t) &= \frac{exp(c_m + \alpha'_m W_t)}{1 + \sum_{n=1}^{M-1} exp(c_n + \alpha'_n W_t)}, m = 1, ..., M-1, \\ P(r_{t+1} = M|W_t) &= \frac{1}{1 + \sum_{n=1}^{M-1} exp(c_n + \alpha'_n W_t)}, \end{aligned} \tag{3.28}$$

where $M = 2$ in our directional prediction accuracy experiments. Maximum likelihood is used to estimate $\theta = \{c_1, \alpha'_1\}$; and the likelihood function is

$$l(\theta) = \prod_t P(r_{t+1} = 1|W_t) \prod_t (1 - P(r_{t+1} = 2|W_t)) \tag{3.29}$$

To maximize this likelihood function, we use the *liblinear* algoritm discussed in Fan et al. (2008).

3.3.4 Linear discriminant analysis

Linear discriminant analysis (LDA) was introduced by Fisher (1936). LDA is useful because it is more stable than logistic regression, when the distribution of predictors, $W_t = \{W_{1,t}, ..., W_{p,t}\}$, is approximately normal. The idea is to model the distribution of W_t from each class of response variable, $r_{j,t+1}$, say (in our experiments, $j = 1, ..., 5$ as discussed above), separately and then use Bayes theorem to update $P(r_{j,t+1} = m|W_t)$. Suppose that π_m represents the prior probability of $r_{j,t+1}$ belonging to the class, m, where $\sum_{m=1}^{M} \pi_m = 1$. The probability density function of W_t belonging to class m is $f_m(W_t)$. Bayes' theorem then implies that:

$$P(r_{j,t+1} = m|W_t) = \frac{f_m(W_t)\pi_m}{\sum_{m=1}^{M} f_m(W_t)\pi_m}, \tag{3.30}$$

Linear discriminant analysis models the density function $f_m(W_t)$ as a multivariate Gaussian process. Namely:

$$f_m(W_t) = \frac{1}{(2\pi)^{(p/2)}|\Sigma_m|^{1/2}} e^{-\frac{1}{2}(W_t-\mu_m)^T\Sigma_m^{-1}(W_t-\mu_m)} \tag{3.31}$$

where μ_m is the mean of W_t for the m^{th} class, and Σ_m is the covariance matrix common to all m classes. Finally, it is worth noting that the log-ratio of the conditional probability density function between two classes is:

$$\begin{aligned} log\frac{P(r_{j,t+1} = l|W_t)}{P(r_{k,t+1} = m|W_t)} &= log\frac{f_l(W_t)}{f_m(W_t)} + log\frac{\pi_l}{\pi_m} \\ &= log\frac{\pi_l}{\pi_m} - \frac{1}{2}(\mu_l + \mu_m)^T\Sigma^{-1}(\mu_l + \mu_m) \\ &+ W_t^T\Sigma^{-1}(\mu_k - \mu_l), \end{aligned} \tag{3.32}$$

which is linear in W_t.

3.3.5 Naive Bayes classifier

The Naive Bayes classifier was first introduced in the pattern recognition field by Duda (1973). More recent machine learning papers in the area include Langley (1993) and Friedman et al. (1997). The naive Bayes model assumes input variables $W_t = \{W_{t,1}, ..., W_{t,p}\}$

are independent in each class of the response variable $r_{j,t+1} = 1, ..., M$. Namely:

$$f_m(W_t) = \prod_{k=1}^{p} f_{mk}(W_{t,k}), \tag{3.33}$$

where $f_m(W_t)$ is the probability density function of W_t in class $r_{j,t+1} = m$. As the LDA model, the log-ratio of the conditional probability density function between two classes is:

$$log\frac{P(r_{t+1}=l|W_t)}{P(r_{t+1}=m|W_t)} = log\frac{\pi_l f_l(W_t)}{\pi_m f_m(W_t)} = log\frac{\pi_l \prod_{k=1}^{p} f_{lk}(W_{t,k})}{\pi_m \prod_{k=1}^{p} f_{mk}(W_{t,k})} = log\frac{\pi_l}{\pi_m} + \sum_{k=1}^{p} log\frac{f_{lk}(W_{t,k})}{f_{mk}(W_{t,k})}. \tag{3.34}$$

3.3.6 Support vector machines

Support vector machines (SVMs) were first proposed by Vapnik and Chervonenkis (1964). A key recent paper in this area is Cortes and Vapnik (1995). A key impetus for this machine learning method is that linearity is a strict assumption, and may yield poor approximations in high-dimensional and high-frequency data environments. This has led to the introduction of various "learning methods", of which SVMs are an example.

SVMs utilize hyperplanes in order to delineate boundaries for the separation of observations into different categories. The idea is to find "optimal" separating boundaries that effectively categorize data and maximize the distance from the closest observations to the boundary. While several other techniques such as the Latent Dirichlet allocation (LDA) also incorporate a similar idea, support vector machine/regression models are interesting because estimation only requires a small percentage of the data (i.e., to construct so-called "support vectors"). Rather than depending on all sample data, as in the case of LDA, optimization hinges on the use of these support vectors, which are easy to construct using big data and are robust to overfitting problems.

Without loss of generality, define a hyperplane for a p-dimensional dataset as:

$$c + \alpha_1 W_{t,1} + \cdots + \alpha_p W_{t,p} = 0,$$

where $W_t = (W_{t,1}, \cdots, W_{t,p})'$, for observations $t = 1, ..., T$. These hyperplanes are used in the following classification rule:

$$If \quad c + \alpha_1 W_{t,1} + \cdots + \alpha_p W_{t,p} > 0, \quad then \quad d_t = 1$$

and

$$If \quad c + \alpha_1 W_{t,1} + \cdots + \alpha_p W_{t,p} < 0, \quad then \quad d_t = -1$$

Optimal separating boundaries are obtained by maximizing the margin around the boundary, say M. Under separability, M denotes a certain "minimal distance" of data points from the boundary. Under nonseparability, a small number of data points may be misclassified, in the sense that they reside on the other side of the boundary. In this setup, ξ_t is defined to be the magnitude of any miscalssification. If there is none, then $\xi_t = 0$. Otherwise, $\xi_t =$ equals the distance from the data points to the hypoerplane, with $\xi_t \geqslant 0$. An additional constraint "$\sum_{i=1}^{N} \xi_t \leqslant constant$" controls the level of missclassification allowed. The objective function used to estimate the parameters in a support vector machine is:

$$\begin{aligned} &\underset{c,\alpha,||\alpha,c||=1}{max} M \\ &s.t. \quad d_t(c + \alpha_1 W_{t,1} + \cdots + \alpha_p W_{t,p}) \geq M(1 - \xi_t), \end{aligned} \tag{3.35}$$

where $||\cdot||$ denotes the Euclidean norm. Support vector regression extends the idea of support vector machines into a regression framework. Rather than focusing on the distance of support vectors to hyperplanes, under, support vector regression minimizes the error between fitted and true observational values. For example, using the simplest linear regression where $f(W_t) = c + \alpha_1 W_{t,1} + \cdots + \alpha_p W_{t,p}$, support vector regression incorporates the error term $r_{j,t+1} - f(W_t)$ into its objective function. Following Cortes and Vapnik (1995), the objective function for support vector regression can be written as:

$$\underset{c,\alpha}{min} \sum_{i=1}^{N} V(r_{j,t+1} - f(W_t)) + \lambda|\alpha|^2 + \lambda c^2 \tag{3.36}$$

where λ is a tuning parameter. Here:

$$V_m(\varepsilon) = \begin{cases} 0, & \text{if } |\varepsilon| < m \\ |\varepsilon| - m, & \text{otherwise.} \end{cases} \tag{3.37}$$

Notably, regardless of the specification of $V_m(\varepsilon)$, solutions for the optimal values of c and α are a linear combination of kernel functions, $K(W_t, W_t') = \sum_{n=1}^{p} h_n(W_t)h_n(W_t')$. We utilize three different kernels in our experiments, including:

- linear kernel: $K(W_t, W_t') = \sum_{n=1}^{p} W_{t,n}W_{t,n}$
- polynomial kernel: $K(W_t, W_t') = (1 + \sum_{n=1}^{p} W_{t,n}W_{t,n})^d$
- radial kernel: $K(W_t, W_t') = exp(-\gamma||W_t - W_t'||^2)$

Therefore, the hyperparameters that we estimate with our training dataset include: 1) λ in linear kernel; 2) λ and d in polynomial kernel; and 3)λ and γ in radial kernel.[7]

3.3.7 Random forest methods

The random forest machine learning method was first introduced by Breiman (2001). Like other tree-based statistical learning techniques, it is based on specifying "subsections" of the predictor space, W_t. Tree-based methods use the mean of response of variable $r_{j,t}$ under each subsection to construct forecasts. The procedure used to develop subsections resembles the structure of the tree, and each subsection of data is therefore also called a tree node.

The first step of tree-based methods involves bootstrapping the sample data. Then, within each bootstrapped sample, the tree-based algorithm tries to find the best "split" of the predictor W_t using the criteria of least squares. Hence, the space of predictors and response variable is partitioned into $m = 1, 2, \cdots,$ M regions/tree nodes. More specifically, for each tree node, $R_1(m, c) = \{W_{t,m}|W_{t,m} < c\}$ and $R_2(m, c) = \{W_{t,m}|W_{t,m} \geq c\}$, the parameters m and c are determined by solving the following problem:

$$min \sum_{t:W_t \in R_1(m,c)} (r_{j,t+1} - \hat{r}_{R_1})^2 + \sum_{t:W_t \in R_2(m,c)} (r_{j,t+1} - \hat{r}_{R_2})^2 \tag{3.38}$$

where $\hat{r}_{R1}$ is predicted value of response variable $r_{j,t+1}$ and equals the mean of $r_{j,t+1}$ in the sample data associated with R_1. Here, $\hat{r}_{R2}$ is defined analogously.

---

[7] All machine learning methods utilized in our experiments involve three distinct sample periods, including: (i) a training dataset (used for the estiamtion of hypoerparameters), (ii) a forecasting model estimation period, and (iii) and ex-ante forecasting period.

The major difference between random forest and other tree-based methods, in particular bagging, is the additional constraint requiring the choice of m from only a subset of the p predictors, which is randomly chosen, and usually consists of $\sqrt{p}$ of the original predictors. This design avoids the problem of correlation among fitted trees, when one or a few predictors dominate other predictors. For further discussion, see Friedman et al. (2001).

The algorithm that we utilize in order to carry out random forest regression is:

1. Draw B bootstrap samples from the data.

2. For each bootstrap sample b, where $b = 1, 2, \cdots, B$:

(a) Choose a subset of variables from the p predictors.

(b) Find the optimal variable m and corresponding cutoff value c that yield the lowest sum squared prediction error.

(c) Partition the data at W_t with cutoff value c.

(d) Recursively repeat the above procedures until a minimal tree node size, say n_{min} is reached.

3. Make predictions based on developed trees, called $T_b(W_t)$, using:

$$\hat{r}^B(W_t) = \frac{1}{B}\sum_{b=1}^{B} T_b(W_t).$$

Within each boostrap, observations are independently drawn from the training sample, with replacement. The number of observations in each boostrap is the same as in the traning sample. We use a validation (training) dataset to conduct cross validation and tune the hyperparameter B, for values of $B = \{100, 200, 300, 400\}$.

In a classification setting, the objective loss function is different than in the regression framework above. Classification models often use an alternative approach based on the "classification error rate", which measures the fraction of training observations which are not classified as belonging to the majority class, within a specific region. However, use of the Gini index for model specification is preferable in our context, because classification error rates are not sufficiently sensitive for our tree-based model. The Gini index is denoted as:

$$L = \sum_{n=1}^{N} \hat{p}_{mn}(1 - \hat{p}_{mn}) \tag{3.39}$$

where $\hat{p}_{mn}$ is the proportion of training observations in the m^{th} region that belongs to the n^{th} class (n = 1, 2, $\cdots$, N). Under the binary classification case, N=2.

Hyperparameters that we estimate for our random forests are the maximum depth of the tree, the minimum number of samples required to split an internal node, the minimum number of bootstrap samples required to be at a tree node, the number of predictors to consider when looking for the best split, and the number of trees.

3.3.8 Gradient tree boosting

The gradient boosting method is developed in Friedman (2001) for regression and classification. The objective loss function[8][9] is:

$$L(\hat{f}) = \sum_{i=1}^{N} L(r_{t+1}, f(W_t)) \tag{3.40}$$

Here we use l_2 penalty for the loss function $L(.)$. $f(.)$ is a sum of regression trees:

$$f_M(W_t) = \sum_{m=1}^{M} K(W_t; \kappa) \tag{3.41}$$

Where each $K(W_t; \kappa)$ represents a regression tree and κ is the parameter in the model. One solution to this loss function is to estimate the tree $K(W_t; \kappa)$ at mth iteration to fit the negative gradient:

$$\hat{\kappa}_m = \underset{\kappa}{argmin}(-g_m - K(W_t; \kappa))^2 \tag{3.42}$$

where the components of the gradient g_m are:

$$g_{im} = [\frac{\partial L(r_{t+1}, f(W_t))}{\partial f(W_t)}]_{f(W_t)=f_{m-1}(W_t)}, i = 1, ..., N \tag{3.43}$$

The following summarizes the gradient tree boosting algorithm:

[8] We use the notation in Friedman et al. (2001)

[9] The loss function for gradient boosting classification is same as in the random forest.

1. Start $f_0(W_t) = argmin_\theta \sum_{i=1}^{N} L(r_{t+1}, \theta)$.

2. For $m = 1$ to M:

   (a). For $i = 1, 2, ..., N$ compute:

$$s_{im} = -[\frac{\partial L(r_{t+1}, f(W_t))}{\partial f(W_t)}]_{f=f_{m-1}} \tag{3.44}$$

   (b). Train a regression tree with target s_{im} to get the terminal regions S_{jm}, $j = 1, 2, ...J_m$.

   (c). For $j = 1, 2, ...J_m$ compute:

$$\hat{\theta}_{jm} = \underset{\theta}{argmin} \sum_{W_t \in S_{jm}} L(r_{t+1}.f_{m-1}(W_t) + \theta) \tag{3.45}$$

   (d). Update $f_m(W_t) = f_{m-1}(W_t) + \lambda \sum_{j=1}^{J_m} \theta_{jm} I(W_t \in S_{jm})$

3. Output $\hat{f}(W_t) = f_M(W_t)$

In the hyperparameter estimation, the learning rate λ shrinks the contribution of each tree. M captures the number of booting stages in the estimation. We also tune other variables including min_sample_split[10], min_samples_leaf[11] and max_depth[12] during the cross-validation.

3.3.9 Neural networks

Neural network models build on a set of nonlinear functions mimicking the neural architecture of brains. The earliest neural networks trace back to Rosenblatt (1958) and McCulloch and Pitts (1943), where propositional logic models and probabilistic models are proposed to describe nervous system activity, information storage, and organization in the brain. Since these early papers, neural networks and their applications have been studied extensively

[10]The minimum number of samples required to split an internal node.

[11]The minimum number of samples required to be at a leaf node.

[12]The maximum number of nodes in the tree.

across numerous disciplines. A key paper in this area is Hornik et al. (1989), who prove that multilayer feed-forward networks are "universal approximators", in the sense that as long as the complexity of the network is allowed to grow (i.e., increasing the number of so-called "hidden units") with the sample size, then a network can estimate an arbitrary function, arbitrarily well. Not surprisingly, given this result, recent work shows that networks with multiple hidden layers are often better approximators than models with one hidden layer (see e.g, He et al. (2016)). IN this paper we utilize the traditional "feed-forward" neural network model of the variety discussed in Hornik et al. (1989).

Let W_t denote the "inputs" to the neural network. A hidden layer is defined as:

$$G_m = f(c + \alpha' W_t), \tag{3.46}$$

where $W_t = (W_{t,1}, ..., W_{t,p})$. The nonlinear function, $f(\cdot)$, is called the activation function, and we utilize four such functions in our experiments, with choice amongst them carried out using cross validation. These include:

Identity Function:$f(W_t) = W_t$

Sigmoid Function:$f(W_t) = 1/(1 + exp(-W_t))$

Hyperbolic Tan Function:$f(W_t) = tanh(W_t)$

Rectified Linear Unit Function:$f(W_t) = max(0, W_t)$

The output $r_{j,t+1}$ given as:

$$r_{j,t+1} = g(\theta_{0s} + \theta_s^M G), \quad m = 1, ..., M \tag{3.47}$$

where $G = (G_1, G_2, ..., G_M)$. We incorporate up to four hidden layers in our experiments, and the number of neurons (variables) in each layer is selected according to the geometric pyramid rule (see Masters (1993)).

In our classification variant of this model, we use cross-entropy, L, as the loss function, where:

$$L = -\sum_{m=1}^{M} \hat{p}_{mn} log \hat{p}_{mn}, \tag{3.48}$$

with $\hat{p}_{mn}$ defined to be the proportion of training observations in the m^{th} region arising from the n^{th} class. The output function, $g(\cdot)$, in our classification method is the *Softmax* function[13]. For parameter estimation, we use least squares with quasi-Newton numerical optimization, and the parameter α is tuned during cross-validation.

3.3.10 K-nearest-neighbor classifiers

The nearest neighbor classification method was first proposed by Cover and Hart (1967) in the field of pattern recognition. This nonparametric method performs clustering based on minimum distance measures. Given an unclassified point, W_0, and k points, W_t, $t = 1, ..., N$, in a training dataset are selected based on the closest distance to W_0, and then the point W_0 is classified. Distance, d_l is measured using the standard Euclidean norm:

$$d_l = \|W_t - W_0\| \tag{3.49}$$

The number of neighbors, k, is dependent on a tuning parameter which is calibrated using cross-validation.

3.3.11 Hybrid machine learning methods

We also explore the usefulness of a hybrid class of models that combines the lasso with neural networks. These hybrid models are based on a two step specification method. In the first step, the lasso is utilized to predict the forecasting target using the model:

$$r_{j,t+1} = c + \alpha' W_t + \epsilon_{t+1} \tag{3.50}$$

with specification achieved by minimizing the following function:

$$L(\lambda, \alpha) = |r_{t+1} - c - \alpha' W_t|^2 + \lambda |\alpha'|^2. \tag{3.51}$$

In the second step, the residual, ϵ_{t+1}, estimated using this minimizer is deployed as our forecasting target, and neural networks are estimated. In this step, we carry out two types

[13] See Bridle (1990).

of experiments, based on the use different input variables, W_t. In the first type, W_t is the same as that used in the lasso. In the second type, Z_t is instead used as the input into the networks, where Z_t is the subset of W_t obtained by utilizing the lasso as a variable selection device.

3.3.12 Experimental setup and forecast evaluation

All forecasting models are estimated using three difference rolling window sizes, and all models and parameters are re-estimated/re-specified prior to the construction of each new daily forecast.[14] Additionally, and as discussed above, monthly forecasts are formed by aggregating daily forecasts. Forecasting performance is evaluated using mean squarefore-cast error (MSFE), where $MSFE = \frac{1}{T}\sum_{t=1}^{T}(r_{j,t} - \hat{r}_{j,t})^2$, with $\hat{r}_{j,t}$ denoting a prediction. Comparative model accuracy is evaluated using the Diebold and Mariano (DM) test (see Diebold and Mariano (1995a)). The null hypothesis of equal predictive accuracy of two forecasting models, say f and g, in this test is:

$$H_o : E[l(\epsilon_{t+h}^{f})] - E[l(\epsilon_{t+h}^{g})] = 0, \tag{3.52}$$

where ϵ_{t+h}^{f} is the prediction error in model f, ϵ_{t+h}^{g} is the prediction error in model g, and $l(\cdot)$ is the quadratic loss function. If we assume there is no parameter estimation error(i.e., $P/R \to 0$, where $P + R = T$, P denotes the number of ex-ante forecasts, and R is the length of the rolling window, or the initial length of the recursive window), and also under an assumption that the models are nonnested, then $DM_P = \frac{\bar{d}}{\hat{\sigma}_{\bar{d}}}$, where $\bar{d} = P^{-1}\sum_{t=1}^{P} d_t$ has a standard normal limiting distribution (here $\hat{\sigma}_{\bar{d}_t}$ is a heteroskedasticity and autocorrelation robust estimator of the standard deviation of $\bar{d}$), and $d_t = (\hat{\epsilon}_t^{f})^2 - (\hat{\epsilon}_t^{g})^2$ are estimates of true forecasting errors ϵ_{t+h}^{f} and ϵ_{t+h}^{g}. Details concerning appropriate critical values for cases in which parameter estiamtion error is not assumed to be negligible, asympotitcally, and/or in which models are nested are contained in Corradi and Swanson (2006c) and McCracken (2000b).

[14] Papers discussing the use of rolling and recursive estimation windows include Clark and McCracken (2009), Rossi and Inoue (2012), and the papers cited therein.

We adopt the Pesaran and Timmermann (1992) test to check the independence of our directional forecasts (namely, we construct classical chi-square tests of independence). In this context, we consider confusion matrices defined as follows:[15]

$$\begin{array}{cc|cc} & & \multicolumn{2}{c}{\text{Predicted}} \\ & & \text{up} & \text{down} \\ \hline \text{Actual} & \text{up} & n_1 & n_2 \\ & \text{down} & n_3 & n_4 \\ \hline \end{array} \tag{3.53}$$

Here n_1 (n_4) is the number of correct forecasts of upward (downward) return forecasts and n_2 (n_3) is the number of incorrect forecasts of upward (downward) movement in returns. Next, define:

$$p_{actu} = \frac{n_1 + n_2}{n_1 + n_2 + n_3 + n_4}, q_{actu} = \frac{p_{actu}(1 - p_{actu})}{n_1 + n_2 + n_3 + n_4}, \tag{3.54}$$

$$p_{pred} = \frac{n_1 + n_3}{n_1 + n_2 + n_3 + n_4}, q_{pred} = \frac{p_{pred}(1 - p_{pred})}{n_1 + n_2 + n_3 + n_4}, \tag{3.55}$$

The null hypothesis of Peseran-Timmermann (PT) test is that the model provides no value in directional forecasting. The test statistic is:

$$PT = \frac{p_{true} - p}{\sqrt{v - w}} \to N(0, 1), \tag{3.56}$$

where

$$p_{true} = \frac{n_1 + n_4}{n_1 + n_2 + n_3 + n_4}, p = p_{actu}p_{pred} + (1 - p_{actu})(1 - p_{pred}) \tag{3.57}$$

and

$$v = \frac{p(1 - p)}{n_1 + n_2 + n_3 + n_4}, w = (2p_{pred} - 1)^2 q_{actu} + (2p_{actu} - 1)^2 q_{pred} + 4q_{actu}q_{pred}. \tag{3.58}$$

The PT test is a one-sided test and the critical region is the upper tail of the standard normal distribution.

In addition, we report point direction forecasting performance using directional predictive accuracy rates (DPAR). The DPAR is defined as:

$$DPAR = \frac{\text{Number of correct forecasts}}{\text{Totol number of forecasts}} = \frac{n_1 + n_4}{n_1 + n_2 + n_3 + n_4} \tag{3.59}$$

[15]See Swanson and White (1997) for details.

We impose a simple “filter” on our forecasts in order to address the occasional occurrence of so-called “nonsense” forecasts, as discussed in Swanson and White (1997). Namely, if the one day change associated with a daily prediction exceeds the 90% percent of the average change observed during the past 22 trading days (i.e., one month), then the forecast from random walk model in Section 3.2 is used in place of the associated model based prediction.

Finally, correlation indices, macro risk factors, and volatility risk factors are 30-day moving averages. This smoothing was found to yield superior results relative to the use of un-smoothed uncertainty measures in our experiments.[16]

3.4 Empirical Results

3.4.1 Data

Our analysis is based on the use of 4 different datasets: 5-minute frequency equity price data, trading volume data, widely used macroeconomic predictors (as detailed in Welch and Goyal (2007)), and additional macroeconomic variables. Table 3.1A and Table 3.1B summarize the predictors and the prediction targets. Technical indicators, correlation indices, volatility risk factor, and all forecasting targets are derived from the 5-minute high-frequency price dataset, which is extracted from Trade and Quote (TAQ) database.[17] The trading volume dataset used to compute technical indicators is obtained from Yahoo Finance. The macroeconomic predictors include book to market ratio of the Dow Jones Industrial Average, net equity expansion and dividend-price ratios for S&P 500 index are from the dataset detailed in Welch and Goyal (2007). Our additional macroeconomic variables, including, for example, default spreads, term spreads, and the consumer price index are obtained from the FRED-MD database of Federal Reserve Bank of St.Louis.

More specifically, the macroeconomic variables that are used to build our macro risk

[16]In the prediction of SPY, all correlation indices are incorporated in the model. In the prediction of XLF, XLK, XLY, and XLV, only corresponding correlation indices are incorporated in the prediction model. See Data Section for further detials.

[17]Data obtained from Wharton Research Data Service (WRDS).

factor, MF_t^{mac}, are obtained from the FRED-MD database of Federal Reserve Bank of St.Louis, and include (1) the daily yield curve spreads, defined as the difference between the 10-year U.S. Treasury bond yield and the 3-month Treasury bill yield; (2) weekly initial claims for unemployment insurance; (3) the monthly number of nonfarm payroll employees; and (4) quarterly gross domestic product. All of these variables are log differenced in all calculations, in order to ensure stationarity, and are then standardized, with the exception of yield spreads, which are standardized, but not log-differenced. The fourth row in Table 3.1A and the third row in Table 3.1B show the transformations used in conjunction with our macroeconomic variables.

The daily financial variable log-returns that make up our target set of variables to be forecasted include: SPY (SPDR S&P 500 ETF Trust), XLF (Financial Select Sector SPDR Fund), XLK (Technology Select Sector SPDR Fund), XLY (Consumer Discretionary SPDR), and XLV (Health Care SPDR). All data cover the period from Jan 03, 2006 to Dec 31, 2017, with high frequency financial variables measured at intra-day and daily frequencies, and macroeconomic data measures at daily, weekly, monthly, and quarterly frequencies.

3.4.2 Forecasting results

Table 3.1C lists all forecasting models in the experiments, both for level and directional prediction. Table 3.2 and Table 3.3 report 1-step-ahead daily relative MSFEs of all forecasting models, using a rolling and a recursive window. The random walk model is used as a benchmark to generate relative MSFEs for all machine learning models. The monthly aggregate relative MSFEs of all forecasting models are tabulated in Table 3.4 (rolling window) and Table 3.5 (recursive window). We calculate the monthly aggregate return by summing over all 1-step-ahead daily return predictions. Table 3.6 (rolling window) and Table 3.7 (recursive window) contain the directional predictive accuracy rate based on 1-step-ahead daily level forecasting results. The directional accuracy rate based on monthly aggregate level forecasting results are shown in Table 3.8 (rolling window) and Table 3.9. Notably, direction forecasting results in Table 3.6-3.9 are directly derived from level prediction results.

For example, returns forecasts are classified as upward signals if forecasts are positive, and are otherwise classified as downward signals. The forecasting period of all tables is from Jun 2009 - Dec 2017 with a total of 2129 observations. We summarize the main empirical findings in the following:

First, machine learning models yields significantly smaller MSFEs and higher DPARs than the benchmark random walk model at monthly frequency. For example, monthly relative MSFEs results in Table 3.4 and Table 3.5 suggest most machine learning models outperform benchmark in all scenarios. Similarly, in Table 3.8 and Table 3.9, in terms of DPARs at a monthly frequency, machine learning models also stand out to be the best "DPARs" model. It is also noteworthy that some entries in Tables 3.4, 3.5, 3.8 and 3.9 are starred, especially for random forest and boosting models, indicating these machine learning models are statistically significantly different from benchmark, based on application of DM test and PT test discussed in Section 3.2. However, in Tables 3.2, 3.3, 3.6 and 3.7 at daily frequency, machine learning models show little improvement in terms MSFEs, and they perform slightly better under the measurement of DPARs, as for each given forecasting target and predictors, the best DPARs models are always machine learning models.

Second, the random forest model "wins" over other machine learning and benchmark models in both level and directional forecasting at monthly frequency. Evidently, almost all entries in bold [18] are listed under the random walk models in Tables 3.4, 3.5, 3.8 and 3.9. In terms of level forecasting, random forest stands out to be the best MSFEs model in 19 of 30 cases (Table 3.4) and 14 of 30 cases (Table 3.5). Also in terms of direction forecasting, random forest dominates other machine learning and benchmark models in 23 of 30 cases (Table 3.8), and 19 over 30 cases (Table 3.9). In particular, the lowest relative MSFEs for random forest model reaches 0.5503 (Table 3.5), and the highest DPARs achieves 0.8350 (Table 3.8). Note that other machine learning models including boosting and support vector regressions also prove to be the MSFEs and DPARs best models a few times in the forecasting "horse race".

[18]The entires with smallest relative MSFEs and largest DPARs in each row are denoted in bold

Third, deep learning models outperform shallow learning models in both level and direction predictions. In Tables 3.2-3.5, deep learning models with 2-4 hidden layers have lower MSFEs than shallow learning models with only one hidden layer across different choices of predictors and forecasting targets. In Tables 3.6-3.9, DPARs of deep learning models with 2-4 hidden layers are significantly higher than the DPARs of shallowing learning models with one hidden layer. Deep learning models are more efficient in capturing the data pattern and more accurate in forecasting the target variable.

Fourth, hybrid machine learning models, which combine lasso and neural network models, outperform individual models in both level and direction forecasting. As shown in Table 3.4 and Table 3.5, at monthly frequency, hybrid models have smaller MSFEs than individual lasso or neural network models in 23 of 30 cases (Table 3.4), and in 17 of 30 cases (Table 3.5). Moreover, directional prediction results show hybrid models "win" over individual lasso or neural network models in 22 of 30 cases (Table 3.8), and 19 of 30 cases (Table 3.9).

Fifth, all three risk factors, including market correlation indices, volatility risk factors, and macro risk factors, have significant marginal predictive content. In Figure 3.8, adding the volatility factor to our forecasting models reduces relative MSFEs by 3.2%-18.3% for different forecasting targets (SPY, XLF, XLK, XLY, and XLV). Adding the macro factor reduces relative MSFEs by 0.8%-22.5%, for the different target variables. Finally, adding the correlation indices leads to relative MSFE reductions of 1.7%-28.8%. These findings hold when monthly aggregate relative directional predictive accuracy rates (DPARs) are analyzed, as shown in Figure 3.9. In Figure 3.9, we see that adding the volatility factor leads to 1.3%-10.3% DPAR increases, while adding the macro factor increases DPARs by 1.3%-11.4%. Finally, adding correlation indices increase DPARs by 1.3%-8.9%, with an exception of SPY, for which no gains are noted.

Sixth, each sector generally has a different sensitivity to the input information. We evaluate the contribution of different inputs by comparing MSFEs and DPARs using the leave-one-out scheme within each type of model. Each round, we leave one of the following five categories of predictors out of the model: 1) macro variables 2) technical variables 3) volatility risk factors 4) macro risk factors and 5) market correlation indices. Details about the predictors in each category can be found in Table 3.1B. First examining Tables 3.4-3.5, under the MSFEs best model- random forest, for the SPY, leaving inputs such as macro variables, factors for uncertainty, market correlation and macroeconomic condition, induces larger MSFEs than original model having all variables, except technical indicators. This evidence is also confirmed in terms of DPARs results shown in Tables 3.8-3.9, with only one exception for correlation index. For XLF and XLY, as shown in Tables 3.4, 3.5, 3.8 and 3.9, leaving any type of inputs yields larger MSFEs and lower DPARs than original model having all variables. For XLK, removing macro variables or macro condition factor leads to larger MSFEs (Tables 3.4-3.5), and removing macro variables, technical variables or macro condition factor yields lower DPARs (Tables 3.8-3.9). Finally, for XLV, leaving macro variables and technical variables leads to larger MSFEs (Tables 3.4-3.5), and removing any type of inputs induces lower DPARs (Tables 3.8-3.9).

Figure 3.1-3.3 show the continuous component correlation indices and the jump component correlation indices of energy sector (XLE) and S&P500 (SPY), finance sector (XLF) and SPY, industrial sector (XLI) and SPY, technology sector (XLK) and SPY, health care sector (XLV) and SPY, and consumer discretionary sector (XLY) and SPY from 2006:01 - 2017:12. The jump correlation index surges during the 2008 and 2011 financial crisis, and drops when the market volatility is low. However, the energy market correlation index behaves differently comparing with the other four sectors. One sensible explanation is energy market depends more on the balance of supply and demand in energy commodities while less related to the financial market condition. Interestingly, the correlation index based on the continuous part behaves oppositely to the correlation index calculated by jump components.

Figure 3.4 depicts the volatility risk factors of the S&P 500 market (MF_t^{TRV}), financial sector (MF_t^{XLF}), technology sector (MF_t^{XLK}), health care sector (MF_t^{XLV}), and consumer discretionary sector (MF_t^{XLY}). During the Great Recession, the financial sector volatility risk factor positions higher than risk factors of all other sectors, and shows a unique two-peak shape corresponding to the period of December 2008 and April 2009. The second peak is extremely contrasting since volatility factors calmed down significantly among all other sectors during the time. These two peaks can match back to historical events, as the first peak points to the big market tumble at the beginning of December 2008, with S&P 500 down 9% and financial sector fell the most by 17%, and the second peak relates to a big bull market rally from March to May 2009 with financial stocks strikingly went up 150%, explaining the unique second peak in the financial section volatility risk factor.

3.5 Concluding Remarks

In this paper, we extensively study the performance of machine learning individual models, as well as hybrid machine learning models, in the sector-level equity return forecasting, including random forest, boosting, support vector machine, penalized regression, logistic regression, latent discriminant analysis, naive Bayes classifier, k-nearest-neighbor classifier, neural network, and hybrid models. The impetus of our study is to analyze a number of new finance and macro-oriented latent measures of uncertainty, and to assess their marginal predictive content. These measures are constructed using high frequency and high dimensional financial data, as well as mixed frequency macroeconomic indicators. Out-of-sample forecasting experiments are carried out for the following financial assets: SPY (SPDR S&P 500 ETF Trust), XLF (Financial Select Sector SPDR Fund), XLK (Technology Select Sector SPDR Fund), XLV (Health Care SPDR), and XLY (Consumer Discretionary SPDR). We analyze both level and directional predictions at daily and monthly frequencies. Results from our empirical experiments are promising. Machine learning models, especially the random forest model, achieve significantly higher directional accuracy rates and lower mean square forecasting errors than the random walk benchmark. Moreover, various of our new latent uncertainty measures deliver significant marginal predictive content, which is particularly useful for forecasting at a monthly frequency. All categories of predictors show contributions to both level and directional forecasting.

Table 3.1A: Predictor Variables*

Predictor Name	Category	Description	X_t	Frequency
T10Y3M	Macro Variables	term spread: 10-year treasury constant maturity minus 3-month treasuty constant maturity	X_t	Daily
defauspr		default spread: the difference between BAA and AAA-rated corporate bond yields	X_t	Daily
b/m		ratio of book value to market value of the Dow Jones Industrial Average	X_t	Monthly
ntis		Net Equity Expansion: the ratio of 12-month moving sums of net issues by NYSE listed stocks divided by the total end-of-year market capitalization of NYSE stocks.	X_t	Monthly
Diff_CPIAUCSL		Consumer Price Index	$ln(X_t) - ln(X_{t-1})$	Monthly
D/P		Dividend Price ratio of S&P 500 index	$x_t = log(D_t) - log(P_t)$	Daily
indi_30_90	Technical Indicators	30-day trading volume indicator	X_t	Daily
indi_90_120		90-day trading volume indicator	X_t	Daily
indi_30_120		120-day trading volume indicator	X_t	Daily
MA_30_90		30-day price trend indicator	X_t	Daily
MA_90_120		90-day price trend indicator	X_t	Daily
MA_30_120		120-day price trend indicator	X_t	Daily
XLECorr_ma	Correlation Index	continuous part price correlation index between enery sector(XLE) and SPY	X_t	Daily
XLFcorr_ma		continuous part price correlation between financial sector(XLF) and SPY	X_t	Daily
XLICorr_ma		continuous part price correlation index between industry sector(XLI) and SPY	X_t	Daily
XLKCorr_ma		continuous part price correlation index between technology sector(XLK) and SPY	X_t	Daily
XLVCorr_ma		continuous part price correlation index between health care sector(XLV) and SPY	X_t	Daily
XLYCorr_ma		continuous part price correlation index between consumer discretionary sector(XLY) and SPY	X_t	Daily
XLEJump_ma		Jump part price correlation index between enery sector(XLE) and SPY	X_t	Daily
XLFJump_ma		Jump part price correlation between financial sector(XLF) and SPY	X_t	Daily
XLIJump_ma		Jump part price correlation index between industry sector(XLI) and SPY	X_t	Daily
XLKJump_ma		Jump part price correlation index between technology sector(XLK) and SPY	X_t	Daily
XLVJump_ma		Jump part price correlation index between health care sector(XLV) and SPY	X_t	Daily
XLYJump_ma		Jump part price correlation index between consumer discretionary sector(XLY) and SPY	X_t	Daily
MF^{TRV}	Volatility Risk Factor	Multi-frequency financial volatility risk factor	X_t	Daily
MF_t^{mac}	Macro Risk Factor	Macroeconomic factor	X_t	Daily
Return_lag	Lag Term	lag one day of prediction target return	X_{t-1}	Daily

*Note: Table 3.1A shows all predictors in the forecasting models for the period 2006:01-2017:12. All predictors are divided into six categories, which is shown in the second column. Data transformations used in forecasting experiments are given in the fourth column of the table. See Section 3.2 and 3.3 for further details.

Table 3.1B: Target Forecast Variables*

Target Name	Description	X_t	Frequency
SPY	SPDR S&P 500 ETF Trust	$ln(X_t) - ln(X_{t-1})$	Daily
XLF	Financial Sector SPDR Fund	$ln(X_t) - ln(X_{t-1})$	Daily
XLK	Technology Sector SPDR Fund	$ln(X_t) - ln(X_{t-1})$	Daily
XLY	Consumer Discretionary SPDR	$ln(X_t) - ln(X_{t-1})$	Daily
XLV	Health Care SPDR	$ln(X_t) - ln(X_{t-1})$	Daily

*Notes: This table reports the prediction targets. Data transformations used in forecasting experiments are given in the third column of the table. See Section 3.3 for further details.

Table 3.1C: Models Used in Forecasting Experiments*

	Method	Description
Level Forecasting	Benchmark	Random Walk
	Linear	Linear regression
	SVR_rbf	Support vector regresion with radial basis function kernel
	SVR_lin	Support vector regresion with linear kernel
	SVR_poly	Support vector regresion with polynomial kernel
	RanForest	Random forest regression
	Boosting	Gradient boosting regression
	Lasso	Lasso regression
	Ridge	Ridge regression
	Nnet1	Neural network regression with one hidden layer
	Nnet2	Neural network regression with twp hidden layers
	Nnet3	Neural network regression with three hidden layers
	Nnet4	Neural network regression with four hidden layers
	Hybrid1	A hybrid mode of Lasso and Nnet1 with selected variables
	Hybrid2	A hybrid mode of Lasso and Nnet2 with selected variables
	Hybrid3	A hybrid mode of Lasso and Nnet3 with selected variables
	Hybrid4	A hybrid mode of Lasso and Nnet4 with selected variables
	Hybrid5	A hybrid mode of Lasso and Nnet1 with All variables
	Hybrid6	A hybrid mode of Lasso and Nnet2 with All variables
	Hybrid7	A hybrid mode of Lasso and Nnet3 with All variables
	Hybrid8	A hybrid mode of Lasso and Nnet4 with All variables
Direction Forecasting	Logit	Logistic regression
	LDA	Linear discriminant analysis
	NB	Naive bayes classifier
	SVC_RBF	Support vector classification with radial basis function kernel
	SVC_lin	Support vector classification with linear kernel
	SVC_poly	Support vector classification with polynomial kernel
	KNN	K-nearest neighbors algorithm
	Boosting	Gradient boosting classification
	RanForest	Random forest classification
	Nnet1	Neural network classification with one hidden layer
	Nnet2	Neural network classification with two hidden layers
	Nnet3	Neural network classification with three hidden layers
	Nnet4	Neural network classification with four hidden layers

*Notes: This table reports the models in forecasting experiments. Complete details for all models are given in Section 3.3.

Table 3.2: 1-Step-Ahead Daily Relative MSFEs of All Forecasting Models (Rolling Window)*

		Linear	SVRrbf	SVRlin	SVRpoly	RanForest	boosting	lasso	ridge	NN1	NN2	NN3	NN4	hybrid1	hybrid2	hybrid3	hybrid4	hybrid5	hybrid6	hybrid7	hybrid8
SPY	All Variables	1.2077***	1.7704***	1.1303***	1.2676***	1.0366***	1.1504***	0.9995	1.0541***	1.5332***	1.3431***	1.0253***	0.9993	1.0037	1.0079	1.0009	**0.9987**	1.1088***	1.0129	1.0046	0.9998
	Drop Macro Variables	1.1528***	1.0984***	1.095***	1.2324***	1.0314**	1.2492***	0.9995	1.0344***	1.2886***	1.2190***	1.0025	1.0007*	1.0124**	1.0089*	1.0061	0.9994	1.1604***	1.0973***	**0.9989**	1.0000
	Drop technical Variables	1.1869***	1.0644***	1.1026***	1.0516***	1.0856***	1.5325***	**0.9995**	1.0326***	1.3756***	1.1605***	1.0085*	1.8679***	1.0029	1.0072	1.0054	1.0018	1.0516***	1.0213**	1.0016	1.2696***
	Drop Volatility factor	1.1987***	1.0897***	1.1193***	1.1592***	1.0200**	1.1308***	**0.9994**	1.0515***	4.2393***	1.279***	1.0138***	1.3581***	1.0132**	1.0060*	1.0069**	1.0068	1.9241***	1.0394***	1.0039	1.0518**
	Drop Macro Factor	1.1991***	1.7923***	1.1525***	1.1309***	1.0183**	1.4545***	1.0002	1.0481***	1.6493***	1.3019***	1.0141***	0.9998	1.0059**	1.0168***	**0.9972**	1.0007*	1.1324***	1.0224***	1.0004	0.9994
	Drop Correlation index	1.1680***	1.1115***	1.0959***	1.06***	1.0223**	1.2094***	0.9995	1.0529***	1.6683***	1.1513***	1.0115**	0.9998	1.0141***	1.0070*	1.0024	1.0003	1.2129***	1.0204**	0.9999	**0.9987**
XLF	All Variables	1.2234***	1.0897***	1.0849***	1.0626***	1.0525***	1.1453***	**1.0017**	1.0781***	1.1496***	1.2958***	1.0463***	1.1017***	1.0227***	1.0275***	1.0062**	1.0084**	1.0524***	1.1137***	1.0214**	1.0226***
	Drop Macro Variables	1.0932***	1.0779***	1.0661***	1.1795***	1.028***	1.1493***	**1.0017**	1.0607***	1.0734***	1.0277***	1.0239***	1.1391***	1.0141***	1.0070**	1.0042*	1.0200***	1.0344***	1.0047	1.0097**	1.0423***
	Drop technical Variables	1.1324***	1.0338***	1.0367***	1.0226***	1.0904***	1.1737***	**1.0005**	1.0312***	1.0756***	1.0713***	1.0023	1.0604***	1.0077*	1.0042	1.0079***	1.0070**	1.0366***	1.0249**	1.0081**	1.0094**
	Drop Volatility factor	1.2243***	1.2762***	1.0803***	1.2180***	1.0309***	1.1586***	1.0003	1.0782***	1.0979***	1.0557***	**0.9996**	1.1466***	1.0157***	1.0100**	1.0010	1.0183***	1.059***	1.0212***	1.0003	1.0211**
	Drop Macro Factor	1.2139***	1.0819***	1.0760***	1.0596***	1.0313***	1.1528***	1.0028**	1.0822***	1.1366***	1.0839***	1.0435***	2.7153**	1.0208***	1.0184***	**1.0005**	1.0182***	1.0301**	1.0330***	1.0295***	1.1065***
	Drop Correlation index	1.2058***	1.0876***	1.0800***	1.0631***	1.0246***	1.1374***	1.0017	1.0763***	1.0698***	1.0231**	1.1251***	1.0272***	1.0106**	1.0055	1.0209***	**0.9978**	1.0325***	1.0106*	1.0187***	1.0193***
XLK	All Variables	1.1195***	1.0637***	1.0737***	1.039***	1.0196***	1.2742***	1.0002	1.0608***	1.2355***	1.207***	1.0064**	**0.9995**	1.0018	1.0006	0.9998	1.0003	1.0265***	1.0750***	1.0012	0.9998
	Drop Macro Variables	1.0652***	1.0489***	1.0421***	1.0294***	1.024***	1.1828***	1.0002	1.0413***	1.0639***	1.0965***	1.0060	1.0000	1.0064	0.9999	**0.9994**	1.0003	1.0100*	1.0114**	1.0013	1.0004
	Drop technical Variables	1.0651***	1.0257***	1.0397***	1.0161**	1.0245***	1.1583***	1.0002	1.0184***	1.0575***	1.0988***	1.0042	1.001	**0.9995**	1.0066**	1.0005	1.0002	1.0111**	1.0356***	1.0054**	0.9999
	Drop Volatility factor	1.1177***	1.0440***	1.0516***	1.0314***	1.0213***	1.2149***	**1.0001**	1.0567***	1.1596***	1.0768***	1.0006	1.0015***	1.0110*	1.0018	1.0017	**1.0001**	1.0219***	1.0174***	1.0047**	**1.0001**
	Drop Macro Factor	1.1126***	1.0573***	1.0731***	1.0392***	1.0175***	1.4152***	1.0000	1.0581***	1.2087***	1.2515***	1.0023	1.1242***	1.0008	1.0012	**0.9995**	1.0022	1.0188***	1.0434***	1.0027**	1.0023
	Drop Correlation index	1.1138***	1.0621***	1.0693***	1.1387***	1.0529***	1.3824***	1.0002	1.0608***	1.1406***	1.1016***	1.0357***	1.1165***	1.0084**	1.0027	**0.9989**	1.0292***	1.0189***	1.0192***	**0.9989**	1.0191***
XLY	All Variables	1.1453***	1.0769***	1.1102***	1.0737***	1.0465***	1.1069***	1.0010	1.0724***	1.2696***	1.1585***	1.0587***	**0.9994**	1.0170***	1.0154***	1.0022	1.0011*	1.1153***	1.0253***	1.0078*	1.0011
	Drop Macro Variables	1.0537***	1.0663***	1.0478***	1.0401***	1.0735***	1.1008***	1.0010	1.0433***	1.0843***	1.0771***	1.0123***	**1.0002**	1.0100**	1.0119***	1.0035*	1.0011*	1.0138**	1.0139***	**1.0002**	1.0012*
	Drop technical Variables	1.0799***	1.0351***	1.0610***	1.0307***	1.0192***	1.0994***	1.0005	1.0246***	1.2612***	1.0887***	1.1074***	**1.0002**	1.0096***	1.0006	1.0140**	1.0004	1.0690***	1.0310***	1.0102**	1.0003
	Drop Volatility factor	1.1477***	1.0642***	1.1014***	1.0615***	1.0253***	1.0991***	1.0006	1.0714***	1.8243***	1.1428***	1.0062**	1.2808***	1.0108*	1.0079***	**0.9998**	1.0518***	1.2371***	1.0157**	1.0016	1.0581***
	Drop Macro Factor	1.1374***	1.0622***	1.1144***	1.0622***	1.0421***	1.0883***	**1.0006**	1.0719***	1.2166***	1.1413***	1.1840***	1.0002	1.0112***	1.0032	1.0010	1.0008	1.0200***	1.0201***	1.0456***	1.0013**
	Drop Correlation index	1.1430***	1.0752***	1.1019***	1.0740***	1.0294***	1.1325***	**1.0010**	1.0725***	1.1388***	1.0933***	1.0139***	1.2876***	1.0065*	1.0105***	1.0035*	1.0100***	1.0317***	1.0179***	1.0054*	1.0477***
XLV	All Variables	1.0904***	1.0499***	1.0596***	1.1656***	1.0251***	1.1955***	1.0000	1.0426***	2.5925***	1.4951***	1.0005	1.0004	1.0012*	1.0011	1.0001	1.0000	**0.9999**	1.0016	1.0001	**0.9999***
	Drop Macro Variables	1.0622***	1.0459***	1.0333***	1.0269***	1.0325***	1.1757***	1.0000	1.038***	1.1325***	1.0622***	1.0155***	1.0003	1.0012*	1.0011	1.0001	1.0000	**0.9999**	1.0016	1.0001	**0.9999***
	Drop technical Variables	1.0584***	1.0238***	1.0176**	1.0154**	1.0301***	1.1875***	1.0000	1.0154***	1.2394***	1.0946***	1.0091**	1.0653***	1.0012*	1.0011	1.0001	1.0000	**0.9999**	1.0016	1.0001	**0.9999***
	Drop Volatility factor	1.0847***	1.0493***	1.036***	1.0339***	1.0555***	1.2246***	1.0000	1.0371***	1.2447***	1.121***	1.0079***	**0.9996**	1.0012*	1.0011	1.0001	1.0000	0.9999	1.0016	1.0001	0.9999*
	Drop Macro Factor	1.0936***	1.0475***	1.0502***	1.022***	1.0355***	1.2327***	1.0000*	1.0397***	1.2465***	2.4683***	1.0082***	1.0007*	1.0012*	1.0011	1.0001	1.0000	**0.9999**	1.0016	1.0001	**0.9999***
	Drop Correlation index	1.0762***	1.2771***	1.0365***	1.1761***	1.0323***	1.2265***	1.0000	1.0415***	1.1636***	1.0452***	1.0191***	1.0465***	1.0012*	1.0011	1.0001	1.0000	**0.9999**	1.0016	1.0001	**0.9999***

*Notes: See notes to Table 3.1A. Table 3.2 reports the 1-step-ahead relative mean square forecasting error (MSFE) of market sector ETFs with rolling window size 500. Forecasts are daily, for the period 2009:6-2017:12. Tabulated relative MSFEs are calculated such that numerical values less than unity indicates the alternative model has lower point MSFE than the random walk benchmark model. Entries in bold denote models with lowest relative MSFE for a given forecasting target and predictors. Starred entries denote rejection of the null of equal predictive accuracy, based on the application of Diebold and Mariano (1995a) (DM) test. All machine learning models are tested against the random walk benchmark, based on MSFE loss. Significance levels for the test are reported as $***p < 0.01$, $**p < 0.05$, and $*p < 0.1$, where p is the $p-$value corresponding to DM test statistics.

Table 3.3: 1-Step-Ahead Daily Relative MSFEs of All Forecasting Models (Recursive Window)*

		Linear	SVRrbf	SVRlin	SVRpoly	RanForest	boosting	lasso	ridge	NN1	NN2	NN3	NN4	hybrid1	hybrid2	hybrid3	hybrid4	hybrid5	hybrid6	hybrid7	hybrid8
SPY	All Variables	1.1616***	1.7658***	1.0666***	1.221***	1.0094*	1.0455***	0.9991	1.0404***	1.3753***	1.2400***	1.0219***	1.0003	1.0016	1.0076**	1.0067*	**0.9989**	1.2382***	1.0969***	1.0181***	0.9993
	Drop Macro Variables	1.0805***	1.0588***	1.0463***	1.1630***	1.0064*	1.0656***	0.9991	1.019***	1.5325***	1.4164***	1.036***	1.0007	1.1077***	1.0455***	1.0081*	**0.9986**	1.2673***	1.2595***	1.0182***	1.0007
	Drop technical Variables	1.1531***	1.0438***	1.0610***	1.0283***	1.0658***	1.4103***	**0.9991**	1.0311***	1.2795***	1.1489***	1.0183***	1.8884***	1.0031	1.0064	1.0085**	1.0099***	1.1486***	1.0604***	1.0174***	1.3809***
	Drop Volatility factor	1.1592***	1.0537***	1.0592***	1.0923***	1.0026	1.0471***	**0.9991**	1.0337***	3.4572***	1.3582***	1.0005	1.1905***	1.0774***	1.0413***	1.0019	1.0404***	2.7528***	1.2012***	1.0167***	1.1285***
	Drop Macro Factor	1.1534***	1.7902***	1.0719***	1.0881***	1.0017	1.3266***	1.0003***	1.0338***	2.1067***	1.2658***	1.0113**	1.0001	1.0030	1.0062**	**0.9990**	1.0001	1.1328***	1.0271***	0.9994	0.9999
	Drop Correlation index	1.1270***	1.0726***	1.0435***	1.0232***	1.0056*	1.0661***	0.9991	1.0358***	1.5577***	1.1924***	1.0251***	0.9990	1.1031***	1.0319***	1.0042	**0.9984**	1.4183***	1.1044***	1.007*	0.9988
XLF	All Variables	1.1840***	1.0586***	1.0534***	1.037***	1.0538***	1.0630***	**1.0018***	1.0689***	1.1587***	1.3453***	1.1048***	1.0638***	1.0108**	1.0253***	1.0045**	1.0176***	1.0348***	1.1682***	1.0190**	1.0188***
	Drop Macro Variables	1.0491***	1.0493***	1.0392***	1.1066***	1.0086***	1.0569***	**1.0016**	1.0429***	1.0961***	1.0614***	1.0389***	1.0918***	1.0023	1.0170***	1.0027	1.0382***	1.0487***	1.0059	1.0103**	1.0124
	Drop technical Variables	1.1118***	1.0261***	1.0252***	1.0182***	1.0813***	1.0998***	**1.0009**	1.0287***	1.1098***	1.0513***	1.0189***	1.0137***	1.0102**	1.0094**	1.0048**	1.0074*	1.0600***	1.0547***	1.0035	1.0043
	Drop Volatility factor	1.1809***	1.2085***	1.0517***	1.2064***	1.0155***	1.0746***	1.0005	1.0665***	1.1476***	1.1086***	1.0027	1.0648***	1.0243***	1.0049	**0.9986**	1.0082	1.0671***	1.0170**	1.0029	1.0493***
	Drop Macro Factor	1.1778***	1.0578***	1.0486***	1.0392***	1.0171***	1.0830***	**1.0025*****	1.0666***	1.1085***	1.1244***	1.0591***	1.3106***	1.0162***	1.0107**	1.0044**	1.0113***	1.0650***	1.0615***	1.0201***	1.1819***
	Drop Correlation index	1.1687***	1.0582***	1.0462***	1.0373***	1.0035	1.0757***	**1.0018***	1.0673***	1.0897***	1.0585***	1.0489**	1.0038	1.0128***	1.0126***	1.0239***	1.0122***	1.0373***	1.0250***	1.0170***	1.0105***
XLK	All Variables	1.0696***	1.0339***	1.0374***	1.0165***	1.0034	1.1304***	**1.0001****	1.0247***	1.1609***	1.4002***	1.0050	1.0009*	1.0015	1.0019	1.0002	1.0003**	1.0195***	1.0514***	1.0013	1.0004*
	Drop Macro Variables	1.0243***	1.0313***	1.0197***	1.0168***	1.0078***	1.0500**	1.0001**	1.0162***	1.0682***	1.1659***	1.0145***	1.0000	1.0027	1.0023	1.0012	1.0002	1.0119***	1.0057	1.0020	**0.9999**
	Drop technical Variables	1.0486***	1.0167**	1.0152**	1.0072	1.0025	1.0658***	**1.0001****	1.0121***	1.0997***	1.0685***	1.0099***	1.0312***	1.0031**	1.0021	1.0013	1.0004	1.0147***	1.0130**	1.0026*	1.0001
	Drop Volatility factor	1.0682***	1.0306***	1.0229***	1.0131**	1.0072***	1.092***	1.0000	1.0237***	1.1590***	1.1779***	1.0169***	1.0014***	1.0022	1.0024**	**0.9992**	1.0003	1.0304***	1.0165***	1.0000	0.9999
	Drop Macro Factor	1.0677***	1.0325***	1.0354***	1.0144***	1.0033	1.2668***	1.0001***	1.0234***	1.1008***	1.1787***	1.0133***	1.087***	1.0012***	1.0005***	0.9998	1.0009**	1.0055*	1.0170*	**0.9994**	1.0095
	Drop Correlation index	1.0689***	1.0370***	1.0318***	1.0834***	1.0576***	1.1796***	**1.0001****	1.0250***	1.1549***	1.0475***	1.0445***	1.0955***	1.0034**	1.0039**	1.0011	1.0118***	1.0126**	1.0119**	1.0051*	1.0149**
XLY	All Variables	1.0860***	1.0493***	1.0604***	1.0311***	1.0443***	1.0381***	1.0005	1.0360***	1.4131***	1.1167***	1.1025***	**0.9993**	1.0144***	1.0046	1.0019	1.0004	1.1698***	1.0172***	1.0154***	1.0009*
	Drop Macro Variables	1.0284***	1.0346***	1.0246**	1.0177**	1.0559***	1.0407***	1.0004	1.0216***	1.1148***	1.1730***	1.0228***	1.0005	1.0065**	0.9997	**0.9984**	1.0003	1.0132**	1.0156**	1.0020	1.0004
	Drop technical Variables	1.0582***	1.0193***	1.0312***	1.0136**	1.0032	1.0400***	1.0004	1.0164***	1.2477***	1.1056***	1.0720***	1.0018***	1.0087***	**0.9996**	1.0027	1.0003	1.0981***	1.0313***	1.0161**	0.9999
	Drop Volatility factor	1.0838***	1.0374***	1.0619***	1.0260***	1.0065	1.0347***	**0.9999**	1.0346***	1.5831***	1.1435***	1.0076***	1.1433***	1.0053	1.0025	1.0009	1.0350***	1.2985***	1.0511***	1.0036*	1.0396***
	Drop Macro Factor	1.0836***	1.0409***	1.0579***	1.0271***	1.0323***	1.0451***	1.0005*	1.0347***	1.1708***	1.1666***	1.4072***	1.0000	1.0017	1.0054**	**1.0004**	1.0006*	1.0248***	1.0046	1.0203***	1.0005
	Drop Correlation index	1.0919***	1.0468***	1.0511***	1.0318***	1.0104**	1.0568***	**1.0005**	1.0416***	1.1296***	1.1119***	1.0087**	1.0701***	1.0041*	1.0019	1.0007	1.0028	1.0332***	1.0134***	1.0084***	1.0429***
XLV	All Variables	1.0513***	1.0231***	1.0237***	1.0894***	1.0058**	1.0761***	**0.9997**	1.0191***	2.0021***	1.5009***	1.0089**	0.9999	**0.9997**	**0.9997**	**0.9997**	**0.9997**	**0.9997**	**0.9997**	**0.9997**	**0.9997**
	Drop Macro Variables	1.0260***	1.0219***	1.0058*	1.0043	1.0114***	1.1138***	**0.9997**	1.0149***	1.1256***	1.1804***	1.0111**	1.0010	**0.9997**	**0.9997**	**0.9997**	**0.9997**	**0.9997**	**0.9997**	**0.9997**	**0.9997**
	Drop technical Variables	1.0440***	1.0060	1.0107*	1.0050	1.0059*	1.0794***	**0.9997**	1.0090**	1.2063***	1.1386***	1.0098**	1.0289***	**0.9997**	**0.9997**	**0.9997**	**0.9997**	**0.9997**	**0.9997**	**0.9997**	**0.9997**
	Drop Volatility factor	1.0489***	1.0153**	1.0192***	1.0131**	1.0520***	1.0786***	**0.9997**	1.0177***	1.1400***	1.1987***	1.0056*	1.0005	**0.9997**	**0.9997**	**0.9997**	**0.9997**	**0.9997**	**0.9997**	**0.9997**	**0.9997**
	Drop Macro Factor	1.0531***	1.0191***	1.0178***	1.0078	1.0167***	1.0814***	**0.9997**	1.0180***	1.1165***	1.8097***	1.0048	0.9998	**0.9997**	**0.9997**	**0.9997**	**0.9997**	**0.9997**	**0.9997**	**0.9997**	**0.9997**
	Drop Correlation index	1.0453***	1.1823***	1.0171**	1.0969***	1.0074*	1.092***	**0.9997**	1.0187***	1.1471***	1.0324***	1.0140***	1.0166***	**0.9997**	**0.9997**	**0.9997**	**0.9997**	**0.9997**	**0.9997**	**0.9997**	**0.9997**

*Notes: See notes in table 3.2. Recursive window size 500.

Table 3.4: Monthly Aggregate Relative MSFEs of All Forecasting Models (Rolling Window)*

		Linear	SVRrbf	SVRlin	SVRpoly	RanForest	boosting	lasso	ridge	NN1	NN2	NN3	NN4	hybrid1	hybrid2	hybrid3	hybrid4	hybrid5	hybrid6	hybrid7	hybrid8
SPY	All Variables	1.7483***	0.8873***	1.1617***	0.9068***	**0.7677*****	0.9405***	0.9748*	1.2326***	1.3868***	1.2175***	1.0273**	1.0027	0.9634	0.9687*	0.9831	0.9967	1.1062**	1.006	1.0035	0.9868
	Drop Macro Variables	1.2639***	0.9889***	0.8629***	0.8879***	0.8048***	**0.5709*****	0.9748*	0.9726***	0.9959***	1.0992***	1.0007	1.0009	0.9801	0.969	1.0213**	0.9831	1.0051***	1.0923	1.0119	0.9849
	Drop technical Variables	1.8221***	1.0555***	1.2425***	1.0566***	**0.6262*****	0.7587***	0.9748*	1.1983***	1.1527***	1.1437***	1.0475	1.0038***	0.9719	1.0121	0.9798	0.9839	1.0158**	0.9678	0.9924	1.1093**
	Drop Volatility factor	1.6141***	0.8772***	1.107***	0.887***	**0.7921*****	0.9105***	0.9638**	1.243***	2.6349***	1.2478***	1.0216**	0.9687***	0.9569**	0.9436	1.0256	0.9671	1.4345***	1.1083*	0.971	0.946
	Drop Macro Factor	1.8599***	0.8122***	1.2382***	0.8328***	0.8202***	**0.7496*****	1.0031**	1.1774**	1.4029***	1.1068***	1.0262	1.0026	1.0061***	1.0353***	1.0025	1.0032***	0.9373***	0.9991**	0.9969*	1.0038
	Drop Correlation index	2.49***	0.9818***	1.3913***	1.0922***	**0.8292***	0.9451***	0.9748*	1.3533***	0.996***	1.1156***	0.9946**	1.0056	0.9942***	0.9271	0.9951	0.994	0.8969***	0.9989*	0.9883	0.9798
XLF	All Variables	3.099***	1.0838***	1.2804***	1.2063***	**0.6448*****	0.8643***	1.0061	1.6261***	1.2243***	1.1551***	1.0017	1.0397**	1.0986**	1.0483	1.0178*	1.0291	1.1412***	1.1386*	1.0066	1.0179*
	Drop Macro Variables	1.4752***	1.1046***	1.1567***	0.8887***	**0.8056***	0.9705***	1.0061	1.3305***	1.0023***	0.994	1.0346	1.0608***	1.0593*	1.0362	1.0157	1.0748**	1.0516**	1.0073*	0.9984	1.0565**
	Drop technical Variables	2.2144***	0.9684***	0.9947**	0.9555***	**0.6043*****	1.0068***	0.9776	1.2519***	0.8767***	1.0276***	1.0008**	1.0098***	0.9876	0.9715	0.9909	0.9724	0.9904	0.9582	0.9904	0.9809
	Drop Volatility factor	3.1517***	0.9783***	1.2784***	0.9667***	**0.7103*****	0.9502***	0.9790	1.6939***	1.1423***	1.0099***	1.0039	0.9828***	1.0868*	0.998	0.9864	0.9849	1.153***	0.9668	0.9855	0.9998
	Drop Macro Factor	3.1485***	0.9962***	1.1717***	1.1858***	**0.6957****	1.0179***	1.0377**	1.7524***	0.964***	1.0581**	1.014**	1.7497***	1.208**	1.0697*	1.0347	1.2017***	1.112	1.0969	1.0252*	1.3216***
	Drop Correlation index	3.2623***	1.0604***	1.2862***	1.2242***	**0.7691*****	0.9230***	1.0061	1.6155***	1.0802***	0.9242	1.0045**	1.0002***	1.0539	1.0408*	1.0637	1.0263	1.081*	1.0503	1.0011**	1.0024*
XLK	All Variables	1.7044***	0.8886***	1.1048***	1.0398***	0.8423*	**0.7639*****	0.9876	1.3904***	1.1089***	0.9331***	0.9848	0.9982***	0.9888	0.9924	0.9958	0.9895	1.0169***	1.0592**	0.9884	0.9879
	Drop Macro Variables	1.3333***	1.0526***	1.1428***	1.0932***	0.9527***	**0.7625*****	0.9876	1.2627***	1.1250**	1.0593***	1.013***	0.9998	1.0101**	0.9924	1.0018*	0.9925	1.0034**	0.9802	0.9960	0.9884
	Drop technical Variables	1.4983***	0.8304*	1.0297***	0.9009	0.8569***	**0.7406*****	0.9876	1.0661***	0.9699**	1.0506**	1.0107	1.0004	0.991	0.9879	0.9926	0.9914	1.0336**	0.9594	0.9949	0.9888
	Drop Volatility factor	1.7496***	0.982***	1.0792***	1.0228***	0.8017***	**0.7876*****	0.9856	1.4733***	1.2542***	1.0244**	0.9946	0.9998	0.9837	0.9621	0.986	0.9919	1.0268***	0.9702*	1.006**	0.989
	Drop Macro Factor	1.726***	0.9079***	1.1431***	1.0093***	0.8521*	**0.7989*****	1.0011	1.4353***	1.0806***	1.1249***	1.0051	0.9868**	1.0016	1.0012	0.9965*	1.0081	0.9406*	0.9953***	1.0022**	0.9906
	Drop Correlation index	1.7010***	0.9152***	1.2148***	0.9635***	**0.6808*****	0.8260***	0.9876	1.4094***	1.0847***	1.0001***	1.0045**	1.0239***	0.9813	0.9854	0.9857	0.9690	1.0026	0.9793*	0.9772*	1.0147**
XLY	All Variables	1.9525***	1.2327***	1.7358***	1.615**	**0.6864*****	0.9469**	1.0039**	1.7071***	1.1996***	0.992***	1.0255***	1.0007	1.0687*	1.0094**	1.0011	0.9992	0.9918***	1.0293**	0.9817	0.9973
	Drop Macro Variables	1.2435***	1.0929**	1.3126**	1.2753**	**0.801*****	0.8964***	1.0039**	1.3240**	1.0950***	1.097**	1.0039*	1.0009	1.0005	0.9798	1.0096	1.0029	1.0126	0.9716	1.0019	0.9972**
	Drop technical Variables	1.6683***	1.0035***	1.3086***	1.1205***	**0.8449**	0.9455***	0.9932*	1.2081*	1.2139***	0.9679	1.015***	0.9998	0.9873	0.9916	1.0011**	0.9933	0.9897***	0.9406**	0.9889	0.9947
	Drop Volatility factor	2.1121***	1.323***	1.6931***	1.5638**	**0.8077**	0.9571*	0.9968	1.8705***	1.7701***	1.1164***	1.0002	1.1238***	1.0383*	0.9849	0.9884	1.0456	1.1911***	0.9968*	0.9942	1.0756**
	Drop Macro Factor	2.063***	1.2558***	1.8644***	1.5697***	**0.7739****	0.9409***	1.0091*	1.7579***	1.4218***	0.9893***	1.0016***	1.0018	1.0572	1.0295	1.0113	1.0053	1.0099	1.0521*	0.9775	1.0076
	Drop Correlation index	2.0376***	1.2363***	1.6822***	1.5612**	**0.7667***	0.9291***	1.0039**	1.6987***	1.076***	1.0547***	1.0017	1.1186***	1.0254	0.9952*	0.9899	1.0743*	1.0587***	0.9467*	0.9868	0.9763***
XLV	All Variables	1.3759***	0.8621***	1.1091***	1.0236***	0.8885***	**0.8131*****	1.0000*	1.1223***	2.3288***	1.4233***	1.0000	0.9972	1.0017**	0.9998**	0.9999	1.0001	0.997	1.0176**	0.9999**	0.9998
	Drop Macro Variables	1.2741***	0.91***	1.0541***	1.0096***	0.8939***	**0.6793*****	1.0000*	1.1034***	1.0395***	1.0102***	1.0114	1.0026	1.0017**	0.9998**	0.9999	1.0001	0.997	1.0176**	0.9999**	0.9998
	Drop technical Variables	1.3406***	0.9177*	1.011**	0.9671**	0.8529***	**0.7487*****	1.0000*	1.0365**	1.1495***	1.0012**	0.9864*	1.0092	1.0017**	0.9998**	0.9999	1.0001	0.9970	1.0176**	0.9999**	0.9998
	Drop Volatility factor	1.3607***	0.9791***	0.9702**	1.0172**	**0.7275*****	0.8929***	1.0000*	1.1463***	1.1242***	1.1446***	0.9932	0.9995	1.0017**	0.9998**	0.9999	1.0001	0.9970	1.0176**	0.9999**	0.9998
	Drop Macro Factor	1.4677***	0.8734***	1.1071***	0.8639	**0.7806****	0.8591***	1.0000*	1.1205***	1.1533***	1.1805***	1.0100*	1.006**	1.0017**	0.9998**	0.9999	1.0001	0.9970	1.0176**	0.9999**	0.9998
	Drop Correlation index	1.3237***	0.8424***	0.9701***	1.0138***	0.8536*	**0.7947*****	1.0000*	1.111***	1.0608***	1.0251***	1.0011**	1.0068	1.0017**	0.9998**	0.9999	1.0001	0.9970	1.0176**	0.9999**	0.9998

*Notes: See notes in Table 3.2. We calcualte the one-month-ahead forecasting results by aggregating daily forecasts during each month, i.e., summing all daily forecasting results within the same month to generate monthly forecasts. Table 3.4 reports aggregate monthly relative MSFEs of all forecasting models, comparing with the random walk benchmark.

Table 3.5: Monthly Aggregate Relative MSFEs of All Forecasting Models (Recursive Window)*

		Linear	SVRrbf	SVRlin	SVRpoly	RanForest	boosting	lasso	ridge	NN1	NN2	NN3	NN4	hybrid1	hybrid2	hybrid3	hybrid4	hybrid5	hybrid6	hybrid7	hybrid8
SPY	All Variables	1.9681***	0.8833***	0.9901***	1.0111***	**0.8542**	0.8708	0.9709	1.3027***	1.2143***	1.1665***	1.001*	1.0025*	0.9742	1.0006	0.9917	0.9815	1.2746***	0.9868***	1.0184*	0.9769
	Drop Macro Variables	1.272***	0.9329***	0.9279***	0.8272***	0.8977	**0.7821****	0.9709	1.0239**	1.2364***	1.1399***	0.9701	0.9986	1.1602***	0.9988**	0.992	0.9877	1.1481***	1.0659***	1.0204**	0.9876
	Drop technical Variables	2.0196***	1.0371***	0.9898***	1.0002**	**0.6678*****	1.1757***	0.9709	1.3048***	0.979***	1.0239***	0.9999	1.4895***	0.9643	1.0265	1.0065	0.9651	1.0006***	0.9985***	0.9485	1.0378***
	Drop Volatility factor	2.1402***	**0.8054*****	1.0078***	0.8201***	0.9139	0.9217*	0.9626	1.2509***	2.4901***	1.2854***	1.0039**	1.0283*	1.1142***	0.9454	0.9808	0.9728	2.2817***	0.9384***	0.9836***	0.9907
	Drop Macro Factor	1.9558***	**0.7929*****	1.0047***	0.8779***	0.9228	0.7932***	1.0031**	1.221**	1.7081***	1.2742***	0.9989***	1.0029	1.0286**	1.0016	1.0007	1.0041***	0.9654**	1.0523**	0.9897	1.0035
	Drop Correlation index	2.5***	**0.8749*****	1.0022**	0.9772**	0.9392*	1.0503	0.9709	1.4531***	1.5625***	1.0748***	1.0135	0.9967	1.0652***	1.0759**	0.9926	0.989	1.0905***	1.0423***	0.9595	0.981
XLF	All Variables	3.4172***	1.1973***	1.2687***	1.2315***	**0.6482*****	0.9754	1.0101	1.9073***	1.328***	1.159***	1.098***	1.018	1.1792**	1.1414***	1.0191	1.0715***	1.1125	1.0738***	1.032*	1.0319
	Drop Macro Variables	1.6088***	1.067***	1.1449***	1.091***	**0.9687**	1.0436***	1.0154	1.5658***	1.1494***	1.0893***	1.082***	0.9522***	1.0553	0.9977	1.0125	1.133*	1.0966***	1.0357**	1.0669	1.0106**
	Drop technical Variables	2.1866***	1.0298**	1.0206**	1.0268***	**0.5503*****	0.9247**	0.9894	1.297***	0.9492**	1.0018**	1.0233	0.9953	1.0091	0.987	0.9995	0.9855*	0.9817***	0.9627**	0.9984	0.9923
	Drop Volatility factor	3.3743***	0.978***	1.3112***	1.088***	**0.7939***	0.8742**	0.9837	1.8959***	1.2773***	1.0765***	0.9932	0.9896**	1.0616**	1.0067	0.985*	1.0005*	1.1601***	1.064**	0.9912	1.0119
	Drop Macro Factor	3.3947***	1.1572***	1.3094***	1.2388***	**0.8363***	1.0022**	1.0303**	1.9457***	1.3884***	1.0384***	1.1048**	1.6947***	1.2049***	1.0367	1.0452**	1.1877**	1.2393***	1.2078***	1.0289	1.3555***
	Drop Correlation index	3.4324***	1.1755***	1.2549***	1.2602***	0.9109	**0.8559*****	1.0101	1.9036***	1.1143***	1.0365***	1.0335	1.0045	1.0371	1.0218	1.0644**	1.0084*	1.0671***	0.9779	1.0396	1.0156
XLK	All Variables	1.5220***	**0.8708****	1.0934***	0.9974	0.948	0.9317***	0.9992	1.2491***	1.1656***	1.1179***	1.0223	1.0013	0.9893	0.9976	0.9981	0.9999	0.9279*	0.9955	0.9938	0.9993*
	Drop Macro Variables	1.2517**	0.9799***	0.9743**	1.0062**	0.9818**	**0.9027*****	0.9992	1.22**	1.1047**	1.0603***	0.9656	1.0005*	1.0032*	0.9942	1.002	0.9995	0.965	0.9955	0.9932	1.0002**
	Drop technical Variables	1.2645***	0.8654**	0.9821	0.8782*	0.9343	**0.8222****	0.9992	1.0819***	1.0475***	1.0613	0.9955	0.9958	0.9948	1.0033	0.9995	0.9988	1.0295**	0.9884**	1.0082	1.0013
	Drop Volatility factor	1.5860***	0.9533***	1.0267***	0.9842**	0.9536	**0.9382*****	0.9980	1.2934***	1.2783***	1.2113***	1.0230***	1.0005**	1.0171*	0.9963	0.9988	0.9982	1.0246***	0.9861	0.9964	0.9993
	Drop Macro Factor	1.7095***	0.9769***	1.0286***	0.9313*	0.9523	**0.7848*****	1.001***	1.3137***	1.1496***	1.1982***	1.0255**	1.0066	1.0084***	1.0020	1.0019	1.0028	1.0180	1.0287	1.0045	1.0048
	Drop Correlation index	1.6708***	0.898**	1.0887	0.8555***	**0.6804*****	0.8291***	0.9992	1.2568***	1.1443***	0.9979	1.0614*	1.0489***	0.999	0.9969	1.0052**	0.9999	1.0279	0.9924	1.001	1.016*
XLY	All Variables	1.8408***	1.247**	1.4264***	1.2182*	**0.7092*****	1.0117	0.996	1.4409***	1.3534***	0.9925***	1.0879***	1.0005	1.0358	1.0055	1.0095*	0.9972	1.3947***	0.9633	0.9973	1.0014**
	Drop Macro Variables	1.4291*	1.1673***	1.3200	1.2383	**0.8173*****	0.9661	0.9971	1.4108*	1.2135***	1.0906***	1.0117	1.0017	1.0008	0.9978	0.9927	0.9974	0.9833	0.9909**	1.0214	0.9957
	Drop technical Variables	1.5498**	0.9578*	1.0905	0.9874	0.9136	**0.9025****	0.9958	1.1403**	1.0781***	1.0088	1.0577***	1.0019*	1.0134	1.0018	0.9843*	0.9937	1.029***	0.9831**	0.9915**	1.0009
	Drop Volatility factor	1.8591***	1.1862	1.4617***	1.2741*	**0.8657**	0.993	0.9886	1.5085***	1.1874***	1.0351***	0.9740	1.1356***	1.0255	0.9942	0.9883	1.0374	1.3065***	1.1439***	1.0038***	0.9725**
	Drop Macro Factor	1.9584***	1.2437***	1.4269***	1.2023*	**0.7152**	1.0108	1.0061	1.4936**	1.1139***	1.1534***	1.1574***	0.9982	1.0154	0.9985	1.0006	1.0027	1.0222***	0.9720	1.1087	1.0042*
	Drop Correlation index	1.8895***	1.2432***	1.2797**	1.1906*	**0.8684**	0.9070	0.9960	1.5130***	1.3423***	1.0424***	1.0063	1.1475***	0.9809	1.0275	1.0024	1.0159	1.1031**	1.0059	1.0163**	1.0321***
XLV	All Variables	1.2423***	**0.8312*****	0.9675*	0.9169***	0.9525	0.8542***	1.0072	1.0179***	1.7498***	1.1977***	1.0334**	1.0066	1.0072	1.0072	1.0072	1.0072	1.0072	1.0072	1.0072	1.0072
	Drop Macro Variables	1.1639***	0.9274***	0.9506	0.9449	0.9414***	**0.8815*****	1.0072	1.0673***	1.0983***	1.2340***	1.0010	1.0017	1.0072	1.0072	1.0072	1.0072	1.0072	1.0072	1.0072	1.0072
	Drop technical Variables	1.2367***	0.8732	0.9604*	**0.8502**	0.9637	0.8748***	1.0072	1.009*	1.0065***	1.0218***	0.9790	1.0077	1.0072	1.0072	1.0072	1.0072	1.0072	1.0072	1.0072	1.0072
	Drop Volatility factor	1.1876***	0.8643***	0.9080	0.8853	**0.7271*****	0.8821***	1.0072	1.009***	1.0662***	1.1857***	0.9798	1.0032	1.0072	1.0072	1.0072	1.0072	1.0072	1.0072	1.0072	1.0072
	Drop Macro Factor	1.3455***	**0.7981*****	0.9466	0.8646	0.8776**	0.9106	1.0072	1.0427***	1.0530***	1.4495***	0.9957	0.9985	1.0072	1.0072	1.0072	1.0072	1.0072	1.0072	1.0072	1.0072
	Drop Correlation index	1.2112***	**0.8075*****	0.8876	0.8921***	0.9425	0.8690	1.0072	1.0343***	1.1448***	1.0012**	1.0105	1.0123	1.0072	1.0072	1.0072	1.0072	1.0072	1.0072	1.0072	1.0072

*Notes: See notes in Table 3.4. Recursive window size 500.

Table 3.6: Directional Predictive Accuracy Rate Based on 1-Step-Ahead Daily Level Forecasting Results (Rolling Window)*

		RandomWalk	Linear	SVRrbf	SVRlin	SVRpoly	RanForest	boosting	lasso	ridge	NN1	NN2	NN3	NN4	hybrid1	hybrid2	hybrid3	hybrid4	hybrid5	hybrid6	hybrid7	hybrid8
SPY	All Variables	0.5303	0.5082	0.4899	0.5233	0.5228	0.5092	0.512	0.542	0.4951	0.4847	0.4974	0.4941	0.5369	0.5378	0.535	0.5369	0.5420	0.5416	0.5355	0.5420	**0.5444**
	Drop Macro Variables	0.5303	0.4941	0.5082	0.5129	0.5106	0.5228	0.4979	0.542	0.5021	0.4984	0.4852	0.5223	0.5317	0.5326	0.5425	0.5359	**0.5430**	0.5270	0.5326	0.5378	0.5383
	Drop technical Variables	0.5303	0.5106	0.5242	0.5294*	0.5355	0.5059	0.5049	**0.5420**	0.4998	0.5106	0.5190*	0.5209	0.5115	0.5364	0.5364	0.5378	0.5416	0.5326	0.5406	0.5355	0.5322
	Drop Volatility factor	0.5303	0.5031	0.5167	0.5195	0.5176	0.5298	0.5153	**0.5402**	0.4913	0.5040	0.4974	0.5078	0.5059	0.5373	**0.5402**	0.5345	0.5383	0.5359	0.5312	0.5392	0.5359
	Drop Macro Factor	0.5303	0.5035	0.4984	0.5214	0.5181	0.5195	0.5035	0.5303	0.5002	0.5012	0.5101	0.5031	0.5233	0.5317	0.5331	**0.5364**	0.5303	0.5355	0.5336	0.5355	0.5298
	Drop Correlation index	0.5303	0.5129	0.5218	0.5364**	0.5322	0.5171	0.5073	**0.5420**	0.4951	0.5049	0.4937	0.504	0.5275	0.5369	0.5373	0.5388	0.5406	0.5411	0.5359	0.5373	0.5402
XLF	All Variables	0.5045	0.5124	0.5012	0.5073	0.5101	0.4782	0.5106	0.5059	0.5007	0.4847	0.4951	0.4913	0.4814	0.5073	0.5031	0.5087	**0.5204**	0.5148	0.5139	0.5129	0.5092
	Drop Macro Variables	0.5045	0.4955	0.4838	**0.5134**	0.5016	0.4988	0.4927	0.5059	0.5007	0.4974	0.5110	0.4918	0.4927	0.5012	0.5059	0.5031	0.5110	0.5101	0.5157	0.5096	0.5106
	Drop technical Variables	0.5045	0.5134	0.5035	0.5049	0.5002	0.4819	0.4979	0.5054	0.5087	0.5031	0.5096	0.5073	0.4965	0.5148	0.5059	0.5045	0.5096	0.5049	**0.5228**	0.5082	0.5054
	Drop Volatility factor	0.5045	0.5162	0.4918	0.5016	0.4984	0.4890	0.5026	0.5049	0.4984	0.4941	0.4960	**0.5308***	0.5021	0.5082	0.5063	0.5087	0.5068	0.5162	0.5129	0.5124	0.5073
	Drop Macro Factor	0.5045	0.4998	0.4890	0.5106	0.5068	0.4852	0.4998	0.5040	0.4885	0.4984	0.5012	0.4922	0.4674	0.4998	0.5054	0.5059	0.5045	**0.5157**	0.5106	0.5134	0.5016
	Drop Correlation index	0.5045	0.5204	0.4937	0.5181	0.5068	0.5040	0.5059	0.5059	0.5035	0.5106	**0.5247**	0.4988	0.5021	0.5026	0.5092	0.5049	0.511	0.5143	0.5195	0.5016	0.5002
XLK	All Variables	0.5406	0.5237*	0.5279	0.5524***	0.5270	0.5200	0.4960	0.5434	0.5298	0.5002	0.5087	0.5171	0.5388	0.5402	0.5458	0.5425	0.5449	**0.5552****	0.543	0.5383	0.5420
	Drop Macro Variables	0.5406	0.5157	0.5186	0.5218	0.5228	0.5115	0.5087	0.5434	0.5233	0.5181	0.4937	0.5237	0.5411	0.5373	0.5416	0.5416	0.5444	0.5439	0.5420	**0.5463**	0.5439
	Drop technical Variables	0.5406	0.5355***	0.5364*	0.5345**	0.5251	0.5195	0.5073	0.5434	0.5171	0.5233	0.5059	0.5350	0.5378	0.5430	0.5416	0.5463	0.5420	0.5463	0.5378	**0.5481**	0.5383
	Drop Volatility factor	0.5406	0.5218	0.5265	0.5317**	0.5176	0.5139	0.5167	0.5434	0.5284*	0.5106	0.5124	0.5247	0.5397	0.5434	0.5406	0.5406	**0.5467**	0.5444	0.5449	0.5420	0.5388
	Drop Macro Factor	0.5406	0.5195	0.5359*	0.5383***	0.5341*	0.5294	0.5092	0.5406	0.5308	0.4927	0.5002	0.5383	0.5322	0.5458	0.5402	0.5434	0.543	0.5481	0.5449	0.5425	**0.5463**
	Drop Correlation index	0.5406	0.5261	0.5265	**0.5486*****	0.5204	0.5223	0.511	0.5434	0.5345*	0.5007	0.5120	0.5143	0.4927	0.5388	0.5425	0.5420	0.5430	**0.5486**	0.5420	0.5463	0.5420
XLY	All Variables	0.5355	0.5303**	0.5176	0.5303*	0.5289	0.5087	0.5096	0.5369	0.5148	0.5176	0.4829	0.5153	0.5355	0.5326	0.5284	0.5355	0.5383	0.5373	0.5308	**0.5392**	0.5341
	Drop Macro Variables	0.5355	0.5218	0.5153	0.519	0.5233	0.5209	0.5101	0.5369	0.5153	0.5073	0.5124	0.5082	0.5308	0.5341	0.5308	0.5373	**0.5392**	0.5294	0.5383	0.5303	0.5378
	Drop technical Variables	0.5355	0.5289**	0.5204	0.519	0.5157	0.5153	0.5167	0.5312	0.5129	0.4941	0.5054	0.5115	0.5373	0.5261	**0.5364**	0.5322	0.5331	0.5326	0.5279	0.5303	0.5317
	Drop Volatility factor	0.5355	0.5223	0.5190	0.5157	0.5312	0.5265	0.5162	**0.5373**	0.5195	0.4998	0.4857	0.5237	0.5002	0.5322	0.5341	**0.5373**	0.5308	0.5336	0.5359	0.535	0.5322
	Drop Macro Factor	0.5355	0.5223*	0.5214	0.5110	0.5218	0.5101	0.5120	0.5378	0.5063	0.4974	0.4979	0.5082	0.5317	0.5359	0.5378	0.5345	**0.5383**	0.5364	0.5364	0.5341	0.5364
	Drop Correlation index	0.5355	0.5326**	0.5218	0.5289	0.5242	0.5162	0.5233*	0.5369	0.5148	0.5101	0.5016	0.5171	0.5016	**0.5359**	0.5336	0.5303	0.5341	0.5270	0.5416	0.5303	0.5303
XLV	All Variables	0.5228	0.5082	0.4866	0.5082	0.4852	0.504	0.5045	0.5228	0.4852	0.4908	0.4861	**0.5359***	0.5148	0.5214	0.5228	0.5228	0.5228	0.5233	0.5223	0.5228	0.5228
	Drop Macro Variables	0.5228	0.4932	0.5035	0.4918	0.5016	0.5045	0.5012	0.5228	0.4941	0.5031	0.5063	0.5124	**0.5242**	0.5214	0.5228	0.5228	0.5228	0.5233	0.5223	0.5228	0.5228
	Drop technical Variables	0.5228	0.5031	0.504	0.5204	0.5068	0.5059	0.5087	0.5228	0.5021	0.5021	0.5059	**0.5279**	0.5186	0.5214	0.5228	0.5228	0.5228	0.5233	0.5223	0.5228	0.5228
	Drop Volatility factor	0.5228	0.5031	0.4998	0.5078	0.4974	0.5007	0.5031	0.5228	0.4904	0.5181	0.5134	0.5148	**0.5298**	0.5214	0.5228	0.5228	0.5228	0.5233	0.5223	0.5228	0.5228
	Drop Macro Factor	0.5228	0.5124	0.4922	0.5082	0.5016	0.5045	0.5162	0.5228	0.4899	0.4974	0.5049	0.5073	0.5195	0.5214	0.5228	0.5228	0.5228	**0.5233**	0.5223	0.5228	0.5228
	Drop Correlation index	0.5228	0.5082	0.5176	0.5148	0.489	0.5068	0.4960	0.5228	0.4838	0.4955	0.5054	0.5195	0.5176	0.5214	0.5228	0.5228	0.5228	**0.5233**	0.5223	0.5228	0.5228

*Notes: Table 3.6 reports the 1-step-head directional predictive accuracy rate (DPAR) of market sector ETFs With rolling window size 500, for the period 2009:6-2017:12. All DPARs are derived from level forecasting results. If the forecasted return is positive, then it is classified as an upward direction, otherwise as a downward direction. Entries in bold denote models with highest directional accuracy rate for a given forecasting target and predictors. Starred entries denote rejection of the null of no information about the direction of change forecasting, based on Pesaran and Timmermann (1992) (PT) test. Significance levels for the test are reported as $***p < 0.01$, $**p < 0.05$, and $*p < 0.1$, where p is the $p-$value corresponding to PT test statistics.

Table 3.7: Directional Predictive Accuracy Rate Based on 1-Step-Ahead Daily Level Forecasting Results (Recursive Window)*

		RandomWalk	Linear	SVRrbf	SVRlin	SVRpoly	RanForest	boosting	lasso	ridge	NN1	NN2	NN3	NN4	hybrid1	hybrid2	hybrid3	hybrid4	hybrid5	hybrid6	hybrid7	hybrid8
SPY	All Variables	0.4988	0.4951	0.4937	0.5242	0.5256	**0.5378**	0.5364*	0.5129	0.4908	0.4946	0.496	0.4852	0.4979	0.5106	0.5218	0.5129	0.5181	0.496	0.5228	0.5214	0.5171
	Drop Macro Variables	0.4988	0.4814	0.5068	0.5265	0.5012	**0.5289**	0.5045	0.5129	0.4946	0.5124*	0.4918	0.4861	0.4843	0.5040	0.5106	0.5181	0.5251	0.5073	0.5092	0.5021	0.5045
	Drop technical Variables	0.4988	0.5059	**0.5388****	0.5355	0.5350	0.5298**	0.5242*	0.5129	0.5082	0.5087	0.4725	0.5035	0.4969	0.5200	0.5214	0.5200	0.4998	0.5209	0.5106	0.5139	0.5171
	Drop Volatility factor	0.4988	0.4819	0.5139	0.5237	0.5181	**0.5331**	0.5326	0.5139	0.4852	0.5007	0.4965	0.511	0.4927	0.5167	0.5063	0.5186	0.5176	0.5298	0.5026	0.5176	0.5157
	Drop Macro Factor	0.4988	0.4937	0.4974	0.5195	0.5237	**0.5388**	0.5218*	0.4988	0.4890	0.4843	0.4955	0.5059	0.5087	0.5021	0.5035	0.5002	0.4984	0.5063	0.5031	0.5073	0.4993
	Drop Correlation index	0.4988	0.4955	0.5026	0.5171	0.5209	0.5153	0.5181	0.5129	0.4767	0.5012	0.5129	0.4730	0.4988	0.5186	0.5143	**0.5270**	0.5214	0.5092	0.5054	0.5190	0.5204
XLF	All Variables	0.4998	0.5021	0.5082	0.5016	0.5049	0.4791	**0.5148**	0.4974	0.4979	0.4852	0.5002	0.4908	0.4918	0.4998	0.5026	0.4965	0.4960	0.5012	0.5026	0.4979	0.5021
	Drop Macro Variables	0.4998	0.5026	0.4951	0.4984	0.5049	0.5035	**0.5134**	0.4974	0.5063	0.4937	0.5035	0.5054	0.4974	0.5007	0.4899	0.4998	0.4960	0.4993	0.5012	0.4871	0.4993
	Drop technical Variables	0.4998	0.5031	0.5087	0.4946	0.4974	0.4829	0.4965	0.4974	0.4904	**0.5068**	0.5026	0.4871	0.4946	0.5026	0.4974	0.5035	0.5007	0.4960	0.5059	0.5040	0.4979
	Drop Volatility factor	0.4998	0.5035	0.5078	0.5049	0.4969	0.5040	**0.5134**	0.4974	0.5002	0.4955	0.5016	0.4960	0.5031	0.5007	0.4988	0.5026	0.4988	0.4941	0.5063	0.4960	0.5026
	Drop Macro Factor	0.4998	0.4998	0.4908	0.5007	0.5031	0.4969	0.5082	0.504	0.5016	**0.5106**	0.4937	0.4984	0.4829	0.4932	0.5059	0.4922	0.4988	0.5002	0.4965	0.5016	0.5021
	Drop Correlation index	0.4998	0.5045	0.5101	0.5045	0.5115	0.5124	0.5078	0.4974	0.5035	**0.5181**	0.4922	0.5016	0.5092	0.4927	0.5016	0.4998	0.4993	0.5082	0.5115	0.4984	0.4965
XLK	All Variables	0.5369	0.5223	0.5256	0.5251	0.5303	0.5477	0.5265	0.5359	0.5218	0.5171	0.4899	0.5204	0.5326	0.5425	0.5430	0.5458	0.5364	0.5416	**0.5500**	0.5463	0.5350
	Drop Macro Variables	0.5369	0.4998	0.5265	0.5355	0.5420	0.5265	0.5228	0.5359	0.5031	0.5063	0.4979	0.5162	0.5289	0.5458	0.5383	0.5420	0.5369	0.5420	**0.5472**	0.5458	0.5364
	Drop technical Variables	0.5369	0.5237	0.5373	0.5406	0.5312	**0.5458**	0.5364	0.5359	0.5124	0.4998	0.5200	0.5242	0.5331	0.5406	0.5378	0.5350	0.5392	0.5449	0.5430	0.5439	0.5350
	Drop Volatility factor	0.5369	0.5294	0.5275	0.5233	0.5115	0.5463	0.5223	0.5359	0.5214	0.4937	0.4876	0.4984	0.5350	0.5392	0.5434	0.5411	0.5373	0.5434	0.5402	**0.5486**	0.5463
	Drop Macro Factor	0.5369	0.5016	0.5317	0.5115	0.5265	**0.5458**	0.4960	0.5369	0.5082	0.5298**	0.5026	0.4974	0.5350	0.5359	0.5359	0.5388	0.5388	0.5411	0.5449	0.5402	0.5383
	Drop Correlation index	0.5369	0.5214	0.527	0.5265	0.5242	0.5251	0.5031	0.5359	0.5157	0.4899	0.5063	0.5120	0.4941	0.5453	0.5359	0.5378	0.5430	0.5481	0.5449	**0.551**	0.5500
XLY	All Variables	0.5294	0.5195	0.5186	0.5200	0.5265	0.5162	0.5383	0.5294	0.5167	0.5031	0.5124	0.5167	0.5279	0.5265	0.5284	0.5279	0.5317	**0.5388**	0.5345	0.5326	0.5275
	Drop Macro Variables	0.5294	0.5303**	0.5228	0.5326	0.5345	0.5284**	0.5294	0.5298	0.5209	0.4941	0.4974	0.4955	0.5265	0.5270	0.5326	0.5308	0.5294	**0.5397**	0.5341	0.5345	0.5275
	Drop technical Variables	0.5294	0.5087	0.5298	0.5298	0.5298	0.5341	0.5336	0.5294	0.5035	0.5007	0.4988	0.5284	0.5190	0.5350	0.5265	**0.5350**	0.5303	0.5308	0.5284	**0.5350**	0.5312
	Drop Volatility factor	0.5294	0.5204	0.52	0.5223	0.5223	0.5336	0.535	0.5294	0.5124	0.5209*	0.5026	0.5078	0.4908	0.5322	0.5294	0.5303	0.535	0.5303	0.5308	0.527	**0.5406**
	Drop Macro Factor	0.5294	0.5162	0.5096	0.5153	0.5223	0.5261	0.5373	0.5294	0.5096	0.5031	0.5031	0.5026	0.5233	0.5364	0.5359	0.5308	0.5275	0.5350	**0.5392**	0.5364	0.5303
	Drop Correlation index	0.5294	0.5317**	0.5251	0.5284	0.5317	0.5312	0.5256	0.5294	0.5148	0.5162	0.5261**	0.5181	0.5129	0.5308	0.5242	0.5308	0.5312	**0.5383**	0.5289	0.5317	0.5284
XLV	All Variables	0.5228	0.5045	0.5045	0.5035	0.4955	**0.5275**	0.5261	0.5186	0.5063	0.4969	0.4918	0.5002	0.5115	0.5186	0.5186	0.5186	0.5186	0.5186	0.5186	0.5186	0.5186
	Drop Macro Variables	**0.5228**	0.4857	0.5096	0.4984	0.5040	0.5087	0.5021	0.5186	0.4904	0.4984	0.5101	0.5040	0.5106	0.5186	0.5186	0.5186	0.5186	0.5186	0.5186	0.5186	0.5186
	Drop technical Variables	0.5228	0.5007	0.511	0.5186	0.5242	**0.5247**	0.5214	0.5186	0.5171	0.5035	0.4984	0.5045	0.512	0.5186	0.5186	0.5186	0.5186	0.5186	0.5186	0.5186	0.5186
	Drop Volatility factor	0.5228	0.5026	0.512	0.5134	0.5139	0.5049	0.5214	0.5186	0.504	0.5186	0.5059	**0.5256**	0.5171	0.5186	0.5186	0.5186	0.5186	0.5186	0.5186	0.5186	0.5186
	Drop Macro Factor	0.5228	0.4922	0.5209	0.5101	0.5101	**0.5233**	**0.5233**	0.5186	0.4974	0.5115	0.4927	0.519	0.52	0.5186	0.5186	0.5186	0.5186	0.5186	0.5186	0.5186	0.5186
	Drop Correlation index	0.5228	0.5153	0.5110	**0.5237**	0.5031	0.5181	0.5214	0.5186	0.5087	0.5087	0.5026	0.5143	0.5167	0.5186	0.5186	0.5186	0.5186	0.5186	0.5186	0.5186	0.5186

*Notes: See notes to Table 3.6. Recursive window size 500.

Table 3.8: Directional Predictive Accuracy Rate Based on Monthly Aggregate Level Forecasting Results (Rolling Window)*

		RandomWalk	Linear	SVRrbf	SVRlin	SVRpoly	RanForest	boosting	lasso	ridge	NN1	NN2	NN3	NN4	hybrid1	hybrid2	hybrid3	hybrid4	hybrid5	hybrid6	hybrid7	hybrid8
SPY	All Variables	0.5922	0.6117***	0.7184***	0.6117**	0.6214**	**0.7573*** **	0.6408***	0.6019	0.5825*	0.5049	0.5534	0.6019	0.5922	0.6019	0.6117	0.5922	0.6117	0.5922	0.5825	0.5922	0.6117
	Drop Macro Variables	0.5922	0.6117**	0.6602***	0.6505***	0.6019*	0.6893***	**0.7476*** **	0.6019	0.5825	0.5825*	0.5049	0.6117	0.5922	0.6214	0.6117	0.5922	0.6117	0.5922	0.5534	0.6019	0.6019
	Drop technical Variables	0.5922	0.5922**	0.6796***	0.6699***	0.6796***	**0.8058*** **	0.7573***	0.6019	0.6117**	0.5534	0.5243	0.5825	0.4854	0.6117	0.6019	0.6214	0.6117	0.6019	0.6408*	0.5922	0.5728
	Drop Volatility factor	0.5922	0.6117***	0.6602***	0.6214**	0.6311**	**0.7476*** **	0.6602***	0.6117	0.6019**	0.4466	0.5437	0.6214	0.5534	0.6214	0.6019	0.5825	0.6117	0.6019	0.5825	0.6019	0.5922
	Drop Macro Factor	0.5922	0.5922**	0.7379***	0.6602***	0.6602***	0.6796***	**0.7379*** **	0.5922	0.5728	0.5243	0.5146	0.6019	0.6019	0.6019	0.5922	0.5922	0.6019	0.6117	0.5825	0.5922	0.5922
	Drop Correlation index	0.5922	0.6117***	0.6408***	0.5922	0.6019	**0.7864*** **	0.767***	0.6019	0.5534	0.5922	0.5631	0.6019	0.6019	0.6117	0.6117	0.6117	0.6019	0.5922	0.6019	0.5825	0.6019
XLF	All Variables	0.6019	0.6311***	0.6505***	0.5922**	0.534	**0.835*** **	0.7087***	0.6019	0.5825*	0.5437	0.6019*	0.6311**	0.5922	0.6214	0.6019	0.6019	0.5922	0.6214	0.6408*	0.5534	0.6311*
	Drop Macro Variables	0.6019	0.5534	0.6505***	0.5728	**0.6602*** **	0.6408*	**0.6602*** **	0.6019	0.5243	0.5922*	0.5534	0.5534	0.5146	0.6311*	0.6505**	0.5922	0.5825	0.6019	0.6214	0.5922	0.6019
	Drop technical Variables	0.6019	0.5825**	0.699***	0.5825**	0.5825**	**0.7864*** **	0.5922*	0.6019	0.6019**	0.6214**	0.5825	0.6214	0.5534	0.6408**	0.6214	0.6117	0.5728	0.6505**	0.6699***	0.6117	0.6214
	Drop Volatility factor	0.6019	0.6019**	0.6214*	0.5631	0.5922	**0.7573*** **	0.6699***	0.6019	0.5534	0.5922	0.6408**	0.5728	0.6214**	0.6214*	0.5825	0.5922	0.6214*	0.6117	0.6214	0.6019	0.5825
	Drop Macro Factor	0.6019	0.5922**	0.6311**	0.5631	0.5825	**0.7767*** **	0.6893***	0.6117	0.5534	0.5437	0.5825	0.6117	0.4951	0.6602**	0.6408*	0.6117	0.6505**	0.6505**	0.6408*	0.5922	0.6117
	Drop Correlation index	0.6019	0.5728*	0.6311**	0.6019**	0.5534	**0.7670*** **	0.6796***	0.6019	0.6019**	0.5049	0.5825	0.5534	0.5631	0.6505**	0.6214*	0.6214*	0.6214	0.6408**	0.6311*	0.6019	0.6019
XLK	All Variables	0.6117	0.6311***	0.6699***	0.6699***	0.6699***	0.7087***	**0.7184*** **	0.6311	0.6408***	0.6117*	0.5728	0.6505	0.6117	0.6214	0.6214	0.6408	0.6311	0.6408	0.6699*	0.6505	0.6311
	Drop Macro Variables	0.6117	0.5534	0.6408*	0.6214	0.6311*	0.6408*	**0.6990*** **	0.6311	0.5631	0.5534	0.5728	0.6602*	0.6214	0.6311	0.6214	0.6311	0.6311	0.6505	0.6408	0.6408	0.6311
	Drop technical Variables	0.6117	0.6214**	0.6311	0.699***	0.6602***	0.6990***	**0.7184*** **	0.6311	0.5728*	0.5825	0.6311	0.6311	0.6311	0.6311	0.6311	0.6311	0.6311	0.6408	0.6408	0.6117	0.6311
	Drop Volatility factor	0.6117	0.5922**	0.6214*	0.6311***	0.6311***	**0.7184*** **	**0.7184*** **	0.6311	0.5825*	0.5243	0.6214	0.6408	0.6214	0.6311	0.6311	0.6311	0.6311	0.6311	0.6408	0.6408	0.6311
	Drop Macro Factor	0.6117	0.6602***	0.6505**	**0.7184*** **	0.6117	0.699***	0.6408***	0.6214	0.6311*	0.534	0.5146	0.6505	0.5825	0.6214	0.6117	0.6117	0.6311	0.6602	0.6408	0.6311	0.6602
	Drop Correlation index	0.6117	0.6214**	0.6602**	0.6699***	0.6602***	**0.7476*** **	0.6893***	0.6311	0.6311**	0.5728	0.6019	0.5728	0.5825	0.6117	0.6311	0.6214	0.6214	0.6505	0.6602	0.6408	0.6214
XLY	All Variables	0.5922	0.534	0.5534	0.5728	0.534	**0.7573*** **	0.6408***	0.5728	0.534	0.5825	0.5728	0.5631	0.5922	0.5825	0.5922	0.5922	0.5728	0.6117	0.6019	0.5922	0.5922
	Drop Macro Variables	0.5922	0.5243	0.5534	0.6019	0.5825	**0.7184*** **	0.7087***	0.5728	0.5243	0.5922	0.5825	0.6019	0.5922	0.5922	0.5825	0.5825	0.5728	0.5922	0.5825	0.5922	0.5728
	Drop technical Variables	0.5922	0.5728	0.5534	0.5631	0.5534	**0.6893*** **	0.6796***	0.5825	0.5631	0.534	0.5825	0.5049	0.5922	0.5922	0.6117	0.5922	0.5825	0.6019	0.6117	0.6019	0.5825
	Drop Volatility factor	0.5922	0.5437	0.5728	0.534	0.5243	**0.7087*** **	0.6699***	0.5728	0.5049	0.5534	0.5728	0.5922	0.5728	0.5922	0.5922	0.5922	0.5922	0.6019	0.6019	0.5728	0.6311
	Drop Macro Factor	0.5922	0.5922**	0.5437	0.5922*	0.5631	**0.7379*** **	0.6408***	0.5922	0.4951	0.5534	0.5631	0.6019	0.5922	0.5631	0.5922	0.5728	0.5825	0.5728	0.5825	0.5922	0.5825
	Drop Correlation index	0.5922	0.5146	0.5437	0.5437	0.5437	**0.7087*** **	0.6893***	0.5728	0.5243	0.5825	0.5534	0.5728	0.5534	0.5922	0.6019	0.5728	0.5728	0.5825	0.6311	0.6019	0.5825
XLV	All Variables	0.6214	0.6505***	0.6699**	0.6699***	0.6505**	**0.7670*** **	0.7282***	0.6214	0.6311**	0.466	0.5631	0.6505	0.6505	0.6214	0.6214	0.6214	0.6214	0.6214	0.6117	0.6214	0.6214
	Drop Macro Variables	0.6214	0.5243	0.6505	0.6214	0.6602**	**0.6990*** **	0.6602**	0.6214	0.5922	0.6699***	0.6214	0.5631	0.6408	0.6214	0.6214	0.6214	0.6214	0.6214	0.6117	0.6214	0.6214
	Drop technical Variables	0.6214	0.6214*	0.6699*	0.6505*	0.6311	**0.7573*** **	0.7476***	0.6214	0.6602***	0.6311*	0.6408*	0.6505	0.6408	0.6214	0.6214	0.6214	0.6214	0.6214	0.6117	0.6214	0.6214
	Drop Volatility factor	0.6214	0.6311**	0.6311	0.6214	0.5437	**0.7476*** **	0.7282***	0.6214	0.6408**	0.5534	0.5437	0.6214	0.6505	0.6214	0.6214	0.6214	0.6214	0.6214	0.6117	0.6214	0.6214
	Drop Macro Factor	0.6214	0.6117***	0.699***	0.6214*	0.5631	**0.7573*** **	0.7282***	0.6214	0.6214**	0.5631	0.5825*	0.6214	0.6214	0.6214	0.6214	0.6214	0.6214	0.6214	0.6117	0.6214	0.6214
	Drop Correlation index	0.6214	0.6311**	0.699***	0.6408*	0.6311**	**0.7573*** **	0.7087***	0.6214	0.6311**	0.5825	0.5825	0.6117	0.6019	0.6214	0.6214	0.6214	0.6214	0.6214	0.6117	0.6214	0.6214

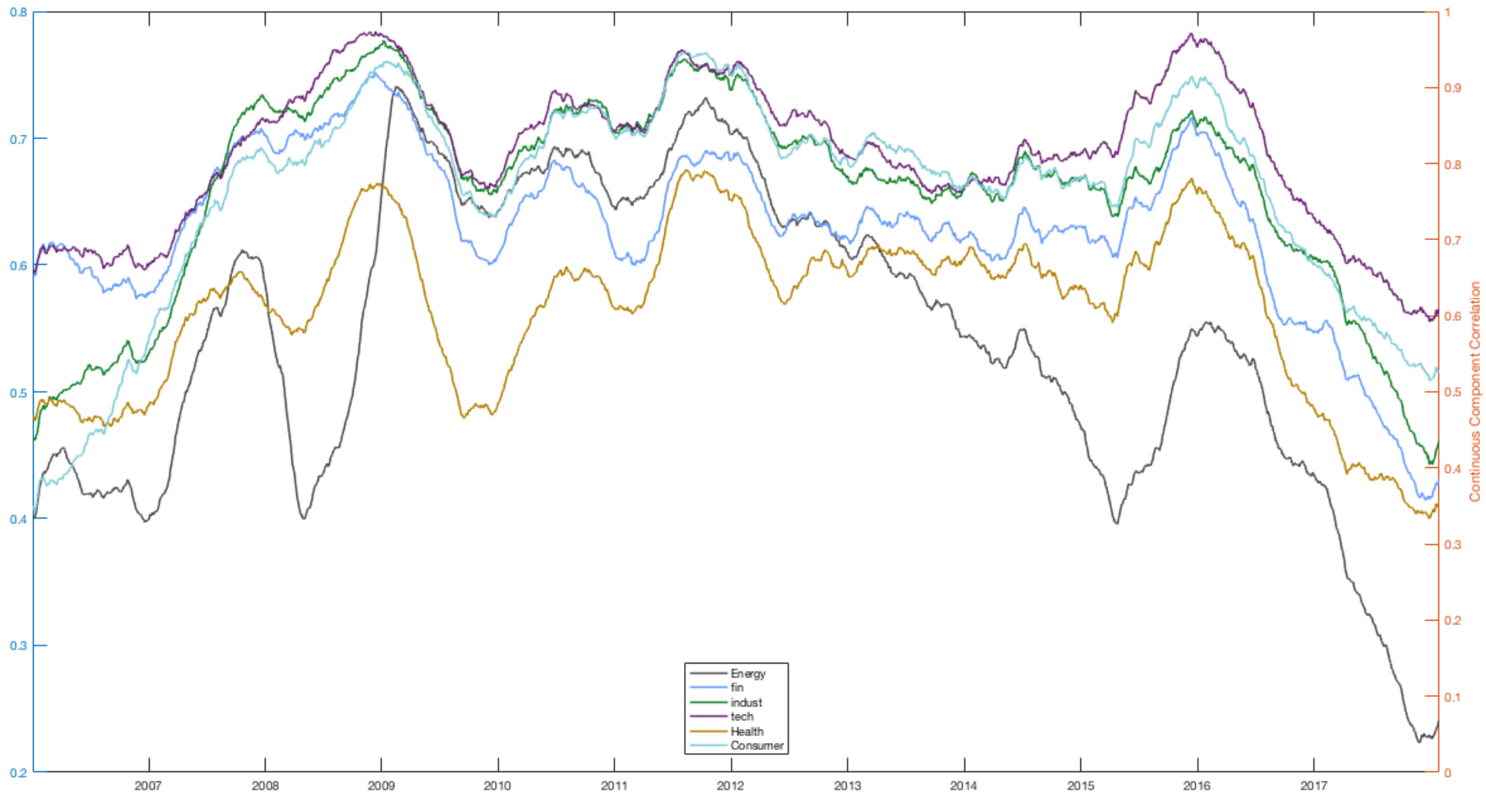
*Notes: See notes to Table 3.6. We calcualte the one-month-ahead forecasting results by aggregating daily forecasts during each month, i.e., summing all daily forecasting results within the same month to generate monthly forecasts. All DPARs are derived from monthly aggregate results. If the monthly aggregate return is positive, then it is classified as an upward direction, otherwise as a downward direction.

Table 3.9: Directional Predictive Accuracy Rate Based on Monthly Aggregate Level Forecasting Results (Recursive Window)*

		RandomWalk	Linear	SVRrbf	SVRlin	SVRpoly	RanForest	boosting	lasso	ridge	NN1	NN2	NN3	NN4	hybrid1	hybrid2	hybrid3	hybrid4	hybrid5	hybrid6	hybrid7	hybrid8
SPY	All Variables	0.5146	0.4951	**0.6893*** **	0.6019	0.6602***	0.6796***	0.6699***	0.5437	0.5728**	0.4757	0.3883	0.5728**	0.5243	0.5728	0.5243	0.5922	0.5631	0.4951	0.5146	0.5049	0.5437
	Drop Macro Variables	0.5146	0.5049	0.6505***	0.6117	0.5922	0.6408**	**0.6796*** **	0.5437	0.5340	0.4951	0.5049	0.5340	0.5049	0.5340	0.5534	0.5146	0.5631	0.5534	0.4951	0.5049	0.5243
	Drop technical Variables	0.5146	0.5146	0.6117*	0.6311*	0.6117	**0.8155*** **	0.6699***	0.5437	0.5922***	0.5825**	0.5049	0.4563	0.4757	0.5825	0.5534	0.5534	0.5728	0.5437	0.4951	0.534	0.5146
	Drop Volatility factor	0.5146	0.4660	**0.6699*** **	0.6311**	0.6505***	0.6408**	**0.6699*** **	0.5437	0.5534**	0.4369	0.4563	0.4757	0.5437	0.5049	0.5049	0.5437	0.5534	0.6019*	0.5728	0.5340	0.5631
	Drop Macro Factor	0.5146	0.4757	**0.7184*** **	0.5922	0.6505**	0.6019	0.6893***	0.5146	0.5728**	0.5437*	0.4951	0.5243	0.5146	0.5146	0.5243	0.5049	0.5146	0.5243	0.5146	0.5146	0.5146
	Drop Correlation index	0.5146	0.4563	**0.6602*** **	0.5728	0.6019	0.6214	0.6408**	0.5437	0.4854	0.4757	0.5437	0.5146	0.5146	0.4757	0.5631	0.5728	0.5534	0.5437	0.5146	0.5243	0.5728
XLF	All Variables	0.5243	0.5243	0.4951	0.5340	0.5437	**0.8058*** **	0.6019	0.5243	0.4466	0.5243	0.6019***	0.5049	0.5728	0.5437	0.5340	0.534	0.5728	0.5146	0.5340	0.5437	0.5437
	Drop Macro Variables	0.5243	0.5049	0.5825	0.5922	**0.6505** **	0.6408***	0.6505***	0.5243	0.4466	0.5049	0.4466	0.4466	0.5728*	0.5340	0.5146	0.5243	0.5049	0.5146	0.5146	0.4951	0.5437
	Drop technical Variables	0.5243	0.5437	0.5631	0.5340	0.5243	**0.7767*** **	0.5243	0.5243	0.534	0.4563	0.4563	0.5146	0.4854	0.5243	0.5631	0.5340	0.5534	0.5728	0.5146	0.5437	0.534
	Drop Volatility factor	0.5243	0.5243	0.6408***	0.5631	0.6602***	**0.699*** **	0.6019	0.5243	0.4369	0.4757	0.4466	0.534	0.5631	0.5437	0.5340	0.534	0.5146	0.5243	0.5243	0.5340	0.5243
	Drop Macro Factor	0.5243	0.5049	0.4563	0.534	0.5631	**0.6893*** **	0.5728	0.5340	0.4272	0.4951	0.5437	0.4563	0.4660	0.5146	0.5437	0.5146	0.5049	0.4854	0.5243	0.5437	0.5049
	Drop Correlation index	0.5243	0.4854	0.5146	0.5243	0.5631*	**0.5922**	0.5825	0.5243	0.4563	0.4854	0.4466	0.5340	0.5922**	0.5437	0.5534	0.5534	0.5146	0.5146	0.5243	0.5534	0.5437
XLK	All Variables	0.6117	0.5437	**0.6699*** **	0.5922	0.6214	0.6602	0.6602**	0.6311	0.4660	0.6019*	0.5340	0.6019	0.6117	0.6408	0.6311	0.6311	0.6311	0.6505	0.6214	0.5922	0.6408
	Drop Macro Variables	0.6117	0.4660	0.6505	0.6602	**0.6699***	0.6505	0.699***	0.6311	0.4272	0.5534	0.4563	0.6408	0.6214	0.6214	0.6117	0.6117	0.6311	0.6311	0.6408	0.6117	0.6214
	Drop technical Variables	0.6117	0.5631	0.6408*	0.6214	0.6311	**0.6893*** **	**0.6893*** **	0.6311	0.5437	0.5922*	0.5243	0.6019	0.6019	0.6311	0.6214	0.6408	0.6214	0.6505	0.6408	0.6214	0.6019
	Drop Volatility factor	0.6117	0.5243	0.6214*	0.5728	0.5728	**0.6699***	0.6117	0.6311	0.4563	0.4369	0.466	0.5922	0.6214	0.6311	0.6408	0.6311	0.6311	0.6408	0.6117	0.6117	0.6408
	Drop Macro Factor	0.6117	0.5049	0.6796***	0.5534	0.6117	0.6699*	**0.6893*** **	0.6117	0.4563	0.5825	0.5825	0.6214	0.5631	0.6117	0.6117	0.6117	0.6214	0.6408	0.6311	0.6117	0.6214
	Drop Correlation index	0.6117	0.4757	0.6505**	0.6117	0.6699***	**0.7670*** **	0.6796***	0.6311	0.4951	0.4854	0.5631	0.5049	0.6408**	0.6408	0.6214	0.6311	0.6408	0.6019	0.6699*	0.6505	0.6408
XLY	All Variables	0.5825	0.5437	0.5243	0.5534	0.5825	**0.767*** **	0.6019	0.5728	0.5049	0.5437	0.5049	0.5340	0.5922	0.5728	0.5631	0.5825	0.5728	0.5728	0.6117	0.5631	0.5631
	Drop Macro Variables	0.5825	0.4369	0.5631	0.5437	0.5534	**0.699*** **	0.6408*	0.5728	0.4175	0.5340	0.5146	0.5534	0.5728	0.5728	0.5631	0.5728	0.5728	0.6117	0.6019	0.5728	0.5631
	Drop technical Variables	0.5825	0.5922*	0.5631	0.5825	0.6117	**0.6505***	0.6214	0.5728	0.5631	0.5340	0.5049	0.5825	0.5631	0.5825	0.5825	0.5728	0.5728	0.5922	0.6214	0.6019	0.5631
	Drop Volatility factor	0.5825	0.5631	0.5146	0.534	0.5534	**0.6311**	0.6019	0.5728	0.4951	0.5631	0.5243	0.6019	0.5049	0.5825	0.5728	0.5534	0.5825	0.5825	0.5728	0.5534	0.5922
	Drop Macro Factor	0.5825	0.5049	0.5049	0.6019	0.5631	**0.7379*** **	0.5534	0.5728	0.5146	0.4563	0.4466	0.4757	0.5728	0.5922	0.5922	0.5825	0.5728	0.6019	0.6019	0.5922	0.5728
	Drop Correlation index	0.5825	0.4854	0.5631	**0.6311***	0.5631	**0.6311**	0.6214	0.5728	0.5049	0.4466	0.5437	0.5631	0.5146	0.5728	0.5534	0.5728	0.5825	0.5825	0.5825	0.5631	0.5631
XLV	All Variables	0.6214	0.6019	0.7087***	0.6408**	0.6602**	**0.7282*** **	0.7087**	0.6117	0.5922	0.5243	0.534	0.6117	0.6019	0.6117	0.6117	0.6117	0.6117	0.6117	0.6117	0.6117	0.6117
	Drop Macro Variables	0.6214	0.5146	0.6505	0.6019	0.5922	0.6602	**0.6699**	0.6117	0.5437	0.5534	0.5146	0.6214	0.6019	0.6117	0.6117	0.6117	0.6117	0.6117	0.6117	0.6117	0.6117
	Drop technical Variables	0.6214	0.5728	0.6408*	0.6602**	0.6311	**0.7087** **	0.699*	0.6117	0.6408**	0.5340	0.5728	0.5922	0.5728	0.6117	0.6117	0.6117	0.6117	0.6117	0.6117	0.6117	0.6117
	Drop Volatility factor	0.6214	0.5922	0.6699**	0.6214	0.5922	**0.7573*** **	0.7087**	0.6117	0.6117*	0.5825	0.6311*	0.6602*	0.6311	0.6117	0.6117	0.6117	0.6117	0.6117	0.6117	0.6117	0.6117
	Drop Macro Factor	0.6214	0.5437**	0.7282***	0.6214**	0.6019	**0.7282*** **	0.7184***	0.6117	0.6214***	0.5631	0.4660	0.5922	0.6117	0.6117	0.6117	0.6117	0.6117	0.6117	0.6117	0.6117	0.6117
	Drop Correlation index	0.6214	0.6408*	0.6505**	0.6796***	0.6602**	0.6990*	**0.7087** **	0.6117	0.6214	0.5146	0.5631	0.6214	0.6117	0.6117	0.6117	0.6117	0.6117	0.6117	0.6117	0.6117	0.6117

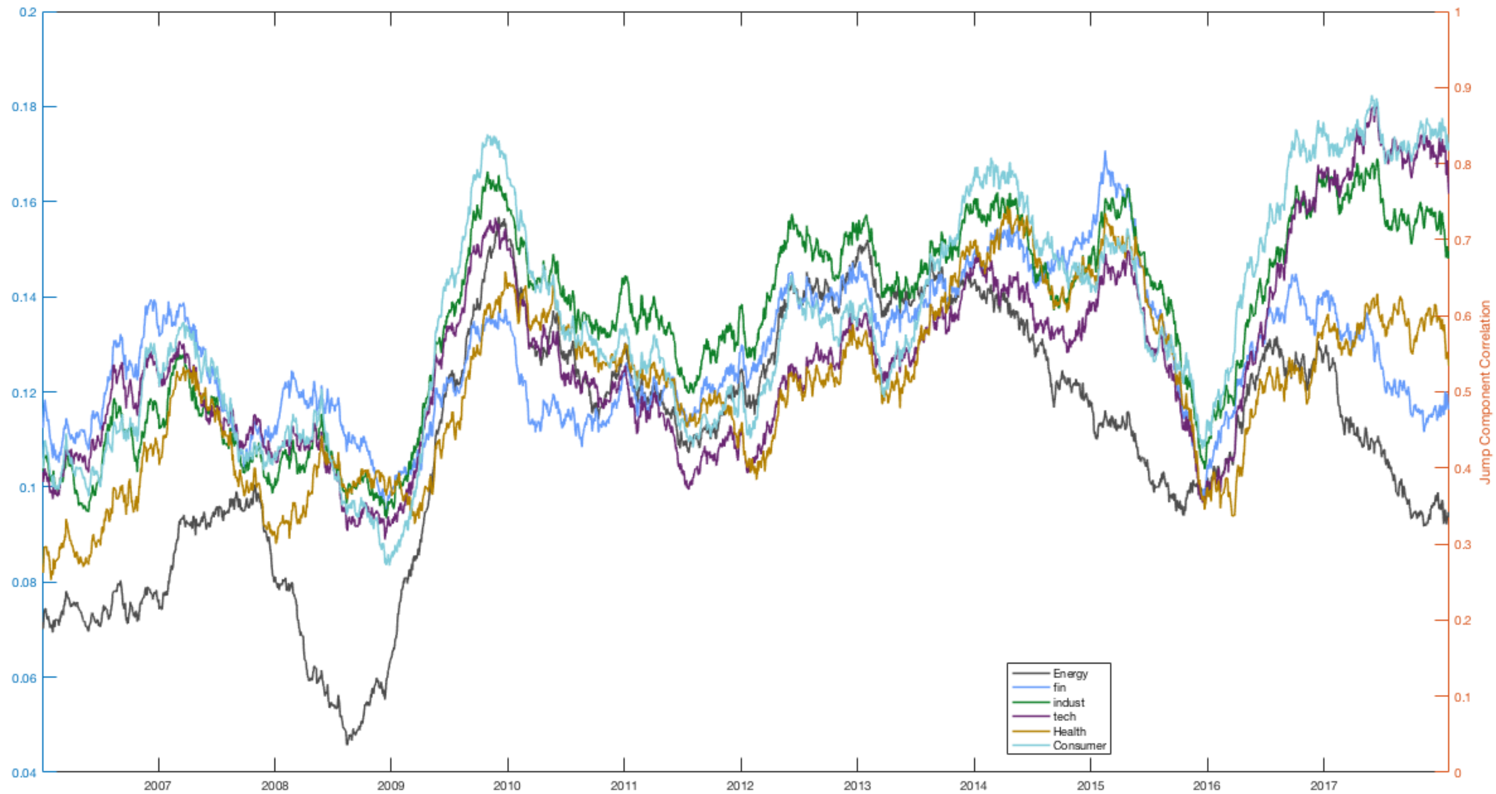
*Notes: See notes to Table 3.8. Recursive window size 500.

Figure 3.1: Sector Continuous Component Correlation Index with S&P 500 Market*



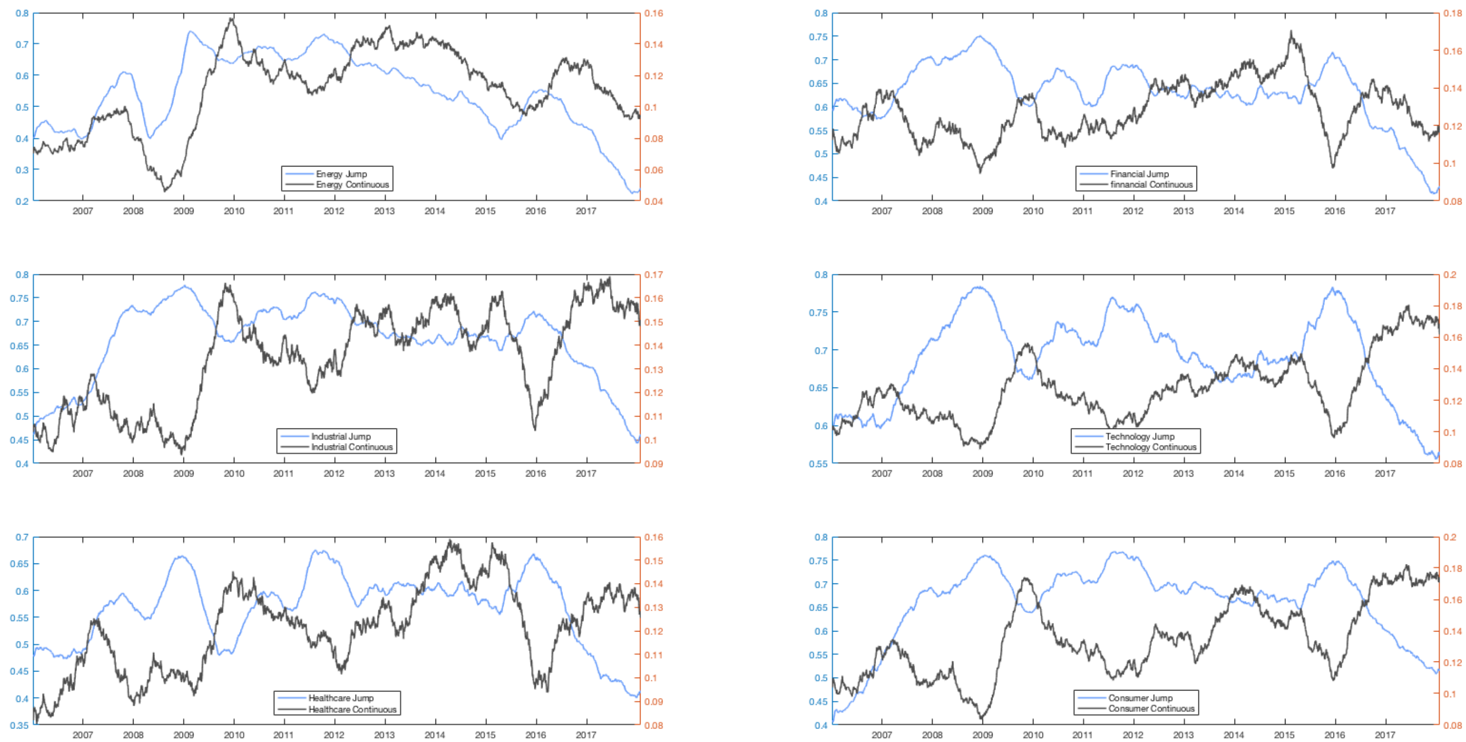
*Notes: This figure depicts the continuous component correlation indices of energy sector (XLE) and S&P 500(SPY), finance sector (XLF) and S&P 500(SPY), industrial sector (XLI) and S&P 500(SPY), technology sector (XLK) and S&P 500(SPY), health care sector (XLV) and S&P 500(SPY), and consumer discretionary sector (XLY) and S&P 500(SPY) from 2006:01 - 2017:12. All correlation indices are measured within [0,1] scale.

Figure 3.2: Sector Jump Component Correlation Index with S&P 500 Market*



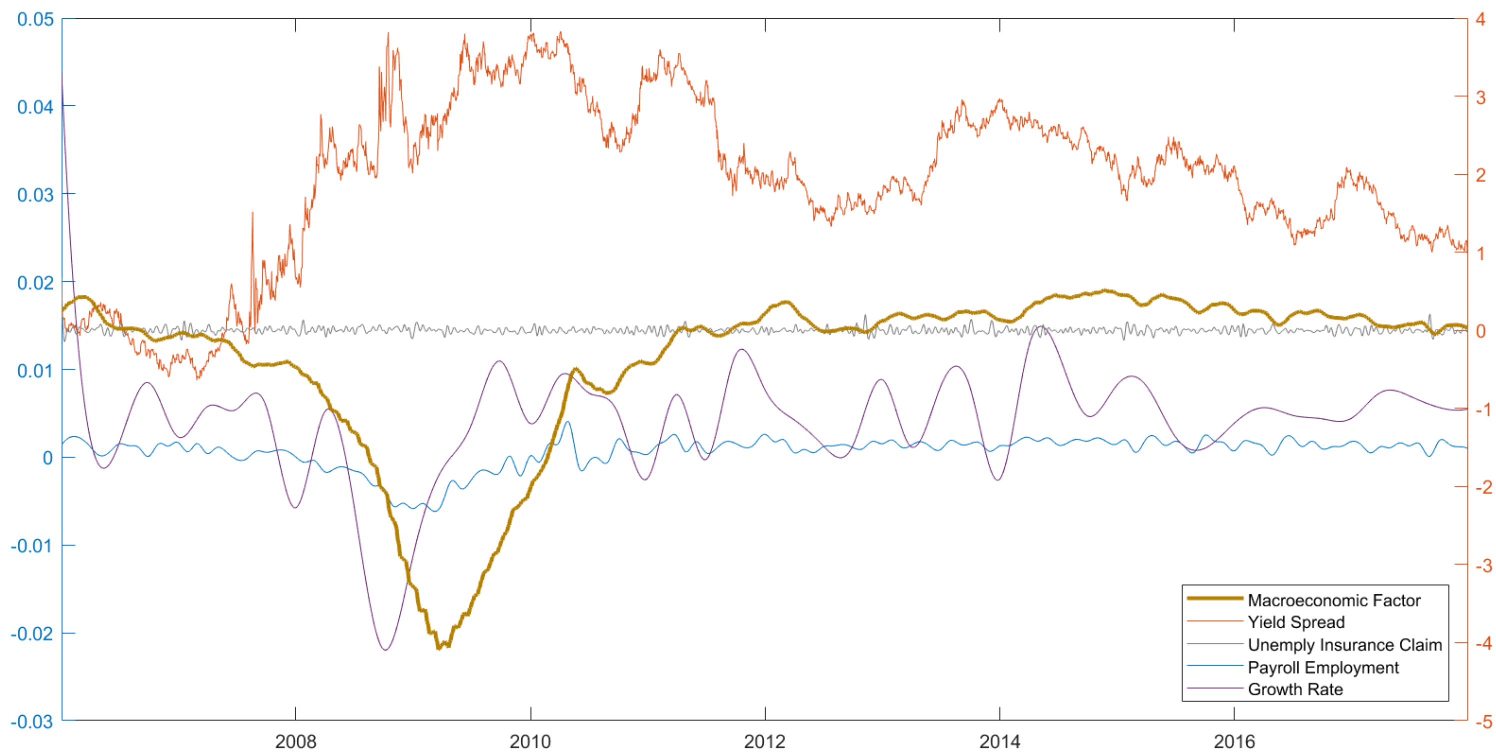
*Notes: See notes in Figure 3.1. Figure 3.2 depicts the jump component correlation indices of energy sector (XLE) and S&P 500(SPY), finance sector (XLF) and S&P 500(SPY), industrial sector (XLI) and S&P 500(SPY), technology sector (XLK) and S&P 500(SPY), health care sector (XLV) and S&P 500(SPY), and consumer discretionary sector (XLY) and S&P 500(SPY) from 2006:01 - 2017:12.

Figure 3.3: Comparison between jump part correlation index and continuous part correlation index*



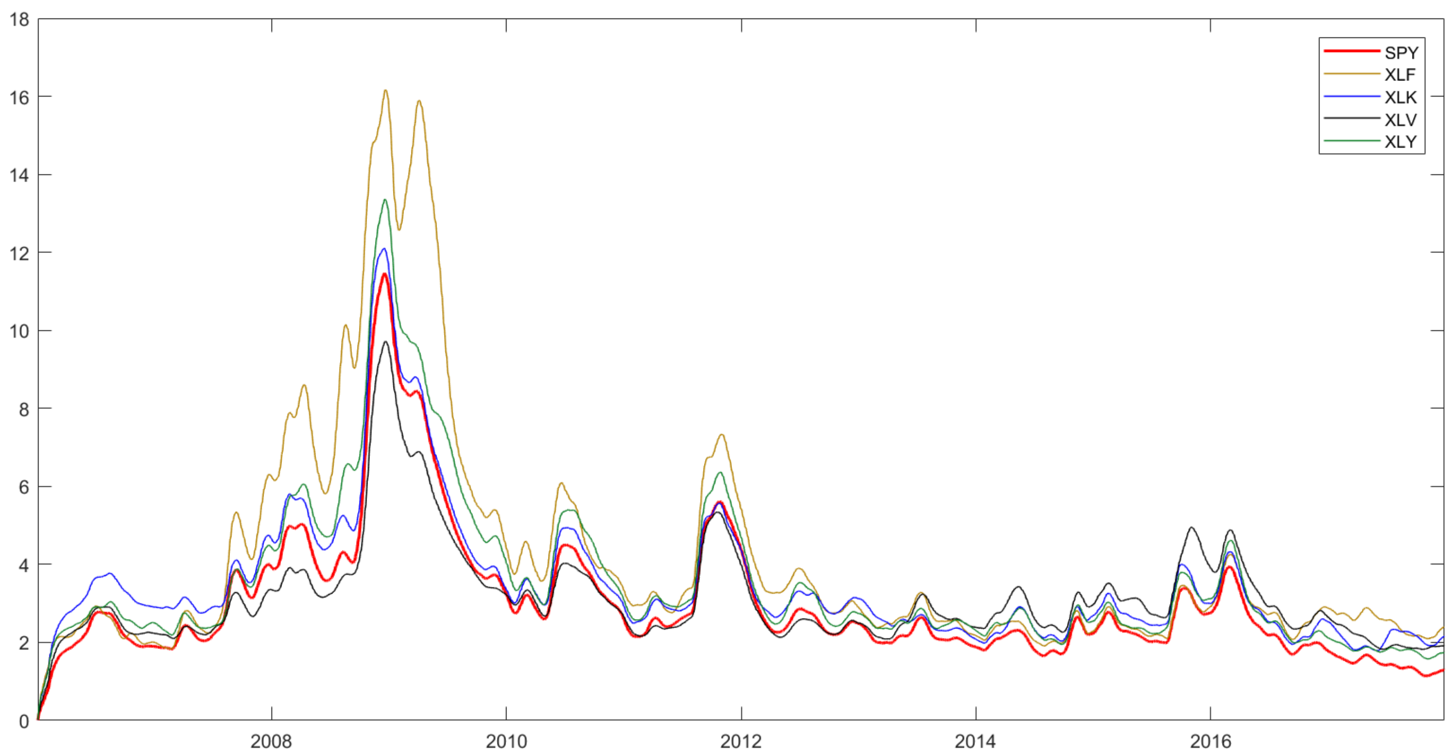
*Notes: This figure compares the jump part correlation index with continuous part correlation index of energy sector (XLE) and S&P500(SPY), finance sector (XLF) and S&P500(SPY), industrial sector (XLI) and S&P500(SPY), technology sector (XLK) and S&P500(SPY), health care sector (XLV) and S&P500(SPY), and consumer discretionary sector (XLY) and S&P500(SPY) from 2006:01 - 2017:12.

Figure 3.4: Macroeconomic Factor and Macroeconomic Series*



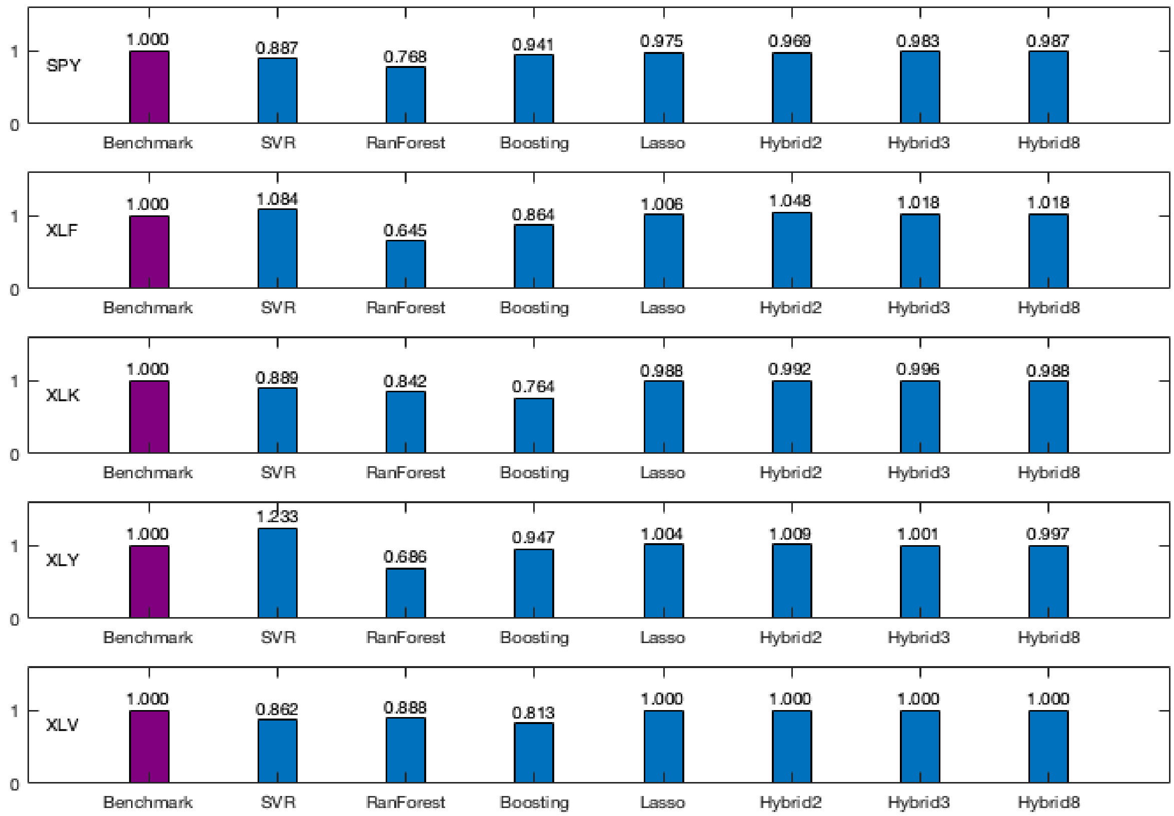
*Notes: We plot logged differenced value of original unemployment insurance claim, payroll employment and growth rate data. Y-axis on the right-hand side corresponds to the macroeconomic risk factor (MF_t^{mac}), yield spread and unemployment insurance claim. Y-axis on the right-hand side is for payroll employment and growth rate.

Figure 3.5: Volatility Factors*



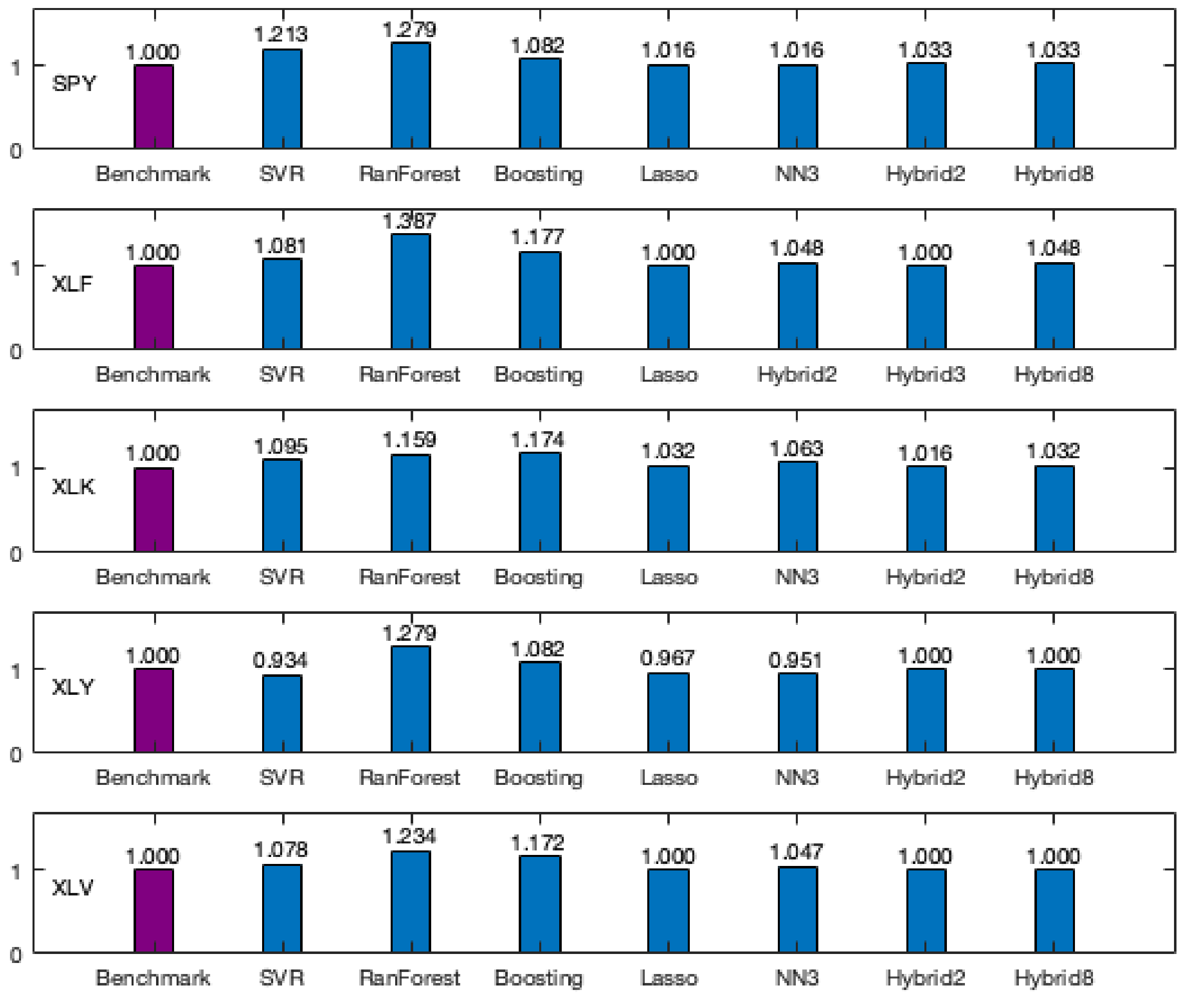
*Notes: We plot volatility risk factors for the S&P 500 market (MF_t^{TRV}), financial sector (MF_t^{XLF}), technology sector (MF_t^{XLK}), health care sector (MF_t^{XLV}), and consumer discretionary sector (MF_t^{XLY}).

Figure 3.6: Monthly Aggregate Relative MSFEs For Machine Leaning Models (Rolling Window Size)*



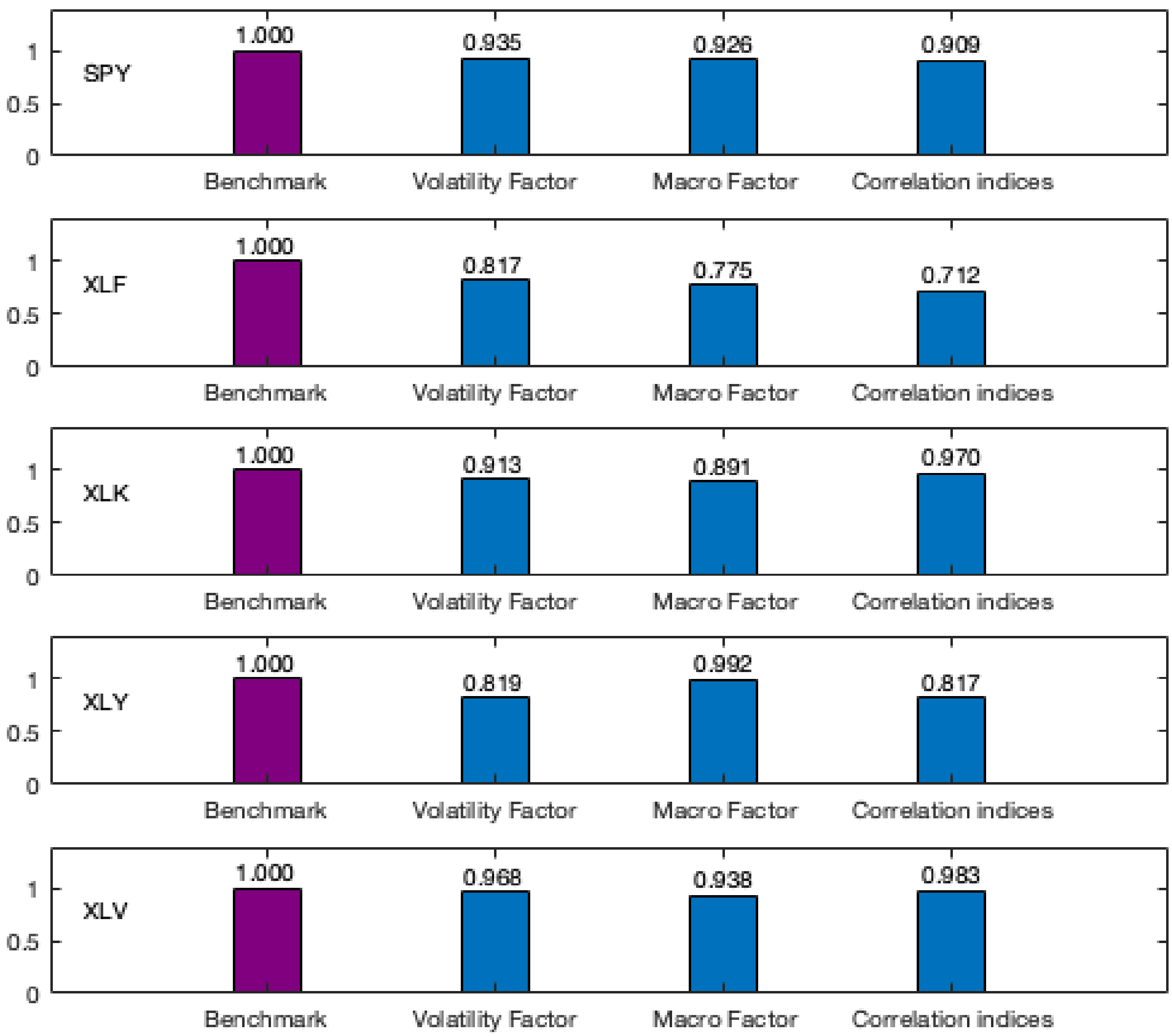
*Notes: Figure 3.6 shows the relative mean square forecasting error (MSFE) for machine learning models. Relative MSFEs are calculated such that numerical values less than unity indicates the alternative model has lower point MSFE than the random walk benchmark model. The panels from top to bottom display different forecasting targets including SPY (S&P 500 ETF), XLF (financial sector ETF), XLK (technology sector ETF), XLY (consumer discretionary sector ETF), and XLV (health care sector ETF). Results in each panel are obtained in Table 3.4 under the row of All Variables.

Figure 3.7: Monthly Aggregate Relative DPARs for Machine Leaning Models (Rolling Window Size)*



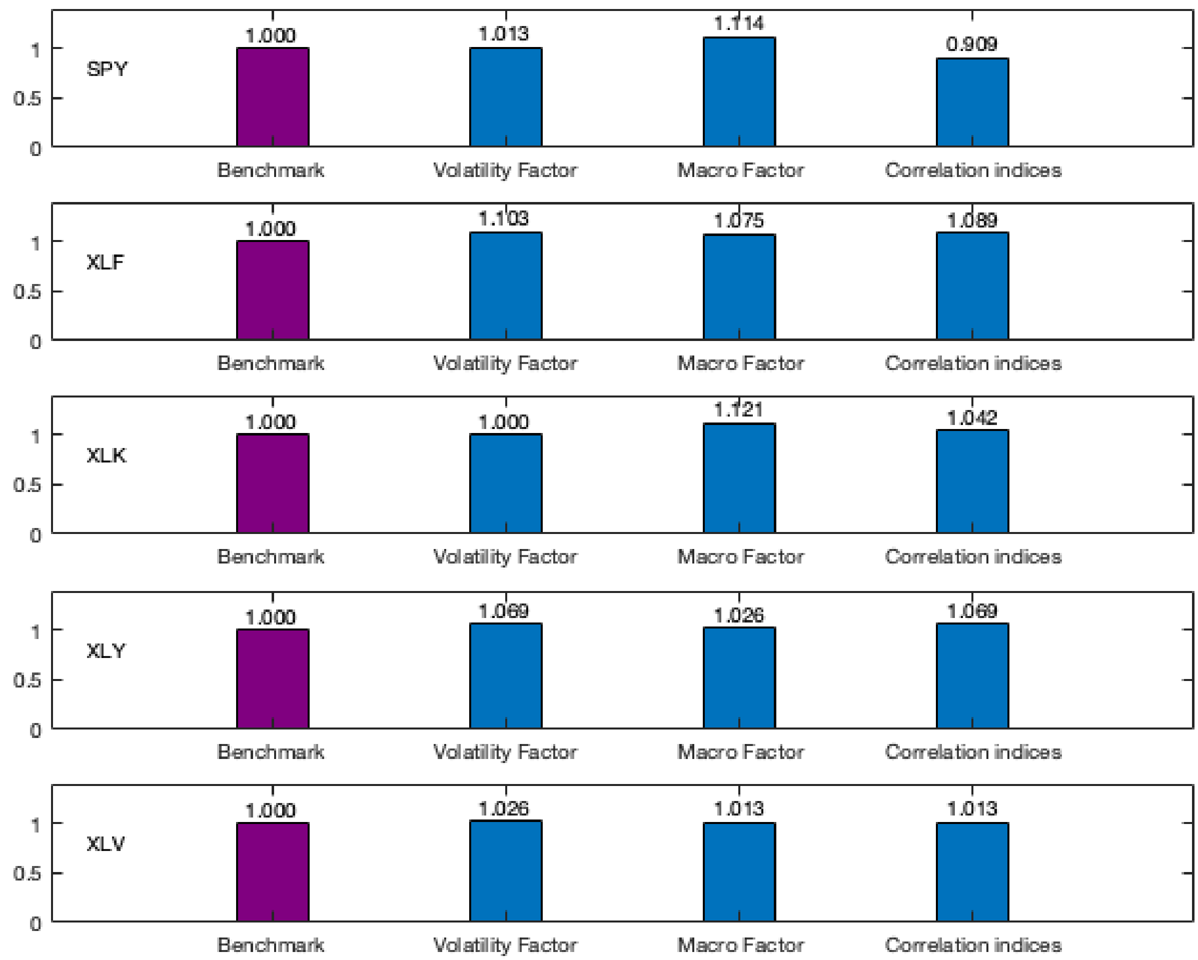
*Notes: Figure 3.7 shows the relative directional prediction accuracy rate (DPARs) for machine learning models. Relative DAPRs are calculated such that numerical values less than unity indicates the alternative model has lower DPAR than the random walk benchmark model. The panels from top to bottom display different forecasting targets including SPY (S&P 500 ETF), XLF (financial sector ETF), XLK (technology sector ETF), XLY (consumer discretionary sector ETF), and XLV (health care sector ETF). Results in each panel are obtained in Table 3.8 under the row of All Variables.

Figure 3.8: Monthly Aggregate Relative MSFEs Due to the Factors (Recursive Window Size)*



*Notes: See notes to Figure ??. Figure 3.8 shows the relative mean square forecasting error (MSFE) of forecasting models with factors. The benchmark model is the same forecasting models but without factors. The results in each panel are obtained from Table 3.5. Within each forecasting target (SPY, XLF, XLK, XLY, and XLV), we navigate to the MSFEs best machine learning model and further analyze the contribution of adding each factor.

Figure 3.9: Monthly Aggregate Relative DPARs Due to the Factors (Rolling Window Size)*



*Notes: See notes in Figure 3.7. Figure 3.9 shows the relative directional prediction accuracy rate (DPARs) of forecasting models with factors. The benchmark model is the same forecasting models but without factors. The results in each panel are obtained from Table 3.9. Within each forecasting target (SPY, XLF, XLK, XLY, and XLV), we navigate to the DPARs best machine learning model and further analyze the contribution of adding each factor.

Chapter 4

Forecast Evaluation

Note: This chapter is coauthored with Mingmian Cheng and Norman Swanson, and published in "Macroeconomic Forecasting in the Era of Big Data" (Fuleky (2020)) as a chapter entitled "Forecasting Evaluation" (Cheng et al. (2020)).

Part I: Forecast Evaluation Using Point Predictive Accuracy Tests

In this section, our objective is to review various commonly used statistical tests for comparing the relative accuracy of point predictions from different econometric models. Four main groups of tests are outlined: (i) tests for comparing two non-nested models, (ii) tests for comparing two nested models, (iii) tests for comparing multiple models, where at least one model is non-nested, and (iv) tests that are consistent against generic alternative models. The papers cited in this section (and in subsequent sections) contain references to a large number of papers that develop alternative related tests.

Of note is that the tests that we discuss in the sequel assume that all competing models are approximations to some unknown underlying data generating process, and are thus potentially misspecified. The objective is to select the "best" model from amongst multiple alternatives, where "best" refers to a given loss function, say.

4.1 Comparison of two non-nested models

The starting point of our discussion is the Diebold-Mariano (DM: Diebold and Mariano (2002)) test for the null hypothesis of equal predictive accuracy between two competing models, given a pre-specified loss function. This test sets the groundwork for many subsequent predictive accuracy tests. The DM test assumes that parameter estimation error is asymptotically negligible by positing that the number of observations used for in-sample model estimation grows faster than the number of observations used in out-of-sample forecast evaluation. Parameter estimation error in DM tests, which are often also called DM-West tests, is explicitly taken into account of in West (1996), although at the cost of requiring that the loss function is differentiable.

To fix ideas and notation, let $u_{i,t+h} = y_{t+h} - f_i(Z_i^t, \theta_i^\dagger)$ be the h-step ahead forecast error associated with the i-th model, $f_i(\cdot, \theta_i^\dagger)$, where the benchmark model is always denoted as "model 0", i.e. $f_0(\cdot, \theta_0^\dagger)$. As $\theta_i^\dagger$ and thus $u_{i,t+h}$ are unknown, we construct test statistics using $\widehat{\theta}_{i,t}$ and $\widehat{u}_{i,t+h} = y_{t+h} - f_i(Z_i^t, \widehat{\theta}_{i,t})$, where $\widehat{\theta}_{i,t}$ is an estimator of $\theta_i^\dagger$ constructed using information in Z_i^t from time periods 1 to t, under a recursive estimation scheme, or from $t-R+1$ to t, under a rolling-window estimation scheme. Hereafter, for notational simplicity, we only consider the recursive estimation scheme, and the rolling-window estimation scheme can be treated in an analogous manner. To do this, split the total sample of T observations into two sub-samples of length R and n, i.e. $T = R + n$, where only the last n observations are used for forecast evaluation. At each step, we first estimate the model parameters as follows,

$$\widehat{\theta}_{i,t} = \arg\min_{\theta_i} \ \frac{1}{t} \sum_{j=1}^{t} q(y_j - f_i(Z_i^{j-1}, \theta_i)), \quad t \geq R \tag{4.1}$$

These parameters are used to parameterize the prediction model, and an h-step-ahead prediction (and prediction error) is constructed. This procedure is repeated by adding one new observation to the original sample, yielding a new h-step-ahead prediction (and prediction error). In such a manner, we can construct a sequence of $(n-h+1)$ h-step ahead prediction errors. For a given loss function, $g(\cdot)$, the null hypothesis of DM test is specified as,

$$H_0 : E(g(u_{0,t+h}) - g(u_{1,t+h})) = 0$$

against

$$H_A : E(g(u_{0,t+h}) - g(u_{1,t+h})) \neq 0$$

Of particular note here is that the loss function $g(\cdot)$ used for forecast evaluation may not be the same as the loss function $q(\cdot)$ used for model estimation in Equation (4.1). However, if they are the same (e.g. models are estimated by ordinary least square (OLS) and forecasts are evaluated by a quadratic loss function, say), parameter estimation error is asymptotically negligible, regardless of the limiting ratio of n/R, as $T \to \infty$.

Define the following statistic,

$$\widehat{S}_n(0,1) = \frac{1}{\sqrt{n}} \sum_{t=R-h+1}^{T-h} (g(\widehat{u}_{0,t+h}) - g(\widehat{u}_{1,t+h}))$$

then,

$$\begin{aligned}\widehat{S}_n(0,1) - S_n(0,1) = E(\nabla_{\theta_0} g(u_{0,t+h}))\frac{1}{\sqrt{n}} \sum_{t=R-h+1}^{T-h} (\widehat{\theta}_{0,t+h} - \theta_0^{\dagger}) \\ - E(\nabla_{\theta_1} g(u_{1,t+h}))\frac{1}{\sqrt{n}} \sum_{t=R-h+1}^{T-h} (\widehat{\theta}_{1,t+h} - \theta_1^{\dagger}) + o_p(1)\end{aligned} \tag{4.2}$$

The limiting distribution of the right-hand side of Equation (4.2) is given by Lemma 4.1 and Theorem 4.1 in West (1996). From Equation (4.2), we can immediately see that if $g(\cdot) = q(\cdot)$, then $E(\nabla_{\theta_i} g(u_{i,t+h})) = 0$ by the first order conditions, and parameter estimation error is asymptotically negligible. Another situation in which parameter estimation error vanishes asymptotically is when $n/R \to 0$, as $T \to \infty$.

Without loss of generality, consider the case of $h = 1$. All results carry over to the case when $h > 1$. The DM test statistic is given by,

$$\widehat{DM}_n = \frac{1}{\sqrt{n}} \frac{1}{\widehat{\sigma}_n} \sum_{t=R}^{T-1} (g(\widehat{u}_{0,t+1}) - g(\widehat{u}_{1,t+1}))$$

with

$$\begin{aligned}\widehat{\sigma}_n =& \widehat{S}_{gg} + 2\Pi \widehat{F}_0' \widehat{A}_0 \widehat{S}_{h_0h_0} + 2\Pi \widehat{F}_1' \widehat{A}_1 \widehat{S}_{h_1h_1} \widehat{A}_1 \widehat{F}_1 \\ & - 2\Pi(\widehat{F}_1' \widehat{A}_1 \widehat{S}_{h_1h_0} \widehat{A}_0 \widehat{F}_0 + \widehat{F}_0' \widehat{A}_0 \widehat{S}_{h_0h_1} \widehat{A}_1 \widehat{F}_1) \\ & + \Pi(\widehat{S}_{gh_1}' \widehat{A}_1 \widehat{F}_1 + \widehat{F}_1' \widehat{A}_1 \widehat{S}_{gh_1})\end{aligned}$$

where for $i, j = 0, 1$, $\Pi = 1 - \frac{\ln(1+\pi)}{\pi}$, and $q_t(\widehat{\theta}_{i,t}) = q(y_t - f_i(Z_i^{t-1}, \widehat{\theta}_{i,t}))$,

$$\widehat{S}_{h_ih_j} = \frac{1}{n} \sum_{\tau=-l_n}^{l_n} w_\tau \sum_{t=R+l_n}^{T-l_n} \nabla_\theta q_t(\widehat{\theta}_{i,t}) \nabla_\theta q_{t+\tau}(\widehat{\theta}_{j,t})'$$

$$\begin{aligned}
\widehat{S}_{gh_i} =& \frac{1}{n}\sum_{\tau=-l_n}^{l_n} w_\tau \sum_{t=R+l_n}^{T-l_n}\left(g(\widehat{u}_{0,t}) - g(\widehat{u}_{1,t}) - \frac{1}{n}\sum_{t=R}^{T-1}(g(\widehat{u}_{0,t+1}) - g(\widehat{u}_{1,t+1}))\right)\\
&\times \nabla_\theta q_{t+\tau}(\widehat{\theta}_{i,t})'\\
\widehat{S}_{gg} =& \frac{1}{n}\sum_{\tau=-l_n}^{l_n} w_\tau \sum_{t=R+l_n}^{T-l_n}\left(g(\widehat{u}_{0,t}) - g(\widehat{u}_{1,t}) - \frac{1}{n}\sum_{t=R}^{T-1}(g(\widehat{u}_{0,t+1}) - g(\widehat{u}_{1,t+1}))\right)\\
&\times \left(g(\widehat{u}_{0,t+\tau}) - g(\widehat{u}_{1,t+\tau}) - \frac{1}{n}\sum_{t=R}^{T-1}(g(\widehat{u}_{0,t+1}) - g(\widehat{u}_{1,t+1}))\right)
\end{aligned}$$

with $w_\tau = 1 - \frac{\tau}{l_n - 1}$, and

$$\widehat{F}_i = \frac{1}{n}\sum_{t=R}^{T-1}\nabla_{\theta_i} g(\widehat{u}_{i,t+1}), \qquad \widehat{A}_i = \left(-\frac{1}{n}\sum_{t=R}^{T-1}\nabla^2_{\theta_i} q(\widehat{\theta}_{i,t})\right)^{-1}$$

Assumption 1.1: (y_t, Z^{t-1}), with y_t scalar and Z^{t-1} an $\Re^\zeta$-valued $(0 < \zeta < \infty)$ vector, is a strictly stationary and absolutely regular β-mixing process with size $-4(4+\psi)/\psi$, $\psi > 0$.

Assumption 1.2: (i) $\theta^\dagger$ is uniquely identified (i.e. $E(q(y_t, Z^{t-1}, \theta_i))) > E(q(y_t, Z^{t-1}, \theta_i^\dagger)))$ for any $\theta_i \neq \theta_i^\dagger$); (ii) $q(\cdot)$ is twice continuously differentiable on the interior of Θ, and for Θ a compact subset of $\Re^\varrho$; (iii) the elements of $\nabla_\theta q$ and $\nabla^2_\theta q$ are p-dominated on Θ, with $p > 2(2+\psi)$, where ψ is the same positive constant as defined in Assumption 1.1; and (iv) $E(-\nabla^2_\theta q)$ is negatively definite uniformly on Θ.

PROPOSITION 1.1 (From Theorem 4.1 in West (1996)): With Assumptions 1.1 and 1.2, also, assume that $g(\cdot)$ is continuously differentiable, then, if as $n \to \infty$, $l_n \to \infty$ and $l_n/n^{1/4} \to 0$, then as $n,\ R \to \infty$, under H_0,

$$\widehat{DM}_n \xrightarrow{d} N(0,1)$$

Under H_A,

$$\Pr(n^{-1/2}|\widehat{DM}_n| > \epsilon) \to 1, \quad \forall \epsilon > 0$$

It is immediate to see that if either $g(\cdot) = q(\cdot)$ or $n/R \to 0$, as $T \to \infty$, the estimator $\widehat{\sigma}_n$ collapses to $\widehat{S}_{gg}$. Note that the limiting distribution of DM test obtains only for the case of short-memory series. Corradi et al. (2001) extends the DM test to the case of co-integrated variables and Rossi (2005) to the case of series with high persistence. Finally, note that the two competing models are assumed to be non-nested. If they are nested, then $u_{0,t+h} = u_{1,t+h}$ under the null, and both $\sum_{t=R-h+1}^{T-h}(g(\widehat{u}_{0,t+h}) - g(\widehat{u}_{1,t+h}))$ and $\widehat{\sigma}_n$ converge in probability to zero at the same rate if $n/R \to 0$. Therefore the DM test statistic does not converge in distribution to a standard normal variable under the null. Comparison of nested models is introduced in the next section.

4.2 Comparison of two nested models

There are situations in which we may be interested in comparing forecasts from nested models. For instance, one of the driving forces behind the literature on out-of-sample comparison of nested models is the seminal paper by Meese and Rogoff (1983), who find that no models driven by economic fundamentals can beat a simple random walk model, in terms of out-of-sample predictive accuracy, when forecasting exchange rates. The models studied in this paper are nested, in the sense that parameter restrictions can be placed on the more general models that reduce these models to the random walk benchmark studied by these authors. When testing out-of-sample Granger causality, alternative models are also nested. Since the DM test discussed above is valid only when the competing models are non-nested, we introduce alternative tests that address testing among nested models.

4.2.1 Clark and McCracken tests for nested models

Clark and McCracken (2001) (CMa) and Clark and McCracken (2003) (CMb) propose several tests for nested linear models, under the assumption that prediction errors follow martingale difference sequences (this rules out the possibility of dynamic misspecification under the null for these particular tests), where CMa tests are tailored for the case of one-step-ahead forecasts, and CMb tests for the case of multi-step-ahead forecasts.

Consider the following two nested models. The restricted model is,

$$y_t = \sum_{j=1}^{q} \beta_j y_{t-j} + \epsilon_t$$

and the unrestricted model is,

$$y_t = \sum_{j=1}^{q} \beta_j y_{t-j} + \sum_{j=1}^{k} \alpha_j x_{t-j} + u_t \tag{4.3}$$

The null hypothesis of CMa tests is formulated as,

$$H_0 : E(\epsilon_t^2) - E(u_t^2) = 0$$

against

$$H_A : E(\epsilon_t^2) - E(u_t^2) > 0$$

We can immediately see from the null and the alternative hypotheses that CMa tests implicitly assume that the restricted model cannot beat the unrestricted model. This is the case when the models are estimated by OLS and the quadratic loss function is employed for evaluation.

CMa propose the following three different test statistics,

$$ENC - T = (n-1)^{1/2} \frac{\bar{c}}{(n^{-1} \sum_{t=R}^{T-1} (c_{t+1} - \bar{c}))^{1/2}}$$

$$ENC - REG = (n-1)^{1/2} \frac{n^{-1} \sum_{t=R}^{T-1} (\widehat{\epsilon}_{t+1} (\widehat{\epsilon}_{t+1} - \widehat{u}_{t+1}))}{(n^{-1} \sum_{t=R}^{T-1} (\widehat{\epsilon}_{t+1} - \widehat{u}_{t+1})^2 n^{-1} \sum_{t=R}^{T-1} \widehat{\epsilon}_{t+1}^2 - \bar{c})^{1/2}}$$

$$ENC - NEW = n \frac{\bar{c}}{n^{-1} \sum_{t=1} \widehat{u}_{t+1}^2}$$

where $c_{t+1} = \widehat{\epsilon}_{t+1}(\widehat{\epsilon}_{t+1} - \widehat{u}_{t+1})$, $\bar{c} = n^{-1} \sum_{t=R}^{T-1} c_{t+1}$, and $\widehat{\epsilon}_{t+1}$ and $\widehat{u}_{t+1}$ are OLS residuals.

Assumption 2.1: (y_t, x_t) are strictly stationary and strong mixing processes, with size $\frac{-4(4+\delta)}{\delta}$, for some $\delta > 0$, and $E(y_t^8)$ and $E(x_t^8)$ are both finite.

Assumption 2.2: Let $z_t = (y_{t-1}, ..., y_{t-q}, x_{t-1}, ..., x_{t-q})$ and $E(z_t u_t | \mathcal{F}_{t-1}) = 0$, where $\mathcal{F}_{t-1}$ is the σ-field up to time $t-1$, generated by $(y_{t-1}, y_{t-2}, ..., x_{t-1}, x_{t-2}, ...)$. Also,

$E(u_t^2|\mathcal{F}_{t-1}) = \sigma_u^2$.

Note that Assumption 2.2 assumes that the unrestricted model is dynamically correct and that u_t is conditionally homoskedastic.

PROPOSITION 2.1 (From Theorem 3.1–3.3 in Clark and McCracken (2001)): With Assumptions 2.1 and 2.2, under the null, (i) if as $T \to \infty$, $n/R \to \pi > 0$, then $ENC-T$ and $ENC-REG$ converge in distribution to Γ_1/Γ_2 where $\Gamma_1 = \int_{(1+\pi)^{-1}}^{1} s^{-1}W_s^{'}dW_s$ and $\Gamma_2 = \int_{(1+\pi)^{-1}}^{1} W_s^{'}W_s ds$, with W_s a k-dimensional standard Brownian motion (here k is the number of restrictions or the number of extra regressors in the unrestricted model). $ENC-NEW$ converges in distribution to Γ_1. (ii) If as $T \to \infty$, $n/R \to 0$, then $ENC-T$ and $ENC-REG$ converge in distribution to $N(0,1)$. $ENC-NEW$ converges in probability to 0.

Therefore, as $T \to \infty$ and $n/R \to \pi > 0$, all three test statistics have non-standard limiting distributions. Critical values are tabulated for different k and π in CMa. Also note that the above proposition is valid only when $h=1$, i.e. the case of one-step ahead forecasts, since Assumption 2.2 is violated when $h > 1$. For this case, CMb propose a modified test statistic for which $MA(h-1)$ errors are allowed. Namely, they propose using the following statistic:

$$\begin{aligned} ENC-T^{'} =& (n-h+1)^{1/2}\times \\ & \frac{(n-h+1)^{-1}\sum_{t=R}^{T-h}\widehat{c}_{t+h}}{((n-h+1)^{-1}\sum_{j=-\bar{j}}^{\bar{j}}\sum_{t=R+j}^{T-h}K(\frac{j}{M})(\widehat{c}_{t+h}-\bar{c})(\widehat{c}_{t+h-j}-\bar{c}))^{1/2}}, \end{aligned}$$

where $K(\cdot)$ is a kernel and $0 \leq K(\frac{j}{M}) \leq 1$, with $K(0)=1$ and $M = o(n^{1/2})$, and $\bar{j}$ does not grow with the sample size. Therefore, the denominator of $ENC-T^{'}$ is a consistent estimator of the long-run variance when $E(c_t c_{t+|k|}) = 0$ for all $|k| > h$. Of particular note is that although $ENC-T^{'}$ allows for $MA(h-1)$ errors, dynamic misspecification under the null is still not allowed. Also note that, when $h=1$, $ENC-T^{'}$ is equivalent to $ENC-T$.

Another test statistic suggested in CMb is a DM-type test with nonstandard critical values that are needed in order to modify the DM test in order to allow for the comparison of nested models. The test statistic is:

$$MSE - T' = (n-h+1)^{1/2} \times \frac{(n-h+1)^{-1}\sum_{t=R}^{T-h}\widehat{d}_{t+h}}{((n-h+1)^{-1}\sum_{j=-\bar{j}}^{\bar{j}}\sum_{t=R+j}^{T-h}K(\frac{j}{M})(\widehat{d}_{t+h}-\overline{d})(\widehat{d}_{t+h-j}-\overline{d}))^{1/2}}$$

where $\widehat{d}_{t+h} = \widehat{u}^2_{t+h} - \widehat{\epsilon}^2_{t+h}$ and $\overline{d} = (n-h+1)^{-1}\sum_{t=R}^{T-h}\widehat{d}_{t+h}$.

Evidently, this test is a standard DM test, although it should be stressed that the critical values used in the application of this variant of the test are different. The limiting distributions of the $ENC-T'$ and $MSE-T'$ are provided in CMb, and are non-standard. Moreover, for the case of $h > 1$, the limiting distributions contain nuisance parameters, so that critical values cannot be tabulated directly. Instead, CMb suggest a modified version of the bootstrap method in Kilian (1999) to carry out statistical inference. For this test, the block bootstrap can also be used to carry out inference (see Corradi and Swanson (2007) for details.)

4.2.2 Out-of-sample tests for Granger causality

CMa and CMb tests do not take dynamic misspecification into account under the null. Chao et al. (2001) (CCS) propose out-of-sample tests for Granger causality allowing for possible dynamic misspecification and conditional heteroskedascity. The idea is very simple. If the coefficients $\alpha_j, j = 1, ..., k$ in Equation (4.3) are all zeros, then residuals ϵ_{t+1} are uncorrelated with lags of x. As a result, including regressors $x_{t-j}, j = 1, ..., k$ does not help improve predictive accuracy, and the unrestricted model does not outperform the restricted model.

Hereafter, for notational simplicity, we only consider the case of $h = 1$. All results can be generalized to the case of $h > 1$. Formally, the test statistic is,

$$m_n = n^{-1/2}\sum_{t=R}^{T-1}\widehat{\epsilon}_{t+1}X_t,$$

where $X_t = (x_t, x_{t-1}, ..., x_{t-k-1})^{'}$. The null hypothesis and the alternative hypothesis are formulated as,

$$H_0 : E(\epsilon_{t+1}x_{t-j}) = 0, \quad j = 0, 1, ..., k-1$$

$$H_A : E(\epsilon_{t+1}x_{t-j}) \neq 0, \quad \text{for some } j$$

Assumption 2.3: (y_t, x_t) are strictly stationary and strong mixing processes, with size $\frac{-4(4+\delta)}{\delta}$, for some $\delta > 0$, and $E(y_t^8)$ and $E(x_t^8)$ are both finite. $E(\epsilon_t y_{t-j}) = 0, j = 1, 2, ..., q$.

PROPOSITION 2.2 (From Theorem 1 in Chao et al. (2001)): With Assumption 2.3, as $T \to \infty$, $n/R \to \pi, 0 \leq \pi < \infty$, (i) under the null, for $0 < \pi < \infty$,

$$m_n \xrightarrow{d} N(0, \Xi)$$

with

$$\begin{aligned}\Xi =& S_{11} + 2(1 - \pi^{-1}\ln(1+\pi))F^{'}MS_{22}MF - \\ &(1 - \pi^{-1}\ln(1+\pi))(F^{'}MS_{12} + S_{12}^{'}MF)\end{aligned}$$

where $F = E(Y_t X_t^{'})$, $M = \text{plim}(\frac{1}{t}\sum_{j=q}^{t} Y_j Y_j^{'})^{-1}$, and $Y_j = (y_{j-1}, ..., y_{j-q})^{'}$. Furthermore,

$$S_{11} = \sum_{j=-\infty}^{\infty} E((X_t\epsilon_{t+1} - \mu)(X_{t-j}\epsilon_{t-j+1} - \mu)^{'})$$

$$S_{22} = \sum_{j=-\infty}^{\infty} E((Y_{t-1}\epsilon_t)(Y_{t-j-1}\epsilon_{t-j})^{'})$$

$$S_{12} = \sum_{j=-\infty}^{\infty} E((\epsilon_{t+1}X_t - \mu)(Y_{t-j-1}\epsilon_{t-j})^{'})$$

where $\mu = E(X_t\epsilon_{t+1})$. In addition, for $\pi = 0$,

$$m_n \xrightarrow{d} N(0, S_{11})$$

(ii) Under the alternative,

$$\lim_{n\to\infty} \Pr(|n^{-1/2}m_n| > 0) = 1$$

COROLLARY 2.1 (From Corollary 2 in Chao et al. (2001)): With Assumption 2.3, as $T \to \infty$, $n/R \to \pi, 0 \leq \pi < \infty, l_T \to \infty, l_T/T^{1/4} \to 0$, (i) under the null, for $0 < \pi < \infty$,

$$m_n' \widehat{\Xi}^{-1} m_n \xrightarrow{d} \chi_k^2$$

with

$$\begin{aligned} \widehat{\Xi} = \widehat{S}_{11} &+ 2(1 - \pi^{-1}\ln(1+\pi))\widehat{F}' \widehat{M} \widehat{S}_{22} \widehat{M} \widehat{F} \\ &- (1 - \pi^{-1}\ln(1+\pi))(\widehat{F}' \widehat{M} \widehat{S}_{12} + \widehat{S}_{12}' \widehat{M} \widehat{F}) \end{aligned}$$

where $\widehat{F} = n^{-1} \sum_{t=R}^{T} Y_t X_t'$, $\widehat{M} = (n^{-1} \sum_{t=R}^{T-1} Y_t Y_t')r^{-1}$, and

$$\begin{aligned} \widehat{S}_{11} =& \frac{1}{n} \sum_{t=R}^{T-1} (\widehat{\epsilon}_{t+1} X_t - \widehat{\mu}_1)(\widehat{\epsilon}_{t+1} X_t - \widehat{\mu}_1)' \\ &+ \frac{1}{n} \sum_{t=\tau}^{l_T} w_\tau \sum_{t=R+\tau}^{T-1} (\widehat{\epsilon}_{t+1} X_t - \widehat{\mu}_1)(\widehat{\epsilon}_{t+1-\tau} X_{t-\tau} - \widehat{\mu}_1)' \\ &+ \frac{1}{n} \sum_{t=\tau}^{l_T} w_\tau \sum_{t=R+\tau}^{T-1} (\widehat{\epsilon}_{t+1-\tau} X_{t-\tau} - \widehat{\mu}_1)(\widehat{\epsilon}_{t+1} X_t - \widehat{\mu}_1)' \end{aligned}$$

$$\begin{aligned} \widehat{S}_{12} =& \frac{1}{n} \sum_{\tau=0}^{l_T} w_\tau \sum_{t=R+\tau}^{T-1} (\widehat{\epsilon}_{t+1-\tau} X_{t-\tau} - \widehat{\mu}_1)(Y_{t-1} \widehat{\epsilon}_t)' \\ &+ \frac{1}{n} \sum_{\tau=1}^{l_T} w_\tau \sum_{t=R+\tau}^{T-1} (\widehat{\epsilon}_{t+1} X_t - \widehat{\mu}_1)(Y_{t-1-\tau} \widehat{\epsilon}_{t-\tau})' \end{aligned}$$

$$\begin{aligned} \widehat{S}_{22} =& \frac{1}{n} \sum_{t=R}^{T-1} (Y_{t-1} \widehat{\epsilon}_t)(Y_{t-1} \widehat{\epsilon}_t)' \\ &+ \frac{1}{n} \sum_{\tau=1}^{l_T} w_\tau \sum_{t=R+\tau}^{T-1} (Y_{t-1} \widehat{\epsilon}_t)(Y_{t-1-\tau} \widehat{\epsilon}_{t-\tau})' \\ &+ \frac{1}{n} \sum_{\tau=1}^{l_T} w_\tau \sum_{t=R+\tau}^{T-1} (Y_{t-1-\tau} \widehat{\epsilon}_{t-\tau})(Y_{t-1} \widehat{\epsilon}_t)' \end{aligned}$$

with $w_\tau = 1 - \frac{\tau}{l_T+1}$. In addition, for $\pi = 0$,

$$m_n' \widehat{S}_{11}^{-1} m_n \xrightarrow{d} \chi_k^2$$

(ii) Under the alternative, $m_n' \widehat{S}_{11}^{-1} m_n$ diverges at rate n.

Note that a "nonlinear" variant of the above CCS test has also been developed by the same authors. In this generic form of the test, one can test for nonlinear Granger causality, for example, where the alternative hypothesis is that some (unknown) function of the x_t can be added to the benchmark linear model that contains no x_t in order to improve predictive accuracy. This alternative test is thus consistent against generic nonlinear alternatives. Complete details of this test are given in the next section.

4.3 A predictive accuracy test that is consistent against generic alternatives

The test discussed in the previous subsection is designed to have power against a given (linear) alternative; and while it may have power against other alternatives, it is not designed to do so. Thus, it is not consistent against generic alternatives. Tests that are consistent against generic alternatives are sometimes called portmanteau tests, and it is this sort of extension of the out-of-sample Granger causality test discussed above that we now turn our attention to. Broadly speaking, the above consistency has been studied in the consistent specification testing literature (see Bierens (1990), Bierens and Ploberger (1997), De Jong (1996), Hansen (1996a), Lee et al. (1993) and Stinchcombe and White (1998)).

Corradi and Swanson (2002) draw on both the integrated conditional moment (ICM) testing literature of Bierens (1990) and Bierens and Ploberger (1997) and on the predictive accuracy testing literature; and propose an out-of-sample version of the ICM test that is consistent against generic nonlinear alternatives. This test is designed to examine whether there exists an unknown (possibly nonlinear) alternative model with better predictive power than the benchmark model, for a given loss function. A typical example is the case in which the benchmark model is a simple autoregressive model and we want to know whether including some unknown functions of the past information can produce more accurate forecasts. This is the case of nonlinear Granger causality testing discussed above. Needless to say, this test can be applied to many other cases. One important feature of this test is that the same loss function is used for in-sample model estimation and out-of-sample predictive evaluation (see Granger (1993) and Weiss (1996)).

Consider the following benchmark model,

$$y_t = \theta_1^{\dagger} y_{t-1} + u_t,$$

where $\theta_1^{\dagger} = \arg\min_{\theta_1 \in \Theta_1} E(q(y_t - \theta_1 y_{t-1}))$. The generic alternative model is,

$$y_t = \theta_{2,1}^{\dagger}(\gamma) y_{t-1} + \theta_{2,2}^{\dagger}(\gamma)\omega(Z^{t-1}, \gamma) + \upsilon_t$$

where

$$\theta_2^{\dagger}(\gamma) = (\theta_{2,1}^{\dagger}(\gamma), \theta_{2,2}^{\dagger}(\gamma))' = \arg\min_{\theta_2 \in \Theta_2} E(q(y_t - \theta_{2,1}(\gamma) y_{t-1} - \theta_{2,2}(\gamma)\omega(Z^{t-1}, \gamma))).$$

The alternative model is "generic" due to the term $\omega(Z^{t-1}, \gamma)$, where the function $\omega(\cdot)$ is a generically comprehensive function, as defined in Bierens (1990) and Bierens and Ploberger (1997). The test hypotheses are:

$$H_0 : E(g(u_t) - g(\upsilon_t)) = 0$$

$$H_A : E(g(u_t) - g(\upsilon_t)) > 0$$

By definition, it is clear that the benchmark model is nested within the alternative model. Thus the former model can never outperform the latter. Equivalently, the hypotheses can be restated as,

$$H_0 : \theta_{2,2}^{\dagger}(\gamma) = 0$$

$$H_A : \theta_{2,2}^{\dagger}(\gamma) \neq 0$$

Note that, given the definition of $\theta_2^{\dagger}(\gamma)$, we have that

$$E\left(g'(\upsilon_t) \times \left(-y_t, -\omega(Z^{t-1}, \gamma)\right)'\right) = 0$$

Hence, under the null, we have that $\theta_{2,2}^{\dagger}(\gamma) = 0$, $\theta_{2,1}^{\dagger}(\gamma) = \theta_1^{\dagger}$ and $E(g'(u_t)\omega(Z^{t-1}, \gamma)) = 0$. As a result, the hypotheses can be once again be restated as,

$$H_0 : E(g'(u_t)\omega(Z^{t-1}, \gamma)) = 0$$

$$H_A : E(g^{'}(u_t)\omega(Z^{t-1}, \gamma)) \neq 0$$

The test statistic is given by

$$M_n = \int m_n(\gamma)^2 \phi(\gamma) d\gamma$$

with

$$m_n(\gamma) = n^{-1/2} \sum_{t=R}^{T-1} g^{'}(\widehat{u}_t + 1)\omega(Z^t, \gamma)$$

where $\int \phi(\gamma)d\gamma = 1$, $\phi(\gamma) \geq 0$, and $\phi(\gamma)$ is absolutely continuous with respect to Lebesgue measure.

Assumption 4.1: (i) (y_t, Z^t) is a strictly stationary and absolutely regular strong mixing sequence with size $-4(4+\psi)/\psi, \psi > 0$; (ii) $g(\cdot)$ is three times continuously differentiable in θ, over the interior of Θ, and $\nabla_\theta g$, $\nabla^2_\theta g$, $\nabla_\theta g^{'}$, $\nabla^2_\theta g^{'}$ are $2r$-dominated uniformly in Θ, with $r \geq 2(2+\psi)$; (iii) $E(-\nabla^2_\theta g(\theta))$ is negative definite, uniformly in Θ; (iv) $\omega(\cdot)$ is a bounded, twice continuously differentiable function on the interior of Γ and $\nabla_\gamma \omega(Z^t, \gamma)$ is bounded uniformly in Γ; (iv) $\nabla_\gamma \nabla_\theta g^{'}(\theta)\omega(Z^t, \gamma)$ is continuous on $\Theta \times \Gamma$, Γ a compact subset of $\Re^d$ and is $2r$-dominated uniformly in $\Theta \times \Gamma$, with $r \geq 2(2+\psi)$.

Assumption 4.2: (i) $E(g^{'}(y_t - \theta_1^\dagger y_{t-1})) < E(g^{'}(y_t - \theta_1 y_{t-1})), \forall \theta \neq \theta^\dagger$; (ii) $\inf_\gamma E(g^{'}(y_t - \theta^\dagger_{2,1}(\gamma)y_{t-1} + \theta^\dagger_{2,2}(\gamma)\omega(Z^{t-1}, \gamma))) < E(g^{'}(y_t - \theta_{2,1}(\gamma)y_{t-1} + \theta_{2,2}(\gamma)\omega(Z^{t-1}, \gamma))), \forall \theta \neq \theta^\dagger(\gamma)$.

Assumption 4.3: $T = R + n$, and as $T \to \infty$, $n/R \to \pi$, with $0 \leq \pi < \infty$.

PROPOSITION 4.1 (From Theorem 1 in Corradi and Swanson (2002)): With Assumptions 4.1–4.3, the following results hold: (i) Under the null,

$$M_n \xrightarrow{d} \int Z(\gamma)^2 \phi(\gamma) d\gamma$$

where Z is a Gaussian process with covariance structure,

$$\begin{aligned} K(\gamma_1, \gamma_2) = {} & S_{gg}(\gamma_1, \gamma_2) + 2\Pi \mu_{\gamma_1} A^\dagger S_{hh} A^\dagger \mu_{\gamma_2} \\ & + \Pi \mu^{'}_{\gamma_1} A^\dagger S_{gh}(\gamma_2) + \Pi \mu^{'}_{\gamma_2} A^\dagger S_{gh}(\gamma_1) \end{aligned}$$

with $\mu_{\gamma_1} = E(\nabla_{\theta_1}(g^{'}(u_t)\omega(Z^t, \gamma_1)))$, $A^{\dagger} = (-E(\nabla^2_{\theta_1} q(u_t)))^{-1}$, and

$$S_{gg}(\gamma_1, \gamma_2) = \sum_j E(g^{'}(u_{s+1})\omega(Z^s, \gamma_1)g^{'}(u_{s+j+1})\omega(Z^{s+j}, \gamma_1))$$

$$S_{hh} = \sum_j E(\nabla_{\theta_1} q(u_s)\nabla_{\theta_1} q(u_{s+j})^{'})$$

$$S_{gh}(\gamma_1) = \sum_j E(g^{'}(u_{s+1})\omega(Z^s, \gamma_1)\nabla_{\theta_1} q(u_{s+j})^{'})$$

and γ, γ_1 and γ_2 are generic elements of Γ.

(ii) Under the alternative, for $\epsilon > 0$ and $\delta < 1$,

$$\lim_{n\to\infty} \Pr\left(n^{-\delta}\int m_n(\gamma)^2\phi(\gamma)d\gamma > \epsilon\right) = 1$$

The limiting distribution under the null is a Gaussian process with a covariance structure that reflects both the time dependence and the parameter estimation error. Therefore the critical values cannot be tabulated. Valid asymptotic critical values can be constructed by using the block bootstrap for recursive estimation schemes, as detailed in Corradi and Swanson (2007). In particular, define,

$$\widetilde{\theta}^*_{1,t} = \arg\min_{\theta_1} \frac{1}{t}\sum_{j=2}^{t}[g(y^*_j - \theta_1 y^*_{j-1}) - \theta_1^{'}\frac{1}{T}\sum_{i=2}^{T}\nabla_{\theta} g(y_i - \widehat{\theta}_1 y_{i-1})]$$

Then the bootstrap statistic is,

$$M^*_n = \int m^*_n(\gamma)^2\phi(\gamma)d\gamma$$

where

$$m^*_n(\gamma) = n^{-1/2}\sum_{t=R}^{T-1}\left(g^{'}(u^*_t)\omega(Z^{*,t}, \gamma) - T^{-1}\sum_{i=1}^{T-1} g^{'}(\widehat{u}_t)\omega(Z^i, \gamma)\right)$$

Assumption 4.4: For any t, s and $\forall i, j, k = 1, 2$, and for $\Delta < \infty$,

$$\text{(i)} \quad E(\sup_{\theta,\gamma,\gamma^+} |g^{'}(\theta)\omega(Z^{t-1}, \gamma)\nabla^k_{\theta} g^{'}(\theta)\omega(Z^{s-1}, \gamma^+)|^4) < \Delta$$

where $\nabla_\theta^k(\cdot)$ denotes the k-th element of the derivative of its argument with respect to θ.

$$\text{(ii)} \quad E(\sup_\theta |\nabla_\theta^k(\nabla_\theta^i g(\theta))\nabla_\theta^j g(\theta)|^4) < \Delta$$

and

$$\text{(iii)} \quad E(\sup_{\theta,\gamma} |g^{'}(\theta)\omega(Z^{t-1},\gamma)\nabla_\theta^k(\nabla_\theta^j g(\theta))|^4) < \Delta$$

PROPOSITION 4.2 (From Proposition 5 in Corradi and Swanson (2007)): With Assumptions 4.1–4.4, also assume that as $T \to \infty$, $l \to \infty$, and $l/T^{1/4} \to 0$, then as $T, n, R \to \infty$,

$$\Pr\left(\sup_\delta \left|\overset{*}{\Pr}(\int m_n^*(\gamma)^2\phi(\gamma)d\gamma \leq \delta) - Pr(\int m_n(\gamma)^2\phi(\gamma)d\gamma \leq \delta)\right| > \epsilon\right) \to 0$$

The above proposition justifies the bootstrap procedure. For all samples except a set with probability measure approaching zero, M_n^* mimics the limiting distribution of M_n under the null, ensuring asymptotic size equal to α. Under the alternative, M_n^* still has a well defined limiting distribution, while M_n explodes, ensuring unit asymptotic power.

In closing, note that $\widetilde{\theta}_{1,t}^*$ can be replaced with $\theta_{1,t}^*$ if parameter estimation error is assumed to be asymptotically negligible. In this case, critical values are constructed via standard application of the block bootstrap.

4.4 Comparison of multiple models

The predictive accuracy tests that we have introduced to this point are all used to choose between two competing models. However, an even more common situation is when multiple (more than two) competing models are available, and the objective is to assess whether there exists at least one model that outperforms a given "benchmark" model. If we sequentially compare each of the alternative models with the benchmark, we induce the so-called "data snooping" problem, where sequential test bias results in the size of our test increasing to unity, so that the null hypothesis is rejected with probability one, even when the null is true. In this subsection, we review several tests for comparing multiple models and addressing the issue of data snooping.

4.4.1 A reality check for data snooping

White (2000) proposes a test called the "reality check", which is suitable for comparing multiple models. We use the same notation as that used when discussing the DM test, except that there are now multiple alternative models, i.e. model $i = 0, 1, 2, ..., m$. Recall that $i = 0$ denotes the benchmark model. Define the following test statistic,

$$\widehat{S}_n = \max_{i=1,...,m} \widehat{S}_n(0, i) \tag{4.4}$$

where

$$\widehat{S}_n(0, i) = \frac{1}{\sqrt{n}} \sum_{t=R}^{T-1} (g(\widehat{u}_{0,t+1}) - g(\widehat{u}_{i,t+1})), \quad i = 1, ..., m$$

The reality check tests the following null hypothesis:

$$H_0 : \max_{i=1,...,m} E(g(u_{0,t+1}) - g(u_{i,t+1})) \leq 0$$

against

$$H_A : \max_{i=1,...,m} E(g(u_{0,t+1}) - g(u_{i,t+1})) > 0$$

The null hypothesis states that no competing model amongst the set of m alternatives yields more accurate forecasts than the benchmark model, for a given loss function; while the alternative hypothesis states that there is at least one alternative model that outperforms the benchmark model. By jointly considering all alternative models, the reality check controls the family-wise error rate (FWER), thus circumventing the issue of data snooping, i.e. sequential test bias.

Assumption 3.1: (i) $f_i(\cdot, \theta_i^\dagger)$ is twice continuously differentiable on the interior of Θ_i and the elements of $\nabla_{\theta_i} f_i(Z^t, \theta_i)$ and $\nabla^2_{\theta_i} f_i(Z^t, \theta_i)$ are p-dominated on Θ_i, for $i = 1, ..., m$, with

$p > 2(2+\psi)$, where ψ is the same positive constant defined in Assumption 1.1; (ii) $g(\cdot)$ is positively valued, twice continuously differentiable on Θ_i, and $g(\cdot)$, $g^{'}(\cdot)$, and $g^{''}(\cdot)$ are p-dominated on Θ_i, with p defined in (i); and

(iii) let $c_{ii} = \lim_{T\to\infty} \mathrm{Var}\left(T^{-1/2}\sum_{t=1}^{T}(g(u_{0,t+1}) - g(u_{i,t+1}))\right)$, $i = 1, ..., m$, define analogous covariance terms, c_{ji}, $j, i = 1, ..., m$, and assume that c_{ji} is positive semi-definite.

PROPOSITION 3.1 (Parts (i) and (iii) are from Proposition 2.2 in White (2000)): With Assumptions 1.1, 1.2 and 3.1, then under the null,

$$\max_{i=1,...,m}\left(\widehat{S}_n(0,i) - \sqrt{n}E\left(g(u_{0,t+1}) - g(u_{i,t+1})\right)\right) \xrightarrow{d} \max_{i=1,...,m} S(0,i)$$

where $S = (S(0,1), ..., S(0,m))^{'}$ is a zero mean Gaussian process with covariance matrix given by V, with V an $m\times m$ matrix, and: (i) If parameter estimation error vanishes, then for $i = 0, ..., m$,

$$V = S_{g_i g_i} = \sum_{\tau=-\infty}^{\infty} E\left(g(u_{0,1}) - g(u_{i,1})\right)\left(g(u_{0,1+\tau}) - g(u_{i,1+\tau})\right)$$

(ii) If parameter estimation error does not vanish, then

$$\begin{aligned} V =& S_{g_i g_i} + 2\Pi\mu_0^{'}A_0^{\dagger}C_{00}A_0^{\dagger}\mu_0 + 2\Pi\mu_i^{'}A_i^{\dagger}C_{ii}A_i^{\dagger}\mu_i \\ &- 4\Pi\mu_0^{'}A_0^{\dagger}C_{0i}A_i^{\dagger}\mu_i + 2\Pi S_{g_i q_0}A_0^{\dagger}\mu_0 - 2\Pi S_{g_i q_i}A_i^{\dagger}\mu_i \end{aligned}$$

where

$$C_{ii} = \sum_{\tau=-\infty}^{\infty} E(\nabla_{\theta_i}q_i(y_{1+s}, Z^s, \theta_i^{\dagger}))(\nabla_{\theta_i}q_i(y_{1+s+\tau}, Z^{s+\tau}, \theta_i^{\dagger}))^{'}$$

$$S_{g_i q_i} = \sum_{\tau=-\infty}^{\infty} E\left((g(u_{0,1}) - g(u_{i,1}))\right)\left(\nabla_{\theta_i}q_i(y_{1+s+\tau}, Z^{s+\tau}, \theta_i^{\dagger})\right)^{'}$$

$A_i^{\dagger} = (E(-\nabla^2_{\theta_i}q_i(y_t, Z^{t-1}, \theta_i^{\dagger})))^{-1}$, $\mu_i = E(\nabla_{\theta_i}g(u_{i,t+1}))$, and $\Pi = 1 - \pi^{-1}\ln(1+\pi)$. (iii) Under the alternative, $\Pr(n^{-1/2}|S_n| > \epsilon) \to 1$ as $n\to\infty$.

Of particular note is that since the maximum of a Gaussian process is not Gaussian, in general, the construction of critical values for inference is not straightforward. White

(2000) proposes two alternatives. The first is a simulation-based approach starting from a consistent estimator of V, say $\widehat{V}$. With $\widehat{V}$, for each simulation $s = 1, ..., S$, one realization is drawn from m-dimensional $N(0, \widehat{V})$ and the maximum value over $i = 1, ..., m$ is recorded. Repeat this procedure for S times, with a large S, and use the $(1-\alpha)$-percentile of the empirical distribution of the maximum values. A main drawback to this approach is that we need to first estimate the covariance structure V. However, if m is large and the prediction errors exhibit a high degree of heteroskedasticity and time dependence, the estimator of V becomes imprecise and thus the inference unreliable, especially in finite samples. The second approach relies on bootstrap procedures to construct critical values, which overcomes the problem of the first approach. We resample blocks of $g(\widehat{u}_{0,t+1}) - g(\widehat{u}_{i,t+1})$, and for each bootstrap replication $b = 1, ..., B$, we calculate

$$\widehat{S}_n^{*(b)}(0,i) = n^{-1/2} \sum_{t=R}^{T-1} (g^*(\widehat{u}_{0,t+1}) - g^*(\widehat{u}_{i,t+1})) \tag{4.5}$$

and the bootstrap statistic is given by

$$S_n^* = \max_{i=1,\dots,m} |\widehat{S}_n^{*(b)}(0,i) - \widehat{S}_n(0,i)|$$

the $(1-\alpha)$-percentile of the empirical distribution of B bootstrap statistics is then used for inference. Note that in White (2000), parameter estimation error is assumed to be asymptotically negligible. In light of this, Corradi and Swanson (2007) suggest a "re-centering" bootstrap procedure in order to explicitly handle the issue of non-vanishing parameter estimation error, when constructing critical values for this test. The new bootstrap statistic is defined as,

$$S_n^{**} = \max_{i=1,\dots,m} S_n^{**}(0,i)$$

where

$$\begin{aligned} S_n^{**}(0,i) =& n^{-1/2} \sum_{t=R}^{T-1} [(g(y_{t+1}^* - f_0(Z^{*,t}, \widetilde{\theta}_{0,t}^*)) - g(y_{t+1}^* - f_i(Z^{*,t}, \widetilde{\theta}_{i,t}^*))) \\ & - \frac{1}{T} \sum_{j=1}^{T-1} (g(y_{j+1} - f_0(Z^j, \widehat{\theta}_{0,t})) - g(y_{j+1} - f_i(Z^j, \widehat{\theta}_{i,t})))] \end{aligned}$$

Note that $S_n^{**}(0,i)$ is different from the standard bootstrap statistic in Equation (4.5), which is defined as the difference between the statistic constructed using original samples and that using bootstrap samples. The $(1-\alpha)$-percentile of the empirical distribution of S_n^{**} can be used to construct valid critical values for inference in the case of non-vanishing parameter estimation error. Proposition 2 in Corradi and Swanson (2007) establishes the first order validity for the recursive estimation scheme and Corradi and Swanson (2006a) outline the approach to constructing valid bootstrap critical values for the rolling window estimation scheme. Finally, note that Corradi and Swanson (2007) explain how to use the simple block bootstrap for constructing critical values when parameter estimation error is assumed to be asymptotically negligible. This procedure is perhaps the most obvious method to use for constructing critical values as it involves simply resampling the original data, carrying out the same forecasting procedures as used using the original data, and then constructing bootstrap statistics. These bootstrap statistics can be used (after subtracting the original test statistic from each of them) to form an empirical distribution which mimics the distribution of the test statistic under the null hypothesis. Finally, the empirical distribution can be used to construct critical values, which are the $(1-\alpha)$-quantiles of said distribution.

From Equation (4.4) and Proposition 3.1, it is immediate to see that the reality check can be rather conservative when a many alternative models are strictly dominated by the benchmark model. This is because those "bad" models do not contribute to the test statistic, simply because they are ruled out by the maximum, but contribute to the bootstrap statistics. Therefore, when many inferior models are included, the probability of rejecting the null hypothesis is actually smaller than α. Indeed, it is only for the least favorable case, in which $E(g(u_{0,t+1}) - g(u_{i,t+1})) = 0, \forall i$, that the distribution of $\widehat{S}_n$ coincides with that of

$$\max_{i=1,\ldots,m} \left(\widehat{S}_n(0,i) - \sqrt{n} E\left(g(u_{0,t+1}) - g(u_{i,t+1})\right) \right)$$

We introduce two approaches for addressing the conservative nature of this test below.

4.4.2 A test for superior predictive ability

Hansen (2005) proposes a modified reality check called the superior predictive ability (SPA) test that controls the FWER and addresses the inclusion of inferior models. The SPA test statistic is defined as,

$$T_n = \max\left\{0, \max_{i=1,\ldots,m} \frac{\widehat{S}_n(0,i)}{\sqrt{\widehat{\nu}_{i,i}}}\right\}$$

where $\widehat{\nu}_{i,i} = \frac{1}{B}\sum_{b=1}^{B}\left(\frac{1}{n}\sum_{t=R}^{T-1}((g(\widehat{u}_{0,t+1}) - g(\widehat{u}_{i,t+1})) - (g(\widehat{u}^*_{0,t+1}) - g(\widehat{u}^*_{i,t+1})))^2\right)$.

The bootstrap statistic is then defined as,

$$T_n^{*(b)} = \max\left\{0, \max_{i=1,\ldots,m}\{\frac{n^{-1/2}\sum_{t=R}^{T-1}(\widehat{d}_{i,t}^{*(b)} - \widehat{d}_{i,t}\mathbf{1}_{\{\widehat{d}_{i,t}\geq -A_{T,i}\}})}{\sqrt{\widehat{\nu}_{i,i}}}\}\right\}$$

where $\widehat{d}_{i,t}^{*(b)} = g(\widehat{u}^*_{0,t+1}) - g(\widehat{u}^*_{i,t+1})$, $\widehat{d}_{i,t} = g(\widehat{u}_{0,t+1}) - g(\widehat{u}_{i,t+1})$, and $A_{T,i} = \frac{1}{4}T^{-1/4}\sqrt{\widehat{\nu}_{i,i}}$.

The idea behind the construction of SPA bootstrap critical values is that when a competing model is too slack, the corresponding bootstrap moment condition is not re-centered, and the bootstrap statistic is not affected by this model. Therefore, the SPA test is less conservative than the reality check. Corradi and Distaso (2011) derive a general class of SPA tests using the generalized moment selection approach of Andrews and Soares (2010) and show that Hansen's SPA test belongs to this class. Romano and Wolf (2005) propose a multiple step extension of the reality check which ensures tighter control of irrelevant models.

4.4.3 A test based on sub-sampling

The conservative property of the reality check can be alleviated by using the sub-sampling approach to constructing critical values, at the cost of sacrificing power in finite samples. Critical values are obtained from the empirical distribution of a sequence of statistics constructed using subsamples of size $\widetilde{b}$, where $\widetilde{b}$ grows with the sample size, but at a slower rate (see Politis et al. (1999)).

In the context of the reality check, as $n \to \infty, \widetilde{b} \to \infty$, and $\widetilde{b}/n \to 0$, define

$$S_{n,a,\widetilde{b}} = \max_{i=1,\ldots,m} S_{n,a,\widetilde{b}}(0,i), \quad a = R, ..., T - \widetilde{b} - 1$$

where

$$S_{n,a,\widetilde{b}}(0,i) = \widetilde{b}^{-1/2}\sum_{t=a}^{a+\widetilde{b}-1}(g(\widehat{u}_{0,t+1}) - g(\widehat{u}_{i,t+1}))$$

We obtain the empirical distribution of $T-\widetilde{b}-1$ statistics, $S_{n,a,\widetilde{b}}$, and reject the null if the test statistic $\widehat{S}_n$ is greater than the $(1-\alpha)$-quantile of the empirical distribution. The advantage of the sub-sampling approach over the bootstrap is that the test has correct size when $\max_{i=1,\ldots,m} E(g(\widehat{u}_{0,t+1})-g(\widehat{u}_{i,t+1}))<0$ for some i, while the bootstrap approach delivers a conservative test in this case. However, although the sub-sampling approach ensures that the test has unit asymptotic power, the finite sample power may be rather low, since $S_{n,a,\widetilde{b}}$ diverges at rate $\sqrt{\widetilde{b}}$ instead of $\sqrt{n}$, under the alternative. Finally, note that the sub-sampling approach is also valid in the case of non-vanishing parameter estimation error because each statistic constructed using subsamples properly mimics the distribution of actual statistic.

Part II: Forecast Evaluation Using Density Based Predictive Accuracy Tests

Note: Much for this part are drawn from Corradi and Swanson (2006c), Corradi and Swanson (2004), and Corradi and Swanson (2013). Norman Swanson is one of the coauthors of this chapter.

In Part I, we introduced a variety of tests designed for comparing models based on point forecast accuracy. However, there are many practical situations in which economic decision making crucially depends not only on conditional mean forecasts (e.g. point forecasts), but also on predictive confidence intervals or predictive conditional distributions (also called predictive densities). One such case, for instance, is when value at risk (VaR) measures are used in risk management for assessment of the amount of projected financial losses due to extreme tail behavior, e.g. catastrophic events. Another common case is when economic agents are undertaking to optimize their portfolio allocations, in which case the joint distribution of multiple assets is required to be modeled and fully understood. The purpose of this section is to discuss recent tests for comparing (potentially misspecified) conditional distribution models.

4.5 The Kullback-Leibler information criterion approach

A well-known measure of distributional accuracy is the Kullback-Leibler Information Criterion (KLIC). Using the KLIC involves simply choosing the model which minimizes the KLIC (see, e.g., White (1982), Vuong (1989), Gianni and R (2007), Kitamura (2002)). Of note is that White (1982) shows that quasi maximum likelihood estimators minimize the KLIC, under mild conditions. In order to implement the KLIC, one might choose model 0 over model 1, if

$$E(\ln f_0(y_t|Z^t, \theta_0^\dagger) - \ln f_1(y_t|Z^t, \theta_1^\dagger)) > 0$$

For the i.i.d case, Vuong (1989) suggests using a likelihood ratio test for choosing the conditional density model that is closer to the "true" conditional density, in terms of the KLIC. Gianni and R (2007) suggests using a weighted version of the likelihood ratio test proposed in Vuong (1989) for the case of dependent observations, while Kitamura (2002) employs a KLIC-based approach to select among misspecified conditional models that satisfy given moment conditions. Furthermore, the KLIC approach has recently been employed for the evaluation of dynamic stochastic general equilibrium models (see e.g., Schorfheide (2010), Fernández-Villaverde and Rubio-RamíRez (2004), and Chang et al. (2002)). For example, Fernández-Villaverde and Rubio-RamíRez (2004) show that the KLIC-best model is also the model with the highest posterior probability.

The KLIC is a sensible measure of accuracy, as it chooses the model which on average gives higher probability to events which have actually occurred. Also, it leads to simple likelihood ratio type tests which have a standard limiting distribution and are not affected by problems associated with accounting for parameter estimation error. However, it should be noted that if one is interested in measuring accuracy over a specific region, or in measuring accuracy for a given conditional confidence interval, say, this cannot be done in as straightforward manner using the KLIC. For example, if we want to evaluate the accuracy of different models for approximating the probability that the rate of inflation tomorrow, given the rate of inflation today, will be between 0.5% and 1.5%, say, we can do so quite easily using the square error criterion, but not using the KLIC.

4.6 A predictive density accuracy test for comparing multiple misspecified models

Corradi and Swanson (2005) (CSa) and Corradi and Swanson (2006b) (CSb) introduce a measure of distributional accuracy, which can be interpreted as a distributional generalization of mean square error. In addition, Corradi and Swanson (2005) apply this measure to the problem of selecting amongst multiple misspecified predictive density models. In this section we discuss these contributions to the literature.

Consider forming parametric conditional distributions for a scalar random variable, y_t, given Z^t, where $Z^t = (y_{t-1}, ..., y_{t-s_1}, X_t, ..., X_{t-s_2+1})$, with s_1, s_2 finite. With a little abuse of notation, now we define the group of conditional distribution models, from which one wishes to select a "best" model, as

$$\{F_i(u|Z^t, \theta_i^\dagger)\}_{i=1,...,m},$$

and define the true conditional distribution as

$$F_0(u|Z^t, \theta_0) = \Pr(y_{t+1} \leq u|Z^t)$$

Assume that $\theta_i^\dagger \in \Theta_i$, where Θ_i is a compact set in a finite dimensional Euclidean space, and let $\theta_i^\dagger$ be the probability limit of a quasi maximum likelihood estimator (QMLE) of the parameters of the conditional distribution under model i. If model i is correctly specified, then $\theta_i^\dagger = \theta_0$. If $m > 2$, follow White (2000). Namely, choose a particular conditional distribution model as the "benchmark" and test the null hypothesis that no competing model can provide a more accurate approximation of the "true" conditional distribution, against the alternative that at least one competitor outperforms the benchmark model. Needless to say, pairwise comparison of alternative models, in which no benchmark need be specified, follows as a special case.

In this context, measure accuracy using the above distributional analog of mean square error. More precisely, define the mean square (approximation) error associated with model

i, in terms of the average over U of $E\left((F_i(u|Z^t,\theta_i^\dagger) - F_0(u|Z^t,\theta_0))^2\right)$, where $u \in U$, and U is a possibly unbounded set on the real line, and the expectation is taken with respect to the conditioning variables. In particular, model 1 is more accurate than model 2, if

$$\int_U E((F_1(u|Z^t,\theta_1^\dagger) - F_0(u|Z^t,\theta_0))^2 - (F_2(u|Z^t,\theta_2^\dagger) - F_0(u|Z^t,\theta_0))^2)\phi(u)du < 0$$

where $\int_U \phi(u)du = 1$ and $\phi(u)du \geq 0$, $\forall u \in U \in \Re$.

This measure integrates over different quantiles of the conditional distribution. For any given evaluation point, this measure defines a norm and it implies a standard goodness of fit measure. Note that this measure of accuracy leads to straightforward evaluation of distributional accuracy over a given region of interest, as well as to straightforward evaluation of specific quantiles. A conditional confidence interval version of the above condition which is more natural to use in applications involving predictive interval comparison follows immediately, and can be written as

$$E\Big(((F_1(\bar{u}|Z^t,\theta_1^\dagger) - F_1(\underline{u}|Z^t,\theta_1^\dagger)) - (F_1(\bar{u}|Z^t,\theta_0) - F_1(\underline{u}|Z^t,\theta_0)))^2$$
$$-((F_2(\bar{u}|Z^t,\theta_2^\dagger) - F_2(\underline{u}|Z^t,\theta_2^\dagger)) - (F_1(\bar{u}|Z^t,\theta_0) - F_1(\underline{u}|Z^t,\theta_0)))^2\Big) \leq 0$$

Hereafter, $F_1(\cdot|\cdot,\theta_1^\dagger)$ is taken as the benchmark model, and the objective is to test whether some competitor model can provide a more accurate approximation of $F_0(\cdot|\cdot,\theta_0)$ than the benchmark.The null and the alternative hypotheses are:

$$H_0 : \max_{i=2,\ldots,m} \int_U E((F_1(u|Z^t,\theta_1^\dagger) - F_0(u|Z^t,\theta_0))^2$$
$$-(F_i(u|Z^t,\theta_i^\dagger) - F_0(u|Z^t,\theta_0))^2)\phi(u)du \leq 0$$

versus

$$H_A : \max_{i=2,\ldots,m} \int_U E((F_1(u|Z^t,\theta_1^\dagger) - F_0(u|Z^t,\theta_0))^2$$
$$-(F_i(u|Z^t,\theta_i^\dagger) - F_0(u|Z^t,\theta_0))^2)\phi(u)du > 0,$$

where $\phi(u) \geq 0$ and $\int_U \phi(u) = 1$, $u \in U \in \Re$, U possibly unbounded. Note that for a given u, we compare conditional distributions in terms of their (mean square) distance from the true distribution. We then average over U. As discussed above, a possibly more natural version of the above hypotheses is in terms of conditional confidence intervals evaluation, so that the objective is to "approximate" $\Pr(\underline{u} \leq Y_{t+1} \leq \bar{u}|Z^t)$, and hence to evaluate a region of the predictive density. In that case, the null and alternative hypotheses can be stated as:

$$H_0' : \max_{i=2,\ldots,m} E(((F_1(\overline{u}|Z^t\ ,\ \theta_1^{\dagger}) - F_1(\underline{u}|Z^t\ ,\ \theta_1^{\mathrm{f}}))$$
$$-(F_0(\overline{u}|Z^t,\ \theta_0) - F_0(\underline{u}|Z^t,\ \theta_0)))^2$$
$$-((F_i(\overline{u}|Z^t,\ \theta_i^{\dagger}) - F_i(\underline{u}|Z^t,\ \theta_i^{\dagger}))$$
$$-(F_0(\overline{u}|Z^t,\ \theta_0) - F_0(\underline{u}|Z^t,\ \theta_0)))^2)\ \leq 0$$

versus

$$H_A' : \max_{i=2,\ldots,m} E(((F_1(\overline{u}|Z^t\ ,\ \theta_1^{\dagger}) - F_1(\underline{u}|Z^t\ ,\ \theta_1^{\mathrm{f}}))$$
$$-(F_0(\overline{u}|Z^t,\ \theta_0) - F_0(\underline{u}|Z^t,\ \theta_0)))^2$$
$$-((F_k(\overline{u}|Z^t,\ \theta_i^{\dagger}) - F_i(\underline{u}|Z^t,\ \theta_i^{\dagger}))$$
$$-(F_0(\overline{u}|Z^t,\ \theta_0) - F_0(\underline{u}|Z^t,\ \theta_0)))^2)\ > 0$$

Alternatively, if interest focuses on testing the null of equal accuracy of two conditional distribution models, say F_1 and F_i, we can simply state the hypotheses as:

$$H_0'' : \int_U E((F_1(u|Z^t,\ \theta_1^{\dagger}) - F_0(u|Z^t,\ \theta_0))^2$$
$$-\ (F_i\ (u|Z^t,\ \theta_i^{\dagger}) - F_0(u|Z^t,\ \theta_0))^2)\phi(u)\mathrm{d}u = 0$$

versus

$$H_A'' : \int_U E((F_1(u|Z^t,\ \theta_1^{\dagger}) - F_0(u|Z^t,\ \theta_0))^2$$
$$-\ (F_i\ (u|Z^t,\ \theta_i^{\dagger}) - F_0(u|Z^t,\ \theta_0))^2)\phi(u)\mathrm{d}u \neq 0,$$

or we can write the predictive density (interval) version of these hypotheses.

Of course, we do not know $F_0(u|Z^t)$. However, it is easy to see that

$$\begin{aligned} E((F_1(u|Z^t,\ \theta_1^\dagger) - F_0(u|Z^t,\ \theta_0))^2 - (F_i(u|Z^t,\ \theta_i^\dagger) - F_0(u|Z^t,\ \theta_0))^2) \\ = E((1\{y_{t+1}\ \leq u\} - F_1(u|Z^t,\ \theta_1^\dagger))^2) \\ - E\ ((1\{y_{t+1}\ \leq u\} - F_i\ (u|Z^t,\ \theta_i^\dagger))^2), \end{aligned} \tag{4.6}$$

where the right-hand side of Equation (4.6) does not require any knowledge of the true conditional distribution.

The intuition behind Equation (4.6) is very simple. First, note that for any given u, $E(1\{y_{t+1} \leq u\}|Z^t) = \Pr(y_{t+1} \leq u|Z^t) = F_0(u|Z^t,\ \theta_0)$. Thus, $1\{y_{t+1} \leq u\} - F_i(u|Z^t,\ \theta_i^\dagger)$ can be interpreted as an "error" term associated with computation of the conditional expectation under F_i. Now, for $i = 1, ..., m$:

$$\mu_i^2(u) = E(\ (1\{y_{t+1}\ \leq u\} - F_i\ (u|Z^t,\ \theta_i^\dagger))^2)$$

$$= E(((1\{y_{t+1}\ \leq u\} - F_0(u|Z^t,\ \theta_0)) - (F_i(u|Z^t,\ \theta_i^\dagger) - F_0(u|Z^t,\ \theta_0)))^2)$$

$$= E((1\{y_{t+1}\ \leq u\} - F_0(u|Z^t,\ \theta_0))^2) + E\ ((F_i\ (u|Z^t,\ \theta_i^\dagger) - F_0(u|Z^t,\ \theta_0))^2),$$

given that the expectation of the cross product is zero (which follows because $1\ \{y_{t+1} \leq u\} - F_0(u|Z^t,\ \theta_0)$ is uncorrelated with any measurable function of Z^t). Therefore,

$$\begin{aligned} \mu_1^2(u) - \mu_i^2(u) = E(\ (F_1\ (u|Z^t,\ \theta_1^\dagger) - F_0(u|Z^t,\ \theta_0))^2) \\ - E\ ((F_i\ (u|Z^t,\ \theta_i^\dagger) - F_0(u|Z^t,\ \theta_0))^2) \end{aligned} \tag{4.7}$$

The statistic of interest is

$$Z_{n,j} = \max_{i=2,\ldots m} \int_U Z_{n,u,j}(1,\ i)\phi(u)\mathrm{d}u, \quad j = 1, 2,$$

where for $j = 1$ (rolling estimation scheme),

$$Z_{n,u,1}(1,\ i) = \frac{1}{\sqrt{n}} \sum_{t=R}^{T-1} ((1\{y_{t+1}\ \leq u\} - F_1(u|Z^t,\ \widehat{\theta}_{1,t,\mathrm{rol}}))^2$$

$$- (1\{y_{t+1}\ \leq u\} - F_i(u|Z^t,\ \widehat{\theta}_{i,t,\mathrm{rol}}))^2)$$

and for $j = 2$ (recursive estimation scheme),

$$Z_{n,u,2}(1,\ i) = \frac{1}{\sqrt{n}} \sum_{\mathrm{t}=R}^{T-1} ((1\{y_{t+1}\ \leq u\} - F_1(u|Z^t,\ \widehat{\theta}_{1,t,\mathrm{rec}}))^2$$

$$-\ (1\{y_{t+1}\ \leq u\} - F_i(u|Z^t,\ \widehat{\theta}_{i,t,\mathrm{rec}}))^2),$$

where $\widehat{\theta}_{i,t,\mathrm{rol}}$ and $\widehat{\theta}_{i,t,\mathrm{rec}}$ are defined as:

$$\widehat{\theta}_{i,t,\mathrm{rol}} = \arg\min_{\theta\in\Theta} \frac{1}{R} \sum_{j=t-R+1}^{t} q(y_j, Z^{j-1}, \theta),\ \ R \leq t \leq T-1$$

and

$$\widehat{\theta}_{i,t,\mathrm{rec}} = \arg\min_{\theta\in\Theta} \frac{1}{t} \sum_{j=1}^{t} q(y_j, Z^{j-1}, \theta),\ \ t = R, R+1, R+n-1$$

As shown above and in Corradi and Swanson (2005), the hypotheses of interest can be restated as:

$$H_0 : \max_{i=2,\ldots,m} \int_U (\mu_1^2(u) - \mu_i^2(u))\phi(u)\mathrm{d}u \leq 0$$

versus

$$H_A : \max_{i=2,\ldots,m} \int_U (\mu_1^2(u) - \mu_i^2(u))\phi(u)\mathrm{d}u > 0$$

where $\mu_i^2(u) = E(\ (1\{y_{t+1}\ \leq u\} - F_i\ (u|Z^t,\ \theta_i^\dagger))^2)$.

Assumption 6.1: (i) $\theta_i^\dagger$ is uniquely defined,

$$E(\ln(f_i(y_t, Z^{t-1}, \theta_i))) < E(\ln(f_i(y_t, Z^{t-1}, \theta_i^\dagger))),$$

for any $\theta_i \neq \theta_i^\dagger$; (ii) $\ln f_i$ is twice continuously differentiable on the interior of Θ_i, and $\forall\Theta_i$ a compact subset of $\Re^{\varrho(i)}$; (iii) the elements of $\nabla_{\theta_i} \ln f_i$ and $\nabla^2_{\theta_i} \ln f_i$ are p-dominated on Θ_i, with $p > 2(2+\psi)$, where ψ is the same positive constant as defined in Assumption 1.1; and (iv) $E(-\nabla^2_{\theta_i} \ln f_i)$ is negatively definite uniformly on Θ_i.

Assumption 6.2: $T = R + n$, and as $T \to \infty$, $n/R \to \pi$, with $0 < \pi < \infty$.

Assumption 6.3: (i) $F_i(u|Z^t,\theta_i)$ is continuously differentiable on the interior of Θ_i and $\nabla_{\theta_i}F_i(u|Z^t,\theta_i^\dagger)$ is $2r$-dominated on Θ_i, uniformly in u, $r>2$, $\forall i$;[1] and (ii) let

$$v_{ii}(u) = \text{plim}_{T\to\infty}\text{Var}\Big(\frac{1}{\sqrt{T}}\sum_{t=s}^{T}(((1\{y_{t+1}\leq u\}-F_1(u|Z^t,\theta_1^\dagger))^2-\mu_1^2(u))$$

$$-((1\{y_{t+1}\leq u\}-F_i(u|Z^t,\theta_i^\dagger))^2-\mu_i^2(u))\Big),\ \forall i$$

define analogous covariance terms, $v_{j,i}(u), j,i=2,...,m$, and assume that $[v_{j,i}(u)]$ is positive semi-definite, uniformly in u.

PROPOSITION 6.1 (From Proposition 1 in Corradi and Swanson (2006b)): With Assumptions 1.1, 6.1–6.3, then

$$\max_{i=2,\ldots,m}\int_U (Z_{n,u,j}\ (1\ ,\ i)-\sqrt{n}(\mu_1^2(u)-\mu_i^2(u)))\phi_U(u)\mathrm{d}u$$

$$\xrightarrow{d}\max_{i=2,\ldots,m}\int_U Z_{1,i,j}(u)\phi_U(u)\mathrm{d}u$$

where $Z_{1,i,j}(u)$ is a zero mean Gaussian process with covariance $C_{i,j}(u,\ u')(j=1$ corresponds to rolling and $j=2$ to recursive estimation schemes), equal to:

$$E(\sum_{j=-\infty}^{\infty}((1\{y_{s+1}\ \leq u\}-F_1(u|Z^s,\ \theta_1^\dagger))^2-\mu_1^2(u))\times\ ((1\{y_{s+j+1}\ \leq u'\}$$

$$-F_1(u'|Z^{s+j},\ \theta_1^\dagger))^2-\mu_1^2(u')))+E(\sum_{j=-\infty}^{\infty}((1\{y_{s+1}\ \leq u\}-F_i(u|Z^s,\ \theta_i^\dagger))^2-\mu_i^2(u))$$

$$\times\ ((1\{y_{s+j+1}\ \leq u'\}-F_i(u'|Z^{s+j},\ \theta_i^\dagger))^2-\mu_i^2(u')))-2E(\sum_{j=-\infty}^{\infty}((1\{y_{s+1}\ \leq u\}$$

$$-F_1(u|Z^s,\ \theta_1^\dagger))^2-\mu_1^2(u))\times\ ((1\{y_{s+j+1}\ \leq u'\}-F_i(u'|Z^{s+j},\ \theta_i^\dagger))^2-\mu_i^2(u^{'})))$$

$$+4\Pi_j m_{\theta_1^\dagger}(u)'A(\theta_1^\dagger)\times E(\sum_{j=-\infty}^{\infty}\nabla_{\theta_1}\ln f_1(y_{s+1}|Z^s,\ \theta_1^\dagger)\nabla_{\theta_1}\ln f_1(y_{s+j+1}|Z^{s+j},\ \theta_1^\dagger)')$$

$$\times A(\theta_1^\dagger)m_{\theta_1^\dagger}(u')+4\Pi_j m_{\theta_i^\dagger}(u)'A(\theta_i^\dagger)\times E(\sum_{j=-\infty}^{\infty}\nabla_{\theta_i}\ln f_i(y_{s+1}|Z^s,\ \theta_i^\dagger)$$

---

[1]We require that for $j=1,...,p_i$, $E(\nabla_\theta F_i(u|Z^t,\theta_i^\dagger))_j\geq D_t(u)$, with $\sup_t\sup_{u\in\Re}E(D_t(u)^{2r})<\infty$.

$$\times\nabla_{\theta_i}\ln f_i(y_{s+j+1}|Z^{s+j},\ \theta_i^{\dagger})') \times A(\theta_i^{\dagger})m_{\theta_i^{\dagger}}(u') - 4\Pi_j m_{\theta_1^{\dagger}}(u,\ )'A(\theta_1^{\dagger})$$

$$\times E(\sum_{j=-\infty}^{\infty}\nabla_{\theta_1}\ln f_1(y_{s+1}|Z^s,\ \theta_1^{\dagger})\nabla_{\theta_i}\ln f_i(y_{s+j+1}|Z^{s+j}\times A(\theta_i^{\dagger})m_{\theta_i^{\dagger}}(u')$$

$$-4C\Pi_j m_{\theta_1^{\dagger}}(u)'A(\theta_1^{\dagger})\times E(\sum_{j=-\infty}^{\infty}\nabla_{\theta_1}\ln f_1(y_{s+1}|Z^s,\ \theta_1^{\dagger})\times((1\{y_{s+j+1}\ \leq u\}$$

$$-F_1\ (u|Z^{s+j},\ \theta_1^{\dagger}))^2-\mu_1^2(u)))+4C\Pi_j m_{\theta_1^{\dagger}}(u)'A(\theta_1^{\dagger})\times E(\sum_{j=-\infty}^{\infty}\nabla_{\theta_1}\ln f_1(y_{s+1}|Z^s,\ \theta_1^{\dagger})$$

$$\times((1\{y_{s+j+1}\ \leq u\}-F_i(u|Z^{s+j},\ \theta_i^{\dagger}))^2-\mu_i^2(u)))-4C\Pi_j m_{\theta_i^{\dagger}}(u)'A(\theta_i^{\dagger})$$

$$\times E(\sum_{j=-\infty}^{\infty}\nabla_{\theta_i}\ln f_i(y_{s+1}|Z^s,\theta_i^{\dagger})'\times((1\{y_{s+j+1}\ \leq u\}-F_i(u|Z^{s+j},\ \theta_i^{\dagger}))^2-\mu_i^2(u)))$$

$$+4C\Pi_j m_{\theta_i^{\dagger}}(u)'A(\theta_i^{\dagger})\times E(\sum_{j=-\infty}^{\infty}\nabla_{\theta_i}\ln f_i(y_{s+1}|Z^s,\ \theta_i^{\dagger})'\times\ \ ((1\{y_{s+j+1}\ \leq u\}$$

$$-F_1(u|Z^{s+j},\ \theta_1^{\dagger}))^2-\mu_1^2(u)))$$

with

$$m_{\theta_i^{\dagger}}(u)' = E(\nabla_{\theta_i}F_i(u|Z^t,\ \theta_i^{\dagger})'(1\{y_{t+1}\ \leq u\}-F_i(u|Z^t,\ \theta_i^{\dagger})))$$

and

$$A(\theta_i^{\dagger})\ = A_i^{\dagger} =\ (E(-\nabla^2_{\theta_i}\ln f_i(y_{t+1}|Z^t,\ \theta_i^{\dagger})))^{-1}$$

and for $j=1$ and $n\leq R$, $\Pi_1=(\pi-\frac{\pi^2}{3})$, $C\Pi_1=\frac{\pi}{2}$, and for $n>R$, $\Pi_1=(1-\frac{1}{3\pi})$ and $C\Pi_1=(1-\frac{1}{2\pi})$. Finally, for $j=2$, $\Pi_2=2(1-\pi^{-1}\ln(1+\pi))$ and $C\Pi_2=0.5\Pi_2$.

From this proposition, note that when all competing models provide an approximation to the true conditional distribution that is as (mean square) accurate as that provided by the benchmark (i.e. when $\int_U(\mu_1^2(u)-\mu_i^2(u))\phi(u)\mathrm{d}u=0,\forall i$), then the limiting distribution is a zero mean Gaussian process with a covariance kernel which is not nuisance parameter free. Additionally, when all competitor models are worse than the benchmark, the statistic diverges to minus infinity at rate $\sqrt{n}$. Finally, when only some competitor models are worse than the benchmark, the limiting distribution provides a conservative test, as Z_P will always be smaller than $\max_{i=2,\ldots,m}\int_U(Z_{n,u}\ (1,\ i)-\sqrt{n}(\mu_1^2(u)-\mu_i^2(u)))\phi(u)\mathrm{d}u$, asymptotically. Of course, when H_A holds, the statistic diverges to plus infinity at rate $\sqrt{n}$.

For the case of evaluation of multiple conditional confidence intervals, consider the statistic:

$$V_{n,\tau} = \max_{i=2,\ldots,m} V_{n,\underline{u},\overline{u},\tau}(1,\ i)$$

where

$$V_{n,\underline{u},\overline{u},\tau}(1,\ i) = \frac{1}{\sqrt{n}} \sum_{t=R}^{T-1} ((1\{\underline{u} \le y_{t+1} \ \le \overline{u}\} - (F_1(\overline{u}|Z^t,\ \widehat{\theta}_{1,t,\tau})$$
$$-F_1(\underline{u}|Z^t,\ \widehat{\theta}_{1,t,\tau})))^2 - (1\{\underline{u} \le y_{t+1} \ \le \overline{u}\} - (F_i(\overline{u}|Z^t,\ \widehat{\theta}_{i,t,\tau}) - F_i(\underline{u}|Z^t,\ \widehat{\theta}_{i,t,\tau})))^2)$$

where $s = \max\{s1,\ s2\}, \tau = 1, 2$, and $\widehat{\theta}_{i,t,\tau} = \widehat{\theta}_{i,t,\text{rol}}$ for $\tau = 1$, and $\widehat{\theta}_{i,t,\tau} = \widehat{\theta}_{k,t,\text{rec}}$ for $\tau = 2$.

We then have the following result,

PROPOSITION 6.2 (From Proposition lb in Corradi and Swanson (2006b)): With Assumptions 1.1, 6.1–6.3, then for $\tau = 1$,

$$\max_{i=2,\ldots m} (V_{n,\underline{u},\overline{u},\tau} \ (1\ ,\ i) \ - \ \sqrt{n}(\mu_1^2 - \mu_i^2)) \ \xrightarrow{d} \max_{i=2,\ldots m} V_{n,i,\tau}(\underline{u},\ \overline{u})$$

where $V_{n,i,\tau}(\underline{u},\ \overline{u})$ is a zero mean normal random variable with covariance $c_{ii} = v_{ii} + p_{ii} + cp_{ii}$, where v_{ii} denotes the component of the long-run variance matrix we would have in absence of parameter estimation error, p_{ii} denotes the contribution of parameter estimation error and cp_{ii} denotes the covariance across the two components. In particular:

$$v_{ii} = E \sum_{j=-\infty}^{\infty} (((1\{\underline{u} \le y_{s+1} \ \le \overline{u}\} - (F_1(\overline{u}|Z^s,\ \theta_1^\dagger) - F_1(\underline{u}|Z^s,\ \theta_1^\dagger)))^2 - \mu_1^2)$$
$$\times\ ((1\{\underline{u} \le y_{s+1+j} \ \le \overline{u}\} - (F_1(\overline{u}|Z^{s+j},\ \theta_1^\dagger) - F_1(\underline{u}|Z^{s+j},\ \theta_1^\dagger)))^2 - \mu_1^2))$$
$$+E \sum_{j=-\infty}^{\infty} (((1\{\underline{u} \le y_{s+1} \ \le \overline{u}\} - (F_i(\overline{u}|Z^s,\ \theta_i^\dagger) - F_i(\underline{u}|Z^s,\ \theta_i^\dagger)))^2 - \mu_i^2)$$
$$\times\ ((1\{\underline{u} \le y_{s+1+j} \ \le \overline{u}\} - (F_i(\overline{u}|Z^{s+j},\ \theta_i^\dagger) - F_i(\underline{u}|Z^{s+j},\ \theta_i^\dagger)))^2 - \mu_i^2))$$
$$-2E \sum_{j=-\infty}^{\infty} (((1\{\underline{u} \le y_{s+1} \ \le \overline{u}\} - (F_1(\overline{u}|Z^s,\ \theta_1^\dagger) - F_1(\underline{u}|Z^s,\ \theta_1^\dagger)))^2 - \mu_1^2)$$
$$\times\ ((1\{\underline{u} \le y_{s+1+j} \ \le \overline{u}\} - (F_i(\overline{u}|Z^{s+j},\ \theta_i^\dagger) - F_i(\underline{u}|Z^{s+j},\theta_i^\dagger)))^2 - \mu_i^2))$$

Also,

$$p_{ii} = 4m'_{\theta_1^\dagger} A(\theta_1^\dagger) E(\sum_{j=-\infty}^{\infty} \nabla_{\theta_1} \ln f_i(y_{s+1}|Z^s,\ \theta_1^\dagger) \nabla_{\theta_1} \ln f_i(y_{s+1+j}|Z^{s+j},\ \theta_1^\dagger)') \times A(\theta_1^\dagger) m_{\theta_1^\dagger}$$

$$+4m'_{\theta_i^\dagger} A(\theta_i^\dagger) E(\sum_{j=-\infty}^{\infty} \nabla_{\theta_i} \ln f_i(y_{s+1}|Z^s,\ \theta_i^\dagger) \nabla_{\theta_i} \ln f_i(y_{s+1+j}|Z^{s+j},\ \theta_i^\dagger)') \times A(\theta_i^\dagger) m_{\theta_i^\dagger}$$

$$-8m'_{\theta_1^\dagger} A(\theta_1^\dagger) E(\nabla_{\theta_1} \ln f_1(y_{s+1}|Z^s,\ \theta_1^\dagger) \nabla_{\theta_i} \ln f_i(y_{s+1+j}|Z^{s+j},\ \theta_i^\dagger)') \times A(\theta_i^\dagger) m_{\theta_i^\dagger}$$

Finally,

$$cp_{ii} = -4m'_{\theta_1^\dagger} A(\theta_1^\dagger) E(\sum_{j=-\infty}^{\infty} \nabla_{\theta_1} \ln f_1(y_{s+1}|Z^s,\ \theta_1^\dagger)$$

$$\times\ ((1\{\underline{u} \leq y_{s+j}\ \leq \overline{u}\} - (F_1(\overline{u}|Z^{s+j},\ \theta_1^\dagger) - F_1(\underline{u}|Z^{s+j},\ \theta_1^\dagger)))^2 - \mu_1^2)$$

$$+8m'_{\theta_1^\dagger} A(\theta_1^\dagger) E(\sum_{j=-\infty}^{\infty} \nabla_{\theta_1} \ln f_1(y_s|Z^s,\ \theta_1^\dagger)$$

$$\times\ ((1\{\underline{u} \leq y_{s+1+j}\ \leq \overline{u}\} - (F_i(\overline{u}|Z^{s+j}, \theta_i^\dagger) - F_i(\underline{u}|Z^s,\ \theta_i)))^2 - \mu_i^2))$$

$$-4m'_{\theta_i^\dagger} A(\theta_i^\dagger) E(\sum_{j=-\infty}^{\infty} \nabla_{\theta_i} \ln f_i(y_{s+1}|Z^s,\ \theta_i^\dagger)$$

$$\times\ ((1\{\underline{u} \leq y_{s+j}\ \leq \overline{u}\} - (F_i(\overline{u}|Z^{s+j},\ \theta_i^\dagger)\ - F_i(\underline{u}|Z^{s+j},\ \theta_i^\dagger)))^2 - \mu_i^2))$$

with

$$m'_{\theta_i^\dagger}\ = E(\nabla_{\theta_i}(F_i(\overline{u}|Z^t,\ \theta_i^\dagger)\ - F_i(\overline{u}|Z^t,\ \theta_i^\dagger))$$

$$\times\ (1\{\underline{u} \leq y_t\ \leq \overline{u}\} - (F_i(\overline{u}|Z^t,\ \theta_i^\dagger) - F_i(\overline{u}|Z^t,\ \theta_i^\dagger))))$$

and

$$A(\theta_i^\dagger)\ =\ (E(-\ln \nabla^2_{\theta_i} f_i(y_t|Z^t,\ \theta_i^\dagger)))^{-1}$$

An analogous result holds for the case where $\tau = 2$, and is omitted for the sake of brevity.

Due to the contribution of parameter estimation error, simulation error, and the time series dynamics to the covariance kernel (see Proposition 6.1), critical values cannot be directly tabulated. As a result, block bootstrap techniques are used to construct valid critical values for statistical inference. In order to show the first order validity of the bootstrap, the authors derive the limiting distribution of appropriately formed bootstrap statistics and show that they coincide with the limiting distribution given in Proposition 6.1.

Recalling that as all candidate models are potentially misspecified under both hypotheses, the parametric bootstrap is not generally applicable in our context. Instead, we must begin by resampling b blocks of length $l, bl = T-1$. Let $\mathrm{Y}_t^* = (\Delta \log X_t^*,\ \Delta \log X_{t-1}^*)$ be the resampled series, such that Y_2^*,..., Y_{l+1}^*, Y_{l+2}^*,..., Y_{T-l+2}^*,..., Y_T^* equals Y_{I_1+1},..., Y_{I_1+l}, Y_{I_2+1},..., Y_{I_b+1},..., Y_{I_b+T}, where I_j, $i=1$,..., b are independent, discrete uniform random variates on 1,..., $T-1+1$. That is, $I_j = i, i = 1$,..., $T-l$ with probability $1/(T-l)$. Then, use Y_t^* to compute $\widehat{\theta}_{j,T}^*$ and plug in $\widehat{\theta}_{j,T}^*$ in order to simulate a sample under model $j, j=1$,..., m. Let $\mathrm{Y}_{j,n}(\widehat{\theta}_{j,T}^*)$, $n=2$,..., S denote the series simulated in this manner. At this point, we need to distinguish between the case where $\delta = 0$ (vanishing simulation error) and $\delta > 0$ (non-vanishing simulation error). In the former case, we do not need to resample the simulated series, as there is no need to mimic the contribution of simulation error to the covariance kernel. On the other hand, in the latter case we draw $\widetilde{b}$ blocks of length $\widetilde{l}$ with $\widetilde{b}l = S-1$, and let $\mathrm{Y}_{j,n}^*(\widehat{\theta}_{j,T}^*)$, $j=1$,..., m, $n=2$,..., S denote the resampled series under model j. Notice that $Y_{j,2}^*(\widehat{\theta}_{j,T}^*)$,..., $\mathrm{Y}_{j,l+1}^*(\widehat{\theta}_{j,T}^*)$,..., $\mathrm{Y}_{j,S}^*(\widehat{\theta}_{j,T}^*)$ is equal to $Y_{j,\widetilde{I}_1}(\widehat{\theta}_{j,T}^*)$,..., $Y_{j,\widetilde{I}_1+l}(\widehat{\theta}_{j,T}^*)$ ..., $Y_{j,\widetilde{I}b_1+l}(\widehat{\theta}_{j,T}^*)$ where $\widetilde{I}_i, i=1$,..., $\widetilde{b}$ are independent discrete uniform random variates on 1,..., $S-\widetilde{l}$. Also, note that for each of the m models, and for each bootstrap replication, we draw $\widetilde{b}$ discrete uniform random variates (the $\widetilde{I}_i$) on $1, \ldots, S-\widetilde{l}$, and that draws are independent across models. Thus, in our use of notation, we have suppressed the dependence of $\widetilde{I}_i$ on j.

Thereafter, form bootstrap statistics as follows:

$$Z_{n,\tau}^* \ = \max_{i=2,\ldots m} \int_U Z_{n,u,\tau}^*(1,\ i)\phi(u)\mathrm{d}u,$$

where for $\tau = 1$ (rolling estimation scheme), and for $\tau = 2$ (recursive estimation scheme):

$$Z_{n,u,\tau}^*(1,\ i) = \ \frac{1}{\sqrt{n}} \sum_{t=R}^{T-1} (((1\{y_{t+1}^* \ \leq u\} - F_1(u|Z^{*,t}\widetilde{\theta}_{1,t,\tau}^*))^2$$

$$-\ (1\{y_{t+1}^* \ \leq u\} - F_i(u|Z^{*,t}\widetilde{\theta}_{i,t,\tau}^*))^2)$$

$$-\frac{1}{T}\sum_{j=s+1}^{T-1} ((1\{y_{j+1} \ \leq u\} - F_1(u|Z^i,\ \widehat{\theta}_{1,t,\tau}))^2 - \ (1\{y_{j+1} \ \leq u\} - F_i(u|Z^j,\ \widehat{\theta}_{i,t,\tau}))^2))$$

Note that each bootstrap term, say $1\{y^*_{t+1} \ \leq \ u\} - F_i(u|Z^{*,t}, \ \widetilde{\theta}^*_{i,t,\tau})$, $t \geq R$, is re- centered around the (full) sample mean $\frac{1}{T}\sum_{j=s+1}^{T-1}(1\{y_{j+1} \ \ \leq \ \ u\} - F_i \ (u|Z^j \ , \ \widehat{\theta}_{i,t,\tau}))^2$. This is necessary as the bootstrap statistic is constructed using the last n resampled observations, which in turn have been resampled from the full sample. In particular, this is necessary regardless of the ratio n/R. If $n/R \to 0$, then we do not need to mimic parameter estimation error, and so could simply use $\widehat{\theta}_{1,t,\tau}$ instead of $\widetilde{\theta}^*_{1,t,\tau}$, but we still need to recenter any bootstrap term around the (full) sample mean.

Note that re-centering is necessary, even for first order validity of the bootstrap, in the case of over-identified generalized method of moments (GMM) estimators [see, e.g., Hall and Horowitz (1996), Andrews (2002), Andrews (2004), Inoue and Shintani (2006)]. This is due to the fact that, in the over-identified case, the bootstrap moment conditions are not equal to zero, even if the population moment conditions are. However, in the context of m-estimators using the full sample, re-centering is needed only for higher order asymptotics, but not for first order validity, in the sense that the bias term is of smaller order than $T^{-1/2}$. Namely, in the case of recursive m-estimators the bias term is instead of order $T^{-1/2}$ and so it does contribute to the limiting distribution. This points to a need for re-centering when using recursive estimation schemes.

For the confidence interval case, define:

$$V^*_{n,\tau} = \max_{i=2,\dots m}, V^*_{n_{\underline{u},\overline{u}},\tau}(1, \ i)$$

and

$$\begin{aligned}
V^*_{n,\underline{u},\overline{u},\tau}(1, \ i) = \frac{1}{\sqrt{n}}\sum_{t=R}^{T-1}&((1\{\underline{u} \leq y^*_{t+1} \ \ \leq \overline{u}\} - (F_1(\overline{u}|Z^{*t}, \widetilde{\theta}^*_{1,t,\tau}) - F_1(\underline{u}|Z^{*t}, \widetilde{\theta}^*_{1,t,\tau})))^2 \\
&- (1\{\underline{u} \leq y^*_{t+1} \ \ \leq \overline{u}\} - (F_i(\overline{u}|Z^{*t}, \widetilde{\theta}^*_{i,t,\tau}) - F_1(\underline{u}|Z^{*t}, \widetilde{\theta}^*_{i,t,\tau})))^2) \\
-\frac{1}{T}\sum_{j=s+1}^{T-1}&((1\{\underline{u} \leq y_{i+1} \ \ \leq \overline{u}\} - (F_1(\overline{u}|Z^j, \ \widehat{\theta}_{1,t,\tau}) - F_1(\underline{u}|Z^j, \ \widehat{\theta}_{1,t,\tau})))^2 \\
&- (1\{\underline{u} \leq y_{j+1} \ \ \leq \overline{u}\} - (F_i(\overline{u}|Z^j, \ \widehat{\theta}_{i,t,\mathrm{r}}) - F_1(\underline{u}|Z^j, \ \widehat{\theta}_{i,t,\tau})))^2)
\end{aligned}$$

where, as usual, $\tau = 1, 2$. The following results then hold,

PROPOSITION 6.3 (From Proposition 6 in Corradi and Swanson (2006b)): With Assumptions 1.1, 6.1–6.3, also, assume that as $T \to \infty, l \to \infty$, and that $\frac{l}{T^{1/4}} \to 0$. Then, as T, n and $R \to \infty$, for $\tau = 1, 2$:

$$\Pr\,(\sup_{v \in \Re} |\overset{*}{\underset{T}{\Pr}}(\max_{i=2,\ldots m} \int_U Z^*_{n,u,\tau}(1\ ,\ i)\phi(u)\mathrm{d}u \leq v)$$

$$-\Pr\,(\max_{i=2,\ldots,m} \int_U Z^{\mu}_{n,u,\tau}\ (1\ ,\ i)\phi(u)\mathrm{d}u \leq v)| > \epsilon)\ \to 0,$$

where $Z^{\mu}_{n,u,\tau}(1,\ i) = Z_{n,u,\tau}(1,\ i) - \sqrt{n}(\mu_1^2(u) - \mu_i^2(u))$, and where $\mu_1^2(u) - \mu_i^2(u)$ is defined as in Equation (4.7).

PROPOSITION 6.4 (From Proposition 7 in Corradi and Swanson (2006b)): With Assumptions 1.1, 6.1–6.3, also assume that as $T \to \infty, l \to \infty$, and that $\frac{l}{T^{1/4}} \to 0$. Then, as T, n and $R \to \infty$, for $\tau = 1, 2$:

$$\Pr\,(\sup_{v \in \Re} |\overset{*}{\underset{T}{\Pr}}(\max_{i=2,\ldots m}, V^*_{n,\underline{u},\overline{u},\tau}(1,\ i)\ \leq v)$$

$$-\Pr\,(\max_{i=2,\ldots m}, V^{\mu}_{n,\underline{u},\overline{u},\tau}\ (1,\ i) \leq v)|\ > \epsilon)\ \to 0$$

where $V^{\mu}_{n,\underline{u},\overline{u},\tau}(1\ ,\ i) = V_{n,\underline{u},\overline{u},\tau}(1\ ,\ i) - \sqrt{n}(\mu_1^2(u) - \mu_i^2(u))$.

The above results suggest proceeding in the following manner. For brevity, consider the case of $Z^*_{n,\tau}$. For any bootstrap replication, compute the bootstrap statistic, $Z^*_{n,\tau}$. Perform B bootstrap replications (B large) and compute the quantiles of the empirical distribution of the B bootstrap statistics. Reject H_0, if $Z_{n,\tau}$ is greater than the $(1\ -\alpha)$th-percentile. Otherwise, do not reject. Now, for all samples except a set with probability measure approaching zero, $Z_{n,\tau}$ has the same limiting distribution as the corresponding bootstrapped statistic when $E(\mu_1^2(u)\ \ - \mu_i^2(u)) = 0, \forall i$, ensuring asymptotic size equal to α. On the other hand, when one or more competitor models are strictly dominated by the benchmark, the rule provides a test with asymptotic size between 0 and α. Under the alternative, $Z_{n,\tau}$ diverges to (plus) infinity, while the corresponding bootstrap statistic has a well defined limiting distribution, ensuring unit asymptotic power.

From the above discussion, we see that the bootstrap distribution provides correct asymptotic critical values only for the least favorable case under the null hypothesis; that is, when all competitor models are as good as the benchmark model. When $\max_{i=2,\ldots,m} \int_U (\mu_1^2(u) - \mu_i^2(u))\phi(u)\mathrm{d}u = 0$, but $\int_U (\mu_1^2(u) - \mu_i^2(u))\phi(u)\mathrm{d}u < 0$ for some i, then the bootstrap critical values lead to conservative inference. An alternative to our bootstrap critical values in this case is the construction of critical values based on subsampling, which is briefly discussed in Section 4.3. Heuristically, construct $T - 2b_T$ statistics using subsamples of length b_T, where $b_T/T \to 0$. The empirical distribution of these statistics computed over the various subsamples properly mimics the distribution of the statistic. Thus, subsampling provides valid critical values even for the case where $\max_{i=2,\ldots,m} \int_U (\mu_1^2(u) - \mu_i^2(u))\phi(u)\mathrm{d}u = 0$, but $\int_U (\mu_1^2(u) - \mu_i^2(u))\phi(u)\mathrm{d}u < 0$ for some i. This is the approach used by Whang et al. (2004), for example, in the context of testing for stochastic dominance. Needless to say, one problem with subsampling is that unless the sample is very large, the empirical distribution of the subsampled statistics may yield a poor approximation of the limiting distribution of the statistic. Another alternative approach for addressing the conservative nature of our bootstrap critical values is the Hansen's SPA approach (see Section 4.2 and Hansen (2005)). Hansen's idea is to recenter the bootstrap statistics using the sample mean, whenever the latter is larger than (minus) a bound of order $\sqrt{2T \log\log T}$. Otherwise, do not recenter the bootstrap statistics. In the current context, his approach leads to correctly sized inference when $\max_{i=2,\ldots,m} \int_U (\mu_1^2(u) - \mu_i^2(u))\phi(u)\mathrm{d}u = 0$, but $\int_U (\mu_1^2(u) - \mu_i^2(u))\phi(u)\mathrm{d}u < 0$ for some i. Additionally, his approach has the feature that if all models are characterized by a sample mean below the bound, the null is "accepted" and no bootstrap statistic is constructed.

Part III: Forecast Evaluation Using Density Based Predictive Accuracy Tests That Are Not Loss Function Dependent: The Case of Stochastic Dominance

All predictive accuracy tests outlined in previous two parts of this chapter are loss functions dependent, i.e. loss functions such as mean squared forecast error (MSFE) and mean absolute forecast error (MAFE) must be specified prior to test construction. Evidently, given possible misspecification, model rankings may change under different loss functions. In the following section, we introduce a novel criterion for forecast evaluation that utilizes the entire distribution of forecast errors, is robust to the choice of loss function, and ranks distributions of forecast errors via stochastic dominance type tests.

4.7 Robust forecast comparison

Jin et al. (2017) (JCS) introduce the concepts of general-loss (GL) forecast superiority and convex-loss (CL) forecast superiority and develop tests for GL (CL) superiority that are based on an out-of-sample generalization of the tests introduced by Linton et al. (2005). The JCS tests evaluate the entire forecast error distribution and do not require knowledge or specification of a loss function, i.e. tests are robust to the choice of loss function. In addition, parameter estimation error and data dependence are taken into account, and heterogeneity that is induced by distributional change over time is allowed for.

The concepts of general-loss (GL) forecast superiority and convex-loss (CL) forecast superiority are defined as follow:

(1) For any two sequences of forecast errors $u_{1,t}$ and $u_{2,t}$, $u_{1,t}$ general-loss (GL) outperforms $u_{2,t}$, denoted as $u_1 \succeq_G u_2$, if and only if $E(g(u_{1,t})) \leq E(g(u_{2,t}))$, $\forall g(\cdot) \in GL(\cdot)$, where $GL(\cdot)$ are the set of general loss functions with properties specified in Granger (1999); and

(2) $u_{1,t}$ convex-loss (CL) outperforms $u_{2,t}$, denoted as $u_1 \succeq_C u_2$, if and only if $E(g(u_{1,t})) \leq E(g(u_{2,t}))$, $\forall g(\cdot) \in CL(\cdot)$, where $CL(\cdot)$ are the set of general loss functions which in addition are convex.

These authors also establish linkages between GL(CL) forecast superiority and first(second) order stochastic dominance, allowing for the construction of direct tests for GL(CL) forecast superiority. Define

$$G(x) = (F_2(x) - F_1(x))sgn(x),$$

where $sgn(x) = 1$, if $x \geq 0$, and $sgn(x) = -1$, if $x < 0$. Here, $F_i(x)$ denotes the cumulative distribution function (CDF) of u_i, and

$$C(x) = \int_{-\infty}^{x} (F_1(t) - F_2(t))dt1_{\{x<0\}} + \int_{x}^{\infty} (F_2(t) - F_1(t))dt1_{\{x\geq 0\}}$$

Assumption 7.1: $g(\cdot) : \Re \rightarrow \Re^+$ is continuously differentiable, except for finitely many points, with derivative $\nabla g(\cdot)$, such that $\nabla g(z) \leq 0$, $\forall z \leq 0$ and $\nabla g(z) \geq 0$, $\forall z \geq 0$.

PROPOSITION 7.1 (From Propositions 2.2 and 2.3 in Jin et al. (2017)): With Assumption 7.1, $E(g(u_{1,t})) \leq E(g(u_{2,t}))$, $\forall g(\cdot) \in GL(\cdot)$, if and only if $G(x) \leq 0$, $\forall x \in \mathcal{X}$, where $\mathcal{X}$ is the union of the supports of all forecast errors. Further, if $\int_{-\infty}^{x}(F_1(t) - F_2(t))dt1_{\{x<0\}}$ and $\int_{x}^{\infty}(F_2(t) - F_1(t))dt1_{\{x\geq 0\}}$ are well defined for each $x \in \mathcal{X}$, then $E(g(u_{1,t})) \leq E(g(u_{2,t}))$, $\forall g(\cdot) \in CL(\cdot)$ if and only if $C(x) \leq 0$, $\forall x \in \mathcal{X}$.

The above proposition establishes a clear mapping between GL (CL) forecast superiority and first (second) order stochastic dominance. Intuitively, if we construct a graph that contains a plot of $G(x)$ against x. When $u_1 \succeq_G u_2$, we expect all points lie below or on the zero line. Similarly, if we construct a graph that contains a plot of $C(x)$ against x. When $u_1 \succeq_C u_2$, we expect all points lie below or on the zero line as well.

The hypotheses tested in JCS are:

$$H_0 : \max_{i=1,\ldots,m} E(g(u_{0,t+1}) - g(u_{i,t+1})) \leq 0$$

versus

$$H_A : \max_{i=1,\ldots,m} E(g(u_{0,t+1}) - g(u_{i,t+1})) > 0$$

Given Proposition 7.1, the above hypotheses can be restated as

$$H_0^{TG} = H_0^{TG-} \cap H_0^{TG+} : \big(\max_{i=1,\ldots,m} (F_0(x) - F_i(x)) \leq 0,\ \forall x \leq 0 \big)$$
$$\cap \big(\max_{i=1,\ldots,m} (F_i(x) - F_0(x)) \leq 0,\ \forall x > 0 \big)$$

versus

$$H_A^{TG} = H_A^{TG-} \cup H_A^{TG+} : \big(\max_{i=1,\ldots,m} (F_0(x) - F_i(x)) > 0, \text{ for some } x \leq 0 \big)$$
$$\cup \big(\max_{i=1,\ldots,m} (F_i(x) - F_0(x)) > 0, \text{ for some } x > 0 \big)$$

for the case of GL forecast superiority. Similarly, for the case of CL forecast superiority, we have that:

$$H_0^{TC} = H_0^{TC-} \cap H_0^{TC+} : \big(\max_{i=1,\ldots,m} \int_{-\infty}^{x} (F_0(x) - F_i(x)) \leq 0, \ \forall x \leq 0 \big)$$
$$\cap \big(\max_{i=1,\ldots,m} \int_{x}^{\infty} (F_i(x) - F_0(x)) \leq 0, \ \forall x > 0 \big)$$

versus

$$H_A^{TC} = H_A^{TC-} \cup H_A^{TC+} : \big(\max_{i=1,\ldots,m} \int_{-\infty}^{x} (F_0(x) - F_i(x)) > 0, \text{ for some } x \leq 0 \big)$$
$$\cup \big(\max_{i=1,\ldots,m} \int_{x}^{\infty} (F_i(x) - F_0(x)) > 0, \text{ for some } x > 0 \big)$$

Of note is that the above null (alternative) is the intersection (union) of two different null (alternative) hypotheses because of a discontinuity at zero. The test statistics for GL forecast superiority are constructed as follows:

$$TG_n^+ = \max_{i=1,\ldots,k} \sup_{x \in \mathcal{X}^+} \sqrt{n}\widehat{G}_{i,n}(x)$$

and

$$TG_n^- = \max_{i=1,\ldots,k} \sup_{x \in \mathcal{X}^-} \sqrt{n}\widehat{G}_{i,n}(x)$$

with

$$\widehat{G}_{i,n}(x) = \big(\widehat{F}_{0,n}(x) - \widehat{F}_{i,n}(x)\big) sgn(x)$$

where $\widehat{F}_{i,n}(x)$ denotes the empirical CDF of u_i, with

$$\widehat{F}_{i,n}(x) = n^{-1} \sum_{t=R}^{T} 1_{\{u_{i,t} \leq x\}}$$

Similarly, the test statistics for CL forecast superiority are constructed as follows:

$$TC_n^+ = \max_{i=1,\ldots,k} \sup_{x \in \mathcal{X}^+} \sqrt{n}\widehat{C}_{i,n}(x)$$

and

$$TC_n^- = \max_{i=1,...,k} \sup_{x \in \mathcal{X}^-} \sqrt{n}\widehat{C}_{i,n}(x)$$

with

$$\begin{aligned}\widehat{C}_{i,n}(x) &= \int_{-\infty}^{x} \big(\widehat{F}_{0,n}(x) - \widehat{F}_{i,n}(x)\big) dx 1_{\{x<0\}} - \int_{x}^{\infty} \big(\widehat{F}_{i,n}(x) - \widehat{F}_{0,n}(x)\big) dx 1_{\{x\geq 0\}} \\ &= \frac{1}{n}\sum_{t=1}^{n} \Big\{[(u_{0,t} - x)sgn(x)]_+ - [(u_{i,t} - x)sgn(x)]_+\Big\},\end{aligned}$$

where $[z]_+ = \max\{0, z\}$.

Note that in order to reduce computation time, it may be preferable to construct approximations to the suprema in statistics TG^+, TG^-, TC^+ and TC^- by taking maxima over some smaller grid of points, $\mathcal{X}_N = \{x_1, ..., x_N\}$, where $N < n$. Theoretically, the distribution theory is unaffected by using this approximation, as the set of evaluation points becomes dense in the joint support. We now require the following assumptions.

Assumption 7.2: (i) $\{(y_t, Z_i^t)'\}$ is a strictly stationary and α-mixing sequence with mixing coefficient $\alpha(l) = O(l^{-C_0})$, for some $C_0 > \max\{(q-1)(q+1), 1 + 2/\delta\}$, with $i = 0, ..., m$, where q is an even integer that satisfies $q > 3(g_{\max} + 1)/2$. Here, $g_{\max} = \max\{g_0, ..., g_m\}$ and δ is a positive constant;

(ii) For $i = 0, ..., m,\ f_i(Z_i^t, \theta_i)$ is differentiable a.s. with respect to θ_i in the neighborhood $\Theta_i^\dagger$ of $\theta_i^\dagger$, with $\sup_{\theta \in \Theta_0^\dagger} ||\nabla_\theta f_i(Z_i^t, \theta_i)||_2 < \infty$;

(iii) The conditional distribution of $u_{i,t}$ given Z_i^t has bounded density with respect to the Lebesgue measure a.s., and $||u_{i,t}||_{2+\delta} < \infty,\ \forall i$.

Assumption 7.2*: (i) $\{(y_t, Z_i^t)'\}$ is a strictly stationary and α-mixing sequence with mixing coefficient $\alpha(l) = O(l^{-C_0})$, for some $C_0 > \max\{rq/(r-q), 1 + 2/\delta\}$, with $i = 0, ..., m$, and $r > q > g_{\max} + 1$;

(ii) For $i = 0, ..., m,\ f_i(Z_i^t, \theta_i)$ is differentiable a.s. with respect to θ_i in the neighborhood $\Theta_i^\dagger$ of $\theta_i^\dagger$, with $\sup_{\theta \in \Theta_0^\dagger} ||\nabla_\theta f_i(Z_i^t, \theta_i)||_r < \infty$;

(iii) $||u_{i,t}||_r < \infty,\ \forall i$.

Assumption 7.3: $\forall$ i and t, $\widehat{\theta}_{i,t}$ satisfies $\widehat{\theta}_{i,t} - \theta_i^{\dagger} = B_i(t)H_i(t)$, where $B_i(t)$ is a $n_i \times L_i$ matrix and $H_i(t)$ is $L_i \times 1$, with the following:

(i) $B_i(t) \to B_i$ a.s., where B_i is a matrix of rank n_i;

(ii) $H_i(t) = t^{-1}\sum_{s=1}^{t} h_{i,s}$, $R^{-1}\sum_{s=t-R+1}^{t} h_{i,s}$ and $R^{-1}\sum_{s=1}^{R} h_{i,s}$ for the recursive, rolling and fixed schemes, respectively, where $h_{i,s} = h_{i,s}(\theta_i^{\dagger})$;

(iii) $E(h_{i,s}(\theta_i^{\dagger}) = 0$; and

(iv) $||h_{i,s}(\theta_i^{\dagger})||_{2+\delta} < \infty$, for some $\delta > 0$.

Assumption 7.4: (i) The distribution function of forecast errors, $F_i(x, \theta_i)$ is differentiable with respect to θ_i in a neighborhood $\Theta_i^{\dagger}$ of $\theta_i^{\dagger}$, $\forall i$;

(ii) $\forall i$, and $\forall$ sequences of positive constants $\{\xi_n : n \geq 1\}$,

such that $\xi_n \to 0$, $\sup_{x\in\mathcal{X}} \sup_{\theta:\ ||\theta-\theta_i^{\dagger}||\leq\xi_n} ||\nabla_\theta F_i(x,\theta)sgn(x) - \Delta_i^{\dagger}(x)|| = O(\xi_n^{\eta})$, for some $\eta > 0$, where $\Delta_i^{\dagger}(x) = \nabla_\theta F_i(x, \theta_i^{\dagger})sgn(x)$;

(iii) $\sup_{x\in\mathcal{X}} ||\Delta_i^{\dagger}(x)|| < \infty, \forall i$.

Assumption 7.4*: (i) Assumption 5.4 (i) holds;

(ii) $\forall i$, and $\forall$ sequences of positive constants $\{\xi_n : n \geq 1\}$,

such that $\xi_n \to 0$, $\sup_{x\in\mathcal{X}} \sup_{\theta:\ ||\theta-\theta_i^{\dagger}||\leq\xi_n} ||\nabla_\theta\{\int_{-\infty}^{x} F_i(t,\theta)dt1_{\{x<0\}} + \int_{x}^{\infty}(1-F_i(x,\theta))dt1_{\{x\geq 0\}}\} - \Lambda_i^{\dagger}(x)|| = O(\xi_n^{\eta})$, for some $\eta > 0$, where

$$\Lambda_i^{\dagger}(x) = \nabla_\theta\{\int_{-\infty}^{x} F_i(t,\theta_i^{\dagger})dt1_{\{x<0\}} + \int_{x}^{\infty}(1 - F_i(x,\theta_i^{\dagger}))dt1_{\{x\geq 0\}}\};$$

(iii) $\sup_{x\in\mathcal{X}} ||\Lambda_i^{\dagger}(x)|| < \infty, \forall i$.

Assumptions 7.2* and 7.4* are needed for testing H_0^{TC}. Note that the first and third assumptions parallel those imposed by Linton et al. (2005), with the uniform continuity conditions in Assumptions 7.4 and 7.4* strengthened. Assumption 7.2 is needed in order to verify the stochastic equicontinuity of the empirical process, for a class of bounded functions that appears in the TG_n test. Assumption 7.2* introduces a trade-off between mixing sizes and moment conditions, and is used to verify the stochastic equicontinuity result for the possibly unbounded functions that appear in the TC_n test. For further details, see Hansen (1996b). Assumptions 7.4 and 7.4* differ in the amount of smoothness required. For the CL forecast superiority test, less smoothness is required. Finally, it is worth stressing that Assumptions 7.3 and 7.5 are identical to Assumptions 1 and 2 in McCracken (2000a), respectively.

PROPOSITION 7.2 (From Theorem 3.1 in Jin et al. (2017)): (i) With Assumptions 4.3, 7.2–7.4, under H_0^{TG-},

$$TG_n^- \xrightarrow{d} \max_{i=1,\ldots,m} \sup_{x\in\mathcal{B}_i^{g-}} [\widetilde{g}_i(x) + \Delta_{i0}(x)' B_i \upsilon_{i0} - \Delta_{10}(x)' B_1 \upsilon_{10}],\ \text{if } TG^- = 0$$

$$\to -\infty,\ \text{if } TG^- < 0$$

Under H_0^{TG+},

$$TG_n^+ \xrightarrow{d} \max_{i=1,\ldots,m} \sup_{x\in\mathcal{B}_i^{g+}} [\widetilde{g}_i(x) + \Delta_{i0}(x)' B_i \upsilon_{i0} - \Delta_{10}(x)' B_1 \upsilon_{10}],\ \text{if } TG^+ = 0$$

$$\to -\infty,\ \text{if } TG^+ < 0$$

where $\mathcal{B}_i^{g-} = \{x \in \mathcal{X}^- : F_0(x) = F_i(x)\}$ and $\mathcal{B}_i^{g+} = \{x \in \mathcal{X}^+ : F_0(x) = F_i(x)\}$, and $\left(\widetilde{g}_i(\cdot), \upsilon_{i0}, \upsilon_{10}\right)'$ is a mean zero Gaussian process with covariance function given by

$$\Omega_i^g(x_1, x_2) = \lim_{T\to\infty} E \begin{pmatrix} \upsilon_{i,n}^g(x_1, \theta_i^\dagger) - \upsilon_{0,n}^g(x_1, \theta_0^\dagger) \\ \sqrt{n}\overline{H}_{i,n} \\ \sqrt{n}\overline{H}_{0,n} \end{pmatrix} \begin{pmatrix} \upsilon_{i,n}^g(x_2, \theta_i^\dagger) - \upsilon_{0,n}^g(x_2, \theta_0^\dagger) \\ \sqrt{n}\overline{H}_{i,n} \\ \sqrt{n}\overline{H}_{0,n} \end{pmatrix}'$$

with $\overline{H}_{i,n} = n^{-1}\sum_{t=R}^{T} H_i(t)$, and $\upsilon_{i,n}^g(x, \theta)$ is an empirical process defined as

$$\upsilon_{i,n}^g(x, \theta) = \frac{1}{\sqrt{n}} \sum_{t=R}^{T} \{1_{\{u_{i,t+\tau}(\theta)\leq x\}} - F_i(x, \theta)\} sgn(x)$$

(ii) With Assumptions 7.2*, 7.3, 7.4* and 7.5, under $H_0^{TC^-}$,

$$TC_n^- \xrightarrow{d} \max_{i=1,\dots,m} \sup_{x\in\mathcal{B}_i^{c-}} [\widetilde{c}_i(x) + \Lambda_{i0}(x)' B_i \upsilon_{i0} - \Lambda_{10}(x)' B_1 \upsilon_{10}], \text{ if } TC^- = 0$$

$$\to -\infty, \text{ if } TC^- < 0$$

Under $H_0^{TC^+}$,

$$TC_n^+ \xrightarrow{d} \max_{i=1,\dots,m} \sup_{x\in\mathcal{B}_i^{c+}} [\widetilde{c}_i(x) + \Lambda_{i0}(x)' B_i \upsilon_{i0} - \Lambda_{10}(x)' B_1 \upsilon_{10}], \text{ if } TC^+ = 0$$

$$\to -\infty, \text{ if } TC^+ < 0$$

where $\mathcal{B}_i^{c-} = \{x \in \mathcal{X}^- : \int_{-\infty}^{x} (F_i(x) - F_0(x))dx 1_{\{x<0\}}\}$ and $\mathcal{B}_i^{c+} = \{x \in \mathcal{X}^+ : \int_{x}^{\infty} (F_0(x) - F_i(x))dx 1_{\{x\geq 0\}}\}$. Similarly, $\left(\widetilde{c}_i(\cdot), \upsilon_{i0}, \upsilon_{10}\right)'$ is a mean zero Gaussian process with covariance function given by

$$\Omega_i^c(x_1, x_2) = \lim_{T\to\infty} E \begin{pmatrix} \upsilon_{i,n}^c(x_1, \theta_i^\dagger) - \upsilon_{0,n}^c(x_1, \theta_0^\dagger) \\ \sqrt{n}\overline{H}_{i,n} \\ \sqrt{n}\overline{H}_{0,n} \end{pmatrix} \begin{pmatrix} \upsilon_{i,n}^c(x_2, \theta_i^\dagger) - \upsilon_{0,n}^c(x_2, \theta_0^\dagger) \\ \sqrt{n}\overline{H}_{i,n} \\ \sqrt{n}\overline{H}_{0,n} \end{pmatrix}'$$

where $\upsilon_{i,n}^c(x, \theta)$ is an empirical process defined as

$$\begin{aligned}\upsilon_{i,n}^c(x,\theta) =& \frac{1}{\sqrt{n}} \sum_{t=R}^{T} \Big\{ \int_{-\infty}^{x} [1_{\{u_{i,t+\tau}(\theta)\leq s\}} - F_i(s,\theta)]ds 1_{\{x<0\}} \\ & - \int_{x}^{\infty} [1_{\{u_{i,t+\tau}(\theta)\leq s\}} - F_i(s,\theta)]ds 1_{\{x\geq 0\}} \Big\}\end{aligned}$$

The asymptotic null distributions of TG_n^+ (TG_n^-) and TC_n^+ (TC_n^-) depend on the true model parameters and the distribution functions, $F_i(\cdot),\ i = 1, ..., m$, which implies that the asymptotic critical values for TG_n^+ (TG_n^-) and TC_n^+ (TC_n^-) cannot be tabulated. Therefore, the stationary bootstrap is used to approximate the asymptotic null distributions of our test statistics. (Note that the block bootstrap can also be used, as discussed in subsequent research by Corradi, Sin and Swanson.) The objective is to utilize bootstrap procedure that mimics the asymptotic null distribution in the least favorable case, where $F_0(x) = ... = F_m(x)$, $\forall x \in \mathcal{X}$.

Define the bootstrap statistic as:

$$TG_n^{*+} = \max_{i=1,...,k} \sup_{x \in \mathcal{X}^+} \sqrt{n}\Big(\widehat{G}^*_{i,n}(x) - \widehat{G}_{i,n}(x)\Big)$$

with

$$\widehat{G}^*_{i,n}(x) = \big(\widehat{F}^*_{0,n}(x) - \widehat{F}^*_{i,n}(x)\big) sgn(x)$$

where $\widehat{F}^*_{i,n}(x)$ denotes the empirical CDF of resampled u_i, i.e. u_i^*. TG_n^{*-}, TC_n^{*+} and TC_n^{*-} can be defined analogously.

Assumption 7.6: The smoothing parameter, S_n, determining the mean block length in stationary bootstrap satisfies $0 < S_n < 1$, $S_n \to 0$ and $nS_n^2 \to \infty$, as $n \to \infty$.

Assumption 7.7: For any arbitrary $n_i \times 1$ vector, λ_i, with $\lambda_i^{'}\lambda_i = 1$, and $\forall i$, we have (i)

$$\Pr\Big[\limsup_{t \geq R} n^{1/2} \frac{|\lambda_i^{'}(\widehat{\theta}_{i,t} - \theta_i^{\dagger})|}{(\lambda_i^{'}\Sigma_i\lambda_i \text{loglog}(\lambda_i^{'}\Sigma_i\lambda_i)n)^{1/2}} = 1\Big] = 1$$

for the recursive scheme, where $\Sigma_i = B_i[\lim_{T\to\infty} \text{Var}(n^{-1/2}\sum_{t=R+1}^{T} H_i(t))]B_i^{'}$.

(ii)

$$\Pr\Big[\limsup_{t \geq R} R^{1/2} \frac{|\lambda_i^{'}(\widehat{\theta}_{i,t} - \theta_i^{\dagger})|}{(\lambda_i^{'}\Sigma_i\lambda_i \text{loglog}(\lambda_i^{'}\Sigma_i\lambda_i)R)^{1/2}} = 1\Big] = 1$$

for the rolling scheme, where $\Sigma_i = B_i[\lim_{T\to\infty} \text{Var}(R^{-1/2}\sum_{t=R+1}^{T} H_i(t))]B_i^{'}$.

PROPOSITION 7.3 (From Corollary 3.3 in Jin et al. (2017)): With Assumptions 7.2–7.4, 7.6 and 7.7, and that $(n/R)\text{loglog}R \to 0$, as $T \to \infty$, then

$$\rho\Big(L[\max_{i=1,...,m} \sup_{x \in \mathcal{X}^+} \sqrt{n}(\widehat{G}^*_{i,n}(x) - \widehat{G}_{i,n}(x))|U_1, ..., U_{T+\tau}],$$
$$L[\max_{i=1,...,m} \sup_{x \in \mathcal{X}^+} \sqrt{n}(\widehat{G}_{i,n}(x) - G_i(x))]\Big) \xrightarrow{n} 0$$

and

$$\rho\Big(L[\max_{i=1,\dots,m}\sup_{x\in\mathcal{X}^-}\sqrt{n}(\widehat{G}^*_{i,n}(x)-\widehat{G}_{i,n}(x))|U_1,...,U_{T+\tau}],$$
$$L[\max_{i=1,\dots,m}\sup_{x\in\mathcal{X}^-}\sqrt{n}(\widehat{G}_{i,n}(x)-G_i(x))]\Big)\xrightarrow{n}0$$

where ρ is any metric metrizing weak convergence, $L[\cdot]$ denotes the probability law of the corresponding Hilbert space valued random variable, and $U_t=(y_t,Z_0^t,...,Z_m^t)'$.

Therefore, the asymptotic null distribution of TG_n^+ (TG_n^-) can be approximated by $TG_n^{*+}-TG_n^+$ ($TG_n^{*-}-TG_n^-$). Arguments in favor of using the stationary bootstrap with TC_n^+ and TC_n^- are similar.

To conduct inference, use the following approach due to Holm (1979). Define $q^{G+}_{n,S_n}(1-\alpha)$ and $q^{G-}_{n,S_n}(1-\alpha)$ to be the $(1-\alpha)$-th sample quantile of TG_n^{*+} and TG_n^{*-}, respectively. Then, estimate bootstrap p-values, $p^{G+}_{B,n,S_n}=\frac{1}{B}\sum_{s=1}^{B}(TG_n^{*+}\geq TG_n^+)$, and finally use the following rules.

Rule TG: Reject H_0^{TG} at level α, if $\min\Big\{p^{G+}_{B,n,S_n},\ p^{G-}_{B,n,S_n}\Big\}\leq\alpha/2$;
Rule TC: Reject H_0^{TG} at level α, if $\min\Big\{p^{C+}_{B,n,S_n},\ p^{C-}_{B,n,S_n}\Big\}\leq\alpha/2$;

Note that Holm bounds are equivalent to Bonferroni bounds when there are only two hypotheses. From Proposition 7.3, it follows immediately that this test,when implemented using the stationary bootstrap, has asymptotically correct size only in the least favorable case, under the null, and is asymptotically biased towards certain local alternatives.

PROPOSITION 7.4 (From Theorem 4.1 in Jin et al. (2017)): With Assumptions 4.3, 7.2–7.4, under H_A^{TG},

$$\Pr(TG_n^+>q^{G+}_{n,S_n}(1-\alpha))\to 1,\ as\ T\to\infty$$

and

$$\Pr(TG_n^->q^{G-}_{n,S_n}(1-\alpha))\to 1,\ as\ T\to\infty$$

The above proposition ensures unit asymptotic power under the alternative. Similar arguments apply to TC_n^+ and TC_n^- as well. For details of the power of TG_n^+ (TG_n^-) and TC_n^+ (TC_n^-) tests against a sequence of contiguous local alternatives converging to the null, at rate $n^{-1/2}$, see Jin et al. (2017).

Bibliography

Aït-Sahalia, Y. and Jacod, J. 2014. High Frequency Financial Econometrics. Princeton University Press.

Aït-Sahalia, Y. and Xiu, D. 2016. Increased correlation among asset classes: Are volatility or jumps to blame, or both? *Journal of Econometrics* 194:205–219.

Andersen, T. G., Bollerslev, T., Diebold, F. X., and Labys, P. 2003. Modeling and forecasting realized volatility. *Econometrica* 71:579–625.

Anderson, B. D. and Moore, J. B. 2012. Optimal filtering. Courier Corporation.

Andreou, E., Ghysels, E., and Kourtellos, A. 2013. Should macroeconomic forecasters use daily financial data and how? *Journal of Business & Economic Statistics* 31:240–251.

Andrews, D. 2002. Higher-order improvements of a computationally attractive k-step bootstrap for extremum estimators. *Econometrica* 70:119–162.

Andrews, D. 2004. The blockblock bootstrap: Improved asymptotic refinements. *Econometrica* 72:673–700.

Andrews, D. and Soares, G. 2010. Inference for parameters defined by moment inequalities using generalized moment selection. *Econometrica* 78:119–157.

Aruoba, S., Diebold, F., and Scotti, C. 2009a. Updates on ads index calculation. *Manuscript, www. philadelphiafed. org/research-and-data/real-time-center/business-conditions-index* .

Aruoba, S. B., Diebold, F. X., and Scotti, C. 2009b. Real-time measurement of business conditions. *Journal of Business & Economic Statistics* 27:417–427.

Bachmann, R., Elstner, S., and Sims, E. R. 2013. Uncertainty and economic activity: Evidence from business survey data. *American Economic Journal: Macroeconomics* 5:217–49.

Baker, S. R., Bloom, N., and Davis, S. J. 2016. Measuring economic policy uncertainty. *The Quarterly Journal of Economics* 131:1593–1636.

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

Basu, S. and Bundick, B. 2017. Uncertainty shocks in a model of effective demand. *Econometrica* 85:937–958.

Bierens, H. 1990. A consistent conditional moment test of functional form. *Econometrica* 58:1443–1458.

Bierens, H. and Ploberger, W. 1997. Asymptotic theory of integrated conditional moment tests. *Econometrica* 65:1129–1151.

Bloom, N. 2009. The impact of uncertainty shocks. *Econometrica* 77:623–685.

Breiman, L. 2001. Random forests. *Machine learning* 45:5–32.

Bridle, J. S. 1990. Probabilistic interpretation of feedforward classification network outputs, with relationships to statistical pattern recognition. pp. 227–236.

Brock, W., Lakonishok, J., and LeBaron, B. 1992. Simple technical trading rules and the stochastic properties of stock returns. *The Journal of finance* 47:1731–1764.

Carriero, A., Clark, T. E., and Marcellino, M. 2015. Realtime nowcasting with a bayesian mixed frequency model with stochastic volatility. *Journal of the Royal Statistical Society: Series A (Statistics in Society)* 178:837–862.

Carriero, A., Clark, T. E., and Marcellino, M. 2016. Measuring uncertainty and its impact on the economy. *Review of Economics and Statistics* 100:799–815.

Chang, Y., Gomes, J., and Schorfheide, F. 2002. Learning-by-doing as a propagation mechanism. *American Economic Review* 92:1498–1520.

Chao, J., Corradi, V., and Swanson, N. 2001. Out-of-sample tests for granger causality. *Macroeconomic Dynamics* 5:598–620.

Chauvet, M., Senyuz, Z., and Yoldas, E. 2015. What does financial volatility tell us about macroeconomic fluctuations? *Journal of Economic Dynamics and Control* 52:340–360.

Cheng, M., Swanson, N. R., and Yao, C. 2020. Forecast evaluation, pp. 495–537. *In* Macroeconomic Forecasting in the Era of Big Data. Springer.

Christoffersen, P. F. and Diebold, F. X. 2006. Financial asset returns, direction-of-change forecasting, and volatility dynamics. *Management Science* 52:1273–1287.

Clark, T. and McCracken, M. 2001. Tests of equal forecast accuracy and encompassing for nested models. *Journal of Econometrics* 105:85–110.

Clark, T. and McCracken, M. 2003. Evaluating long horizon forecasts.

Clark, T. E. and McCracken, M. W. 2009. Improving forecast accuracy by combining recursive and rolling forecasts. *International Economic Review* 50:363–395.

Corradi, V. and Distaso, W. 2011. Multiple forecast model evaluation. *The Oxford Handbook of Economic Forecasting, Oxford University Press, USA* pp. 391–414.

Corradi, V. and Swanson, N. 2002. A consistent test for out of sample nonlinear predictive ability. *Journal of Econometrics* 110:353–381.

Corradi, V. and Swanson, N. 2005. A test for comparing multiple misspecified conditional interval models. *Econometric Theory* 21:991–1016.

Corradi, V. and Swanson, N. 2006a. Predictive density and conditional confidence interval accuracy tests. *Journal of Econometrics* 135:187–228.

Corradi, V. and Swanson, N. 2006b. Predictive density and conditional confidence interval accuracy tests. *Journal of Econometrics* 135:187–228.

Corradi, V. and Swanson, N. 2007. Nonparametric bootstrap procedures for predictive inference based on recursive estimation schemes. *International Economic Review* 48:67–109.

Corradi, V., Swanson, N., and Olivetti, C. 2001. Predictive ability with cointegrated variables. *Journal of Econometrics* 104:315–358.

Corradi, V. and Swanson, N. R. 2004. Predictive density accuracy tests. *WP04-16* .

Corradi, V. and Swanson, N. R. 2006c. Predictive density evaluation. *Handbook of Economic Forecasting* 1:197–284.

Corradi, V. and Swanson, N. R. 2013. A survey of recent advances in forecast accuracy comparison testing, with an extension to stochastic dominance, pp. 121–143. *In* Recent Advances and Future Directions in Causality, Prediction, and Specification Analysis. Springer.

Corsi, F. 2009. A simple approxiamte long memory model of realized volatility. *Journal of Financial Econometrics* 7:174–196.

Corsi, F. and Renò, R. 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.

Cortes, C. and Vapnik, V. 1995. Support-vector networks. *Machine learning* 20:273–297.

Cover, T. and Hart, P. 1967. Nearest neighbor pattern classification. *IEEE transactions on information theory* 13:21–27.

Cox, D. R. 1966. Some procedures associated with the logistic qualitative response curve.

De Jong, R. 1996. The bierens test under data dependence. *Journal of Econometrics* 72:1–32.

Diebold, F. and Mariano, R. 2002. Comparing predictive accuracy. *Journal of Business & Economic Statistics* 20:134–144.

Diebold, F. X. and Mariano, R. S. 1995a. Com paring predictive accu racy. *Journal of Business and Economic Statistics* 13:253–263.

Diebold, F. X. and Mariano, R. S. 1995b. Comparing predictive accuracy. *Journal of Business & Economic Statistics* 13:253–263.

Duda, R. O., P. E. H. 1973. Pattern classification and scene analysis.

Fama, E. F. and Blume, M. E. 1966. Filter rules and stock-market trading. *The Journal of Business* 39:226–241.

Fama, E. F. and French, K. R. 1992. The cross-section of expected stock returns. *The Journal of Finance* 47:427–465.

Fama, E. F. and French, K. R. 2015. A five-factor asset pricing model. *Journal of financial economics* 116:1–22.

Fan, R.-E., Chang, K.-W., Hsieh, C.-J., Wang, X.-R., and Lin, C.-J. 2008. Liblinear: A library for large linear classification. *Journal of machine learning research* 9:1871–1874.

Fernández-Villaverde, J. and Rubio-RamíRez, J. 2004. Comparing dynamic equilibrium models to data: A bayesian approach. *Journal of Econometrics* 123:153–187.

Fisher, R. A. 1936. The use of multiple measurements in taxonomic problems. *Annals of eugenics* 7:179–188.

FOMC 2019. Transcript of chair powells press conference opening remarks, september 18, 2019. *Board of Governors of the Federal Reserve System* p. https://www.federalreserve.gov/mediacenter/files/FOMCpresconf20190918.pdf.

Frale, C., Marcellino, M., Mazzi, G. L., and Proietti, T. 2008. A monthly indicator of the euro area gdp. *European University Institute; Series/Number: EUI ECO; 2008/32* .

Friedman, J., Hastie, T., and Tibshirani, R. 2001. The elements of statistical learning, volume 1. Springer series in statistics New York.

Friedman, J. H. 2001. Greedy function approximation: a gradient boosting machine. *Annals of statistics* pp. 1189–1232.

Friedman, N., Geiger, D., and Goldszmidt, M. 1997. Bayesian network classifiers. *Machine learning* 29:131–163.

Fuleky, P. 2020. Macroeconomic forecasting in the era of big data: Theory and practice.

Ghysels, E., Sinko, A., and Valkanov, R. 2007. Midas regressions: Further results and new directions. *Econometric Reviews* 26:53–90.

Giacomini, R. and White, H. L. 2006. Tests of conditional predictive ability. *Econometrica* 74:1545–1578.

Gianni, A. and R, G. 2007. Comparing density forecasts via weighted likelihood ratio tests. *Journal of Business & Economic Statistics* 25:177–190.

Gilchrist, S., Sim, J. W., and Zakrajšek, E. 2014. Uncertainty, financial frictions, and investment dynamics. *NBER Working Papers 20038* .

Granger, C. 1993. On the limitations of comparing mean square forecast errors: A comment. *Journal of Forecasting* 12:651–652.

Granger, C. 1999. Outline of forecast theory using generalized cost function. *Spanish Economic Review* 1:161–173.

Gu, S., Kelly, B., and Xiu, D. 2018. Empirical asset pricing via machine learning. Technical report, National Bureau of Economic Research.

HALL, P. AND HOROWITZ, J. 1996. Bootstrap critical values for tests based on generalized-method-of-moments estimators. *Econometrica* 64:891–916.

HANSEN, B. 1996a. Inference when a nuisance parameter is not identified under the null hypothesis. *Econometrica* 64:413430.

HANSEN, B. 1996b. Stochastic equicontinuity for unbounded dependent heterogeneous arrays. *Econometric Theory* 12:347–359.

HANSEN, P. R. AND TIMMERMANN, A. 2012. Choice of sample split in out-of-sample forecast evaluation. *European University Institute; Series/Number: EUI ECO; 2012/10* .

HANSEN, R. 2005. A test for superior predictive ability. *Journal of Business & Economic Statistics* 23:365–380.

HARVEY, C. R. AND LIU, Y. 2018. Lucky factors. *Available at SSRN 2528780* .

HE, K., ZHANG, X., REN, S., AND SUN, J. 2016. Deep residual learning for image recognition. *In* Proceedings of the IEEE conference on computer vision and pattern recognition, pp. 770–778.

HOERL, A. E. AND KENNARD, R. W. 1970. Ridge regression: Biased estimation for nonorthogonal problems. *Technometrics* 12:55–67.

HOLM, S. 1979. A simple sequentially rejective multiple test procedure. *Scandinavian Journal of Statistics* 6:65–70.

HORNIK, K., STINCHCOMBE, M., AND WHITE, H. 1989. Multilayer feedforward networks are universal approximators. *Neural networks* 2:359–366.

HUTCHINSON, J. M., LO, A. W., AND POGGIO, T. 1994. A nonparametric approach to pricing and hedging derivative securities via learning networks. *The Journal of Finance* 49:851–889.

INOUE, A. AND SHINTANI, M. 2006. Bootstrapping gmm estimators for time series. *Journal of Econometrics* 133:531–555.

Jacod, J. and Protter, P. 2011. Discretization of Processes. Springer.

Jin, S., V., C., and Swanson, N. 2017. Robust forecast comparison. *Econometric Theory* 33:1306–1351.

Jo, S. and Sekkel, R. 2017. Macroeconomic uncertainty through the lens of professional forecasters. *Journal of Business & Economic Statistics* pp. 1–11.

Jurado, K., Ludvigson, S. C., and Ng, S. 2015a. Measuring uncertainty. *American Economic Review* 105:1177–1216.

Jurado, K., Ludvigson, S. C., and Ng, S. 2015b. Measuring uncertainty. *American Economic Review* 105:1177–1216.

Kilian, L. 1999. Exchange rates and monetary fundamentals: What do we learn from longhorizon regressions? *Journal of Applied Econometrics* 14:491–510.

Kim, H. H. and Swanson, N. R. 2016. Mining big data using parsimonious factor, machine learning, variable selection and shrinkage methods. *International Journal of Forecasting* .

Kitamura, Y. 2002. Econometric comparisons of conditional models.

Langley, P. 1993. Induction of recursive bayesian classifiers. *In* European Conference on Machine Learning, pp. 153–164. Springer.

Lee, S. S. and Mykland, P. A. 2007. Jumps in financial markets: A new nonparametric test and jump dynamics. *The Review of Financial Studies* 21:2535–2563.

Lee, T., White, H., and Granger, C. 1993. Testing for neglected nonlinearity in time series models: A comparison of neural network methods and alternative tests. *Journal of Econometrics* 56:269–290.

Linton, O., E., M., and Y.J, W. 2005. Consistent tesing for stochastic dominance: A subsampling approach. *Review of Economic Studies* 72:735765.

Mancini, C. 2009. Non-parametric threshold estimation for models with stochastic diffusion coefficient and jumps. *Scandinavian Journal of Statistics* 36:270–296.

Marcellino, M., Porqueddu, M., and Venditti, F. 2016. Short-term gdp forecasting with a mixed-frequency dynamic factor model with stochastic volatility. *Journal of Business & Economic Statistics* 34:118–127.

Mariano, R. S. and Murasawa, Y. 2003. A new coincident index of business cycles based on monthly and quarterly series. *Journal of Applied Econometrics* 18:427–443.

Masters, T. 1993. Practical neural network recipes in C++. Morgan Kaufmann.

McCracken, M. 2000a. Robust out-of-sample inference. *Journal of Econometrics* 99:195–223.

McCracken, M. W. 2000b. Robust out-of-sample inference. *Journal of Econometrics* 99:195–223.

McCulloch, W. S. and Pitts, W. 1943. A logical calculus of the ideas immanent in nervous activity. *The bulletin of mathematical biophysics* 5:115–133.

Meese, R. and Rogoff, K. 1983. Empirical exchange rate models of the seventies: Do they fit out-of-sample? *Journal of International Economics* 14:3–24.

Neely, C. J., Rapach, D. E., Tu, J., and Zhou, G. 2014. Forecasting the equity risk premium: the role of technical indicators. *Management Science* 60:1772–1791.

Pesaran, M. H. and Timmermann, A. 1992. A simple nonparametric test of predictive performance. *Journal of Business & Economic Statistics* 10:461–465.

Pesaran, M. H. and Timmermann, A. G. 1994. A generalization of the non-parametric henriksson-merton test of market timing. *Economics Letters* 44:1–7.

Politis, D. N., Romano, J. P., and Wolf, M. 1999. Subsampling. Springer Science & Business Media.

RAPACH, D. AND ZHOU, G. 2013. Forecasting stock returns, pp. 328–383. *In* Handbook of economic forecasting, volume 2. Elsevier.

RAPACH, D. E., STRAUSS, J. K., AND ZHOU, G. 2013. International stock return predictability: what is the role of the united states? *The Journal of Finance* 68:1633–1662.

ROMANO, J. AND WOLF, M. 2005. Stepwise multiple testing as formalized data snooping. *Econometrica* 73:1237–1282.

ROSENBLATT, F. 1958. The perceptron: a probabilistic model for information storage and organization in the brain. *Psychological review* 65:386.

ROSSI, B. 2005. Testing long-horizon predictive ability with high persistence, and the meese – rogoff puzzle. *International Economic Review* 46:61–92.

ROSSI, B. AND INOUE, A. 2012. Out-of-sample forecast tests robust to the choice of window size. *Journal of Business & Economic Statistics* 30:432–453.

ROSSI, B. AND SEKHPOSYAN, T. 2011. Understanding models forecasting performance. *Journal of Econometrics* 164:158–172.

SCHORFHEIDE, F. 2010. Loss function-based evaluation of dsge models. *Journal of Applied Econometrics* 15:645–670.

STINCHCOMBE, M. AND WHITE, H. 1998. Consistent specification testing with nuisance parameters present only under the alternative. *Econometric Theory* 14:295–325.

SWANSON, N. R. AND WHITE, H. 1995. A model-selection approach to assessing the information in the term structure using linear models and artificial neural networks. *Journal of Business & Economic Statistics* 13:265–275.

SWANSON, N. R. AND WHITE, H. 1997. A model selection approach to real-time macroeconomic forecasting using linear models and artificial neural networks. *Review of Economics and Statistics* 79:540–550.

Swanson, N. R. and Xiong, W. 2017. Big data analytics in economics: What have we learned so far, and where should we go from here?

Tibshirani, R. 1996. Regression shrinkage and selection via the lasso. *Journal of the Royal Statistical Society: Series B (Methodological)* 58:267–288.

Vapnik, V. and Chervonenkis, A. 1964. A note on class of perceptron. *Automation and Remote Control* 24.

Vuong, Q. 1989. Likelihood ratio tests for model selection and non-nested hypotheses. *Econometrica* 57:307–333.

Weiss, A. 1996. Estimating time series models using the relevant cost function. *Journal of Applied Econometrics* 11:539–560.

Welch, I. and Goyal, A. 2007. A comprehensive look at the empirical performance of equity premium prediction. *The Review of Financial Studies* 21:1455–1508.

West, K. 1996. Asymptotic inference about predictive ability. *Econometrica* 64:1067–1084.

Whang, Y.-J., Maasoumi, E., Linton, O., et al. 2004. Consistent testing for stochastic dominance: A subsampling approach. Technical report, Financial Markets Group.

White, H. 1982. Maximum likelihood estimation of misspecified models. *Econometrica* 50:1–25.

White, H. 2000. A reality check for data snooping. *Econometrica* 68:1097–1126.

Yao, C. 2019. Measuring uncertainty using mixed frequency macroeconomic and financial volatility risk factors. *Working Paper, Rutgers University.* .