Stockholm School of Economics
Department of Finance
Master of Science in Finance
Master Thesis

Factor return predictability: A comparison of multivariate forecasts of the Size, Value, Momentum, and Low Volatility premia

Francesco Chincoli
December 2015

Abstract

This paper studies the predictive performance of multivariate models in forecasting joint returns of portfolios tracking the Size, Value, Momentum, and Low Volatility premia. We run recursive out-of-sample forecasting experiments on a number of linear and regime switching models, and compare the accuracy of their point predictions in a qualitative and quantitative fashion. We find evidence that a pure random walk model is at least as accurate as more complex statistical models in forecasting returns of the Size, Value, and Momentum factors. We also conclude that the introduction of Markov switching regimes, heteroskedasticity, and autoregressive terms provides a significant improvement in the forecast of the Low Volatility anomaly only. These results appear to be independent of the forecasting horizon and robust to changes in the loss function used in statistical evaluations.

Keywords: Risk Premium, Factor Return Predictability, Forecast Accuracy, Regime Switching
Supervisor: Prof. Paolo Sodini

Acknowledgements

I owe my deepest gratitude: to my supervisor, Prof. Paolo Sodini, who supported and patiently supervised me throughout all the steps of the thesis; to Prof. Massimo Guidolin, for his interest in the topic, his mentoring, and the valuable insights he has been willing to share; to my friends, source of motivation, strength, and ambition; to my parents, who have always strived for my happiness, believed in my potential, and inspired me throughout my life. Thank you.

INDEX

1. INTRODUCTION

Risk premia originate some of the most intriguing debates in the finance literature. Their existence disrupts conventional wisdom and consolidated theories, such as the efficient market hypothesis. Furthermore, their persistence fascinates practitioners, attracted by the opportunity of greater returns with comparable, if not lower, volatility. It is indeed possible to invest in factor portfolios which track these anomalies. Their returns have been extensively analyzed, and are strongly cyclical, heterogenous, and varying in their cross section (Ilmanen, 2011). These features make the estimation of their conditional moments, which are essential inputs in any wealth allocation, a particularly challenging task. Producing accurate and reliable forecasts is a matter of utmost importance for the success of any active investment strategy, and our work is in line with this need.

The objective of this study is to detect the statistical model which is the most accurate in forecasting joint returns of the Size, Value, Momentum, and Low Volatility factors. We run a horse-race among a number of linear and non-linear multivariate models, evaluating the accuracy of their out-of-sample point predictions. We consider three different forecasting horizons and the dividend yield, the term spread, and the default spread as prediction variables. We reach three conclusions.

First, we find evidence that a simple random walk model is at least as accurate as more complex statistical models in forecasting returns of the Size, Value, and Momentum factors. Neither the prediction variables considered nor autoregressive terms contribute to an increase in performance. The introduction of regime-switching parameters, instead, lead to overall better measures of predictive accuracy, but the improvement is not big enough to be considered significant.

Second, the Low Volatility factor requires the introduction of Markov switching regimes, regressive coefficients and heteroskedasticity in order to produce unbiased forecasts. This property is likely to be due to the highly non-normal distribution of its returns, which can be captured only by the most flexible model considered. The accuracy of this model in forecasting the Low Volatility premium is also significantly better than all the other models considered, although still not enough to beat a pure random walk.

Third, the unbiasedness and the predictive accuracy of the models appear to be independent of the forecasting horizon. This outcome is true for all the factor portfolios, with few exceptions.

We provide further details on three key aspects of the methodology adopted to obtain the above results. The first aspect regards the choice of forecast models. We analyze two linear and two non-linear categories of models. In the former case, we estimate a constant expected return model and two vector autoregressive models, while in the latter case we specify two threshold vector autoregressive models and three Markov switching vector autoregressive models. More specifically, the two vector autoregressive models are different in the methodology of the forecast: one is direct, while the other is iterated. The same is true for the two threshold models. Theoretically, direct forecast should be less affected by model misspecification, while iterated forecasts should be more efficient if the model is correctly specified (Marcellino, Stock, and Watson, 2006).[1] The three Markov switching models, instead, varies in the selection of the regime-dependent parameters. Hence, we consider a total of eight models. Moreover, when selecting the most appropriate specification for each of them, we use information criteria that balance the trade-off between in-sample fit and out-of-sample predictive accuracy. The second aspect relates to the evaluation of the forecasts. The returns predicted from each model are compared in a qualitative and quantitative fashion. In particular, we test whether forecasts are unbiased through a simple Wald test and the Mincer and Zarnowitz regression. We also calculate several measures of predictive accuracy, and use the Diebold-Mariano test to assess whether differences in performance between models are statistically significant. We rely on two loss functions, square and absolute, in order to give more robustness to the analysis.

---

[1] The Authors also find that the relative performance of iterated forecasts improves for longer forecast horizons and lag specifications.

Finally, this study contributes to the literature for the combination of risk premia and econometric models analyzed. With regard to the risk premia, it is the first paper to forecast joint returns not only of the Size, Value, and Momentum factors, but also of the Low Volatility factor. The choice of the Size, Value, and Momentum anomalies is grounded in the literature. In particular, Fama and French (1993) were the first to affirm the concepts of Size and Value premia, while Jeegadesh and Titman (1993) introduced the Momentum premium. Among the many other possible market anomalies,[2] we select the Low Volatility premium due to its recent success among academics and practitioners (Haugen and Baker, 1991). For what concerns the econometric models, some studies prove that, for the Size, Value, and Momentum premia, models with regime shifts outperform single-state benchmarks not only in in-sample fit, but also in out-of-sample predictive accuracy (Perez Quiros and Timmermann, 2000; Black and McMillan, 2004; Guidolin and Timmermann, 2008; Gulen, Xing and Zhang, 2011). Nevertheless, no previous work has compared models within the regime switching class when forecasting factor returns. We evaluate forecasts from threshold and Markov switching models of increasing complexity, thus allowing for a switching variable which is either observable, in the former case, or latent, in the latter case. This approach extends previous analyses and lays the foundations for more comprehensive comparisons. In fact, we make the strong assumption that the excess market return is the variable driving regime shifts. In the analysis, threshold models rank especially bad against with Markov switching ones. Hence, it is essential to test the performance of models with different threshold variables. The choice of the threshold variable is indeed one of the many extensions of this work. The inclusion of additional models, such as smooth transition vector autoregressive models, and prediction variables, for instance return dispersion (Stivers and Sun, 2010), would be an essential step in order to further test the robustness of my conclusions.

[2] Some examples are Yield (Blume, 1980), Liquidity (Amihud and Mendelson, 1986), and Quality (Sloan, 1996).

The rest of this paper is organized as follows. Section 2 documents the existing literature on factor return predictability, in which this study can be placed. Section 3 describes the multivariate econometric models that are used in the comparison. Section 4 presents data and results of the analysis, as well as potential extensions and limitations. Section 5 concludes.

2. LITERATURE REVIEW

Stock return predictability is a widely discussed topic in the finance literature (Ang and Bekaert, 2007). Academic research in this field follows two distinct directions. On the one hand, some studies aim at detecting prediction variables that are able to predict stock returns (Rapach, Wohar, and Rangvid, 2005). With this regard, some financial and macroeconomic variables, such as the dividend yield and the term spread, appear to be useful for forecasting excess market returns, especially in-sample (Campbell and Cochrane, 1999; Campbell and Yogo, 2005; Rapach and Wohar, 2006; Cochrane, 2008). On the other hand, regardless of the set of explanatory variables considered, predictability may require non-linear features to be modelled (Guidolin and Timmermann, 2006). In particular, there is no doubt that regime switching models lead to a better fit of the in-sample dynamics among asset returns. However, their out-of-sample forecasting performance is mixed.

Studies on factor return predictability follow a similar pattern: they focus either on prediction variables or on the statistical modelling of non-linear predictive patterns. Ilmanen (2011) especially supports the former field. Collecting and expanding previous research on the Size, Value, and Momentum anomalies, he gives rational and behavioral explanations for their behavior. He shows that the performance of factor portfolios is based on firm-specific and macroeconomic drivers which are often different for each risk premium. For this reason, the literature is mainly concerned with univariate forecast models which have the ability to model such drivers. Zakamulin (2013) and Zakamulin (2014) argue that the Size premium can be predicted by a linear regression based on lagged returns, default spread, and term premium.

Considering the Value premium, Cohen, Polk and Vuolteenaho (2003) find forecasting power in the value spread, which they define as the difference in book-to-market ratio between a typical value stock and a typical growth stock. Chen, Petkova and Zhang (2008), instead, estimate the expected value premium from long-run growth rates and expected dividend-price ratios. Chordia and Shivakumar (2002) conclude that Momentum profits are explained by cross-sectional differences in conditional expected returns, which can be in turn predicted by macroeconomic variables.[3] Perez-Quiros and Timmermann (2000) were, instead, the first contributors to the latter field. The Authors demonstrate the outperformance of heteroskedastic Markov switching models in point forecasts of the returns of the Size factor. Guidolin and Nicodano (2009) extend their work by predicting the dynamics in the density of the Size factor returns. Similarly to Perez-Quiros and Timmermann (2000), Gulen, Xing, and Zhang (2011) use a two-state, heteroskedastic regime switching model with time-varying transition probabilities and several macroeconomic predictors in order to forecast the Value premium.[4] Moreover, Cooper, Gutierrez, and Hameed (2004), Kim, Roh, Min, and Byun (2014), and Wang and Xu (2015) show that regime switching models are capable of producing accurate forecasts of the returns of the Momentum factor.

Although the Size, Value and Momentum premia have enjoyed widespread success among academics and practitioners, the Low Volatility anomaly has recently emerged. Ang, Hodrick, Xing, and Zhang (2006) provide evidence that stocks with high idiosyncratic volatility have surprisingly low returns in recent samples,[5] while Blitz and Val Vliet (2007) and Van Vliet, Blitz and Van der Grient (2011) show that stocks with low realized volatility achieves greater returns than stocks with high realized volatility. These findings pose a challenge to the standard concept that there is a positive relationship between expected volatility and

[3] The Authors find that value-weighted market dividend yield, default spread, term spread and T-bill rates are statistically significant.

[4] Transition probabilities are a function of the One-month T-Bill rate, while the predictors mentioned are default premium, change in money stock and dividend yield.

[5] The Authors consolidate their findings in a subsequent paper (Ang, Hodrick, Xing, and Zhang, 2009).

expected returns, at least in the last decade. Although practitioners have increasingly been implementing investment strategies based on the Low Volatility anomaly, it seems that its predictability has not been evaluated.

Several papers attempt to compare predictions on different risk premia. The literature is however scarce when dealing with multivariate forecasting. Lynch (2000) adopts a vector autoregressive model in order to implement asset allocation on the Value and Size factors. He shows that returns of the two factors are predictable, and this predictability leads to a better risk-return profile on the overall investment. Moreover, Stivers and Sun (2010) forecast returns on the Value and Momentum factors using return volatility as key predictor. Guidolin and Timmermann (2008), instead, prove that regime switching models dominate single state benchmarks in out-of-sample forecasting experiments on the Size and Value premia. Sarwar, Mateus and Todorovic (2014) and Angelidis and Tessaromatis (2014) have recently extended their study by introducing the Momentum factor in the analysis. However, no previous work seems to have considered multivariate models that include the Low Volatility factor.

This paper, therefore, is the first to evaluate the predictive accuracy of econometric models forecasting joint returns of the Size, Value, Momentum, and Low Volatility factors. We find evidence that regime shifts lead to better forecast accuracy only for the Low Volatility anomaly. Sophisticated statistical models fail in improving the performance of a simple random walk model for the other risk premia. These results extend previous studies focused on the comparison between linear and regime switching models, and should be interpreted as conditional on the multivariate framework of the analysis.

3. EMPIRICAL METHODOLOGY

We study the performance of eight multivariate econometric models in forecasting returns of the Value, Size, Momentum and Low Volatility factor. This Section covers the theoretical framework behind these forecast models. They are both linear and non-linear, and can be grouped as follows:

Linear models (Section 3.1):

- Constant Expected Return model (CER) (Section 3.1.1)
- Vector Autoregressive model (VAR) (Section 3.1.2)

Regime switching models (Section 3.2):

- Threshold Vector Autoregressive model (TVAR) (Section 3.2.1)
- Markov Switching Vector Autoregressive model (MSVAR) (Section 3.2.2)

Moreover, we divide each Section into 3 subsections, which respectively focus on the key features of the forecast model under analysis (*The model*), on the estimation of its parameters (*Estimation*), and on the methodology adopted to perform direct and iterated forecasts (*Forecast*). We assume that all variables are covariance stationary, unless specified otherwise.

3.1 Linear models

3.1.1 Constant Expected Return model

The model

The Constant Expected Return (CER) model might be considered the most naïve econometric model. It assumes that variables follow a pure random walk process with constant drift and volatility:[6]

$$\begin{aligned} \boldsymbol{y}_t &= \boldsymbol{\mu} + \boldsymbol{\varepsilon}_t \\ \{\boldsymbol{\varepsilon}_t\}_{t=1}^{T} &\sim D(0, \boldsymbol{\Omega}_\varepsilon) \,, \end{aligned} \tag{1}$$

where $\boldsymbol{y}_t$ is a $N \times 1$ vector of random variables, $\boldsymbol{\mu}$ is a $N \times 1$ vector of constants, and $\boldsymbol{\Omega}_\varepsilon$ is the variance-covariance matrix of the errors. It further assumes that each dependent variable i is weakly stationary ($\exists \mu_i, {\sigma_i}^2 \; \forall i$), ergodic ($\mu_{it} = \mu_{is} = \mu_i \; \forall s, t$) and serially uncorrelated ($Cov(r_{it}, r_{is}) = Cov(\varepsilon_{it}, \varepsilon_{is}) = 0 \; \forall s \neq t$).[7] Each variable *i* has a constant expected return μ_i and a constant expected volatility σ_i, while its covariance with another variable *j* is σ_{ij}.

Estimation

The estimate of $\boldsymbol{\mu}$ is the $N \times 1$ vector collecting the sample mean of the variables (calculated on the full sample), i.e.:

$$\widehat{\boldsymbol{\mu}} = \overline{\boldsymbol{y}} \,. \tag{2}$$

In particular, by minimizing the sum of squares of the residuals using *ordinary least squares* (OLS):[8]

---

[6] The CER model is usually presented with returns $\boldsymbol{r}_t$ as dependent variables. In this Section, we use $\boldsymbol{y}_t$ for more general tractability.

[7] Innovations are usually assumed to be Normally distributed, although it is not necessary here.

[8] The equation is linear, therefore the Gauss-Markov theorem proves that the OLS estimator is the Best Linear Unbiased Estimator (BLUE).

$$min \sum_{t=1}^{T} (\boldsymbol{e}_t{}' \boldsymbol{e}_t) = min \sum_{t=1}^{T} (\boldsymbol{y}_t - \widehat{\boldsymbol{\mu}})' (\boldsymbol{y}_t - \widehat{\boldsymbol{\mu}})$$

$$= min \sum_{t=1}^{T} (\boldsymbol{y}_t{}' \boldsymbol{y}_t - 2 \boldsymbol{y}_t{}' \widehat{\boldsymbol{\mu}} + \widehat{\boldsymbol{\mu}}' \widehat{\boldsymbol{\mu}}) \,.$$

Hence:

$$\frac{\partial \sum_{t=1}^{T} (\boldsymbol{e}_t{}' \boldsymbol{e}_t)}{\partial \widehat{\boldsymbol{\mu}}} = \sum_{t=1}^{T} (-2 \boldsymbol{y}_t + 2 \widehat{\boldsymbol{\mu}})$$

$$= -2 \sum_{t=1}^{T} \boldsymbol{y}_t + 2T \widehat{\boldsymbol{\mu}} = \boldsymbol{0} \,.$$

Finally:

$$\widehat{\boldsymbol{\mu}} = \frac{\sum_{t=1}^{T} \boldsymbol{y}_t}{T} = \overline{\boldsymbol{y}} \,.$$

Forecast

In the CER model, direct and iterated forecasts coincide. Furthermore, they are independent of the horizon *h*, with *h* being a positive integer:

$$\mathrm{E}_t [\boldsymbol{y}_{t+h}] = \widehat{\boldsymbol{\mu}}, \qquad \forall h \,. \tag{3}$$

3.1.2 Vector Autoregressive model

The model

Introduced by Sims (1980), the Vector Autoregressive model VAR(*p*) is the multivariate generalization of the univariate Autoregressive model AR(*p*):

$$\boldsymbol{y}_t = \boldsymbol{\mu} + \sum_{j=1}^{p} \mathbf{A}_j \boldsymbol{y}_{t-j} + \boldsymbol{u}_t \,, \tag{4}$$

where $\boldsymbol{y}_t$ is a $N \times 1$ vector of random variables, $\{\mathbf{A}_1, \mathbf{A}_2, \dots, \mathbf{A}_p\}$ are $N \times N$ time-invariant matrices of autoregressive coefficients and *p* is the number of lags. The error $\boldsymbol{u}_t$ is assumed to have the following properties: zero mean ($\mathrm{E}(\boldsymbol{u}_t) = \mathbf{0}$), covariance matrix $\boldsymbol{\Omega}_u$ and no serial correlation ($Cov(\boldsymbol{u}_s, \boldsymbol{u}_t) = \mathrm{E}(\boldsymbol{u}_s \boldsymbol{u}_t') = \mathbf{0},\ \forall s \neq t$). It follows that the N dependent variables are linear functions of their past values. The VAR model is presented in its reduced form, as we assume that there are no contemporaneous effects.

Equation (4) considers only endogenous variables. The following VARX(*p,q*) model, instead, uses exogenous regressors, too:

$$\boldsymbol{y}_t = \boldsymbol{\mu} + \sum_{j=1}^{p} \mathbf{A}_j \boldsymbol{y}_{t-j} + \sum_{j=1}^{q} \mathbf{B}_j \boldsymbol{x}_{t-j} + \boldsymbol{u}_t\,, \tag{1}$$

where, in addition to the elements described earlier, $\boldsymbol{x}_t$ is a $M \times 1$ vector of exogenous random variables, $\{\mathbf{B}_1, \mathbf{B}_2, \dots, \mathbf{B}_q\}$ are $N \times M$ time-invariant matrices of regressive coefficients, and *q* is the number of lags of the exogenous variables. Similarly to equation (4), neither endogenous nor exogenous variables generate contemporaneous effects.[9]

Estimation

Consider the reduced-form VARX(*p,q*) in equation (5). Assume that the model is covariance stationary and there are no restrictions in its parameters. The model is a *seemingly unrelated regression* (SUR) in which all the equations have the same explanatory variables. Therefore, each equation can be estimated separately by OLS without any loss of efficiency with respect to *generalized least squares* (GLS). In particular, it is possible to rewrite the model in the following matrix notation (Lütkepohl, 2005):

$$\mathbf{Y} = \mathbf{C}\mathbf{Z} + \mathbf{U}\,,$$

where:

[9] It is indeed possible to assume otherwise by adding the term $\mathbf{B}_o \boldsymbol{x}_t$ to the RHS of equation (5).

$$\mathbf{Y} = [\boldsymbol{y}_p \quad \boldsymbol{y}_{p+1} \quad \cdots \quad \boldsymbol{y}_T]$$

$$\mathbf{C} = [\boldsymbol{\mu} \quad \mathbf{A}_1 \quad \mathbf{A}_2 \quad \cdots \quad \mathbf{A}_p \quad \mathbf{B}_1 \quad \mathbf{B}_2 \quad \cdots \quad \mathbf{B}_q]$$

$$\mathbf{Z} = \begin{bmatrix} 1 & 1 & \cdots & 1 \\ \boldsymbol{y}_{p-1} & \boldsymbol{y}_p & \cdots & \boldsymbol{y}_{T-1} \\ \boldsymbol{y}_{p-2} & \boldsymbol{y}_{p-1} & \cdots & \boldsymbol{y}_{T-2} \\ \vdots & \vdots & \ddots & \vdots \\ \boldsymbol{y}_0 & \boldsymbol{y}_1 & \cdots & \boldsymbol{y}_{T-p} \end{bmatrix}$$

$$\mathbf{U} = [\boldsymbol{y}_p \quad \boldsymbol{y}_{p+1} \quad \cdots \quad \boldsymbol{y}_T]\,.$$

The multivariate OLS estimate for **C** is:

$$\hat{\mathbf{C}} = \mathbf{YZ}'(\mathbf{ZZ}')^{-\mathbf{1}}\,. \tag{2}$$

The estimator above is consistent and asymptotically efficient. Furthermore, the *maximum likelihood estimator* (MLE) and an equation-by-equation OLS estimator would both yield the same result (Hamilton, 1994).

Forecast

We rely on the VAR for iterated forecasts, while on the VARX for direct forecasts. This choice is due to the need of estimating future values not only of the variables of interest, but also of their predictors, when iterating the forecast. More specifically, considering a forecast horizon $h \geq 1$ (with $h \in \mathbb{N}$), it is straightforward to recognize that the VARX(1,1):

$$\boldsymbol{y}_{t+h} = \boldsymbol{\mu} + \mathbf{A}_h \boldsymbol{y}_t + \mathbf{B}_h \boldsymbol{x}_t + \boldsymbol{u}_{t+h} \tag{3}$$

allows for the following *direct* forecast:

$$\mathrm{E}_t[\boldsymbol{y}_{t+h}] = \widehat{\boldsymbol{\mu}} + \widehat{\mathbf{A}}_h \boldsymbol{y}_t + \widehat{\mathbf{B}}_h \boldsymbol{x}_t\,. \tag{8}$$

Such forecast can be performed at time *t*, as all the required information are available. However, the outcome is different when the forecast is indirect.[10] Iterating the forecast using the VARX(1) in equation (7):

[10] Here we use the terms *indirect* and *iterated* interchangeably.

$$
\begin{aligned}
\mathrm{E}_t[\boldsymbol{y}_{t+h}] &= \mathrm{E}_t[\boldsymbol{\mu} + \mathbf{A}_1\boldsymbol{y}_{t+h-1} + \mathbf{B}_1\boldsymbol{x}_{t+h-1} + \boldsymbol{u}_{t+h}] \\
&= \widehat{\boldsymbol{\mu}} + \widehat{\mathbf{A}}_1\mathrm{E}_t[\boldsymbol{\mu} + \mathbf{A}_1\boldsymbol{y}_{t+h-2} + \mathbf{B}_1\boldsymbol{x}_{t+h-2} + \boldsymbol{u}_{t+h-1}] + \widehat{\mathbf{B}}_1\mathrm{E}_t[\boldsymbol{x}_{t+h-1}] \\
&= \cdots \\
&= \widehat{\boldsymbol{\mu}}\sum_{i=1}^{h}\left(\widehat{\mathbf{A}}_1{}^{i-1}\right) + \left(\widehat{\mathbf{A}}_1{}^{h}\right)\boldsymbol{y}_t + \sum_{i=1}^{h}\left(\widehat{\mathbf{A}}_1{}^{i-1}\right)\widehat{\mathbf{B}}_1\mathrm{E}_t[\boldsymbol{x}_{t+h-i}]\,.
\end{aligned}
$$

The third term of the RHS of the equation is impossible to estimate without an assumption on the behaviour of $\boldsymbol{x}_t$. It is possible to overcome this issue by assuming that the exogenous predictors follow the same VAR(1) structure of the dependent variables. For this reason, we rely on the VAR(p) structure in equation (4) in order to perform iterated forecasts. This leads to:

$$
\mathrm{E}_t[\boldsymbol{y}_{t+h}] = \widehat{\boldsymbol{\mu}}\sum_{i=1}^{h}\left(\widehat{\mathbf{A}}_1{}^{i-1}\right) + \left(\widehat{\mathbf{A}}_1{}^{h}\right)\boldsymbol{y}_t\,, \tag{9}
$$

where $\widehat{\mathbf{A}}_1$ is the $(N+M)\times(N+M)$ time-invariant matrix containing the estimates of the autoregressive coefficients.[11]

3.2 Regime switching models

Financial markets are frequently unstable. In many cases the relationship among variables has a dynamic, rather than static, nature, which manifests itself in the time series changing behaviour over time. In particular, these changes may be permanent or temporary. In the former case, the time series undergoes a *structural break* which abruptly alters its properties once and for all. For instance, a war, a financial crisis or a ground-breaking technological innovation are likely to lead to a structural break. In the latter case, instead, the time series exhibits recurring shifts over time. Seasonality, business cycles, and the so-called calendar

[11] If in the previous VARX(1) had *N* endogenous and *k* exogenous variables, the VAR in equation (9) involves *$(N+M)$* endogenous variables.

effects are typical cases in which regime changes occur.[12] Furthermore, stock returns often alternate periods of high mean returns and low volatility, also called *bull regimes*, with periods of low mean returns and high volatility, called *bear regimes* (Ang and Bakaert, 2002).

The existence of structural breaks or regimes implies that parameters may be unstable over time. Linear models, which are static over the entire sample, fail to capture that the relationship among variables may be unstable and dynamic. If this is the case, linear models are likely to be misspecified. Regime switching models try to cope with this shortcoming: they acknowledge the presence of regimes and allow parameters to vary across them.[13] Furthermore, regime switches can generate many non-linear effects by mixing the conditional distributions under the different regimes. As a result, regime switching models are able to capture typical features of financial time series, including fat tails, skewness, heteroskedasticity, and time-varying correlations, which are overlooked by single-state, linear models (Ang and Timmermann, 2011).

In its most general specification, the stochastic variable S_t drives regime shifts and, thus, the parameters that characterize the time series:[14]

$$\boldsymbol{y}_t = \boldsymbol{\mu}_{S_t} + \sum_{j=1}^{p} \mathbf{A}_{jS_t} \boldsymbol{y}_{t-j} + \boldsymbol{u}_t \,, \tag{10}$$

where $\boldsymbol{u}_t \sim IID\left(\mathbf{0}, \boldsymbol{\Omega}_{uS_t}\right)$. Similarly to the VAR model in Section 3.1.2, $\boldsymbol{y}_t$ is a $N \times 1$ vector of random variables and $\left\{\mathbf{A}_{1k}, \mathbf{A}_{2k}, \dots, \mathbf{A}_{pk}\right\}_{k=1}^{K}$ are *pK*, $N \times N$ matrices of autoregressive coefficients. Although we assume that every parameter in equation (10) varies across regimes, it is also possible to restrict such variations to specific parameters only.

---

12 Some examples of calendar effects are the *January effect, the weekend effect,* and *bank holiday* effects (Jacobs and Levy, 1988).

13 More specifically, they allow conditional moments to be time-varying.

14 The regime switching VARX specification, which allows for exogenous regressors, will also be considered in the following.

The following two models extend a linear VAR by introducing regime-specific parameters. The key difference between the models is the rationale underlying regime switches. More specifically, in threshold models, regimes are determined by the value of an observable variable, also known as threshold variable (Tong, 1983), while, in Markov switching models, they are generated by a latent variable with Markov structure (Hamilton, 1989).

3.2.1 Threshold Vector Autoregressive model

The model

In a Threshold Vector Autoregressive model (TVAR) (Tong, 1983), S_t assumes *K* values depending on the value of a *threshold variable z* at time $t-d$:

$$S_t = \begin{cases} 1 & if\ z_{t-d} \leq z_1^* \\ 2 & if\ z_1^* < z_{t-d} \leq z_2^* \\ \vdots & \vdots \\ K & if\ z_{K-1}^* < z_{t-d} \end{cases}, \tag{11}$$

where $\{z_1^*, z_2^*, \dots, z_{K-1}^*\}$ are $K-1$ thresholds and the positive integer *d* is the delay with which the threshold variable operates. The threshold variable can be either a single variable or a combination of many variables. Furthermore, it can be either endogenous or exogenous. In the former case the model is referred to as *Self-Exciting* Threshold Autoregressive (SETAR). While the threshold variable is observable, threshold values are unobservable. Therefore, $\{z_1^*, z_2^*, \dots, z_{K-1}^*\}$ have to be estimated in order to infer which regime prevails at any given point in time.

Defining the following set of K dummy variables $\{D_{1t}(\cdot), D_{2t}(\cdot), \dots, D_{Kt}(\cdot)\}_{t=1}^T$:

$$D_{1t}(z_{t-d}) = \begin{cases} 1 & z_{t-d} \leq z_1^* \\ 0 & otherwise \end{cases}, \qquad D_{2t}(z_{t-d}) = \begin{cases} 1 & if\ z_1^* < z_{t-d} \leq z_2^* \\ 0 & otherwise \end{cases}, \qquad \cdots,$$

$$D_{Kt}(z_{t-d}) = \begin{cases} 1 & if\ z_{K-1}^* < z_{t-d} \\ 0 & otherwise \end{cases},$$

we can rewrite the model in (10) in the following way:

$$\boldsymbol{y}_t = \sum_{k=1}^{K} D_{kt}(z_{t-d}) \left(\boldsymbol{\mu}_k + \sum_{j=1}^{p} \mathbf{A}_{jk} \boldsymbol{y}_{t-j} \right) + \boldsymbol{u}_t \,, \tag{12}$$

where, in addition to equation (10), $D_{kt}(z)$ is the *k-th* dummy variable of the set above. Since regimes are exhaustive and mutually exclusive, such a specification does not suffer of dummy variable trap. In other words, $\sum_{k=1}^{K} D_{kt}(z) = 1$ for any t.

It is also possible to augment the model with exogenous variables, so as to obtain the following TVARX:

$$\boldsymbol{y}_t = \sum_{k=1}^{K} D_{kt}(z_{t-d}) \left(\boldsymbol{\mu}_k + \sum_{j=1}^{p} \mathbf{A}_{jk} \boldsymbol{y}_{t-j} + \sum_{j=1}^{q} \mathbf{B}_{jk} \boldsymbol{x}_{t-j} \right) + \boldsymbol{u}_t \,, \tag{13}$$

where, in addition to the previous model, $\boldsymbol{x}_t$ is a $k \times 1$ vector of exogenous random variables, q is the number of lags of the exogenous regressors, and $\left\{\mathbf{B}_{1k}, \mathbf{B}_{2k}, \dots, \mathbf{B}_{qk}\right\}_{k=1}^{K}$ are *kq,* $N \times k$ matrices of regressive coefficients.

Estimation

The TVAR model in (13) has *K* sets of parameters $\left\{\boldsymbol{\mu}_k, \mathbf{A}_{1k}, \mathbf{A}_{2k}, \dots, \mathbf{A}_{pk}, \boldsymbol{\Omega}_{uk}\right\}$, one for each regime, and the threshold values $\{z_1^*, z_2^*, \dots, z_{K-1}^*\}$ to be estimated. A two-step procedure is usually employed for this purpose. The first step consists of the estimation of the threshold values $\{z_1^*, z_2^*, \dots, z_{K-1}^*\}$, which are the source of the non-linearity in the model. Afterwards, the second step involves the estimation of the parameters of the resulting *K* VARs through Conditional Least Squares.[15]

The key issue relates to the first step, i.e. the estimation of the threshold values. We describe the procedure adopted by Tsay (1998), which is a multivariate extension of the results in Chan (1993) and Hansen (1996b). We focus on the special case of $K = 2$ (two

[15] This is possible because we assume that the TVAR is in reduced form, and, thus, the VARs in each regime are so as well.

regimes and one threshold), $p = 1$ (only one lag) and no exogenous variables, but the procedure can be easily generalized. With two regimes, we can write the TVAR model as:

$$\boldsymbol{y}_t = \begin{cases} \boldsymbol{\mu}_1 + \mathbf{A}_{11}\boldsymbol{y}_{t-1} + \boldsymbol{\Omega}_1^{1/2}\boldsymbol{v}_t & if\ z_{t-d} < z^* \\ \boldsymbol{\mu}_2 + \mathbf{A}_{12}\boldsymbol{y}_{t-1} + \boldsymbol{\Omega}_2^{1/2}\boldsymbol{v}_t & if\ z_{t-d} \geq z^* \end{cases}, \tag{14}$$

where $\boldsymbol{v}_t \sim IID(\mathbf{0}, \mathbf{I}_N)$ is a white noise and $\boldsymbol{\Omega}_{S_t}^{1/2}$ can be obtained through the Cholesky decomposition of the matrix $\boldsymbol{\Omega}_{S_t}$. We assume the threshold variable z_t to be covariance stationary and continuous, and estimate the parameters of the model $\{\boldsymbol{\mu}_1, \boldsymbol{\mu}_2, \mathbf{A}_{11}, \mathbf{A}_{12}, \boldsymbol{\Omega}_1^{1/2}, \boldsymbol{\Omega}_2^{1/2}, z^*, d\}$ in the following two-step procedure. In the first step, for given z^* and d, the model reduces to two separate VARs, depending on the values of the threshold variable. The two VARs can then be estimated individually by the same procedure described in Section 3.2.2. Among the other estimates, we obtain two sum of squares of the residuals, one for each equation: $SSR_1(z^*, d)$ and $SSR_1(z^*, d)$.[16] We define the total sum of squares of the residuals of the model in equation (14) as:

$$SSR(z^*, d) = SSR_1(z^*, d) + SSR_2(z^*, d)\,.$$

In the second step, the conditional least squares of z^* and d are:

$$(\hat{z}^*, \hat{d}) = \underset{z^*, d}{\operatorname{argmin}}\, SSR(z^*, d)\,.$$

Finally, the estimates of the parameters $\widehat{\boldsymbol{\mu}}_k$, $\widehat{\mathbf{A}}_{1k}$ and $\widehat{\boldsymbol{\Omega}}_k$are a function of $\hat{z}^*$ and $\hat{d}$, and can be easily computed through Conditional Least Squares.

Forecast

Similarly to the VAR model, we rely on the TVAR for iterated forecasts and on the TVARX for direct forecasts. The reasons for this choice are unchanged from Section 3.2.3.

[16] The two sum of squares of the residuals are a function of the selected parameters z^* and d.

With respect to the *direct forecast*, we refer to the TVARX in (13) with the following specification:

$$\boldsymbol{y}_{t+h} = \sum_{k=1}^{K} D_{k(t+h)}(z_{t+h-d})(\boldsymbol{\mu}_k + \mathbf{A}_{hk}\boldsymbol{y}_t + \mathbf{B}_{hk}\boldsymbol{x}_t) + \boldsymbol{u}_{t+h} \,.$$

On a forecasting horizon $h \geq 1$, assuming $d \geq h$:

$$\mathrm{E}_t[\boldsymbol{y}_{t+h}] = \sum_{k=1}^{K} D_{k(t+h)}(z_{t+h-d})\big(\widehat{\boldsymbol{\mu}}_k + \widehat{\mathbf{A}}_{hk}\boldsymbol{y}_t + \widehat{\mathbf{B}}_{hk}\boldsymbol{x}_t\big) \,. \tag{15}$$

In particular, $d \geq h$ implies that z_{t+h-d} and, thus, $D_{k(t+h)}(z_{t+h-d})$ are known.[17]

In order to perform an *iterated forecast*, we must consider all the variables as endogenous. Looking at the specification in equation (12) and setting $p = 1$:

$$\boldsymbol{y}_{t+1} = \sum_{k=1}^{K} D_{k(t+1)}(z_{t+1-d})(\boldsymbol{\mu}_k + \mathbf{A}_{1k}\boldsymbol{y}_t) + \boldsymbol{u}_{t+1} \,. \tag{16}$$

On a forecast horizon $h \geq 1$, assuming $d = 1$, it is straightforward to obtain the one-step ahead forecast $\widehat{\boldsymbol{y}}_{t+1}$:

$$\begin{aligned}\mathrm{E}_t[\boldsymbol{y}_{t+1}] &= \sum_{k=1}^{K} \mathrm{E}_t\big[D_{k(t+1)}(z_t)(\boldsymbol{\mu}_k + \mathbf{A}_{1k}\boldsymbol{y}_t)\big] \\ &= \sum_{k=1}^{K} D_{k(t+1)}(z_t)\big(\widehat{\boldsymbol{\mu}}_k + \widehat{\mathbf{A}}_{1k}\boldsymbol{y}_t\big) \,.\end{aligned}$$

However, when performing a multi-step ahead forecast, the situation is rather different, because z_{t+1} is unknown (and, consequently, is a random variable). More precisely, considering the forecast at time $t + 2$: [18]

---

[17] Without the last restriction, the forecast cannot be direct, because if $d < h$, then $t + h - d > t$ and $\mathrm{E}_t[z_{t+h-d}]$ would have to be modelled.

[18] Given a continuous random variable $X \in \mathbb{R}$ and a Bernoulli random variable $D \sim B(1, p)$: $\mathrm{E}[DX] = \mathrm{E}[DX|D = 1]p + \mathrm{E}[DX|D = 0](1 - p) = \mathrm{E}[X|D = 1]p$.

$$
\begin{aligned}
\mathrm{E}_t[\boldsymbol{y}_{t+2}] &= \sum_{k=1}^{K} \mathrm{E}_t\big[D_{k(t+2)}(z_{t+1})(\boldsymbol{\mu}_k + \mathbf{A}_{1k}\boldsymbol{y}_{t+1})\big] \\
&= \sum_{k=1}^{K} \widehat{Pr}\big(D_{k(t+2)}(z_{t+1}) = 1\big|z_t\big)\mathrm{E}_t[\boldsymbol{\mu}_k + \mathbf{A}_{1k}\boldsymbol{y}_{t+1}] \\
&= \sum_{k=1}^{K} \hat{p}_{k(t+2)}\big(\widehat{\boldsymbol{\mu}}_k + \widehat{\mathbf{A}}_{1k}\mathrm{E}_t[\boldsymbol{y}_{t+1}]\big) \\
&= \sum_{k=1}^{K} \hat{p}_{k(t+2)}\left(\widehat{\boldsymbol{\mu}}_k + \widehat{\mathbf{A}}_{1k}\sum_{k=1}^{K} D_{k(t+1)}(z_t)\big(\widehat{\boldsymbol{\mu}}_k + \widehat{\mathbf{A}}_{1k}\boldsymbol{y}_t\big)\right),
\end{aligned}
$$

where $p_{k(t+2)} = Pr\big(D_{k(t+2)}(z_{t+1}) = 1\big|z_t\big) = Pr(z_{k-1}^* < z_{t+1} \leq z_k^*)$.

Although the expression above looks like a closed form equation, the estimation of $\hat{p}_{k(t+2)}$ is not straightforward, because z_{t+1} is itself a random variable. The most employed methodology involves simulations. In particular, a model for z_{t+1} has to be specified. We assume that *z* is an endogenous variable, and consequently, that it follows the TVAR model in equation (16).[19] If *z* is the *i-th* endogenous variable in the TVAR, it follows that:

$$
z_{t+1} = \sum_{k=1}^{K} D_{k(t+1)}(z_t)\big(\boldsymbol{\mu}_k{}^{(i)} + \mathbf{A}_{1k}{}^{(i)}z_t\big) + \varepsilon_{t+1} \,,
$$

where the superscript *(i)* indicates the *i-th* row of the corresponding matrix. The simulation procedure consists in generating random values for the error ε_{t+1}. Therefore, it involves further assumptions on its distribution, which might lead to model misspecification.[20] In order to minimize this risk, the historical distribution of the residuals of the model is usually applied. Once the behaviour of z_{t+1} is simulated, the estimated observations are ordered and assigned to each regime depending on how they rank with respect to the estimated thresholds

---

[19] If *z* is exogenous, it is necessary specify a model for its behaviour at $t + 1$. The following procedure does not require any change.

[20] So far, we have only assumed that $\boldsymbol{u}_t \sim IID(0, \boldsymbol{\Omega}_u)$, leaving the distribution unspecified.

$\{\hat{z}_1^*, \hat{z}_2^*, \dots, \hat{z}_{K-1}^*\}$. Finally, $\{\hat{p}_{k(t+2)}\}_{k=1}^K$ can be obtained by dividing the number of estimated z_{t+1} in each regime by the total number of simulations.[21] This procedure can be applied recursively until the forecast horizon *h* has been reached, as we have shown that:

$$\mathrm{E}_t[\boldsymbol{y}_{t+i}] = \sum_{k=1}^{K} \hat{p}_{k(t+i)}\left(\widehat{\boldsymbol{\mu}}_k + \widehat{\mathbf{A}}_{1k}\mathrm{E}_t[\boldsymbol{y}_{t+i-1}]\right), \qquad \forall i \geq 2\,. \tag{17}$$

However, for any given number *n* of simulations between each period¸ the total number of simulations needed to perform the iterated forecast is n^{h-1}. This increase is due to the need of estimating $\{p_{k(t+i)}\}_{k=1}^K$ for any *i* from 2 to *h* in order to obtain $\mathrm{E}_t[\boldsymbol{y}_{t+h}]$.

3.2.2 Markov Switching Vector Autoregressive model

The model

Markov Switching models differ from the non-linear models discussed previously, as regimes are determined by an unobservable variable, rather than an observable one. In its most general specification, a random vector $\boldsymbol{y}_t$ follows a *K*-state Markov Switching Vector Autoregressive process with heteroskedastic component, namely MSIVARH(*K,p*), if: [22]

$$\boldsymbol{y}_t = \boldsymbol{\mu}_{S_t} + \sum_{j=1}^{p} \mathbf{A}_{jS_t}\boldsymbol{y}_{t-j} + \boldsymbol{\Omega}_{S_t}^{1/2}\boldsymbol{v}_t\,, \tag{18}$$

where $\boldsymbol{v}_t \sim NID(\mathbf{0}, \mathbf{I}_N)$ is a Gaussian white noise,[23] *p* is the number of lags, $\{\mathbf{A}_{1k}, \mathbf{A}_{2k}, \dots, \mathbf{A}_{pk}\}_{k=1}^K$ is the set of Kp, $N \times N$ regime-dependent matrices of autoregressive

[21] The higher the number, the smaller the standard errors of the estimates $\{\hat{p}_{k(t+2)}\}_{k=1}^K$.

[22] The acronym can be read as follows. MS: Markov Switching; I: intercept is switching; VAR: vector autoregressive; H: heteroskedastic component. Finally, *K* is the number of regimes and *p* the number of lags, as usual.

[23] In the previous model, we did not make any assumption on the distribution of the innovations. We have to make an assumption on their distribution when dealing with Markov Switching models, as they are estimated through (Q)MLE.

coefficients, and $\mathbf{\Omega}_{S_t}^{1/2}$ is a lower triangular matrix obtained through the Choleski decomposition of the state-dependent covariance matrix $\mathbf{\Omega}_{S_t}$. Conditional on S_t, the MSIVARH(*K,p*) model in equation (18) is identical to the VAR(*p*) model in equation(4).

The switching variable S_t drives regime changes and is latent. Furthermore, S_t follows a Markov chain process which is characterized by the following properties: it is discrete, first-order, k-state, time-homogeneous, irreducible and ergodic. More specifically:

1. *Discrete:* it can only take a finite number of regimes (*K*).
2. *First order:* is a (discrete) stochastic process which is characterized by the Markov property. This means that the probability of $S_t = i$ only depends on $S_{t-1} = j$, with $i, j = 1, 2, \dots, K$:
   $$Pr(S_t = i \,|S_{t-1}, S_{t-2}, \dots, \boldsymbol{y}_t, \boldsymbol{y}_{t-1}, \dots) = Pr(S_t = i \,|S_{t-1} = j) = p_{ij}.$$
   With this notation, p_{ij} is the probability of switching from regime *i* to regime *j*, also called *transition probability*. It is also worth recalling that any *m-th* order, *k* state Markov chain can be rewritten as a first order, *km*-state Markov chain.
3. *Time-homogeneous:* transition probabilities are constant over time. In other words, we assume that the following $K \times K$ matrix, containing the set of transition probabilities from any regime to another, is not time-varying:
   $$\mathbf{P} = \begin{bmatrix} p_{11} & p_{12} & \cdots & p_{1K} \\ p_{21} & p_{22} & \cdots & p_{2K} \\ \vdots & \vdots & \ddots & \vdots \\ p_{K1} & p_{K2} & \cdots & p_{KK} \end{bmatrix}.$$
   This assumption, albeit strong, can be relaxed in favour of time-heterogeneous Markov switching models, in which transition probabilities become a function of time. [24]
4. *Ergodic:* the Markov chain is ergodic if one of the eigenvalues of $\mathbf{P}'$ is 1 and if all the other eigenvalues lie inside the unit circle.[25] We assume that these conditions hold in

[24] In a time-heterogeneous specification, the model itself becomes unstable.

[25] $\mathbf{P}'$ is simply the transpose of the matrix of transition probabilities $\mathbf{P}$ which we mentioned in the previous point. It has *K* rows and *K* columns.

the MS model.[26] Hence, calling $\bar{\boldsymbol{\xi}}$ the $K \times 1$ eigenvector associated with the unit eigenvalue of the matrix $\mathbf{P}'$, we have that $\bar{\boldsymbol{\xi}} = \mathbf{P}'\bar{\boldsymbol{\xi}}$ and $\bar{\boldsymbol{\xi}}\boldsymbol{\iota}_K = \boldsymbol{\iota}_K$. This implies that the Markov chain has long-run, unconditional state probabilities in $\bar{\boldsymbol{\xi}}$, and, consequently, that the model eventually reaches a steady-state. This assumption implies that, if the set of current state probabilities equal to $\bar{\boldsymbol{\xi}}$, the prediction for the state probabilities at any forward period is $\bar{\boldsymbol{\xi}}$ itself.[27] The vector of ergodic probabilities also indicates the unconditional probability of the occurrence of each regime.

5. *Irreducible:* There is no absorbing state, i.e. no regime in which the Markov chain can be trapped. Formally:
$$\nexists q : Pr(S_t = i \,|S_{t-1} = q) = 0, \qquad \forall i = 1, 2, \dots, q-1, q+1, \dots, K.$$
It follows that $\bar{\boldsymbol{\xi}} > \mathbf{0}$.

We have defined two types of probabilities so far. $\mathbf{P}$ is the matrix of transition probabilities, and contains the time-homogeneous probabilities of switching between regimes; $\bar{\boldsymbol{\xi}}$ is the vector of ergodic probabilities, and collects the unconditional probability of occurrence of each regime. Since S_t is latent, they are both unobservable and must be estimated. Accordingly, we can at most obtain data-driven inferences on the nature of $\{S_t\}_{t=1}^{T}$, rather than its exact value over time. The probability of occurrence of each regime at any given point in time can be collected in the following $K \times 1$ *state vector*: [28]

$$\boldsymbol{\xi}_t = \begin{bmatrix} D(S_t = 1) \\ D(S_t = 2) \\ \vdots \\ D(S_t = K) \end{bmatrix}. \tag{19}$$

---

[26] It is possible to prove that, if these conditions are satisfied, then $\mathbf{P}\boldsymbol{\iota}_K = \boldsymbol{\iota}_K$, i.e. the elements in the rows of $\mathbf{P}$ sum to 1 ($\boldsymbol{\iota}_K$ is a $K \times 1$ vector of ones).

[27] Formally, it is possible to prove that $\lim_{h \to \infty} (\mathbf{P}')^h = \bar{\boldsymbol{\xi}}\boldsymbol{\iota}_K'$, i.e. the predicted transition probabilities in the long-run collapse to the ergodic probabilities.

[28] Similarly to Section 3.2.1, $D(S_t = \mathrm{k})$ is a dummy variable with value 1 if the regime *k* prevails at time *t,* and 0 otherwise.

Similarly to S_t, the series $\{\boldsymbol{\xi}_t\}_{t=1}^{T}$ is unobservable. Its estimate is:

$$\hat{\boldsymbol{\xi}}_t = \begin{bmatrix} \widehat{Pr}(S_t = 1) \\ \widehat{Pr}(S_t = 2) \\ \vdots \\ \widehat{Pr}(S_t = K) \end{bmatrix}.$$

$\hat{\boldsymbol{\xi}}_t$ depends on the set of information used. In particular, defining $\Im_t$ as the set of information available at time t, it is possible to differentiate between filtered and smoothed probabilities. *Filtered probabilities* are the estimate of the discrete probability distribution of the regimes at time *t,* conditional to the information available at that time: $\hat{\boldsymbol{\xi}}_{t|t} = \widehat{Pr}(\xi_t|\Im_t)$. Since they are solely based on real time information, $\{\hat{\boldsymbol{\xi}}_{t|t}\}_{t=1}^{T}$ gives the best measure of the inference on the current regime that could have been performed at each time period. *Smoothed probabilities,* instead, are the estimate of the discrete probability distribution of the regimes at time *t,* conditional to the entire sample: $\hat{\boldsymbol{\xi}}_{t|T} = \widehat{Pr}(\boldsymbol{\xi}_t|\Im_T)$. They represent an ex-post measure of the state of the model at time *t,* and, because they use all the information available, they are the most accurate estimate of $\{\hat{\boldsymbol{\xi}}_{t|t}\}_{t=1}^{T}$. Filtered and smoothed probabilities are used with different purposes. The former probabilities are usually employed for recursive inference and forecast, while the latter ones for backtesting. However, when making inference at time *T*, there is no difference between the two because the data sets coincide.

Markov Switching models are capable of capturing non-linear characteristics of the time series due to the differences among conditional moments across regimes. In particular, differences in means not only influence higher moments including variance, skewness and kurtosis, but also generate autocorrelation in first moments. Moreover, differences in means and differences in variance both affect persistence in volatility, which is proportional to the combined persistence of the regimes (Ang and Timmermann, 2011).

Similarly to the previous models, exogenous variables can easily be added to obtain a MSIVARXH(*K,p,q*):

$$\boldsymbol{y}_t = \boldsymbol{\mu}_{S_t} + \sum_{j=1}^{p} \mathbf{A}_{jS_t} \boldsymbol{y}_{t-j} + \sum_{j=1}^{q} \mathbf{B}_{jS_t} \boldsymbol{x}_{t-j} + \boldsymbol{\Omega}_{S_t}^{1/2} \boldsymbol{v}_t \, , \tag{20}$$

where, in addition to equation (18), $\boldsymbol{x}_t$ is a $M \times 1$ vector of exogenous random variables and $\left\{\mathbf{B}_{1k}, \mathbf{B}_{2k}, \dots, \mathbf{B}_{qk}\right\}_{k=1}^{K}$ are $N \times M$ matrices of regime-dependent regressive coefficients. Both the specifications in equations (18) and (20) are very complex to handle, as they require the estimation of:

$$K\left(N + pN^2 + qM^2 + \frac{N(N+1)}{2} + K - 1\right)$$

parameters. The complexity of the model, therefore, increases with the number of endogenous and exogenous variables (*N* and *M*, respectively), the number of states *K* and the number of lags *p*. An alternative specification with a smaller number of parameters is the MSIH(*K,* 0) model:

$$\boldsymbol{y}_t = \boldsymbol{\mu}_{S_t} + \boldsymbol{\Omega}_{S_t}^{1/2} \boldsymbol{v}_t, \tag{21}$$

which has *K* regimes, no autoregressive terms ($p, q = 0$), and regime-dependent covariance matrix $\boldsymbol{\Omega}_{S_t}$. It is also possible to keep the VARX structure and assume that the autoregressive coefficients are regime-independent. We obtain a MSIH(*K,p,q*) with VARX structure:

$$\boldsymbol{y}_t = \boldsymbol{\mu}_{S_t} + \sum_{j=1}^{p} \mathbf{A}_j \boldsymbol{y}_{t-j} + \sum_{j=1}^{q} \mathbf{B}_j \boldsymbol{x}_{t-j} + \boldsymbol{\Omega}_{S_t}^{1/2} \boldsymbol{v}_t, \tag{22}$$

where $\left\{\mathbf{A}_1, \mathbf{A}_2, \dots, \mathbf{A}_p\right\}$ and $\left\{\mathbf{B}_1, \mathbf{B}_2, \dots, \mathbf{B}_q\right\}$ are regime-independent.

Estimation

Markov Switching models are estimated through the Expectation-Maximization (EM) algorithm (Dempster et al, 1977; Hamilton 1990). Grounded in a frequentist framework, it consists of first estimating the unknown parameters of the model by Maximum Likelihood

Estimation (MLE) (Krolzig, 1997),[29] and then, conditional on the parameter estimates, making inferences on the state variable S_t. The procedure is repeated until the point estimates of the parameters converge (Guidolin and Timmermann, 2008).

More specifically, it is possible to rewrite the model in its state-space form (Guidolin, 2012):

Measurement: $$\boldsymbol{y}_t = \mathbf{Y}_t \boldsymbol{\Psi}(\boldsymbol{\xi}_t \otimes \boldsymbol{\iota}_N) + \boldsymbol{\Sigma}^*(\boldsymbol{\xi}_t \otimes \mathbf{I}_N)\boldsymbol{\varepsilon}_t \tag{23}$$

Transition: $$\boldsymbol{\xi}_{t+1} = \mathbf{P}'\boldsymbol{\xi}_t + \boldsymbol{u}_{t+1}\,, \tag{24}$$

where $\mathbf{Y}_t$ is a $N \times (Np+1)$ vector with structure $[1 \quad \boldsymbol{y}'_{t-1} \quad \cdots \quad \boldsymbol{y}'_{t-p}] \otimes \boldsymbol{\iota}_N$, $\boldsymbol{\Psi}$ is a $(Np+1) \times NK$ matrix containing the parameters of the VAR:

$$\boldsymbol{\Psi} = \begin{bmatrix} \boldsymbol{\mu}'_1 & \boldsymbol{\mu}'_2 & \cdots & \boldsymbol{\mu}'_K \\ \mathbf{A}_{11} & \mathbf{A}_{12} & \cdots & \mathbf{A}_{1K} \\ \vdots & \vdots & \ddots & \vdots \\ \mathbf{A}_{p1} & \mathbf{A}_{p2} & \cdots & \mathbf{A}_{pK} \end{bmatrix},$$

and $\boldsymbol{\Sigma}^*$ is a $N \times NK$ matrix containing all the *K* Choleski factors $\{\boldsymbol{\Sigma}_1, \boldsymbol{\Sigma}_2, \dots, \boldsymbol{\Sigma}_K\}$ for which $\boldsymbol{\Sigma}^*(\boldsymbol{\xi}_t \otimes \mathbf{I}_N)(\boldsymbol{\xi}_t \otimes \mathbf{I}_N)'(\boldsymbol{\Sigma}^*)' = \boldsymbol{\Omega}_{S_t}$. With regard to the innovations of the two equations, $\boldsymbol{\varepsilon}_t \sim NID(\mathbf{0}, \mathbf{I}_N)$, while $\boldsymbol{u}_{t+1}$ is a zero-mean discrete martingale difference sequence vector, uncorrelated not only with $\boldsymbol{\varepsilon}_{t+1}$, but also with $\boldsymbol{\xi}_{t-j}$, $\boldsymbol{\varepsilon}_{t-j}$, and $\boldsymbol{y}_{t-j}$ $\forall j \geq 0$.

The Expectation and the Maximization steps are based on the state-space form in equations (23) and (24). We first define the probability density function (henceforth: PDF) of $\boldsymbol{y}_t$, conditional on the realization of regime *k* and the information up to $t-1$ ($\Im_{t-1}$), which is Normal:[30]

[29] The estimation is actually performed by *Quasi* Maximum Likelihood Estimation (QMLE), as the distribution of the errors is unknown. QMLE estimates are consistent if the conditional mean function and the conditional variance function are correctly specified (Bollerslev and Wooldridge, 2012).

[30] This is due to the assumption of Normally distributed innovations in the measurement equation in (23).

$$f(\boldsymbol{y}_t|S_t = k, \Im_{t-1}) = \ln\frac{1}{\sqrt{2\pi}}\ln\frac{1}{\sqrt{|\boldsymbol{\Omega}_k|}}exp\left(\left(\boldsymbol{y}_t - \overline{\boldsymbol{y}}_{k,t}\right)'\boldsymbol{\Omega}_k^{-1}\left(\boldsymbol{y}_t - \overline{\boldsymbol{y}}_{k,t}\right)\right). \tag{25}$$

If we assume that the regime at time t is unknown, but the regime at time $t-1$ ($S_{t-1} = i$) is observable, the PDF of $\boldsymbol{y}_t$, conditional on the information available at time $t-1$ ($\Im_{t-1}$), is the following mixture of Normal distributions:

$$f(\boldsymbol{y}_t|S_{t-1} = i, \Im_{t-1}) = \sum_{j=1}^{K} p_{ij}\left(\ln\frac{1}{\sqrt{2\pi}}\ln\frac{1}{\sqrt{|\boldsymbol{\Omega}_j|}}exp\left(\left(\boldsymbol{y}_t - \overline{\boldsymbol{y}}_{j,t}\right)'\boldsymbol{\Omega}_j^{-1}\left(\boldsymbol{y}_t - \overline{\boldsymbol{y}}_{j,t}\right)\right)\right),$$

where p_{ij} is the probability of switching from regime *i* to regime *j*.

Then, we make inference on the state variable S_t and, thus, on the Markov chain. We have, therefore, to estimate the vector $\{\boldsymbol{\xi}_t\}_{t=1}^{T}$. From the transition equation in (24), it follows that:[31]

$$\mathrm{E}_t[\boldsymbol{\xi}_{t+1}|\Im_t] = \hat{\boldsymbol{\xi}}_{t+1|t} = \mathbf{P}'\hat{\boldsymbol{\xi}}_{t|t}\,. \tag{26}$$

If, more realistically, we assume that S_{t-1} is not observable, the PDF of $\boldsymbol{y}_t$, conditional solely on the information set $\Im_{t-1}$, can be written as:

$$f(\boldsymbol{y}_t|\Im_{t-1}) = \begin{bmatrix} f(\boldsymbol{y}_t|S_t = 1, \Im_{t-1}) \\ f(\boldsymbol{y}_t|S_t = 2, \Im_{t-1}) \\ \vdots \\ f(\boldsymbol{y}_t|S_t = K, \Im_{t-1}) \end{bmatrix}' \hat{\boldsymbol{\xi}}_{t|t-1} = \boldsymbol{\eta}_t'\mathbf{P}'\hat{\boldsymbol{\xi}}_{t-1|t-1}\,, \tag{27}$$

where $\boldsymbol{\eta}_t'$ collects the *K* PDF in equation (25), which are conditional on S_t and $\Im_{t-1}$. Finally, the joint PDF of the whole sample $\{\boldsymbol{y}_t\}_{t=1}^{T}$, conditional on all the information available ($\Im_T$), is:

$$f(\{\boldsymbol{y}_t\}_{t=1}^{T}|\{\boldsymbol{\xi}_t\}_{t=1}^{T}, \Im_T) = \sum_{\{\xi_t\}_{t=1}^{T}}\prod_{t=1}^{T} f(\boldsymbol{y}_t|\boldsymbol{\xi}_t, \Im_{t-1})Pr(\boldsymbol{\xi}_t|\boldsymbol{\xi}_1)\,. \tag{28}$$

[31] Similarly, for any horizon $h \geq 1$, where $h \in \mathbb{N}$: $\mathrm{E}_t[\boldsymbol{\xi}_{t+h}|\Im_t] = \hat{\boldsymbol{\xi}}_{t+h|t} = (\mathbf{P}')^h\boldsymbol{\xi}_t$.

The PDF in equation (28) is the likelihood function which is maximized in the maximization step.

The EM algorithm works in the following way. Calling $\boldsymbol{\theta}$ the vector collecting all the parameters of the model, we initialize the iteration with arbitrary $\boldsymbol{\xi}_1$ and $\widetilde{\boldsymbol{\theta}}^0$.[32] Conditional on these parameters, we make inference on the filtered and smoothed probabilities, respectively $\{\tilde{\boldsymbol{\xi}}^1_{t|t}\}_{t=1}^T$ and $\{\tilde{\boldsymbol{\xi}}^1_{t|T}\}_{t=1}^T$ (expectation step). Given these estimated smoothed probabilities, we use appropriate first-order conditions to maximize the likelihood function in equation (28) (maximization step). We obtain new estimates of the parameters $\widetilde{\boldsymbol{\theta}}^1$ in the maximization step. These estimates are then used to make new inference on the filtered and smoothed probabilities. The EM procedure involves the iteration of the expectation and the maximization steps until convergence.[33] Under standard regularity conditions, such as identifiability and stability, the ML estimates are consistent and asymptotically normal (Hamilton, 1989; Leroux, 1993).

Forecast

Similarly to the previous models, we rely on a VARX structure for direct forecasts, and on a VAR structure for iterated ones. The reasons for this choice are unchanged from Section 3.2.3.

With respect to the *direct forecast*, we refer to the following MSIVARXH:

$$\boldsymbol{y}_{t+h} = \boldsymbol{\mu}_{S_{t+h}} + \mathbf{A}_{hS_{t+h}}\boldsymbol{y}_t + \mathbf{B}_{hS_{t+h}}\boldsymbol{x}_t + \boldsymbol{\Omega}^{1/2}_{S_{t+h}}\boldsymbol{\upsilon}_{t+h}\,.$$

On a forecast horizon $h \geq 1$, assuming $d \geq h$:

$$\mathrm{E}_t[\boldsymbol{y}_{t+h}] = \mathrm{E}_t\left[\boldsymbol{\mu}_{S_{t+h}} + \mathbf{A}_{hS_{t+h}}\boldsymbol{y}_t + \mathbf{B}_{hS_{t+h}}\boldsymbol{x}_t\right]$$

---

[32] In the following, the subscript indicates the iteration number, while the accent $\sim$ indicates that the estimate is still under an iteration algorithm.

[33] Given two subsequent iteration having estimates $\boldsymbol{\theta}'$ and $\boldsymbol{\theta}$, convergence is achieved if $\boldsymbol{\theta}' - \boldsymbol{\theta} < \bar{\varepsilon}$, with $\bar{\varepsilon}$ being a small number (usually e^{-4}). In other words, when the two subsequent iterations leave the estimated parameters approximately unaltered.

$$= \sum_{k=1}^{K} \left(\widehat{\boldsymbol{\mu}}_k + \widehat{\mathbf{A}}_{hk}\boldsymbol{y}_t + \widehat{\mathbf{B}}_{hk}\boldsymbol{x}_t\right)\widehat{Pr}(S_{t+h} = k|\Im_t)$$

$$= \sum_{k=1}^{K} \left(\widehat{\boldsymbol{\mu}}_k + \widehat{\mathbf{A}}_{hk}\boldsymbol{y}_t + \widehat{\mathbf{B}}_{hk}\boldsymbol{x}_t\right)\widehat{\boldsymbol{\xi}}'_{t+h|t}\boldsymbol{e}_k$$

$$= \sum_{k=1}^{K} \left(\widehat{\boldsymbol{\mu}}_k + \widehat{\mathbf{A}}_{hk}\boldsymbol{y}_t + \widehat{\mathbf{B}}_{hk}\boldsymbol{x}_t\right)\widehat{\boldsymbol{\xi}}'_{t|t}\widehat{\mathbf{P}}^h\boldsymbol{e}_k \ ,$$

where $\widehat{\boldsymbol{\xi}}_{t+h|t} = (\mathbf{P}')^h\boldsymbol{\xi}_{t|t}$ stems from the transition equation in (24). We can estimate the coefficients of the *K* VARX $\{\boldsymbol{\mu}_k, \mathbf{A}_{hk}, \mathbf{B}_{hk}\}_{k=1}^{K}$, the matrix of transition probabilities **P**, and the state price vector $\boldsymbol{\xi}_t$ through the EM algorithm is the previous subsection. Hence, we have obtained a closed form solution.

With regard to the *iterated forecast*, instead, we refer to the following MSIVARH:

$$\boldsymbol{y}_{t+1} = \boldsymbol{\mu}_{S_{t+1}} + \mathbf{A}_{1S_{t+1}}\boldsymbol{y}_t + \boldsymbol{\Omega}_{S_{t+1}}^{1/2}\boldsymbol{v}_{t+1} \ . \tag{29}$$

The one-step ahead forecast is:

$$\mathrm{E}[\boldsymbol{y}_{t+1}|\Im_t] = \sum_{k=1}^{K} \left(\widehat{\boldsymbol{\mu}}_k + \widehat{\mathbf{A}}_{1k}\boldsymbol{y}_t\right)\widehat{\boldsymbol{\xi}}'_{t+1|t}\boldsymbol{e}_k$$

$$= \sum_{k=1}^{K} \left(\widehat{\boldsymbol{\mu}}_k + \widehat{\mathbf{A}}_{1k}\boldsymbol{y}_t\right)\widehat{\boldsymbol{\xi}}'_{t|t}\mathbf{P}\boldsymbol{e}_k \ .$$

However, a multi-step ahead is less straightforward. In particular, on a horizon $h > 1$, $\{\boldsymbol{y}_{t+h-1}, \boldsymbol{y}_{t+h-2}, \dots, \boldsymbol{y}_{t+1}\}$ are unknown and have to be modelled. Their prediction, conditional on the information set $\Im_t$, depends, in turn, on the filtered probabilities $\{\widehat{\boldsymbol{\xi}}_{t+h-1|t}, \widehat{\boldsymbol{\xi}}_{t+h-2|t}, \dots, \widehat{\boldsymbol{\xi}}_{t+1|t}\}$, which are autocorrelated due to regime switching.[34] We follow the suggestion of Doan, Littermann, and Sims (1984), who substitute $\{\widehat{\mathrm{E}}[\boldsymbol{y}_{t+h-1}|\Im_t], \widehat{\mathrm{E}}[\boldsymbol{y}_{t+h-2}|\Im_t], \dots, \ \widehat{\mathrm{E}}[\boldsymbol{y}_{t+1}|\Im_t]\}$ for

[34] Autocorrelation stems from the transition equation in (24).

$\{\mathrm{E}[\boldsymbol{y}_{t+h-1}|\Im_{t+h-2}], \mathrm{E}[\boldsymbol{y}_{t+h-2}|\Im_{t+h-3}], \dots, \mathrm{E}[\boldsymbol{y}_{t+1}|\Im_{t}]\}$. Although this procedure incorrectly assumes that state price vectors are not autocorrelated, it is customary in applied studies (Guidolin, 2011). With this approximation, we can then use a simulation procedure to model $\{\boldsymbol{y}_{t+h-1}, \boldsymbol{y}_{t+h-2}, \dots, \boldsymbol{y}_{t+1}\}$. More specifically, given the estimates of the parameters $\left\{\boldsymbol{\mu}_k, \mathbf{A}_{1k}, \mathbf{B}_k, \boldsymbol{\Omega}_k^{1/2}\right\}_{k=1}^{K}$, of the matrix of transition probabilities **P**, and of the state price vector $\boldsymbol{\xi}_t$, and assuming that $\boldsymbol{v}_t \sim NID(\mathbf{0}, \mathbf{I}_N)$, we can simulate $\boldsymbol{y}_{t+1}$ in the following way:

$$\widetilde{\boldsymbol{y}}_{t+1}^{(i)} = \sum_{k=1}^{K} \left(\widehat{\boldsymbol{\mu}}_k + \widehat{\mathbf{A}}_{1k} \boldsymbol{y}_t\right) \widehat{\boldsymbol{\xi}}'_{t|t} \mathbf{P} \boldsymbol{e}_k + \widehat{\boldsymbol{\Omega}}_{t+1|t}^{1/2} \widetilde{\boldsymbol{v}}_{t+1}^{(i)} ,$$

where, running *n* simulations, the superscript *(i)* indicates the simulation number. In particular, the matrix $\widehat{\boldsymbol{\Omega}}_{t+1|t}$ does not equal to the average of the covariance matrices of the *K* regimes weighted by the filtered probabilities, because differences in the conditional means affect higher moments, including variance, and correlations among variables. The algorithm returns the set $\left\{\widetilde{\boldsymbol{y}}_{t+1}^{(i)}\right\}_{i=1}^{n}$. We can iterate the procedure for each simulation *i* until the horizon *h* is reached:

$$\widetilde{\boldsymbol{y}}_{t+h}^{(i)} = \sum_{k=1}^{K} \left(\widehat{\boldsymbol{\mu}}_k + \widehat{\mathbf{A}}_{1k} \widetilde{\boldsymbol{y}}_{t+h-1}^{(i)}\right) \widehat{\boldsymbol{\xi}}'_{t|t} \mathbf{P}^h \boldsymbol{e}_k + \widehat{\boldsymbol{\Omega}}_{t+h|t+h-1}^{1/2} \widetilde{\boldsymbol{v}}_{t+h}^{(i)} .$$

Finally, the iterated forecast of $\boldsymbol{y}_{t+h}$ is the average of the simulated values at $t+h$:

$$\mathrm{E}[\boldsymbol{y}_{t+h}|\Im_t] = \frac{1}{n} \sum_{i=1}^{n} \widetilde{\boldsymbol{y}}_{t+h}^{(i)} .$$

4. DATA AND RESULTS

4.1 The Data

We study monthly returns on five US stock portfolios over the sample 1929:01 – 2012:12, which comprises a total of 1008 observations. The portfolios are constructed in order to track excess returns on the market portfolio and five risk premia in the US market. We consider NYSE, AMEX and NASDAQ listed firms, with data from the value-weighted Center for Research in Security Prices (CRSP) index. This Section is organized as follows. Subsection 4.1.1 covers the methodology used to create the five factor portfolios, whose returns are the object of this study. Subsection 4.1.2 provides more details on the predictors considered. Finally, Subsection 4.1.3 shows some descriptive statistics on the previous variables.

4.1.1 Factor Portfolios

We collect continuously compounded returns of the following five portfolios: MKT-RF (excess market), SMB (small minus big), HML (high minus low), MOM (momentum) and LOWVOL (low volatility), with data from Kenneth French's data repository.[35] In the following, we describe the standard procedure adopted for their construction.

The first portfolio, MKT, is the market portfolio. It is constituted by all the NYSE, AMEX and NASDAQ listed firms, value-weighted. As the following portfolios are zero net investment, the return on the MKT portfolio is net of the risk-free rate, which we proxy using the 1-month T-Bill rate (Guidolin and Timmermann, 2008). Henceforth, we refer to these excess returns as MKT-RF. The other four portfolios are constructed with the aim of capturing the Size, Value, Momentum, and Low Volatility anomalies. Portfolios are value-weighted and rebalanced yearly in the month of June.

---

[35] More details on portfolio construction are available from the following website: http://mba.tuck.dartmouth.edu/pages/faculty/ken.french/data_library.html.

Similarly to Fama and French (1993), the SMB and the HML portfolios track the Size and the Value premium. The Authors rank stocks according to firm size, measured by market equity (ME), and book-to-market equity, which is defined as the ratio between book equity and market equity (BE/ME). With regard to the former measure, stocks with market capitalization lower than the NYSE median enter the *Small* portfolio, while stocks with market capitalization higher than the NYSE median enter the *Big* portfolio. With regard to the latter measure, they sort stocks into three portfolios, *Value, Neutral* and *Growth,* with breakpoints on the 30th and the 70th percentile of the book-to-market ratio.[36] Finally, they create six portfolios based on the intersection of the two size and the three book-to-market portfolios.[37] The SMB portfolio is long in small firms and short in big firms, independently of their book-to-market equity:

$$r_t^{SMB} = \frac{1}{3}(Small\,Value + Small\,Neutral + Small\,Growth) \\ - \frac{1}{3}(Big\,Value + Big\,Neutral + Big\,Growth).$$

The HML portfolio, instead, is long in firms with high book-to-market equity and short in firms with low book-to-market equity, independently of their size:[38]

$$r_t^{HML} = \frac{1}{2}(Small\,Value + Big\,Value) - \frac{1}{2}(Small\,Growth + Big\,Growth).$$

We use the MOM (Momentum) portfolio in order to capture the Momentum premium, following Fama and French (2010). The Authors consider all the stocks with prior return, i.e. for which data on stock returns from $t-2$ to $t-12$ is available,[39] and rank them on the basis

[36] Book equity at time *t* is calculated at the end of the fiscal year $t-1$, while market equity at the end of December $t-1$.

[37] As mentioned earlier, the following procedure is repeated every twelve months in the month of June.

[38] The *Small Neutral* and the *Big Neutral* portfolios do not enter the equation.

[39] We drop the return at time $t-1$, following a standard procedure in the momentum literature.

of size and prior returns. They use the same methodology of Fama and French (1993) in order to make the two size groups. Furthermore, they consider the average of the 11 returns between $t-2$ and $t-12$ when sorting for prior returns at time t. Breakpoints at the 30th and 70th percentile are then used to form the *Low, Medium* and *High Momentum* portfolios. Finally, they create six portfolios based on the intersection of the two size and the three momentum portfolios. The MOM portfolio is long in firms with high prior returns and short in firms with low prior returns, independently of their size:[40]

$$r_t^{MOM} = \frac{1}{2}(Small\ High\ Momentum + Big\ High\ Momentum) \\ - \frac{1}{2}(Small\ Low\ Momentum + Big\ Low\ Momentum)\ .$$

The LOWVOL portfolio aims at measuring the Low Volatility premium. We sort stocks on the basis of the realized variance of their returns in the previous 12 months. Then, we construct the portfolio as the difference between the tenth decile (lowest realized variance) and the first decile (highest realized variance). Therefore, it is long in low volatility stocks and short in high volatility stocks.

4.1.2 Predictors

Empirical studies on general return predictability often rely on exogenous predictors in order to improve the accuracy of forecast models. We use lagged values of the *dividend yield, term spread* and *default spread* in order to achieve the same objective. These variables not only enjoy broad success in the general forecasting literature (Fama and French, 1988; Fama, 1990; Guidolin and Ono, 2006), but are also the prevailing choice for forecasting risk premia, due to their predictive power (Perez-Quiros and Timmermann, 2000; Chordia and Shivakumar, 2002; Gulen, Xing and Zhang, 2011; Zakamulin, 2013).

---

[40] Similarly to the previous case, the *Small Medium Momentum* and the *Big Medium Momentum* do not enter the equation.

We take data from the CRSP. The *dividend yield* (DY) is calculated as the logarithm of the ratio between the aggregate dividends on the value-weighted CRSP index over the last 12 months and the end-of-month price of the index. The *term spread* (TMS) is defined as the difference between the CRSP long-term bond yield, which corresponds to a 20-year Treasury bond rate,[41] and the 1-month T-bill rate.[42] Finally, the *default spread* (DFY) is the difference between the average Moody's Bbb and Aaa seasoned corporate bonds with comparable maturities. We convert term spread and default spread to continuously compounding yield in order to make them consistent with factor returns.

4.1.3 Summary Statistics

In this Section, we describe observations on the five factor portfolios, namely Excess Market Return (MKT-RF), Size (SMB), Value (HML), Momentum (MOM), and Low Volatility (LOWVOL), and on the three predictors, i.e. dividend yield (DY), term spread (TMS), and default spread (DFY). Observations on factor portfolios are monthly returns while observations on predictors are monthly yields. They are both continuously compounded.

Table 1 contains summary statistics on the full sample available (1929:01 – 2012:12). In particular, Panel A displays annualized mean return, standard deviation, skewness, and kurtosis of the variables. The Market, Size, Value, and Momentum portfolios, albeit zero investment, achieve positive returns, on average, consistently with Carhart (1997). In particular, the mean monthly excess return of the market portfolio is 0.43% (5.16% per annuum),[43] which is a typical value in the equity premium literature (Fama and French, 2002),

---

[41] More precisely, the long-term bond is the closest active bond having a remaining maturity of 20 years. For more detail, see http://www.crsp.com/products/documentation/crsp-select-file-specifications.

[42] Several definitions of term spread can be found in the finance literature. They differ in the choice of the two yields. 1 month, 3 months and 12 months are common choices for the short-term rate, while 5 years, 10 years, 20 years and 30 years for the long-term rate.

[43] We always refer to continuously compounded returns, i.e. log returns. Annualized returns are therefore calculated in the following way: $\bar{r}^{annualized} = 12\bar{r}^{monthly}$.

with an annualized volatility of 19.03%.[44] The Size premium factor has a lower mean return, 0.18% (2.20% per annuum), with a lower volatility of 11.03% annualized; the Value premium factor had a monthly return of 0.50% (5.99% per annuum), with an annualized volatility of 11.90%. The Momentum premium is the most aggressive factor, as it had the highest average monthly return, 0.53% (6.31% per annuum), with an annualized volatility of 18.61%. The Low Volatility portfolio, instead, performed poorly over the whole sample: it had a significantly negative average return (-5.48% monthly), with the highest volatility (21.21% monthly). Although surprising, this finding is consistent with the finance literature, which ascertained the existence of the Low Volatility anomaly only in recent times (Blitz and Val Vliet, 2007). Furthermore, Figure 1, plotting the histogram of each variable against the Normal distribution (top graph), and the Quantile-Quantile (QQ) plots (bottom graph), shows that the distribution of factor returns is likely to be non-normal. The last column of Panel A (Table 1), which contains t-statistics from the Jarque-Bera test, confirms this impression. More specifically, the null hypothesis of Gaussian distribution is rejected for all factors and predictors, proving that they have highly non-normal distributions with skewness and fat tails.[45] In Panel B, we report correlations among returns of factors and predictors. Interestingly, sample correlations among factor returns are quite low: excluding the Low Volatility factor, the highest figure (in absolute value) is 31%. Sample correlations with the Low Volatility factor are, instead, the highest, ranging between 10% and 50%. Finally, Momentum and Low Volatility factor returns seem to be negatively correlated with Market, Size, and Value factor returns. Negative correlation is a particularly desirable property, as it allows for the construction of portfolios with low overall volatility.

---

[44] We calculate annualized volatility in the following way: $\sigma^{annualized} = \sqrt{12}\sigma^{monthly}$. Such computation is based on the assumption that returns *IID*. Although the underlying hypothesis is strong, the measure is still useful for the comparison of summary statistics.

[45] The Jarque-Bera test involves a joint (null) hypothesis that the skewness and the excess kurtosis of the empirical distribution of the data match the values of a normal distribution.

In Table 2, we collect summary statistics on the forecasting window 1980:01 – 2012:12 in order to assess whether they are similar to the ones in the full sample. Panel A shows that the moments of factor returns, although different in values, are consistent with the previous findings. All variables have non-normal distributions in this sample, too. The only element of inconsistency relates to the Value premium: its correlation with the other variables changes in sign, while its distribution has lower skewness and thinner tails, although still non-normal.

4.2 Empirical Results

We study the performance of multivariate econometric models in forecasting monthly returns of factor portfolios over the sample 1980:01 – 2012:12. The exercise is recursive and out-of-sample, as we only use information that were available at the time in which the forecast, either direct or iterated, is performed.[46] However, we rely on the entire sample (1929:01 – 2012:12) when making inference on the goodness of fit of the models. For this reason, the finance literature usually terms these estimates *pseudo* out-of-sample (Inoue and Kilian, 2005).[47]

We consider three forecast horizons, i.e. 1-month, 3-month, and 12-month, which are standard not only in the forecasting literature, but also in research on portfolio optimization. In addition, we compare four categories of multivariate models of increasing complexity. The simplest model is the Constant Expected Return, which serves as a benchmark for the other models, as it rules out any predictability in factor returns. Vector Autoregressive models introduce predictability from lagged returns and predictors. In this class of models, we rely on the VARX for direct forecasts and the VAR for iterated forecasts. The key difference in the two specifications relates to predictors, which are exogenous variables in the former case, while exogenous in the latter. Furthermore, Threshold and Markov Switching models extend the previous specifications by allowing for regime switches in order to capture breaks and non-

---

[46] In other words, when making a forecast from time t to time $t+h$, we only use data up to time t.

[47] In order to have strictly out-of-sample forecasts, the goodness of fit of the models should be tested at each time period.

normal features of factor returns. In particular, threshold models count numerous possible specifications, depending on the lag structure, the threshold variable(s), and the delay parameter. This study focuses on a TVARX and a TVAR for better comparability with the VARX and the VAR models. Their threshold variable is the excess market return, with a delay equal to the smallest lag chosen. This choice is arbitrary, but consistent with previous findings in the finance literature, which stresses the persistence of bull and bear regimes. With regard to Markov switching models, we consider several specification. First, we extend the simple random walk model by introducing regime changes in the mean parameter (MSI model). Furthermore, we allow volatility to be regime dependent, thus taking heteroskedasticity into account (MSIH model). Finally, we add lagged returns and predictors to the previous specification, and obtain a complete MSIVARH model. Therefore, we test a number of Markov switching models, rather than simply focusing on the model that fits the data best. This procedure allows for easier comparability, and is more robust to misspecifications in the model selection phase. As a result, we study the forecast accuracy of the following eight models: a pure random walk model (CER); two vector autoregressive models with direct and iterated forecasts (VARX and VAR, respectively); two threshold vector autoregressive models with direct and iterated forecasts (TVARX and VAR, respectively); three Markov switching models with switching intercept, of which one has no autoregressive terms and homoskedastic errors (MSI), one has no autoregressive terms and heteroskedastic errors (MSIH), and, finally, one with lag structure, predictors and heteroskedastic errors (MSIVARH).

The rest of the Section is organized as follows. In Subsection 4.2.1 (*Model Selection*), we first assess the optimal specification for each model through information criteria. In Subsection 4.2.2 (*Forecast Performance*), we compare the forecast models in both a qualitative and a quantitative fashion. Furthermore, in Subsection 4.2.3 (*Testing for Differential Forecasting Accuracy*) we test whether differences in predictive accuracy between models are statistically significant. Finally, in Subsection 4.2.4 (*Final Remarks*), we collect and discuss the key results of the exercise, and highlight possible extensions and limitations.

4.2.1 Model Selection

The first key task in any empirical application involves the selection of the most appropriate model for the data set. Model adequacy is inevitably linked with the purpose of the analysis, as a misspecified model is unlikely to produce unbiased and accurate forecast. Therefore, for each model considered, we study the specification which fits the data best. With this regard, we use three standard criteria to sort through the candidate models: the *Akaike Information Criterion* (AIC), the *Schwartz Criterion* (SC), and the *Hannan-Quinn Criterion* (HQ). For a general model, the relative statistics can be computed as follows:

$$\mathrm{AIC} = \frac{1}{T}\left(-2logL(\widehat{\boldsymbol{\theta}}) + 2dim(\widehat{\boldsymbol{\theta}})\right)$$

$$\mathrm{SC} = \frac{1}{T}\left(-2logL(\widehat{\boldsymbol{\theta}}) + dim(\widehat{\boldsymbol{\theta}})\ln T\right)$$

$$\mathrm{HQ} = \frac{1}{T}\left(-2logL(\widehat{\boldsymbol{\theta}}) + 2dim(\widehat{\boldsymbol{\theta}})\ln(\ln T)\right),$$

where $-logL(\cdot)$ is the minimum of the logLikelihood function of the model, $\widehat{\boldsymbol{\theta}}$ is the vector containing the estimates of the parameters, and T is the sample size. Each criterion selects the model that minimizes the corresponding statistic.

We rely on these measures because they trade-off in-sample fit with out-of-sample forecast accuracy, which is the key focus of this study (Guidolin and Ono, 2006). Furthermore, they penalize models with a large number of parameters, and therefore privilege parsimonious specifications. The AIC, SC, and HQ criteria are well established in the literature and widely used to compare models with a different number of regimes, as the estimators do not suffer of parameter nuisance (Roeder and Wassermann, 1997).

We analyze a large number of specifications which are fitted on the whole sample (1929:01 – 2012:12) and vary in the number of lags, p, and regimes, K. We consider models with up to three lags ($p \leq 3$), three regimes ($K \leq 3$),[48] and with a saturation ratio, calculated on the full sample, higher than 20.[49] For each specification, Table 3 collects the number of parameters, the saturation ratio, the log-likelihood function, and the statistics of three information criteria, in each column. We divide the forecast models between Panel A, which includes models in which predictors are exogenous variables (CER, VARX, TVARX, MSI, and MSIH), and Panel B, whose models have endogenous predictors (VAR, TVAR, and MSIVARH). We look at the information criteria in order to choose the best specification for each model. The AIC, the SC, and the HQ criterion often point at different specifications, for instance in the case of the VARX (Panel A) and the VAR (Panel B). In particular, the AIC prefers a greater number of lags, while the SC and the HQ criteria are more parsimonious, favouring models with a smaller number of parameters. This is due to the fact that, differently from the SC and the HQ criteria, the AIC is not asymptotically consistent and is biased towards more parametrized models (Enders, 2005).[50] For this reason, we rely on the Hannan-Quinn criterion and select the VARX(1,1) and the VAR(2). The selection of the TVARX(2,1,1) and the TVAR(2,2) follows the same logic. With regard to Markov switching models, instead, the decision is straightforward, as the three criteria always pick at the same model. we, therefore, use the MSI(3), the MSIH(3), and the MSIAH(3)-VAR(1) models in the analysis.

---

[48] The analysis of forecast accuracy of MSI and MSIH models with up to 6 regimes does not lead to any different outcome. In particular, their prediction errors rank very similarly to the corresponding models with 3 regimes.

[49] The *saturation ratio* of a model is defined as the ratio between the number of observations available and the number of parameters to be estimated. The lower the saturation ratio, the higher the standard errors of the estimates of the parameters. Values of the saturation ratio between 15 and 20 are sometimes accepted in the literature. We set 20 as a threshold because, in running the out-of-sample forecasts, we estimate models on expanding subsamples, in which a smaller number of observation. The earliest window of data in the analysis (1929:01 – 1979:12), for instance, includes 612 observations. A saturation ratio of 20 on the full sample is reduced to 12.14 for this subsample.

[50] Given a sample size *n*, an estimator $\hat{p}$ of the true specification of the model is *consistent* if: $\lim_{n\to\infty} Pr(\hat{p}_{sel} = p_{true}) = 1$, where $\hat{p}_{sel}$ is the specification chosen by the estimator and p_{true} is the true, but unknown specification.

Interestingly, the optimal number of regimes in threshold and Markov switching models seems to be different. Information criteria deem threshold models with more than two regimes to be suboptimal, but support the choice of richly parametrized Markov switching models. With this regard, they are especially useful to balance the trade-off implicit in the selection of the appropriate number of regimes. In particular, models with many regimes are more likely to fit complex dynamics and non-normal features of the time series. However, more states demand more parameters. Their estimates, in turn, have higher standard errors, which can adversely affect the prediction accuracy of the model, especially if the saturation ratio is low. The size of the sample, which includes 1008 monthly observations, is unfortunately too small to extend the study to MS specifications with more than three regimes.

4.2.2 Forecast Performance

We study out-of-sample, point forecasts of the returns of the five portfolios as follows. We recursively estimate the eight multivariate models on an expanding window of monthly observations, which starts from 1929:01 – 1979:12 and proceeds up to 2012:12. The last sample is, therefore, 1929:01 – 2012:11. Then, we use the models to make forecasts in the resulting window 1980:01 – 2012:12. In particular, each model, in every month t of the forecasting window, leads to three predictions, conditionally on the information available at time $t-1$, $t-3$, and $t-12$. Eventually, for the 5 factor portfolios, we obtain 3 estimates of returns for 396 months, coming from 8 different models, for a total of 47’520 predictions. The goal of the exercise is to find the most accurate model in forecasting factor returns. With this purpose, we first test whether forecast models are unbiased through a simple Wald test. Furthermore, we evaluate six measures of predictive accuracy in order to compare the performance of the models with respect to each factor portfolio and time horizon considered. Finally, we rely on the Mincer and Zarnowitz (1969) regression to support and extend the previous results.

We define the forecast error from model $\mathcal{M}_i$, at time *t*, on a horizon *h*, and on factor *j*, as:[51]

$$\hat{e}_{t,t+h}^{j,\mathcal{M}_i} = r_{t+h}^{j} - \hat{r}_{t,t+h}^{j,\mathcal{M}_i},$$

where r_{t+h}^{j} is the actual return of factor *j* at time $t+h$, and $\hat{r}_{t,t+h}^{j,\mathcal{M}_i}$ is the forecast from model $\mathcal{M}_i$ for time $t+h$, conditional on the information available at time t. Figure 3-10 plot the prediction errors (top graph) and their distributions (bottom graph) for each model, factor, and horizon. The charts highlight that the distribution of the forecast errors is non-normal for all the models, factors, and horizons. In particular, asymmetries and fat tails are due to some large outliers, which relate to periods in which the models fail to capture the dynamics among risk premia. This is especially evident for the Low Volatility premium, whose prediction errors have a u-shaped form.

We test whether forecasts are unbiased. More specifically, forecast model $\mathcal{M}_i$ is unbiased on a horizon *h* if:[52]

$$\hat{e}_{t,t+h}^{j,\mathcal{M}_i} = r_{t+h}^{j} - \hat{r}_{t,t+h}^{j,\mathcal{M}_i} = 0\,, \qquad \forall t\,.$$

We use the Forecast Error Bias (FEB), which is defined as the mean of the estimated forecast errors, as an estimator for the expected prediction error. Therefore, for any $\mathcal{M}_i$, *h*, and *j*, we test whether the FEB is statistically different from 0. If this is true, there is empirical evidence that model $\mathcal{M}_i$ is biased in forecasting returns of factor *j*, on a horizon *h*. Table 4 collects the outcome of the test. Notably, for the majority of factors, the TVAR(2,2) model is biased in forecasting returns on the 3-month and the 12-month horizon both. In addition, the MSIVARH(3,1) is the only model to be unbiased in the prediction of the returns of the Low Volatility factor at all horizons, with the only exception being $h = 12$ months, where the null hypothesis of zero expected forecast error is rejected at the 10% significance level. The

[51] Alternatively, $\mathrm{E}_t\left(e_{t+h}^{j,\mathcal{M}_i}\right) = r_{t+h}^{j} - \mathrm{E}_t\left(r_{t+h}^{j,\mathcal{M}_i}\right)$.

[52] The *bias* of an estimator $\hat{\theta}$ is defined as $B(\hat{\theta}) = \mathrm{E}(\hat{\theta}) - \theta$. An estimator $\hat{\theta}$ is *unbiased* if and only if its bias $B(\hat{\theta})$ is 0, i.e. $\mathrm{E}(\hat{\theta}) = \theta$.

remaining models produce unbiased forecasts for the Market, Size, Value, and Momentum factors, with some deviations on the 12-month horizon.

If a forecast model is unbiased on a horizon *h*, its predictions at time *t* for $t+h$ are, on average, centred on the actual values of the variable at $t+h$. Unbiasedness, however, gives no indication on how accurate the predictions are. Therefore, in addition to the FEB, we compute five measures of prediction accuracy, which are standard in the forecasting literature (Guidolin et al, 2009). These measures are the Root Mean Squared of Forecast Error (RMSFE), the Forecast Error Variance (FEV), the Mean Absolute Forecast Error (MAFE), the Mean Percent Forecast Error (MPFE), and the Success Ratio (SR). In particular, for each factor *j*, horizon *h*, and model $\mathcal{M}_i$:

$$\mathrm{RMSFE}^{j,h,\mathcal{M}_i} = \sqrt{\frac{1}{T}\sum_{t=1}^{T}\left(\hat{e}_{t,t+h}^{j,\mathcal{M}_i}\right)^2}$$

$$\mathrm{FEV}^{j,h,\mathcal{M}_i} = \frac{1}{T}\sum_{t=1}^{T}\left(\hat{e}_{t,t+h}^{j,\mathcal{M}_i} - \bar{e}^{j,\mathcal{M}_i}\right)^2$$

$$\mathrm{MAFE}^{j,h,\mathcal{M}_i} = \frac{1}{T}\sum_{t=1}^{T}\left|\hat{e}_{t,t+h}^{j,\mathcal{M}_i}\right|$$

$$\mathrm{MPFE}^{j,h,\mathcal{M}_i} = \frac{100\%}{T}\sum_{t=1}^{T}\frac{\hat{e}_{t,t+h}^{j,\mathcal{M}_i}}{r_{t+h}^{j}}$$

$$\mathrm{SR}^{j,h,\mathcal{M}_i} = \frac{1}{T}\left(\#r_{t+h}^{j} * \hat{r}_{t,t+h}^{j,\mathcal{M}_i} \geq 0\right),$$

where *T* is the size of the sample, and $\#r_{t+h}^{j} * \hat{r}_{t,t+h}^{j,\mathcal{M}_i} \geq 0$ indicates the number of times in which the forecast return and the actual return have the same sign. Table 5-9 contains the six measures, calculated for each model $\mathcal{M}_i$ and forecast horizon *h*, given factor *j*. Table 10 and Table 11 collect the top three models in each measure. Each panel focuses on a different forecast horizon *h*, in Table 10, and a different factor portfolio *j*, in Table 11. Finally, Table 12 focuses on the number of times (and percentage, in italic) in which the models rank among the top three for each measure considered, conditionally on the forecasting horizon *h* (Panel

A) and the factor portfolio *j* (Panel B). The table especially highlights the CER, MSI(3), and MSIH(3), as they appear among the top three models 22%, 23%, and 21%, respectively. This result is remarkable because it shows that, although with some exceptions, linear and non-linear models without autoregressive terms have a better performance than models with a lag structure.[53] In particular, the MSI(3) and the MSIH(3) models produce small RMSFE, MAFE, and SR on all the horizons considered. This is mainly due to their ability to minimize the Forecast Error Variance (FEV), and is in line with the results in Guidolin et al. (2009), who prove the superior predictive accuracy of Markov Switching models. However, their bias, measured by the FEB, is not among the lowest, except when forecasting excess market returns. The good performance of the CER model is more surprising, and suggests that factor returns follow a random walk with drift, thus supporting the idea that there exists no predictability in factor returns. This is partially consistent with general findings on stock return predictability with monthly data. Establishing a ranking among the CER, the MSI(3), and the MSIH(3) models is instead hard, as they provide very similar measures. In particular, the MSI(3) and the MSIH(3) models look marginally better when forecasting the Size, Value, and Momentum premia. Finally, threshold models not only are outperformed by Markov switching models in many instances, but also seem unable to improve the forecast accuracy of their corresponding linear models, as they often rank worse in almost all measures.

The performance of the models seems to be independent of the forecast horizon. Although the accuracy of the models generally worsens as the horizon lengthens, with the RMSFE and the MAFE increasing and the SR decreasing, the rankings in Table 10-11 are, in fact, not affected by changes in *h*. Furthermore, the models rank differently depending on the factor being considered. For instance, the dominance of the CER, the MSI(3), and the MSIH(3) model is even more evident in the case of excess market returns, for which they rank in the top class 31%, 30%, and 31% of the times, respectively, while their outperformance is less

[53] This is similar to Guidolin and Timmermann (2008). The Authors recognize that the best specification for the forecast of the Size and Value premia does not include autoregressive terms.

tangible for the Value and Momentum factors. The Low Volatility premium, instead, appears to be the toughest to predict: RMSFE and MAFE are 2 to 4 times higher on all models, while the success ratio is as low as 63%. The MSIVARH(3,1) model, which is the most flexible among the eight considered, is particularly successful in capturing its movements (24% of the times, against an average of 6% on the other factors). In particular, it is always among the top picks for minimizing the RMSFE, the MAFE and the FEV on the Low Volatility factor, while still producing forecasts having the smallest possible bias, as we expected from the previous findings.

We also test the point forecasts using the Mincer and Zarnowitz (1969, MZ) regression, which is a standard tool in the literature. We estimate the following:

$$r_{t+h}^{j} = \beta_{h,0}^{j} + \beta_{h,1}^{j} \hat{r}_{t,t+h}^{j,\mathcal{M}_i} + \varepsilon_{t,t+h}^{j,\mathcal{M}_i} \, .$$

The model $\mathcal{M}_i$ is unbiased in forecasting returns of factor *j* on a horizon *h*, if $\beta_{h,0}^{j} = 0$ and $\beta_{h,1}^{j} = 1$.[54] Therefore, we test the two assumptions individually, through Wald t-tests, and jointly, through F-tests. Furthermore, the higher the R^2 of the regression, the more accurate the forecast model. We run the regression for each combination of model $\mathcal{M}_i$, factor *j*, and horizon *h*, and sort the results by forecasting horizon, in Table 13, and forecast model, in Table 14. The hypothesis that $\beta_{h,0}^{j} = 0$ and $\beta_{h,1}^{j} = 1$, simultaneously, is strongly rejected for the majority of the forecast models, as it can be seen from the p-value of the F-statistics (Column 3). Therefore, there is statistical evidence in order to argue that forecasts might be biased for these models. The R^2s of the MZ regressions (Column 4), support this result for the TVAR(2,2) model, on all factors, and for the Low Volatility factor, independently of the forecast model. The CER, the MSI(3), and the MSIH(3) model, although statistically unbiased for all factors

[54] An identical approach involves the following regression: $\hat{e}_{t,t+h}^{j,\mathcal{M}_i} = \gamma_{0,h}^{j} + \gamma_{1,h}^{j} \hat{r}_{t,t+h}^{j,\mathcal{M}_i} + \xi_{t,t+h}^{j,\mathcal{M}_i}$. The forecast model is unbiased if $\gamma_{0,h}^{j} = 0$ and $\gamma_{1,h}^{j} = 0$.

except the Low Volatility one, have incredibly low R^2s (often smaller than 1%), which undermine their reliability in absolute terms.

In fact, the analysis has been conducted on a purely relative basis so far. We have proved the statistical unbiasedness of some forecast models, and found out which ones are more accurate in a recursive, out-of-sample forecast of the returns on the factor portfolios. However, the conclusion that one model is likely to perform better than another does not imply that such forecast are good enough to be used in empirical applications. Looking at Table 5-9 more closely, all the forecast models are quite inefficient. For instance, the MAFE, which gives an intuitive indication of the average forecast error, ranges between 2,0% and 3,5% for the Market, Size, Value, and Momentum premia, while their monthly average return is only between 0,1% and 0,6%. The R^2s of the MZ regressions stress the same concept, and casts further doubt on the actual predictability in factor returns.

4.2.3 Testing for Differential Forecasting Accuracy

In the previous Section, we tested whether the eight forecast models are unbiased, and we ranked them according to their forecast accuracy. However, a simple comparison among them cannot be comprehensive: the fact that model $\mathcal{M}_i$ ranks better than model $\mathcal{M}_g$ does not imply that the former is statistically more accurate than the latter. In other words, the measures and tools used so far do not provide evidence that the difference between $\mathcal{M}_i$ and $\mathcal{M}_g$ is statistically different from 0. In this Section, we apply Diebold and Mariano's (1995, DM) test of equal predictive accuracy, which exactly serves this purpose.

The DM statistics tests whether the mean loss function values obtained from two alternative forecast models $\mathcal{M}_i$ and $\mathcal{M}_g$ are statistically different. Given a loss function $L\left(\varepsilon_{t,t+h}^{j,\mathcal{M}_i}\right)$, we define the difference of the loss function of the two competing models in the following way:

$$diff_{t,j,h}^{\mathcal{M}_i,\mathcal{M}_g} \equiv L\left(\varepsilon_{t,t+h}^{j,\mathcal{M}_i}\right) - L\left(\varepsilon_{t,t+h}^{j,\mathcal{M}_g}\right). \tag{4.1}$$

We test the null hypothesis that the mean of these differences is 0, i.e. $\mathrm{E}\left(diff_{t,j,h}^{\mathcal{M}_i,\mathcal{M}_g}\right) = 0$, through the following DM statistics:

$$\mathrm{DM}_{t,j,h}^{\mathcal{M}_i,\mathcal{M}_g} = \frac{\frac{1}{T-h}\sum_{t=1}^{T} diff_{t,j,h}^{\mathcal{M}_i,\mathcal{M}_g}}{\hat{\sigma}\left(diff_{t,j,h}^{\mathcal{M}_i,\mathcal{M}_g}\right)},$$

where $\hat{\sigma}\left(diff_{t,j,h}^{\mathcal{M}_i,\mathcal{M}_g}\right)$ is the estimate of the standard error of the loss differential between $\mathcal{M}_i$ and $\mathcal{M}_g$, and is usually computed by the Newey-West estimator.[55] The DM statistic, under the assumption that the loss differential is covariance stationary, has an asymptotically normal distribution (Diebold and Mariano, 1995): $\mathrm{DM}_{t,j,h}^{\mathcal{M}_i,\mathcal{M}_g} \sim N(0,1)$. If the null hypothesis is rejected, we can conclude that model $\mathcal{M}_i$ leads to an expected loss higher than the loss from model $\mathcal{M}_g$. Hence, the latter model should be preferred.

The loss function indicates the *cost* or *disutility* associated with the differences between forecast return and actual return, and is, therefore, a function of the forecast error. The appropriate loss function depends on the utility function and the decision environment of the user (Diebold, 1993). This study is aimed at general applications, therefore we consider two standard loss functions: square and absolute.[56] In the former case, the loss associated with each prediction is the squared error, $L\left(\varepsilon_{t,t+h}^{j,\mathcal{M}_i}\right) = \left(\varepsilon_{t,t+h}^{j,\mathcal{M}_i}\right)^2$, while in the latter case it is the absolute error, $L\left(\varepsilon_{t,t+h}^{j,\mathcal{M}_i}\right) = \left|\varepsilon_{t,t+h}^{j,\mathcal{M}_i}\right|$. The choice of these two functions is supported by Diebold and Lopez (1996), who highlight that optimal forecast errors enjoy desirable statistical properties under a squared loss function.[57] In addition, the absolute loss function, although

[55] Forecast errors in a *h*-period forecast follow $MA(h-1)$ processes. Diebold and Mariano (1995) use the Newey-West estimator because it corrects for the resulting autocorrelation.

[56] Hong and Lee (2003) study two additional loss functions: trading return and correct direction. They argue that squared and absolute loss may not be appropriate for financial returns, as investors eventually aim at maximizing profits, rather than forecast accuracy.

[57] Under mean squared loss, if the dependent variable and its predictors are joint covariance stationary, optimal forecast errors have 0 mean (i.e. forecasts are unbiased), have variance which is

not differentiable in $\varepsilon_{t,t+h}^{j,\mathcal{M}_i} = 0$, is more robust to outliers. The square and the absolute loss functions are both symmetric, thus assigning the same weight, or cost, to positive and negative errors. In empirical applications, the actual economic loss of the agent should be evaluated in order to define the appropriate function. Its utility function may, for instance, require an asymmetric loss function, thus depending on the entire distribution of forecast returns, rather than simply on their point estimate. It is also possible to specify a loss function which is different for the returns of each factor portfolio. This may be the case of a mutual fund consisting of a pool of investors with unequal amount of wealth allotted in the strategies.

The DM test only require loss differential be covariance stationary. Using Augmented Dickey-Fuller tests, we verify whether this assumption is empirically correct for all the combinations of prediction errors, under both square and absolute loss function. The null hypothesis of unit root is rejected at the 1% significance level in all instances. Hence, we proceed in calculating the DM statistics. Table 15-18 display the result of the DM tests for each factor portfolio. The three Panels (A, B, and C) refer to the forecasting horizons considered ($h = 1, 3, 12$), and are structured in the following way. The cells *below* the main diagonal contain the *p*-values of the DM test under the square loss function. The null hypothesis is rejected if the MSFE of two forecast models are statistically different. Similarly, the cells *above* the main diagonal contain the *p*-values of the DM test under the absolute loss function. In this case, the null hypothesis is rejected if the MAFE of two forecast models are statistically different. Values lower than 5% are marked in bold. In either case, if the null hypothesis is not rejected, there is not enough statistical evidence to argue that the performance of a model is significantly better than the other. Similarly to what we have previously noticed, the performance of the forecast models is strongly dependent on the factor under consideration, although very similar on the three horizons. For this reason, we mention values in Panel A of the five Tables, unless stated otherwise.

---

non-decreasing function of the time horizon h, and are at most serially correlated of order $h - 1$. Finally, single-period forecast errors are serially uncorrelated (Diebold and Lopez, 1996).

Under square loss function, the CER, MSI(3), and MSIH(3) models provide significantly better performance than the other models only for the excess market returns, where the DM statistic is significant at the 1% and 5% levels. Considering the absolute loss function, this is confirmed and extended to the Momentum premium. However, the DM tests do not validate the impression that the MSI(3) and the MSIH(3) have better performance than the CER model for the Size, Value, and Momentum premia, as the three models rarely show significant differences in predictive accuracy. Furthermore, the underperformance of the TVARX(2,1,1) and, especially, of the TVAR(2,2), against Markov switching models is clear from the low *p*-values. Results are however mixed when comparing the two threshold models with their corresponding linear specification. The dominance of the MSIVARH(3,1) in forecasting the Low Volatility premium is remarked against all models but the CER and the MSI(3), while it outperforms the MSIH(3) only for $h = 1$. Finally, clear differences between the two loss functions can be noticed only with regard to the Momentum factor, which, excluding the Low Volatility factor, has the widest range of returns.

4.3 Final Remarks

4.3.1 Key Results

We ran a horse-race among eight multivariate forecast models for risk premia, and tested the properties of their *pseudo* out-of-sample predictions in a qualitative and quantitative fashion. We found evidence that a pure random walk model is at least as accurate as more complex statistical models in forecasting returns of the Size, Value, and Momentum factors. Furthermore, the introduction of regime shifts, heteroskedasticity, and autoregressive terms provides a significant improvement in the forecast of the Low Volatility premium only. The key results of the analysis are in the following.

1. *Among the eight forecast models considered, no one is statistically the best for all the factor portfolios.* The purpose of this study is to find the multivariate model which is

the most accurate in forecasting returns of all the factor portfolios. With this regard, there is no clear winner among the eight models.

2. *The CER, MSI(3), and MSIH(3) models generally show significant outperformance against forecast models with autoregressive terms.* This is especially true with regard to their predictive accuracy, as they successfully minimize the MSFE and the MAFE, but less distinct for their forecast bias. The better performance of models without a lag structure casts doubts on the effectiveness of the predictors considered.
3. *There is no statistical evidence to argue that the MSI(3) and the MSIH(3) models are significantly better than the CER model.* The introduction of regimes, in the MSI model, and heteroskedasticity, in MSIH model, allows for a better and more accurate fit of in-sample data. Pseudo out-of-sample predictions also have lower RMSFE and MAFE. Nevertheless, there is not enough empirical evidence to prove that the two Markov switching model lead to significantly more accurate predictions (see the DM tests).
4. *Among the regime switching class, the specifications of threshold models considered fail in improving the forecast performance or the corresponding linear models.* This should not be interpreted as a general shortcoming of threshold models, but rather as a failure of the specifications adopted. We have indeed made restrictive, *a priori* assumptions on the threshold variable and the delay parameter. We consider only the excess market return as threshold variable, and it would be useful to test whether other variables lead to more accurate predictions, similarly to Guidolin et al. (2009).
5. *The predictive accuracy of the forecast models is strongly influenced by the risk premium under analysis, and it appears to be independent on the forecast horizon h.* In particular, the ranking on the modes strictly depends on the variables considered, while there exist negligible differences on the three horizons.
6. *The Low Volatility premium has the most complex dynamics to capture.* Forecasts on the Low Volatility premium always have errors with the greatest bias (FEB) and the highest volatility (FEV). The MSIVARH(3,1), thanks to its flexibility, is particularly good at minimizing them.

7. *Any inference on the optimal model is conditional on the loss function itself.* We have considered the square and the absolute loss function in this study. Results under the two measures have negligible differences. However, they may dramatically change if different loss functions were to be preferred. Hence, any empirical application should accurately tailor the loss function on the preferences (i.e. utility function) of the user before making inference on the model.
8. *In absolute terms, all the forecast models seem quite inefficient at forecasting factor returns.* This is evident in the high values of the RMSFE, the FEV, and the MAFE, as well as in the low R^2 of the MZ regressions.

4.3.2 Limitations and Extensions

We showed that, under a multivariate framework, predictions are unbiased, but inaccurate (point 8). We also reached the conclusion that, under the loss functions considered, there is no optimal forecast model for all factor portfolios (point 1). This may lead to believe that a shortcoming lies in the choice of adopting a multivariate, rather than univariate, approach. Previous studies have indeed focused on forecasting returns of each factor individually, rather than jointly. Moving from a multivariate to a univariate framework especially increases the accuracy of regime switching models, as regimes would model the dynamics of each factor specifically. However, multivariate modelling is needed in any application that requires estimates of the correlations or, more generally, of the dynamics among factor returns. The multivariate nature of this study is, therefore, useful for a number of applications. For instance, any portfolio optimization involving risk premia has to deal with the estimation of their expected return and correlations.[58] With this regard, Ang et al. (2010) show how performance evaluation and active asset allocation change when multiple factor portfolios are considered. Investors and practitioners should also be interested in managing the risks

[58] The use of Markov switching forecast models does not pose an issue. With this regard, Buckley, Sanders, and Seco (2008) consider portfolio optimization problems where asset returns are distributed as a mixture of multivariate normals.

embedded in their portfolios. Standard risk management models, such as the parametric Value at Risk (VaR), are based on expected returns, volatilities and correlations. The reliability of these models, consequently, depends on the accuracy of such estimates.

We outline several extensions of the analysis. First, additional factors and predictors can be considered. Although the dividend yield, the term spread and the default spread do not contribute to the predictive accuracy of the forecast models considered, there may be others with significant explanatory power.[59] It may also be useful to evaluate the forecast accuracy of other models, such as Smooth Transition Vector Autoregressive models, or introduce conditional heteroskedasticity in the linear specifications. Second, several aspects of the methodology can be developed. We have run a forecasting experiment on expanding windows of data, which may favour regime switching models against single state ones. Comparing out-of-sample forecasts on rolling windows of data may change the outcome of the analysis. In addition, we have relied purely on the DM test when evaluating differences in the loss differential between models. Other tests of forecast accuracy, such as the van Dijk and Franses's (2003) weighted test of equal prediction accuracy and the Giacomini and White's (2006) test, would give further robustness to the analysis. Third, results are strictly dependent on the measures of bias and accuracy used in the comparison. In turn, they are based on a loss function, which we considered either square or absolute (point 7). The actual cost incurred by the user could well be different. Hence, comparing the results under additional cost functions can provide more information on the appropriateness of forecast models under diverse decision making environments.

[59] For instance, return dispersion (Stivers and Sun, 2010) or other macroeconomic variables (Sarwar et al., 2014).

5. CONCLUSIONS

This paper showed that sophisticated statistical models fail in improving the predictive accuracy of a simple random walk model when forecasting joint returns of the Size, Value, Momentum, and Low Volatility factors. Although we observed good predictive performance of Markov switching models, the improvements against the naïve random walk are not statistically significant. This result implies that practitioners should not look beyond the sample mean when estimating expected returns of factor portfolios: there is no guarantee that econometric models produce more accurate predictions. The Low Volatility anomaly constitutes the only exception, as it seems to require a Markov switching model with three regimes, heteroskedasticity, autoregressive terms, and lagged prediction variables (dividend yield, term spread, and default spread) for unbiased and accurate forecasts.

These findings should not be taken for granted, because they are limited to the range of models and predictors considered. Furthermore, they are conditional on the specification of a loss function, which we considered either square or absolute. Hence, they may not be reliable in empirical applications if the user is subject to an economic loss which is, for instance, asymmetric or different across factor portfolios. Similarly, evaluating interval forecasts rather than point forecasts may prove the dominance of one of the models that this study considered suboptimal. Finally, the multivariate framework of the analysis lays the foundations for forecasting experiments that involves not only moments of higher order, but also correlations among factors. The flexibility of regime switching models is particularly useful in these comparisons, and may shed light on the limitations of the naïve random walk model, which performed surprisingly well in this study.

REFERENCES

Amihud, Y., & Mendelson, H. (1986). Asset pricing and the bid-ask spread. *Journal of financial Economics, 17*(2), 223-249.

Ang, A., & Bekaert, G. (2007). Stock return predictability: Is it there?. *Review of Financial studies, 20*(3), 651-707.

Ang, A., & Timmermann, A. (2011). Regime changes and financial markets (No. w17182). National Bureau of Economic Research.

Ang, A., Goetzmann, W. N., & Schaefer, S. M. (2010). The efficient market theory and evidence: implications for active investment management. *Foundations and Trends in Finance, 5*(3).

Ang, A., Hodrick, R. J., Xing, Y., & Zhang, X. (2006). The cross-section of volatility and expected returns. *The Journal of Finance, 61*(1), 259-299.

Ang, A., Hodrick, R. J., Xing, Y., & Zhang, X. (2009). High idiosyncratic volatility and low returns: International and further US evidence. *Journal of Financial Economics, 91*(1), 1-23.

Angelidis, T., & Tessaromatis, N. (2014, July). Global portfolio management under state dependent multiple risk premia. In Proceedings of Economics and Finance Conferences (No. 0400966). International Institute of Social and Economic Sciences.

Asness, C. S., Friedman, J. A., Krail, R. J., & Liew, J. M. (2000). Style timing: Value versus growth. *The Journal of Portfolio Management, 26*(3), 50-60.

Asness, C., & Frazzini, A. (2013). *The devil in HML's details. Journal of Portfolio Management, 39*(4), 49-68.

Baker, M., Bradley, B., & Wurgler, J. (2011). Benchmarks as limits to arbitrage: Understanding the low-volatility anomaly. *Financial Analysts Journal, 67*(1), 40-54.

Barroso, P., & Santa-Clara, P. (2015). Momentum has its moments. *Journal of Financial Economics, 116*(1), 111-120.

Black, A. J., & McMillan, D. G. (2004). Non-linear Predictability of Value and Growth Stocks and Economic Activity. *Journal of Business Finance & Accounting, 31*(3-4), 439-474.

Blitz, D., & Van Vliet, P. (2007). The volatility effect: Lower risk without lower return. *Journal of Portfolio Management*, 102-113.

Blume, M. E. (1980). Stock returns and dividend yields: Some more evidence. *The Review of Economics and Statistics*, 567-577.

Brooks, C. (2014). Introductory econometrics for finance. Cambridge university press.

Buckley, I., Saunders, D., & Seco, L. (2008). Portfolio optimization when asset returns have the Gaussian mixture distribution. *European Journal of Operational Research, 185*(3), 1434-1461.

Campbell, J. Y., & Cochrane, J. H. (1999). Explaining the poor performance of consumption-based asset pricing models (No. w7237). National bureau of economic research.

Campbell, J. Y., & Thompson, S. B. (2008). Predicting excess stock returns out of sample: Can anything beat the historical average?. *Review of Financial Studies, 21*(4), 1509-1531.

Campbell, J. Y., & Yogo, M. (2005). Implementing the econometric methods in "Efficient tests of stock return predictability". Unpublished working paper. University of Pennsylvania.

Campbell, J. Y., Chan, Y. L., & Viceira, L. M. (2003). A multivariate model of strategic asset allocation. *Journal of financial economics, 67*(1), 41-80.

Carhart, M. M. (1997). On persistence in mutual fund performance. *The Journal of finance, 52*(1), 57-82.

Chen, L., Petkova, R., & Zhang, L. (2008). The expected value premium. *Journal of Financial Economics, 87*(2), 269-280.

Chordia, T., & Shivakumar, L. (2002). Momentum, business cycle, and time-varying expected returns. *Journal of Finance*, 985-1019.

Christoffersen, P. F., & Diebold, F. X. (1996). Further results on forecasting and model selection under asymmetric loss. *Journal of applied econometrics, 11*(5), 561-571.

Cochrane, J. H. (2008). The dog that did not bark: A defense of return predictability. *Review of Financial Studies, 21*(4), 1533-1575.

Cohen, R. B., Polk, C., & Vuolteenaho, T. (2003). The value spread. *The Journal of Finance, 58*(2), 609-642.

Cooper, M. J., Gutierrez, R. C., & Hameed, A. (2004). Market states and momentum. *The Journal of Finance, 59*(3), 1345-1365.

Daniel, K. D., & Moskowitz, T. J. (2013). Momentum crashes. *Swiss Finance Institute Research Paper*, (13-61), 14-6.

Diebold, F. X. (2015). Comparing predictive accuracy, twenty years later: A personal perspective on the use and abuse of Diebold–Mariano tests. *Journal of Business & Economic Statistics, 33*(1), 1-1.

Diebold, F. X., & Lopez, J. A. (1996). Forecast evaluation and combination.

Diebold, F. X., & Mariano, R. S. (2012). Comparing predictive accuracy. *Journal of Business & economic statistics*.

Dijk, D. V., Teräsvirta, T., & Franses, P. H. (2002). Smooth transition autoregressive models—a survey of recent developments. *Econometric Reviews, 21*(1), 1-47.

Fama, E. F. (1990). Stock returns, expected returns, and real activity. *The Journal of Finance, 45*(4), 1089-1108.

Fama, E. F., & French, K. R. (1988). Dividend yields and expected stock returns. *Journal of financial economics, 22*(1), 3-25.

Fama, E. F., & French, K. R. (1993). Common risk factors in the returns on stocks and bonds. *Journal of financial economics, 33*(1), 3-56.

Fama, E. F., & French, K. R. (1998). Value versus growth: The international evidence. *Journal of finance*, 1975-1999.

Fama, E. F., & French, K. R. (2002). The equity premium. *The Journal of Finance, 57*(2), 637-659.

Fama, E. F., & French, K. R. (2010). Luck versus skill in the cross-section of mutual fund returns. *The Journal of Finance, 65*(5), 1915-1947.

Filippou, I., & Garcia-Ares, P. A. (2015). Value, Momentum and Market Timing. Available at SSRN 2565913.

Guidolin, M., & Nicodano, G. (2009). Small caps in international equity portfolios: the effects of variance risk. *Annals of Finance, 5*(1), 15-48.

Guidolin, M., & Ono, S. (2006). Are the dynamic linkages between the macroeconomy and asset prices time-varying?. *Journal of Economics and Business, 58*(5), 480-518.

Guidolin, M., & Timmermann, A. (2008). International asset allocation under regime switching, skew, and kurtosis preferences. *Review of Financial Studies, 21*(2), 889-935.

Guidolin, M., & Timmermann, A. (2008). Size and value anomalies under regime shifts. *Journal of Financial Econometrics, 6*(1), 1-48.

Gulen, H., Xing, Y., & Zhang, L. (2011). Value versus Growth: Time-Varying Expected Stock Returns. *Financial management, 40*(2), 381-407.

Hamilton, J. D. (1990). Analysis of time series subject to changes in regime. *Journal of econometrics, 45*(1), 39-70.

Hamilton, J. D. (1994). Time series analysis (Vol. 2). Princeton: Princeton university press.

Haugen, R. A., & Baker, N. L. (1991). The efficient market inefficiency of capitalization-weighted stock portfolios. *The Journal of Portfolio Management, 17*(3), 35-40.

Heidari, M. (2015). Momentum Crash Management. Available at SSRN 2578296.

Herzog, F., Dondi, G., Geering, H. P., & Schumann, L. (2004, March). Continuous-time multivariate strategic asset allocation. In *Proceedings of the 11th Annual Meeting of the German Finance Association*, Session 2B (pp. 1-34).

Hong, Y., & Lee, T. H. (2003). Inference on predictability of foreign exchange rates via generalized spectrum and nonlinear time series models. *Review of Economics and Statistics, 85*(4), 1048-1062.

Hubrich, K., & Teräsvirta, T. (2013). Thresholds and Smooth Transitions in Vector Autoregressive Models☆☆ The views expressed in this article are those of the authors and should not be interpreted as reflecting the views of the European Central Bank. In *VAR Models in Macroeconomics–New Developments and Applications: Essays in Honor of Christopher A. Sims* (pp. 273-326). Emerald Group Publishing Limited.

Inoue, A., & Kilian, L. (2005). In-sample or out-of-sample tests of predictability: Which one should we use?. *Econometric Reviews, 23*(4), 371-402.

Jacobs, B. I., & Levy, K. N. (1988). Calendar anomalies: Abnormal returns at calendar turning points. *Financial Analysts Journal, 44*(6), 28-39.

Jang, J., Kang, J., & Lee, C. (2013). State-Dependent Variations in Expected Illiquidity Premium. *한국재무학회 학술대회*, 1710-1762.

Jegadeesh, N., & Titman, S. (1993). Returns to buying winners and selling losers: Implications for stock market efficiency. *The Journal of finance, 48*(1), 65-91.

Johnson, T. C. (2002). Rational momentum effects. *The Journal of Finance, 57*(2), 585-608.

Kim, D. (2012). Value Premium Across Countries. *The Journal of Portfolio Management, 38*(4), 75-86.

Kim, D., Roh, T. Y., Min, B. K., & Byun, S. J. (2014). Time-varying expected momentum profits. *Journal of Banking & Finance, 49*, 191-215.

Larsen, L. S., & Munk, C. (2012). The costs of suboptimal dynamic asset allocation: General results and applications to interest rate risk, stock volatility risk, and growth/value tilts. *Journal of Economic Dynamics and Control, 36*(2), 266-293.

Lee, D. W. (2013). The role of trading volume in the “volatility puzzle”.

Li, W. K., & Lam, K.. (1995). Modelling Asymmetry in Stock Returns by a Threshold Autoregressive Conditional Heteroscedastic Model. *Journal of the Royal Statistical Society. Series D (the Statistician), 44*(3), 333–341. http://doi.org/10.2307/2348704

Li, Y., Ng, D. T., & Swaminathan, B. (2014). Predicting time-varying value premium using the implied cost of capital: Implications for countercyclical risk, mispricing and style investing (No. 2168360). SSRN Working Paper.

Lütkepohl, H. (2005). New introduction to multiple time series analysis. Springer Science & Business Media.

Lynch, A. (2000). Portfolio choice and equity characteristics: Characterizing the hedging demands induced by return predictability. *Journal of Financial Economics*, 67-130.

Marcellino, M., Stock, J. H., & Watson, M. W. (2006). A comparison of direct and iterated multistep AR methods for forecasting macroeconomic time series. *Journal of econometrics, 135*(1), 499-526.

Merville, L. J., & Xu, Y. (2002). The changing factor structure of equity returns. University of Texas.

Mincer, J. A., & Zarnowitz, V. (1969). The evaluation of economic forecasts. In Economic Forecasts and Expectations: Analysis of Forecasting Behavior and Performance (pp. 3-46). NBER.

Pagan, A. R., & Sossounov, K. A. (2003). A simple framework for analysing bull and bear markets. *Journal of Applied Econometrics, 18*(1), 23-46.

Panopoulou, E., & Plastira, S. (2014). Fama French factors and US stock return predictability. *Journal of Asset Management, 15*(2), 110-128.

Patton, A. J., & Timmermann, A. (2007). Properties of optimal forecasts under asymmetric loss and nonlinearity. *Journal of Econometrics, 140*(2), 884-918.

Perez-Quiros, G., & Timmermann, A. (2000). Firm size and cyclical variations in stock returns. *The Journal of Finance, 55*(3), 1229-1262.

Pettersson, J. E. (2014). Time Series Momentum and Volatility States. Available at SSRN 2515685.

Rapach, D. E., & Wohar, M. E. (2006). In-sample vs. out-of-sample tests of stock return predictability in the context of data mining. *Journal of Empirical Finance, 13*(2), 231-247.

Rapach, D. E., Wohar, M. E., & Rangvid, J. (2005). Macro variables and international stock return predictability. *International journal of forecasting, 21*(1), 137-166.

Ross, S. A. (1976). The arbitrage theory of capital asset pricing. *Journal of economic theory, 13*(3), 341-360.

Rytchkov, O. (2010). Expected returns on value, growth, and HML. *Journal of Empirical Finance, 17*(4), 552-565.

Sagi, J. S., & Seasholes, M. S. (2007). Firm-specific attributes and the cross-section of momentum. *Journal of Financial Economics, 84*(2), 389-434.

Sarwar, G., Mateus, C., & Todorovic, N. (2015). Macroeconomic Determinants of Cyclical Variations in Value, Size and Momentum Premiums in the UK. Size and Momentum Premiums in the UK (March 20, 2015).

Sims, C. A. (1980). Macroeconomics and reality. *Econometrica: Journal of the Econometric Society*, 1-48.

Sloan, R. (1996), "Do stock prices fully reflect information in accruals and cash flows about future earnings?" *The Accounting Review 71*: 289-315.

Stivers, C., & Sun, L. (2010). Cross-sectional return dispersion and time variation in value and momentum premiums.

Tsay, R. S. (1998). Testing and modeling multivariate threshold models. *Journal of the American Statistical Association, 93*(443), 1188-1202.

Van Vliet, P., Blitz, D., & Van der Grient, B. (2011). Is the relation between volatility and expected stock returns positive, flat or negative?.

Wang, K. Q., & Xu, J. (2015). Market volatility and momentum. *Journal of Empirical Finance, 30*, 79-91.

Watanabe, A., & Watanabe, M. (2008). Time-varying liquidity risk and the cross section of stock returns. *Review of Financial Studies, 21*(6), 2449-2486.

Yoldas, E. (2012). Threshold Asymmetries in Equity Return Distributions: Statistical Tests and Investment Implications. *Studies in Nonlinear Dynamics & Econometrics, 16*(5).

Zakamulin, V. (2014). Predictable Dynamics in the Small Stock Premium. *Economics Research International*, 2014.

Zakamulin, Valeriy. "Forecasting the size premium over different time horizons." *Journal of Banking & Finance 37.3* (2013): 1061-107

TABLES AND FIGURES

Table 1
Summary statistics for factor portfolios and predictors (full sample)

Panel A presents the main statistics for factors and predictors over the full-sample of the analysis (1929:01 - 2012:12). Return and yield data are annualized and continuously compounded (for instance, 1,00 stands for 1,00%). Panel B displays sample correlations.

Panel A: Main statistics

	Mean	Median	Max	Min	Std. Dev.	Skewness	Kurtosis	Jarque-Bera
MKT-RF	5,17	11,23	393,87	-413,19	19,03	-0,50	9,56	1852
SMB	2,20	0,84	375,14	-214,95	11,03	1,29	15,53	6869
HML	5,99	3,36	364,21	-124,57	11,90	1,81	15,40	7010
MOM	6,31	10,04	202,58	-887,28	18,61	-5,19	60,28	142338
LOWVOL	-65,79	-29,57	299,37	-4889,24	73,46	-10,82	172,19	1221854
Dividend yield	-4011,72	-3981,15	-2295,48	-5437,59	158,50	-0,35	2,84	21
Term spread	20,38	20,94	53,39	-44,62	4,43	-0,30	3,34	20
Default spread	13,64	10,81	65,84	3,83	2,43	2,32	10,65	3361

Panel B: Sample correlations

	MKT_RF	SMB	HML	MOM	LOWVOL	Dividend yield	Term spread	Default spread
MKT-RF	1,000							
SMB	0,328	1,000						
HML	0,155	0,148	1,000					
MOM	-0,335	-0,157	-0,363	1,000				
LOWVOL	-0,339	-0,562	-0,361	0,384	1,000			
Dividend yield	0,049	0,034	0,028	-0,071	-0,193	1,000		
Term spread	0,058	0,102	0,019	-0,077	-0,141	-0,089	1,000	
Default spread	-0,035	0,078	0,069	-0,139	-0,360	0,400	0,307	1,000

Table 2
Summary statistics for factor portfolios and predictors (forecasting window)

Panel A presents the main statistics for factors and predictors over the forecasting window (1980:01 - 2012:12). Return and yield data are annualized and continuously compounded (for instance, 1,00 stands for 1,00%). Panel B displays sample correlations.

Panel A: Main statistics

	Mean	Median	Max	Min	Std. Dev.	Skewness	Kurtosis	Jarque-Bera
MKT-RF	5,68	12,95	141,02	-317,38	16,15	-1,01	6,41	259
SMB	1,11	-0,72	238,82	-214,95	10,65	0,32	10,089	835
HML	4,74	3,95	155,97	-124,57	10,57	0,25	4,97	68
MOM	6,14	9,03	202,58	-511,41	16,94	-2,39	21,36	5940
LOWVOL	-38,33	-31,21	266,63	-1099,60	35,16	-2,72	20,62	5610
Dividend yield	-4426,94	-4456,50	-3301,64	-5437,59	150,56	0,14	2,05	16
Term spread	26,52	28,40	53,39	-44,62	5,08	-0,85	4,00	65
Default spread	13,43	11,76	39,89	6,58	1,67	1,67	6,34	369

Panel B: Sample correlations

	MKT_RF	SMB	HML	MOM	LOWVOL	Dividend yield	Term spread	Default spread
MKT-RF	1,000							
SMB	0,248	1,000						
HML	-0,314	-0,115	1,000					
MOM	-0,119	0,031	-0,045	1,000				
LOWVOL	-0,503	-0,516	0,129	0,318	1,000			
Dividend yield	0,067	-0,010	-0,082	-0,025	-0,027	1,000		
Term spread	0,045	0,125	-0,054	-0,051	-0,151	-0,012	1,000	
Default spread	-0,018	0,082	-0,089	-0,142	-0,028	0,585	0,051	1,000

Table 3
Model selection

The table reports the statistics used to select the most appropriate specification for each class of model. Panel A refers to models in which predictors are exogenous variables, while Panel B contains models having endogenous predictors. For the AIC, SC and HQ criteria, values in bold indicate the specification selected by each criterion.

Panel A: exogenous predictors

Model	Number of parameters	Saturation ratio	Log-likelihood	AIC	SC	HQ
Base model: VARX(*p*,*q*)						
VARX(0,0)	20	251,8	7728,85	-15,4486	**-15,3503**	-15,4112
VARX(1,1)	45	111,9	7809,85	-15,5608	-15,3396	**-15,4767**
VARX(2,2)	70	71,9	7849,22	-15,5896	-15,2455	-15,4588
VARX(3,3)	95	53,0	7899,32	**-15,6399**	-15,1729	-15,4624
Base model: TVARX(*K*,*p*,*q*)						
TVARX(2,1,1)	122	41,3	8241,69	-16,1105	**-15,5155**	**-15,8845**
TVARX(2,2,2)	202	24,9	8374,39	-16,2151	-15,2300	-15,8408
TVARX(2,3,3)	282	17,8	8486,55	**-16,2789**	-14,9036	-15,7564
TVARX(3,1,1)	186	27,1	8266,28	-16,0323	-15,1252	-15,6877
TVARX(3,2,2)	306	16,4	8404,35	-16,0682	-14,5759	-15,5012
TVARX(3,3,3)	426	11,8	8469,17	-15,9587	-13,8812	-15,1694
Base model: MSI(*K*)						
MSI(2)	27	186,7	8110,62	-16,0389	-15,9073	-15,9889
MSI(3)	36	140,0	8404,15	**-16,6035**	**-16,4279**	**-16,5368**
Base model: MSIH(*K*)						
MSIH(2)	42	120,0	9438,18	-18,6432	-18,4384	-18,5654
MSIH(3)	66	76,4	9725,54	**-19,1658**	**-18,8439**	**-19,0435**

Table 3 (Continued)
Model selection

See above.

Panel B: endogenous predictors

Model	Number of parameters	Saturation ratio	Log-likelihood	AIC	SC	HQ
Base model: VAR(*p*)						
VAR(0,0)	44	183,1	13709,57	-27,4581	**-27,4187**	-27,4431
VAR(1)	108	74,6	20233,75	-40,4043	-40,0504	**-40,2698**
VAR(2)	172	46,8	20601,53	-41,0131	-40,3446	-40,7590
VAR(3)	236	34,1	20707,04	**-41,0963**	-40,1132	-40,7226
Base model: TVAR(*K*,*p*)						
TVAR(2,1)	218	37,0	20696,71	-40,6324	**-39,5692**	**-40,2285**
TVAR(2,2)	346	23,3	21196,95	-41,3709	-39,6836	-40,7299
TVAR(2,3)	474	17,0	21338,03	**-41,3969**	-39,0853	-40,5187
TVAR(3,1)	330	24,4	20765,16	-40,5459	-38,9366	-39,9345
TVAR(3,2)	522	15,4	21228,95	-41,0852	-38,5396	-40,1181
TVAR(3,3)	714	11,3	21374,84	-40,9937	-37,5118	-39,6709
Base model: MSIAH(*K*) - VAR(*p*)						
MSIAH(2) - VAR(1)	218	37,0	22837,16	-44,9239	-43,8599	-44,5196
MSIAH(2) - VAR(2)	346	23,3	23146,36	-45,3287	-43,6387	-44,6866
MSIAH(2) - VAR(3)	474	17,0	23248,21	-45,3218	-43,0048	-44,4414
MSIAH(3) - VAR(1)	330	24,4	**23366,38**	**-45,7525**	**-44,1419**	**-45,1406**

Table 4
Forecast errors: test of unbiasedness

For each model, factor, and forecast horizon, we test that the expected value of the forecast error is 0. If the null hypothesis is rejected, then there is evidence to argue that the forecast model is biased. This table reports the *p*-values associated with each test. *,**, and *** indicate that the null hypothesis is rejected at the 10%, 5%, and 1% confidence level, respectively.

Model	Factor	Horizon		
		1-month	3-month	12-month
CER	MKT-RF	0,8683	0,8678	0,8657
	SMB	0,4929	0,4918	0,4873
	HML	0,3635	0,3623	0,3556
	MOM	0,7871	0,7873	0,7887
	LOWVOL	0,0000 ***	0,0000 ***	0,0000 ***
VARX(1,1)	MKT-RF	0,1687	0,2177	0,0190 **
	SMB	0,8907	0,6906	0,5013
	HML	0,4628	0,6010	0,3441
	MOM	0,1411	0,2281	0,9831
	LOWVOL	0,0189 **	0,0934 *	0,0820 *
VAR(2)	MKT-RF	0,3011	0,2947	0,2706
	SMB	0,9560	0,9822	0,9894
	HML	0,5843	0,5659	0,6467
	MOM	0,1190	0,1167	0,0981 *
	LOWVOL	0,0973 *	0,0882 *	0,0701 *
TVARX(1,1)	MKT-RF	0,1724	0,1230	0,0973 *
	SMB	0,6517	0,8370	0,5813
	HML	0,9726	0,3081	0,2335
	MOM	0,1009	0,5638	0,6938
	LOWVOL	0,0103 **	0,0104 **	0,0829 *

Table 4 (Continued)
Forecast errors: test of unbiasedness
See above.

Model	Factor	Horizon					
		1-month		3-month		12-month	
TVAR(2)	MKT-RF	0,8152		0,0006	***	0,0000	***
	SMB	0,4589		0,2303		0,0000	***
	HML	0,6393		0,2288		0,0000	***
	MOM	0,0452	**	0,0158	**	0,1906	
	LOWVOL	0,5143		0,0102	**	0,0000	***
MSI(3)	MKT-RF	0,7524		0,8387		0,8560	
	SMB	0,7130		0,5426		0,5026	
	HML	0,6886		0,4303		0,3752	
	MOM	0,5078		0,7152		0,7652	
	LOWVOL	0,0001	***	0,0000	***	0,0000	***
MSIH(3)	MKT-RF	0,8330		0,8548		0,8547	
	SMB	0,6806		0,4814		0,5100	
	HML	0,5932		0,3605		0,3799	
	MOM	0,5486		0,7730		0,7681	
	LOWVOL	0,0013	***	0,0000	***	0,0000	***
MSIVARH(3,1)	MKT-RF	0,3049		0,2676		0,2983	
	SMB	0,9083		0,1658		0,0938	*
	HML	0,2556		0,3813		0,3758	
	MOM	0,0304	**	0,0008	***	0,0004	***
	LOWVOL	0,6673		0,2001		0,0685	*

Table 5
Measures of forecast accuracy: Excess market returns

The table presents six measures of predictive accuracy, calculated for each model. From the first to the last column, they are: the Root Mean Square Forecast Error (RMSFE), the Forecast Error Bias (FEB), the Forecast Error Variance (FEV), the Mean Absolute Forecast Error (MAFE), the Mean Percent Forecast Error (MPFE), and the Success Ratio (SR). Panel A, B, and C cover different forecast horizons: respectively, 1-month, 3-month, and 12-month.

Panel A: 1-month forecast horizon

	RMSFE	FEB	FEV	MAFE	MPFE	SR
CER	0,0466	0,0004	0,0022	0,0350	91,79%	98,23%
VARX(1,1)	0,0471	0,0033	0,0022	0,0360	97,45%	95,96%
VAR(2)	0,0482	0,0025	0,0023	0,0367	93,96%	95,45%
TVARX(2,1,1)	0,0479	0,0033	0,0023	0,0365	95,98%	94,44%
TVAR(2,2)	0,0551	0,0006	0,0030	0,0395	110,89%	87,63%
MSI(3)	0,0465	0,0007	0,0022	0,0351	91,24%	97,73%
MSIH(3)	0,0467	0,0005	0,0022	0,0351	92,98%	95,96%
MSIVARH(3,1)	0,0484	0,0025	0,0023	0,0367	97,19%	93,69%

Panel B: 3-month forecast horizon

	RMSFE	FEB	FEV	MAFE	MPFE	SR
CER	0,0466	0,0004	0,0022	0,0350	91,77%	98,23%
VARX(1,1)	0,0478	0,0030	0,0023	0,0366	99,71%	94,44%
VAR(2)	0,0482	0,0025	0,0023	0,0367	93,77%	95,20%
TVARX(2,1,1)	0,0477	0,0037	0,0023	0,0366	101,50%	93,43%
TVAR(2,2)	0,0484	0,0082	0,0023	0,0370	107,36%	95,45%
MSI(3)	0,0465	0,0005	0,0022	0,0350	92,00%	97,98%
MSIH(3)	0,0466	0,0004	0,0022	0,0351	91,46%	97,98%
MSIVARH(3,1)	0,0470	0,0026	0,0022	0,0357	99,00%	97,22%

Panel C: 12-month forecast horizon

	RMSFE	FEB	FEV	MAFE	MPFE	SR
CER	0,0466	0,0004	0,0022	0,0350	91,92%	97,98%
VARX(1,1)	0,0468	0,0055	0,0022	0,0359	98,22%	96,97%
VAR(2)	0,0484	0,0027	0,0023	0,0368	95,07%	94,70%
TVARX(2,1,1)	0,0488	0,0041	0,0024	0,0367	97,62%	96,21%
TVAR(2,2)	0,0584	0,0276	0,0027	0,0468	135,66%	85,61%
MSI(3)	0,0466	0,0004	0,0022	0,0351	91,99%	97,98%
MSIH(3)	0,0466	0,0004	0,0022	0,0350	92,36%	97,98%
MSIVARH(3,1)	0,0466	0,0024	0,0022	0,0353	92,76%	97,98%

Table 6
Measures of forecast accuracy: Size factor returns
See notes for Table 5.

Panel A: 1-month forecast horizon

	RMSFE	FEB	FEV	MAFE	MPFE	SR
CER	0,0307	-0,0011	0,0009	0,0220	99,52%	96,46%
VARX(1,1)	0,0307	-0,0002	0,0009	0,0216	119,50%	93,18%
VAR(2)	0,0316	-0,0001	0,0010	0,0222	124,53%	89,90%
TVARX(2,1,1)	0,0319	-0,0007	0,0010	0,0214	124,04%	91,41%
TVAR(2,2)	0,0384	-0,0014	0,0015	0,0280	117,36%	80,81%
MSI(3)	0,0306	-0,0006	0,0009	0,0219	98,95%	97,73%
MSIH(3)	0,0316	-0,0007	0,0010	0,0223	99,50%	98,23%
MSIVARH(3,1)	0,0355	-0,0002	0,0013	0,0235	117,60%	89,39%

Panel B: 3-month forecast horizon

	RMSFE	FEB	FEV	MAFE	MPFE	SR
CER	0,0307	-0,0011	0,0009	0,0220	99,43%	96,46%
VARX(1,1)	0,0308	0,0006	0,0010	0,0220	98,53%	94,44%
VAR(2)	0,0314	0,0000	0,0010	0,0221	124,24%	89,65%
TVARX(2,1,1)	0,0303	0,0003	0,0009	0,0219	98,26%	94,19%
TVAR(2,2)	0,0344	0,0021	0,0012	0,0251	145,06%	81,31%
MSI(3)	0,0308	-0,0009	0,0009	0,0221	99,29%	96,21%
MSIH(3)	0,0307	-0,0011	0,0009	0,0220	98,38%	96,97%
MSIVARH(3,1)	0,0316	-0,0022	0,0010	0,0228	104,15%	92,93%

Panel C: 12-month forecast horizon

	RMSFE	FEB	FEV	MAFE	MPFE	SR
CER	0,0307	-0,0011	0,0009	0,0220	99,28%	96,46%
VARX(1,1)	0,0311	0,0011	0,0010	0,0223	80,49%	89,39%
VAR(2)	0,0314	0,0000	0,0010	0,0223	124,75%	89,39%
TVARX(2,1,1)	0,0314	0,0009	0,0010	0,0224	65,01%	87,12%
TVAR(2,2)	0,0332	0,0114	0,0010	0,0235	83,37%	81,57%
MSI(3)	0,0307	-0,0010	0,0009	0,0220	99,27%	96,46%
MSIH(3)	0,0308	-0,0010	0,0009	0,0221	99,47%	96,72%
MSIVARH(3,1)	0,0309	-0,0026	0,0009	0,0224	97,24%	93,69%

Table 7
Measures of forecast accuracy: Value factor returns
See notes for Table 5.

Panel A: 1-month forecast horizon

	RMSFE	FEB	FEV	MAFE	MPFE	SR
CER	0,0305	-0,0014	0,0009	0,0223	57,68%	89,39%
VARX(1,1)	0,0311	-0,0011	0,0010	0,0228	23,22%	84,34%
VAR(2)	0,0311	-0,0009	0,0010	0,0228	-5,00%	84,34%
TVARX(2,1,1)	0,0311	-0,0001	0,0010	0,0225	107,54%	85,10%
TVAR(2,2)	0,0345	-0,0008	0,0012	0,0250	-143,33%	84,09%
MSI(3)	0,0303	-0,0006	0,0009	0,0223	67,77%	92,42%
MSIH(3)	0,0296	-0,0008	0,0009	0,0216	80,15%	93,94%
MSIVARH(3,1)	0,0318	0,0018	0,0010	0,0234	94,60%	89,65%

Panel B: 3-month forecast horizon

	RMSFE	FEB	FEV	MAFE	MPFE	SR
CER	0,0305	-0,0014	0,0009	0,0224	57,38%	89,39%
VARX(1,1)	0,0311	-0,0008	0,0010	0,0226	52,88%	90,66%
VAR(2)	0,0312	-0,0009	0,0010	0,0227	-9,13%	84,60%
TVARX(2,1,1)	0,0303	-0,0016	0,0009	0,0222	47,14%	89,14%
TVAR(2,2)	0,0321	-0,0019	0,0010	0,0235	-71,43%	88,64%
MSI(3)	0,0305	-0,0012	0,0009	0,0223	59,39%	89,65%
MSIH(3)	0,0304	-0,0014	0,0009	0,0222	59,44%	89,39%
MSIVARH(3,1)	0,0313	0,0014	0,0010	0,0232	72,42%	92,68%

Panel C: 12-month forecast horizon

	RMSFE	FEB	FEV	MAFE	MPFE	SR
CER	0,0306	-0,0014	0,0009	0,0224	56,52%	89,39%
VARX(1,1)	0,0310	-0,0015	0,0010	0,0227	31,04%	85,86%
VAR(2)	0,0314	-0,0007	0,0010	0,0229	-16,78%	84,09%
TVARX(2,1,1)	0,0310	-0,0019	0,0010	0,0226	43,48%	85,61%
TVAR(2,2)	0,0318	-0,0065	0,0010	0,0237	37,12%	80,05%
MSI(3)	0,0306	-0,0014	0,0009	0,0224	56,95%	89,39%
MSIH(3)	0,0305	-0,0013	0,0009	0,0223	63,37%	89,39%
MSIVARH(3,1)	0,0311	0,0014	0,0010	0,0229	28,14%	93,43%

Table 8
Measures of forecast accuracy: Momentum factor returns
See notes for Table 5.

Panel A: 1-month forecast horizon

	RMSFE	FEB	FEV	MAFE	MPFE	SR
CER	0,0489	-0,0007	0,0024	0,0309	-3,37%	89,90%
VARX(1,1)	0,0488	-0,0036	0,0024	0,0318	-20,61%	79,80%
VAR(2)	0,0490	-0,0038	0,0024	0,0316	-50,72%	79,80%
TVARX(2,1,1)	0,0502	-0,0041	0,0025	0,0327	-80,61%	79,29%
TVAR(2,2)	0,0578	-0,0058	0,0033	0,0348	-113,39%	79,29%
MSI(3)	0,0482	-0,0016	0,0023	0,0308	-37,11%	87,63%
MSIH(3)	0,0480	-0,0014	0,0023	0,0307	-67,22%	85,86%
MSIVARH(3,1)	0,0515	-0,0056	0,0026	0,0327	-75,53%	85,35%

Panel B: 3-month forecast horizon

	RMSFE	FEB	FEV	MAFE	MPFE	SR
CER	0,0489	-0,0007	0,0024	0,0309	-3,28%	89,90%
VARX(1,1)	0,0489	-0,0030	0,0024	0,0321	-119,17%	83,08%
VAR(2)	0,0489	-0,0038	0,0024	0,0315	-49,35%	79,55%
TVARX(2,1,1)	0,0503	-0,0015	0,0025	0,0327	-42,04%	84,85%
TVAR(2,2)	0,0522	-0,0063	0,0027	0,0333	-230,67%	84,34%
MSI(3)	0,0488	-0,0009	0,0024	0,0308	-11,85%	89,90%
MSIH(3)	0,0488	-0,0007	0,0024	0,0309	-9,50%	89,14%
MSIVARH(3,1)	0,0496	-0,0082	0,0024	0,0313	-142,49%	83,33%

Panel C: 12-month forecast horizon

	RMSFE	FEB	FEV	MAFE	MPFE	SR
CER	0,0489	-0,0007	0,0024	0,0309	-1,81%	89,90%
VARX(1,1)	0,0489	-0,0001	0,0024	0,0313	-23,03%	88,89%
VAR(2)	0,0492	-0,0041	0,0024	0,0317	-51,49%	79,29%
TVARX(2,1,1)	0,0492	0,0010	0,0024	0,0317	-28,43%	90,15%
TVAR(2,2)	0,0493	-0,0032	0,0024	0,0317	-150,18%	84,34%
MSI(3)	0,0489	-0,0007	0,0024	0,0309	-2,83%	89,90%
MSIH(3)	0,0489	-0,0007	0,0024	0,0309	-5,83%	90,15%
MSIVARH(3,1)	0,0499	-0,0087	0,0024	0,0314	-120,59%	82,58%

Table 9
Measures of forecast accuracy: Low Volatility factor returns
See notes for Table 5.

Panel A: 1-month forecast horizon

	RMSFE	FEB	FEV	MAFE	MPFE	SR
CER	0,1056	0,0291	0,0103	0,0737	280,79%	63,13%
VARX(1,1)	0,1307	0,0153	0,0169	0,0929	1057,85%	78,54%
VAR(2)	0,1432	0,0119	0,0204	0,0975	1607,95%	77,53%
TVARX(2,1,1)	0,1293	0,0166	0,0165	0,0902	907,90%	74,49%
TVAR(2,2)	0,1496	0,0049	0,0224	0,1037	1604,09%	70,45%
MSI(3)	0,1035	0,0207	0,0103	0,0679	222,42%	71,46%
MSIH(3)	0,1342	0,0214	0,0176	0,0799	207,64%	80,30%
MSIVARH(3,1)	0,1125	-0,0024	0,0127	0,0712	361,69%	84,09%

Panel B: 3-month forecast horizon

	RMSFE	FEB	FEV	MAFE	MPFE	SR
CER	0,1057	0,0292	0,0103	0,0738	281,43%	63,13%
VARX(1,1)	0,1161	0,0098	0,0134	0,0805	625,91%	74,75%
VAR(2)	0,1441	0,0123	0,0207	0,0985	1606,61%	77,78%
TVARX(2,1,1)	0,1187	0,0152	0,0139	0,0822	662,80%	74,49%
TVAR(2,2)	0,1329	-0,0170	0,0174	0,0958	617,90%	80,56%
MSI(3)	0,1051	0,0271	0,0103	0,0727	272,68%	65,66%
MSIH(3)	0,1058	0,0294	0,0104	0,0743	316,37%	65,66%
MSIVARH(3,1)	0,1048	0,0067	0,0110	0,0711	370,71%	78,79%

Panel C: 12-month forecast horizon

	RMSFE	FEB	FEV	MAFE	MPFE	SR
CER	0,1058	0,0295	0,0104	0,0740	285,55%	63,13%
VARX(1,1)	0,1304	0,0114	0,0169	0,0931	1171,85%	80,05%
VAR(2)	0,1497	0,0136	0,0223	0,1024	1723,92%	76,77%
TVARX(2,1,1)	0,1294	0,0112	0,0167	0,0927	990,85%	76,77%
TVAR(2,2)	0,1515	-0,0453	0,0210	0,1089	1052,90%	83,08%
MSI(3)	0,1056	0,0289	0,0103	0,0737	284,02%	63,38%
MSIH(3)	0,1057	0,0290	0,0104	0,0738	294,08%	63,89%
MSIVARH(3,1)	0,1050	0,0096	0,0110	0,0721	480,28%	79,04%

Table 10
Model ranking: Forecasting horizon

The table presents the top three forecast model for each factor portfolio, according to the six measures of forecast accuracy in the previous tables. Panel A, B, and C cover different forecasting horizons: respectively, 1-month, 3-month, and 12-month.

Panel A: 1-month forecast horizon

		RMSFE	FEB	FEV	MAFE	MPFE	SR
MKT-RF	1.	MSI(3)	CER	MSI(3)	CER	MSI(3)	CER
	2.	CER	MSIH(3)	CER	MSI(3)	CER	MSI(3)
	3.	MSIH(3)	TVAR(2,2)	MSIH(3)	MSIH(3)	MSIH(3)	VARX(1,1)
SMB	1.	MSI(3)	TVAR(2,2)	MSI(3)	TVARX(2,1,1)	MSI(3)	MSIH(3)
	2.	VARX(1,1)	CER	VARX(1,1)	VARX(1,1)	MSIH(3)	MSI(3)
	3.	CER	TVARX(2,1,1)	CER	MSI(3)	CER	CER
HML	1.	MSIH(3)	CER	MSIH(3)	MSIH(3)	TVAR(2,2)	MSIH(3)
	2.	MSI(3)	VARX(1,1)	MSI(3)	MSI(3)	VAR(2)	MSI(3)
	3.	CER	VAR(2)	CER	CER	VARX(1,1)	MSIVARH(3,1)
MOM	1.	MSIH(3)	TVAR(2,2)	MSIH(3)	MSIH(3)	TVAR(2,2)	CER
	2.	MSI(3)	MSIVARH(3,1)	MSI(3)	MSI(3)	TVARX(2,1,1)	MSI(3)
	3.	VARX(1,1)	TVARX(2,1,1)	VARX(1,1)	CER	MSIVARH(3,1)	MSIH(3)
LOWVOL	1.	MSI(3)	MSIVARH(3,1)	MSI(3)	MSI(3)	MSIH(3)	MSIVARH(3,1)
	2.	CER	TVAR(2,2)	CER	MSIVARH(3,1)	MSI(3)	MSIH(3)
	3.	MSIVARH(3,1)	VAR(2)	MSIVARH(3,1)	CER	CER	VARX(1,1)

Panel B: 3-month forecast horizon

		RMSFE	FEB	FEV	MAFE	MPFE	SR
MKT-RF	1.	MSI(3)	CER	MSI(3)	MSI(3)	MSIH(3)	CER
	2.	CER	MSIH(3)	CER	CER	CER	MSI(3)
	3.	MSIH(3)	MSI(3)	MSIH(3)	MSIH(3)	MSI(3)	MSIH(3)
SMB	1.	TVARX(2,1,1)	MSIVARH(3,1)	TVARX(2,1,1)	TVARX(2,1,1)	TVARX(2,1,1)	MSIH(3)
	2.	MSIH(3)	MSIH(3)	MSIH(3)	MSIH(3)	MSIH(3)	CER
	3.	CER	CER	CER	VARX(1,1)	VARX(1,1)	MSI(3)
HML	1.	TVARX(2,1,1)	TVAR(2,2)	TVARX(2,1,1)	MSIH(3)	TVAR(2,2)	MSIVARH(3,1)
	2.	MSIH(3)	TVARX(2,1,1)	MSIH(3)	TVARX(2,1,1)	VAR(2)	VARX(1,1)
	3.	MSI(3)	CER	MSI(3)	MSI(3)	TVARX(2,1,1)	MSI(3)
MOM	1.	MSI(3)	MSIVARH(3,1)	MSI(3)	MSI(3)	TVAR(2,2)	CER
	2.	MSIH(3)	TVAR(2,2)	VAR(2)	CER	MSIVARH(3,1)	MSIH(3)
	3.	CER	VAR(2)	VARX(1,1)	MSIH(3)	VARX(1,1)	MSIH(3)
LOWVOL	1.	MSIVARH(3,1)	TVAR(2,2)	MSI(3)	MSIVARH(3,1)	MSI(3)	TVAR(2,2)
	2.	MSI(3)	MSIVARH(3,1)	CER	MSI(3)	CER	MSIVARH(3,1)
	3.	CER	VARX(1,1)	MSIH(3)	CER	MSIH(3)	VAR(2)

Table 10 (Continued)
Model ranking: Forecasting horizon

See above.

Panel C: 12-month forecast horizon

		RMSFE	FEB	FEV	MAFE	MPFE	SR
MKT-RF	1.	MSIH(3)	CER	VARX(1,1)	MSIH(3)	CER	CER
	2.	MSI(3)	MSI(3)	MSIH(3)	CER	MSI(3)	MSI(3)
	3.	CER	MSIH(3)	MSIVARH(3,1)	MSI(3)	MSIH(3)	MSIH(3)
SMB	1.	MSI(3)	MSIVARH(3,1)	MSI(3)	MSI(3)	TVARX(2,1,1)	MSIH(3)
	2.	CER	CER	CER	CER	VARX(1,1)	CER
	3.	MSIH(3)	MSI(3)	MSIH(3)	MSIH(3)	TVAR(2,2)	MSI(3)
HML	1.	MSIH(3)	TVAR(2,2)	MSIH(3)	MSIH(3)	VAR(2)	MSIVARH(3,1)
	2.	MSI(3)	TVARX(2,1,1)	MSI(3)	MSI(3)	MSIVARH(3,1)	CER
	3.	CER	VARX(1,1)	CER	CER	VARX(1,1)	MSI(3)
MOM	1.	VARX(1,1)	MSIVARH(3,1)	VARX(1,1)	MSI(3)	TVAR(2,2)	TVARX(2,1,1)
	2.	CER	VAR(2)	CER	CER	MSIVARH(3,1)	MSIH(3)
	3.	MSI(3)	TVAR(2,2)	MSI(3)	MSIH(3)	VAR(2)	CER
LOWVOL	1.	MSIVARH(3,1)	TVAR(2,2)	MSI(3)	MSIVARH(3,1)	MSI(3)	TVAR(2,2)
	2.	MSI(3)	MSIVARH(3,1)	CER	MSI(3)	CER	VARX(1,1)
	3.	MSIH(3)	TVARX(2,1,1)	MSIH(3)	MSIH(3)	MSIH(3)	MSIVARH(3,1)

Table 11
Model ranking: Factor portfolios

Similarly to Table 10, the table presents the top three forecast model for each factor portfolio, according to the six measures of forecast accuracy in the previous tables. Each panel covers a factor portfolio.

Panel A: Excess market returns

		RMSFE	FEB	FEV	MAFE	MPFE	SR
h=1	1.	MSI(3)	CER	MSI(3)	CER	MSI(3)	CER
	2.	CER	MSIH(3)	CER	MSI(3)	CER	MSI(3)
	3.	MSIH(3)	TVAR(2,2)	MSIH(3)	MSIH(3)	MSIH(3)	VARX(1,1)
h=3	1.	MSI(3)	CER	MSI(3)	MSI(3)	MSIH(3)	CER
	2.	CER	MSIH(3)	CER	CER	CER	MSI(3)
	3.	MSIH(3)	MSI(3)	MSIH(3)	MSIH(3)	MSI(3)	MSIH(3)
h=12	1.	MSIH(3)	CER	VARX(1,1)	MSIH(3)	CER	CER
	2.	MSI(3)	MSI(3)	MSIH(3)	CER	MSI(3)	MSI(3)
	3.	CER	MSIH(3)	MSIVARH(3,1)	MSI(3)	MSIH(3)	MSIH(3)

Panel B: Size factor returns

		RMSFE	FEB	FEV	MAFE	MPFE	SR
h=1	1.	MSI(3)	TVAR(2,2)	MSI(3)	TVARX(2,1,1)	MSI(3)	MSIH(3)
	2.	VARX(1,1)	CER	VARX(1,1)	VARX(1,1)	MSIH(3)	MSI(3)
	3.	CER	TVARX(2,1,1)	CER	MSI(3)	CER	CER
h=3	1.	TVARX(2,1,1)	MSIVARH(3,1)	TVARX(2,1,1)	TVARX(2,1,1)	TVARX(2,1,1)	MSIH(3)
	2.	MSIH(3)	MSIH(3)	MSIH(3)	MSIH(3)	MSIH(3)	CER
	3.	CER	CER	CER	VARX(1,1)	VARX(1,1)	MSI(3)
h=12	1.	MSI(3)	MSIVARH(3,1)	MSI(3)	MSI(3)	TVARX(2,1,1)	MSIH(3)
	2.	CER	CER	CER	CER	VARX(1,1)	CER
	3.	MSIH(3)	MSI(3)	MSIH(3)	MSIH(3)	TVAR(2,2)	MSI(3)

Panel C: Value factor returns

		RMSFE	FEB	FEV	MAFE	MPFE	SR
h=1	1.	MSIH(3)	CER	MSIH(3)	MSIH(3)	TVAR(2,2)	MSIH(3)
	2.	MSI(3)	VARX(1,1)	MSI(3)	MSI(3)	VAR(2)	MSI(3)
	3.	CER	VAR(2)	CER	CER	VARX(1,1)	MSIVARH(3,1)
h=3	1.	TVARX(2,1,1)	TVAR(2,2)	TVARX(2,1,1)	MSIH(3)	TVAR(2,2)	MSIVARH(3,1)
	2.	MSIH(3)	TVARX(2,1,1)	MSIH(3)	TVARX(2,1,1)	VAR(2)	VARX(1,1)
	3.	MSI(3)	CER	MSI(3)	MSI(3)	TVARX(2,1,1)	MSI(3)
h=12	1.	MSIH(3)	TVAR(2,2)	MSIH(3)	MSIH(3)	VAR(2)	MSIVARH(3,1)
	2.	MSI(3)	TVARX(2,1,1)	MSI(3)	MSI(3)	MSIVARH(3,1)	CER
	3.	CER	VARX(1,1)	CER	CER	VARX(1,1)	MSI(3)

Table 11 (Continued)
Model ranking: Factor portfolios

See above.

Panel D: Momentum factor returns

		RMSFE	FEB	FEV	MAFE	MPFE	SR
h=1	1.	MSIH(3)	TVAR(2,2)	MSIH(3)	MSIH(3)	TVAR(2,2)	CER
	2.	MSI(3)	MSIVARH(3,1)	MSI(3)	MSI(3)	TVARX(2,1,1)	MSI(3)
	3.	VARX(1,1)	TVARX(2,1,1)	VARX(1,1)	CER	MSIVARH(3,1)	MSIH(3)
h=3	1.	MSI(3)	MSIVARH(3,1)	MSI(3)	MSI(3)	TVAR(2,2)	CER
	2.	MSIH(3)	TVAR(2,2)	VAR(2)	CER	MSIVARH(3,1)	MSIH(3)
	3.	CER	VAR(2)	VARX(1,1)	MSIH(3)	VARX(1,1)	MSIH(3)
h=12	1.	VARX(1,1)	MSIVARH(3,1)	VARX(1,1)	MSI(3)	TVAR(2,2)	TVARX(2,1,1)
	2.	CER	VAR(2)	CER	CER	MSIVARH(3,1)	MSIH(3)
	3.	MSI(3)	TVAR(2,2)	MSI(3)	MSIH(3)	VAR(2)	CER

Panel E: Low Volatility factor returns

		RMSFE	FEB	FEV	MAFE	MPFE	SR
h=1	1.	MSI(3)	MSIVARH(3,1)	MSI(3)	MSI(3)	MSIH(3)	MSIVARH(3,1)
	2.	CER	TVAR(2,2)	CER	MSIVARH(3,1)	MSI(3)	MSIH(3)
	3.	MSIVARH(3,1)	VAR(2)	MSIVARH(3,1)	CER	CER	VARX(1,1)
h=3	1.	MSIVARH(3,1)	TVAR(2,2)	MSI(3)	MSIVARH(3,1)	MSI(3)	TVAR(2,2)
	2.	MSI(3)	MSIVARH(3,1)	CER	MSI(3)	CER	MSIVARH(3,1)
	3.	CER	VARX(1,1)	MSIH(3)	CER	MSIH(3)	VAR(2)
h=12	1.	MSIVARH(3,1)	TVAR(2,2)	MSI(3)	MSIVARH(3,1)	MSI(3)	TVAR(2,2)
	2.	MSI(3)	MSIVARH(3,1)	CER	MSI(3)	CER	VARX(1,1)
	3.	MSIH(3)	TVARX(2,1,1)	MSIH(3)	MSIH(3)	MSIH(3)	MSIVARH(3,1)

Table 12
Model ranking: Statistics

The table shows the number of times and the percentage (in italic) with which models rank among the top three candidates in Table 10 and Table 11. Panel A sorts figures by forecasting horizon, while Panel B by factor portfolio. Totals are displayed in the last column.

Panel A: Forecasting horizon

	Horizon			
	1-month	3-month	12-month	Total
CER	21	18	20	59
	23,3%	*20,0%*	*22,2%*	*21,9%*
VARX(1,1)	9	6	7	22
	10,0%	*6,7%*	*7,8%*	*8,1%*
VAR(2)	3	4	3	10
	3,3%	*4,4%*	*3,3%*	*3,7%*
TVARX(2,1,1)	4	9	4	17
	4,4%	*10,0%*	*4,4%*	*6,3%*
TVAR(2,2)	6	6	6	18
	6,7%	*6,7%*	*6,7%*	*6,7%*
MSI(3)	22	18	21	61
	24,4%	*20,0%*	*23,3%*	*22,6%*
MSIH(3)	17	21	19	57
	18,9%	*23,3%*	*21,1%*	*21,1%*
MSIVARH(3,1)	8	8	10	26
	8,9%	*8,9%*	*11,1%*	*9,6%*

Table 12 (Continued)
Model ranking: Statistics
See above.

Panel B: Factor portfolios

	Factor portfolio					
	MKT-RF	SMB	HML	MOM	LOWVOL	Total
CER	17	14	9	9	10	59
	31,5%	*25,9%*	*16,7%*	*16,7%*	*18,5%*	*21,9%*
VARX(1,1)	2	6	5	6	3	22
	3,7%	*11,1%*	*9,3%*	*11,1%*	*5,6%*	*8,1%*
VAR(2)	0	0	4	4	2	10
	0,0%	*0,0%*	*7,4%*	*7,4%*	*3,7%*	*3,7%*
TVARX(2,1,1)	0	7	6	3	1	17
	0,0%	*13,0%*	*11,1%*	*5,6%*	*1,9%*	*6,3%*
TVAR(2,2)	1	2	4	6	5	18
	1,9%	*3,7%*	*7,4%*	*11,1%*	*9,3%*	*6,7%*
MSI(3)	16	11	12	10	12	61
	29,6%	*20,4%*	*22,2%*	*18,5%*	*22,2%*	*22,6%*
MSIH(3)	17	12	10	10	8	57
	31,5%	*22,2%*	*18,5%*	*18,5%*	*14,8%*	*21,1%*
MSIVARH(3,1)	1	2	4	6	13	26
	1,9%	*3,7%*	*7,4%*	*11,1%*	*24,1%*	*9,6%*

Table 13
Mincer and Zarnowitz regression: Factor portfolios

This table collects results from the Mincer and Zarnowitz (1969) regression, sorted by factor portfolio. The columns *Beta 0* and *Beta 1* contain the *p*-values of the simple Wald tests that the intercept is significantly different from 0 and the coefficient is significantly different from 1. The column *F-test* displays the *p*-values of the joint test. The last column includes the R-squared of the regressions. *, **, *** indicate that the null hypothesis is rejected at the 10%, 5%, and 1% confidence level, respectively.

Panel A: 1-month forecasting horizon

Factor	Model	Beta 0		Beta 1		F-test		R-squared
MKT-RF	CER	0,0301	**	0,0305	**	0,0948	*	1,18%
	VARX(1,1)	0,0560	*	0,0066	***	0,0097	***	1,86%
	VAR(2)	0,0274	**	0,0000	***	0,0000	***	6,88%
	TVARX(2,1,1)	0,0438	**	0,0000	***	0,0000	***	5,11%
	TVAR(2,2)	0,0347	**	0,0000	***	0,0000	***	28,71%
	MSI(3)	0,6071		0,6820		0,8749		0,04%
	MSIH(3)	0,2579		0,1816		0,4002		0,45%
	MSIVARH(3,1)	0,0606	*	0,0000	***	0,0000	***	7,41%
SMB	CER	0,1937		0,1506		0,2812		0,52%
	VARX(1,1)	0,8229		0,0110	**	0,0388	**	1,63%
	VAR(2)	0,6679		0,0000	***	0,0000	***	6,60%
	TVARX(2,1,1)	0,7144		0,0000	***	0,0000	***	7,76%
	TVAR(2,2)	0,6099		0,0000	***	0,0000	***	36,18%
	MSI(3)	0,9659		0,4797		0,7279		0,13%
	MSIH(3)	0,1764		0,0000	***	0,0000	***	7,16%
	MSIVARH(3,1)	0,4675		0,0000	***	0,0000	***	25,64%
HML	CER	0,1820		0,1678		0,2556		0,48%
	VARX(1,1)	0,1525		0,0000	***	0,0002	***	4,19%
	VAR(2)	0,1508		0,0000	***	0,0000	***	5,08%
	TVARX(2,1,1)	0,1087		0,0000	***	0,0000	***	5,00%
	TVAR(2,2)	0,0159	**	0,0000	***	0,0000	***	22,01%
	MSI(3)	0,7701		0,4151		0,6620		0,17%
	MSIH(3)	0,5696		0,8207		0,8454		0,01%
	MSIVARH(3,1)	0,0235	**	0,0000	***	0,0000	***	8,28%
MOM	CER	0,1084		0,1047		0,2582		0,67%
	VARX(1,1)	0,8598		0,0031	***	0,0043	***	2,20%
	VAR(2)	0,8071		0,0008	***	0,0011	***	2,80%
	TVARX(2,1,1)	0,4147		0,0000	***	0,0000	***	5,82%
	TVAR(2,2)	0,0619	*	0,0000	***	0,0000	***	27,77%
	MSI(3)	0,3175		0,4524		0,6057		0,14%
	MSIH(3)	0,5411		0,8217		0,8148		0,01%
	MSIVARH(3,1)	0,1587		0,0000	***	0,0000	***	8,91%
LOWVOL	CER	0,4063		0,2489		0,0000	***	0,34%
	VARX(1,1)	0,0000	***	0,0000	***	0,0000	***	39,03%
	VAR(2)	0,0000	***	0,0000	***	0,0000	***	49,52%
	TVARX(2,1,1)	0,0000	***	0,0000	***	0,0000	***	37,90%
	TVAR(2,2)	0,0000	***	0,0000	***	0,0000	***	54,31%
	MSI(3)	0,4054		0,0000	***	0,0000	***	6,90%
	MSIH(3)	0,0000	***	0,0000	***	0,0000	***	42,14%
	MSIVARH(3,1)	0,0000	***	0,0000	***	0,0000	***	19,70%

Table 13 (Continued)
Mincer and Zarnowitz regression: Factor portfolios
See above.

Panel B: 3-month forecasting horizon

Factor	Model	Beta 0		Beta 1		F-test		R-squared
MKT-RF	CER	0,0171	**	0,0173	**	0,0579	*	1,43%
	VARX(1,1)	0,0404	**	0,0000	***	0,0000	***	4,69%
	VAR(2)	0,0277	**	0,0000	***	0,0000	***	6,81%
	TVARX(2,1,1)	0,0535	*	0,0000	***	0,0000	***	4,50%
	TVAR(2,2)	0,0278	**	0,0000	***	0,0000	***	5,05%
	MSI(3)	0,9554		0,9337		0,9762		0,00%
	MSIH(3)	0,1903		0,1925		0,4201		0,43%
	MSIVARH(3,1)	0,0500	*	0,0160	**	0,0297	**	1,46%
SMB	CER	0,1701		0,1308		0,2520		0,58%
	VARX(1,1)	0,5998		0,0356	**	0,1012		1,12%
	VAR(2)	0,6827		0,0000	***	0,0000	***	5,67%
	TVARX(2,1,1)	0,7631		0,2834		0,5503		0,29%
	TVAR(2,2)	0,5252		0,0000	***	0,0000	***	20,26%
	MSI(3)	0,2513		0,1427		0,2835		0,54%
	MSIH(3)	0,8382		0,9526		0,7797		0,00%
	MSIVARH(3,1)	0,2444		0,0000	***	0,0000	***	5,28%
HML	CER	0,0226	**	0,0200	**	0,0441	**	1,37%
	VARX(1,1)	0,0021	***	0,0000	***	0,0001	***	4,70%
	VAR(2)	0,1457		0,0000	***	0,0000	***	5,20%
	TVARX(2,1,1)	0,7095		0,0208	**	0,0412	**	1,35%
	TVAR(2,2)	0,0029	***	0,0000	***	0,0000	***	10,28%
	MSI(3)	0,8497		0,7378		0,6934		0,03%
	MSIH(3)	0,0599	*	0,0853	*	0,1497		0,75%
	MSIVARH(3,1)	0,0051	***	0,0000	***	0,0000	***	5,42%
MOM	CER	0,0886	*	0,0854	*	0,2192		0,75%
	VARX(1,1)	0,7575		0,0347	**	0,0520	*	1,13%
	VAR(2)	0,8554		0,0011	***	0,0014	***	2,67%
	TVARX(2,1,1)	0,1035		0,0000	***	0,0000	***	5,52%
	TVAR(2,2)	0,0384	**	0,0000	***	0,0000	***	11,09%
	MSI(3)	0,2160		0,2354		0,4625		0,36%
	MSIH(3)	0,9651		0,9771		0,9590		0,00%
	MSIVARH(3,1)	0,9696		0,1943		0,0017	***	0,43%
LOWVOL	CER	0,2702		0,1533		0,0000	***	0,52%
	VARX(1,1)	0,0001	***	0,0000	***	0,0000	***	24,94%
	VAR(2)	0,0000	***	0,0000	***	0,0000	***	50,14%
	TVARX(2,1,1)	0,0002	***	0,0000	***	0,0000	***	28,16%
	TVAR(2,2)	0,0000	***	0,0000	***	0,0000	***	40,94%
	MSI(3)	0,7520		0,1923		0,0000	***	0,43%
	MSIH(3)	0,7080		0,0337	**	0,0000	***	1,14%
	MSIVARH(3,1)	0,0015	***	0,0000	***	0,0000	***	6,31%

Table 13 (Continued)
Mincer and Zarnowitz regression: Factor portfolios
See above.

Panel C: 12-month forecasting horizon

Factor	Model	Beta 0		Beta 1		F-test		R-squared
MKT-RF	CER	0,0215	**	0,0218	**	0,0709	*	1,33%
	VARX(1,1)	0,0299	**	0,2359		0,0325	**	0,36%
	VAR(2)	0,0298	**	0,0000	***	0,0000	***	7,46%
	TVARX(2,1,1)	0,0426	**	0,0000	***	0,0000	***	8,35%
	TVAR(2,2)	0,0817	*	0,0000	***	0,0000	***	18,22%
	MSI(3)	0,0270	**	0,0275	**	0,0864	*	1,23%
	MSIH(3)	0,7747		0,7930		0,9501		0,02%
	MSIVARH(3,1)	0,1963		0,4405		0,4331		0,15%
SMB	CER	0,2939		0,2357		0,3889		0,36%
	VARX(1,1)	0,5338		0,0006	***	0,0021	***	2,97%
	VAR(2)	0,6801		0,0000	***	0,0000	***	5,78%
	TVARX(2,1,1)	0,5554		0,0000	***	0,0001	***	4,82%
	TVAR(2,2)	0,0513	*	0,0001	***	0,0000	***	4,09%
	MSI(3)	0,2948		0,2371		0,3970		0,35%
	MSIH(3)	0,0855	*	0,0538	*	0,1252		0,94%
	MSIVARH(3,1)	0,9735		0,1701		0,0965	*	0,48%
HML	CER	0,0010	***	0,0009	***	0,0026	***	2,77%
	VARX(1,1)	0,1198		0,0003	***	0,0008	***	3,32%
	VAR(2)	0,1013		0,0000	***	0,0000	***	6,16%
	TVARX(2,1,1)	0,1785		0,0002	***	0,0004	***	3,52%
	TVAR(2,2)	0,0480	**	0,0000	***	0,0000	***	4,63%
	MSI(3)	0,0011	***	0,0010	***	0,0029	***	2,73%
	MSIH(3)	0,2226		0,1967		0,2956		0,42%
	MSIVARH(3,1)	0,0082	***	0,0001	***	0,0004	***	3,70%
MOM	CER	0,1793		0,1739		0,3823		0,47%
	VARX(1,1)	0,4206		0,2074		0,4511		0,40%
	VAR(2)	0,7737		0,0002	***	0,0003	***	3,40%
	TVARX(2,1,1)	0,1420		0,0159	**	0,0504	*	1,47%
	TVAR(2,2)	0,4172		0,0125	**	0,0187	**	1,57%
	MSI(3)	0,1640		0,1587		0,3537		0,50%
	MSIH(3)	0,2593		0,2443		0,4858		0,34%
	MSIVARH(3,1)	0,3529		0,0312	**	0,0002	***	1,17%
LOWVOL	CER	0,1343		0,0662	*	0,0000	***	0,85%
	VARX(1,1)	0,0000	***	0,0000	***	0,0000	***	39,16%
	VAR(2)	0,0000	***	0,0000	***	0,0000	***	53,76%
	TVARX(2,1,1)	0,0000	***	0,0000	***	0,0000	***	38,15%
	TVAR(2,2)	0,0000	***	0,0000	***	0,0000	***	51,20%
	MSI(3)	0,1959		0,1049		0,0000	***	0,67%
	MSIH(3)	0,1410		0,0638	*	0,0000	***	0,87%
	MSIVARH(3,1)	0,0006	***	0,0000	***	0,0000	***	6,08%

Table 14
Mincer and Zarnowitz regression: Forecast model

Similarly to Table 13, this table collects results from the Mincer and Zarnowitz (1969) regression, sorted by forecast model. See notes in Table 13 for details.

Panel A: 1-month forecasting horizon

Factor	Model	Beta 0		Beta 1		F-test		R-squared
CER	MKT-RF	0,0301	**	0,0305	**	0,0948	*	1,18%
	SMB	0,1937		0,1506		0,2812		0,52%
	HML	0,1820		0,1678		0,2556		0,48%
	MOM	0,1084		0,1047		0,2582		0,67%
	LOW VOL	0,4063		0,2489		0,0000	***	0,34%
VARX(1,1)	MKT-RF	0,0560	*	0,0066	***	0,0097	***	1,86%
	SMB	0,8229		0,0110	**	0,0388	**	1,63%
	HML	0,1525		0,0000	***	0,0002	***	4,19%
	MOM	0,8598		0,0031		0,0043		2,20%
	LOW VOL	0,0000	***	0,0000	***	0,0000	***	39,03%
VAR(2)	MKT-RF	0,0274	**	0,0000	***	0,0000	***	6,88%
	SMB	0,6679		0,0000	***	0,0000	***	6,60%
	HML	0,1508		0,0000	***	0,0000	***	5,08%
	MOM	0,8071		0,0008	***	0,0011	***	2,80%
	LOW VOL	0,0000	***	0,0000	***	0,0000	***	49,52%
TVARX(2,1,1)	MKT-RF	0,0438	**	0,0000	***	0,0000	***	5,11%
	SMB	0,7144		0,0000	***	0,0000	***	7,76%
	HML	0,1087		0,0000		0,0000		5,00%
	MOM	0,4147		0,0000	***	0,0000	***	5,82%
	LOW VOL	0,0000	***	0,0000	***	0,0000	***	37,90%
TVAR(2,2)	MKT-RF	0,0347	**	0,0000	***	0,0000	***	28,71%
	SMB	0,6099		0,0000	***	0,0000	***	36,18%
	HML	0,0159	**	0,0000	***	0,0000	***	22,01%
	MOM	0,0619	*	0,0000	***	0,0000	***	27,77%
	LOW VOL	0,0000	***	0,0000	***	0,0000	***	54,31%
MSI(3)	MKT-RF	0,6071		0,6820		0,8749		0,04%
	SMB	0,9659		0,4797		0,7279		0,13%
	HML	0,7701		0,4151		0,6620		0,17%
	MOM	0,3175		0,4524		0,6057		0,14%
	LOW VOL	0,4054		0,0000	***	0,0000	***	6,90%
MSIH(3)	MKT-RF	0,2579		0,1816		0,4002		0,45%
	SMB	0,1764		0,0000	***	0,0000	***	7,16%
	HML	0,5696		0,8207		0,8454		0,01%
	MOM	0,5411		0,8217		0,8148		0,01%
	LOW VOL	0,0000	***	0,0000	***	0,0000	***	42,14%
MSIVARH(3,1)	MKT-RF	0,0606		0,0000		0,0000		7,41%
	SMB	0,4675		0,0000	***	0,0000	***	25,64%
	HML	0,0235	**	0,0000	***	0,0000	***	8,28%
	MOM	0,1587		0,0000	***	0,0000	***	8,91%
	LOW VOL	0,0000	***	0,0000	***	0,0000	***	19,70%

Table 14 (Continued)
Mincer and Zarnowitz regression: Forecast model
See above.

Panel B: 3-month forecasting horizon

Factor	Model	Beta 0		Beta 1		F-test		R-squared
CER	MKT-RF	0,0171	**	0,0173	**	0,0579	*	1,43%
	SMB	0,1701		0,1308		0,2520		0,58%
	HML	0,0226	**	0,0200	**	0,0441	**	1,37%
	MOM	0,0886	*	0,0854	*	0,2192		0,75%
	LOW VOL	0,2702		0,1533		0,0000	***	0,52%
VARX(1,1)	MKT-RF	0,0404	**	0,0000	***	0,0000	***	4,69%
	SMB	0,5998		0,0356	**	0,1012		1,12%
	HML	0,0021	***	0,0000	***	0,0001	***	4,70%
	MOM	0,7575		0,0347		0,0520		1,13%
	LOW VOL	0,0001	***	0,0000	***	0,0000	***	24,94%
VAR(2)	MKT-RF	0,0277	**	0,0000	***	0,0000	***	6,81%
	SMB	0,6827		0,0000	***	0,0000	***	5,67%
	HML	0,1457		0,0000	***	0,0000	***	5,20%
	MOM	0,8554		0,0011	***	0,0014	***	2,67%
	LOW VOL	0,0000	***	0,0000	***	0,0000	***	50,14%
TVARX(2,1,1)	MKT-RF	0,0535	*	0,0000	***	0,0000	***	4,50%
	SMB	0,7631		0,2834		0,5503		0,29%
	HML	0,7095		0,0208		0,0412		1,35%
	MOM	0,1035		0,0000	***	0,0000	***	5,52%
	LOW VOL	0,0002	***	0,0000	***	0,0000	***	28,16%
TVAR(2,2)	MKT-RF	0,0278	**	0,0000	***	0,0000	***	5,05%
	SMB	0,5252		0,0000	***	0,0000	***	20,26%
	HML	0,0029	***	0,0000	***	0,0000	***	10,28%
	MOM	0,0384	**	0,0000	***	0,0000	***	11,09%
	LOW VOL	0,0000	***	0,0000	***	0,0000	***	40,94%
MSI(3)	MKT-RF	0,9554		0,9337		0,9762		0,00%
	SMB	0,2513		0,1427		0,2835		0,54%
	HML	0,8497		0,7378		0,6934		0,03%
	MOM	0,2160		0,2354		0,4625		0,36%
	LOW VOL	0,7520		0,1923		0,0000	***	0,43%
MSIH(3)	MKT-RF	0,1903		0,1925		0,4201		0,43%
	SMB	0,8382		0,9526		0,7797		0,00%
	HML	0,0599	*	0,0853	*	0,1497		0,75%
	MOM	0,9651		0,9771		0,9590		0,00%
	LOW VOL	0,7080		0,0337	**	0,0000	***	1,14%
MSIVARH(3,1)	MKT-RF	0,0500		0,0160		0,0297		1,46%
	SMB	0,2444		0,0000	***	0,0000	***	5,28%
	HML	0,0051	***	0,0000	***	0,0000	***	5,42%
	MOM	0,9696		0,1943		0,0017	***	0,43%
	LOW VOL	0,0015	***	0,0000	***	0,0000	***	6,31%

Table 14 (continued)
Mincer and Zarnowitz regression: Forecast model
See above.

Panel C: 12-month forecasting horizon

Factor	Model	Beta 0		Beta 1		F-test		R-squared
CER	MKT-RF	0,0215	**	0,0218	**	0,0709	*	1,33%
	SMB	0,2939		0,2357		0,3889		0,36%
	HML	0,0010	***	0,0009	***	0,0026	***	2,77%
	MOM	0,1793		0,1739		0,3823		0,47%
	LOW VOL	0,1343		0,0662	*	0,0000	***	0,85%
VARX(1,1)	MKT-RF	0,0299	**	0,2359		0,0325	**	0,36%
	SMB	0,5338		0,0006	***	0,0021	***	2,97%
	HML	0,1198		0,0003	***	0,0008	***	3,32%
	MOM	0,4206		0,2074		0,4511		0,40%
	LOW VOL	0,0000	***	0,0000	***	0,0000	***	39,16%
VAR(2)	MKT-RF	0,0298	**	0,0000	***	0,0000	***	7,46%
	SMB	0,6801		0,0000	***	0,0000	***	5,78%
	HML	0,1013		0,0000	***	0,0000	***	6,16%
	MOM	0,7737		0,0002	***	0,0003	***	3,40%
	LOW VOL	0,0000	***	0,0000	***	0,0000	***	53,76%
TVARX(2,1,1)	MKT-RF	0,0426	**	0,0000	***	0,0000	***	8,35%
	SMB	0,5554		0,0000	***	0,0001	***	4,82%
	HML	0,1785		0,0002		0,0004		3,52%
	MOM	0,1420		0,0159	**	0,0504	*	1,47%
	LOW VOL	0,0000	***	0,0000	***	0,0000	***	38,15%
TVAR(2,2)	MKT-RF	0,0817	*	0,0000	***	0,0000	***	18,22%
	SMB	0,0513	*	0,0001	***	0,0000	***	4,09%
	HML	0,0480	**	0,0000	***	0,0000	***	4,63%
	MOM	0,4172		0,0125	**	0,0187	**	1,57%
	LOW VOL	0,0000	***	0,0000	***	0,0000	***	51,20%
MSI(3)	MKT-RF	0,0270	**	0,0275	**	0,0864	*	1,23%
	SMB	0,2948		0,2371		0,3970		0,35%
	HML	0,0011	***	0,0010	***	0,0029	***	2,73%
	MOM	0,1640		0,1587		0,3537		0,50%
	LOW VOL	0,1959		0,1049		0,0000	***	0,67%
MSIH(3)	MKT-RF	0,7747		0,7930		0,9501		0,02%
	SMB	0,0855	*	0,0538	*	0,1252		0,94%
	HML	0,2226		0,1967		0,2956		0,42%
	MOM	0,2593		0,2443		0,4858		0,34%
	LOW VOL	0,1410		0,0638	*	0,0000	***	0,87%
MSIVARH(3,1)	MKT-RF	0,1963		0,4405		0,4331		0,15%
	SMB	0,9735		0,1701		0,0965	*	0,48%
	HML	0,0082	***	0,0001	***	0,0004	***	3,70%
	MOM	0,3529		0,0312	**	0,0002	***	1,17%
	LOW VOL	0,0006	***	0,0000	***	0,0000	***	6,08%

Table 15
Diebold-Mariano test of equal predictive accuracy: Excess market returns

The table presents *p*-values for Diebold and Mariano's (1995, DM) test. Null hypothesis: no difference in forecast accuracy. *p*-values in bold are below the 5% threshold. The cells below the main diagonal contain *p*-values based on the squared loss function, while the cells below show *p*-values based on the absolute loss function. Each panel focuses on a different forecasting horizon.

Panel A: 1-month forecast horizon

	CER	VARX(1,1)	VAR(2)	TVARX(2,1,1)	TVAR(2,2)	MSI(3)	MSIH(3)	MSIVARH(3,1)
CER		**0,0097**	**0,0007**	**0,0040**	**0,0000**	0,7301	0,7380	**0,0288**
VARX(1,1)	0,2191		0,0599	0,1504	**0,0012**	**0,0120**	**0,0244**	0,3361
VAR(2)	**0,0029**	**0,0072**		0,7180	**0,0059**	**0,0011**	**0,0013**	0,9268
TVARX(2,1,1)	**0,0297**	0,0548	0,5666		**0,0089**	**0,0046**	**0,0084**	0,8649
TVAR(2,2)	**0,0025**	**0,0058**	**0,0142**	**0,0129**		**0,0000**	**0,0001**	**0,0291**
MSI(3)	0,7933	0,2003	**0,0027**	**0,0211**	**0,0025**		0,9248	**0,0389**
MSIH(3)	0,7560	0,2671	**0,0041**	**0,0347**	**0,0034**	0,7163		**0,0250**
MSIVARH(3,1)	0,0716	0,1384	0,8578	0,5939	**0,0225**	0,0770	0,0592	

Panel B: 3-month forecast horizon

	CER	VARX(1,1)	VAR(2)	TVARX(2,1,1)	TVAR(2,2)	MSI(3)	MSIH(3)	MSIVARH(3,1)
CER		**0,0004**	**0,0029**	**0,0012**	**0,0099**	0,7195	0,4181	**0,0252**
VARX(1,1)	**0,0267**		0,9033	0,9464	0,6296	**0,0004**	**0,0009**	**0,0202**
VAR(2)	**0,0051**	0,4713		0,9260	0,7176	**0,0025**	**0,0037**	**0,0279**
TVARX(2,1,1)	**0,0412**	0,8649	0,4221		0,6125	**0,0010**	**0,0022**	**0,0222**
TVAR(2,2)	0,0682	0,5137	0,8398	0,4529		**0,0092**	**0,0139**	0,0892
MSI(3)	0,1151	**0,0218**	**0,0038**	**0,0334**	0,0607		0,4107	**0,0225**
MSIH(3)	0,6829	**0,0319**	**0,0060**	**0,0483**	0,0735	0,3531		**0,0382**
MSIVARH(3,1)	0,2377	0,1136	**0,0112**	0,1395	0,1337	0,1891	0,2621	

Panel C: 12-month forecast horizon

	CER	VARX(1,1)	VAR(2)	TVARX(2,1,1)	TVAR(2,2)	MSI(3)	MSIH(3)	MSIVARH(3,1)
CER		0,0745	**0,0064**	**0,0010**	**0,0001**	0,0658	0,4914	0,2706
VARX(1,1)	0,6342		**0,0215**	0,1158	**0,0002**	0,0756	0,0635	**0,0361**
VAR(2)	**0,0050**	**0,0022**		0,7274	**0,0004**	**0,0064**	**0,0049**	**0,0032**
TVARX(2,1,1)	**0,0196**	0,0576	0,5893		**0,0011**	**0,0010**	**0,0008**	**0,0071**
TVAR(2,2)	**0,0029**	**0,0019**	**0,0055**	**0,0182**		**0,0001**	**0,0001**	**0,0001**
MSI(3)	0,8396	0,6331	**0,0050**	**0,0196**	**0,0029**		0,3917	0,2773
MSIH(3)	0,3652	0,5852	**0,0038**	**0,0183**	**0,0028**	0,3610		0,2281
MSIVARH(3,1)	0,8684	0,5008	**0,0010**	**0,0300**	**0,0022**	0,8666	0,7632	

Table 16
Diebold-Mariano test of equal predictive accuracy: Size factor returns

The table presents *p*-values for Diebold and Mariano's (1995, DM) test comparing forecasts of Size factor returns. Null hypothesis: no difference in forecast accuracy. *p*-values in bold are below the 5% threshold. The cells below the main diagonal contain *p*-values based on the squared loss function, while the cells below show *p*-values based on the absolute loss function. Each panel focuses on a different forecasting horizon.

Panel A: 1-month forecast horizon

	CER	VARX(1,1)	VAR(2)	TVARX(2,1,1)	TVAR(2,2)	MSI(3)	MSIH(3)	MSIVARH(3,1)
CER		0,2432	0,7862	0,2043	**0,0000**	0,2501	0,2463	0,0507
VARX(1,1)	0,9161		**0,0480**	0,5200	**0,0000**	0,4835	0,0831	**0,0015**
VAR(2)	0,2623	0,1166		0,0671	**0,0000**	0,5303	0,8189	0,0557
TVARX(2,1,1)	0,5634	0,4797	0,9000		**0,0000**	0,3552	0,0659	**0,0005**
TVAR(2,2)	**0,0000**	**0,0000**	**0,0000**	**0,0011**		**0,0000**	**0,0000**	**0,0000**
MSI(3)	0,5196	0,9405	0,2088	0,5334	**0,0000**		**0,0076**	**0,0248**
MSIH(3)	0,1162	0,1050	0,9942	0,8689	**0,0000**	0,0812		0,0774
MSIVARH(3,1)	0,0515	**0,0302**	0,1029	**0,0044**	0,2023	**0,0477**	0,0528	

Panel B: 3-month forecast horizon

	CER	VARX(1,1)	VAR(2)	TVARX(2,1,1)	TVAR(2,2)	MSI(3)	MSIH(3)	MSIVARH(3,1)
CER		0,9553	0,8851	0,6553	**0,0002**	0,9192	0,1558	**0,0028**
VARX(1,1)	0,7877		0,8552	0,5133	**0,0001**	0,9478	0,9084	**0,0272**
VAR(2)	0,4049	0,4431		0,6661	**0,0001**	0,8893	0,8111	0,1436
TVARX(2,1,1)	0,5453	0,3995	0,3852		**0,0000**	0,6476	0,7549	0,0669
TVAR(2,2)	**0,0004**	**0,0001**	**0,0112**	**0,0000**		**0,0002**	**0,0002**	**0,0044**
MSI(3)	0,5900	0,8301	0,4184	0,5286	**0,0005**		0,1935	**0,0031**
MSIH(3)	0,3414	0,7143	0,3779	0,5905	**0,0005**	0,2213		**0,0014**
MSIVARH(3,1)	**0,0404**	0,0632	0,8062	0,2188	**0,0111**	**0,0454**	**0,0328**	

Panel C: 12-month forecast horizon

	CER	VARX(1,1)	VAR(2)	TVARX(2,1,1)	TVAR(2,2)	MSI(3)	MSIH(3)	MSIVARH(3,1)
CER		0,3654	0,7138	0,3709	0,0727	0,2900	0,3760	0,1539
VARX(1,1)	**0,0031**		0,9141	0,6587	0,0761	0,3602	0,3920	0,8754
VAR(2)	0,4725	0,6698		0,7129	0,2223	0,7103	0,7346	0,8389
TVARX(2,1,1)	0,0883	0,2360	0,9297		0,1774	0,3673	0,3934	0,9312
TVAR(2,2)	**0,0298**	**0,0328**	0,0548	0,0638		0,0719	0,0740	0,2169
MSI(3)	0,4361	**0,0030**	0,4715	0,0871	**0,0296**		0,2926	0,1524
MSIH(3)	0,1585	**0,0113**	0,4849	0,0968	**0,0303**	0,1140		0,1750
MSIVARH(3,1)	0,3245	0,3636	0,5718	0,2587	0,0522	0,3227	0,3981	

Table 17
Diebold-Mariano test of equal predictive accuracy: Value factor returns

The table presents *p*-values for Diebold and Mariano's (1995, DM) test comparing forecasts of Value factor returns. Null hypothesis: no difference in forecast accuracy. *p*-values in bold are below the 5% threshold. The cells below the main diagonal contain *p*-values based on the squared loss function, while the cells below show *p*-values based on the absolute loss function. Each panel focuses on a different forecasting horizon.

Panel A: 1-month forecast horizon

	CER	VARX(1,1)	VAR(2)	TVARX(2,1,1)	TVAR(2,2)	MSI(3)	MSIH(3)	MSIVARH(3,1)
CER		0,2876	0,3665	0,7451	**0,0003**	0,8794	**0,0420**	0,0679
VARX(1,1)	0,4271		0,9999	0,4454	**0,0002**	0,3320	**0,0253**	0,2569
VAR(2)	0,3937	0,8652		0,5187	**0,0001**	0,3898	**0,0491**	0,3148
TVARX(2,1,1)	0,6240	0,9838	0,9575		**0,0004**	0,7066	0,0947	0,1192
TVAR(2,2)	**0,0021**	**0,0028**	**0,0031**	**0,0134**		**0,0004**	**0,0000**	**0,0252**
MSI(3)	0,5817	0,4184	0,3845	0,5845	**0,0032**		**0,0081**	0,0802
MSIH(3)	0,0829	0,1275	0,1229	0,2912	**0,0014**	0,0535		**0,0057**
MSIVARH(3,1)	0,1943	0,2678	0,3981	0,3534	**0,0295**	0,2201	0,0803	

Panel B: 3-month forecast horizon

	CER	VARX(1,1)	VAR(2)	TVARX(2,1,1)	TVAR(2,2)	MSI(3)	MSIH(3)	MSIVARH(3,1)
CER		0,3416	0,4169	0,5992	**0,0012**	0,2094	**0,0222**	**0,0185**
VARX(1,1)	0,1201		0,7182	0,2371	**0,0109**	0,2496	0,1268	0,1283
VAR(2)	0,2195	0,9212		0,2844	0,1690	0,3620	0,2374	0,4062
TVARX(2,1,1)	0,5830	0,2061	0,1657		**0,0017**	0,7244	0,9193	**0,0358**
TVAR(2,2)	**0,0055**	0,0664	0,2256	**0,0022**		**0,0010**	**0,0003**	0,5886
MSI(3)	0,5073	0,0899	0,1817	0,6385	**0,0053**		0,1095	**0,0111**
MSIH(3)	0,1198	0,0934	0,1226	0,8044	**0,0024**	0,2972		**0,0080**
MSIVARH(3,1)	0,0722	0,6233	0,7952	0,1363	0,2240	0,0558	**0,0468**	

Panel C: 12-month forecast horizon

	CER	VARX(1,1)	VAR(2)	TVARX(2,1,1)	TVAR(2,2)	MSI(3)	MSIH(3)	MSIVARH(3,1)
CER		0,4506	0,1266	0,4405	**0,0004**	0,1079	0,3539	0,1861
VARX(1,1)	0,4671		0,5650	0,7823	0,0738	0,4391	0,4202	0,5278
VAR(2)	0,1312	0,5203		0,4823	0,2054	0,1209	0,1116	0,9291
TVARX(2,1,1)	0,4449	0,8891	0,2685		**0,0282**	0,4268	0,4041	0,4131
TVAR(2,2)	**0,0009**	0,2885	0,5014	0,2779		**0,0004**	**0,0004**	0,1402
MSI(3)	0,4358	0,4606	0,1272	0,4387	**0,0010**		0,5012	0,1791
MSIH(3)	0,1994	0,4479	0,1191	0,4250	**0,0008**	0,3211		0,1691
MSIVARH(3,1)	0,2693	0,9089	0,5589	0,9503	0,2474	0,2612	0,2530	

Table 18

Diebold-Mariano test of equal predictive accuracy: Momentum factor returns

The table presents *p*-values for Diebold and Mariano's (1995, DM) test comparing forecasts of Momentum factor returns. Null hypothesis: no difference in forecast accuracy. *p*-values in bold are below the 5% threshold. The cells below the main diagonal contain *p*-values based on the squared loss function, while the cells below show *p*-values based on the absolute loss function. Each panel focuses on a different forecasting horizon.

Panel A: 1-month forecast horizon

	CER	VARX(1,1)	VAR(2)	TVARX(2,1,1)	TVAR(2,2)	MSI(3)	MSIH(3)	MSIVARH(3,1)
CER		0,9701	0,9041	0,3991	0,1153	0,5068	0,4459	0,1208
VARX(1,1)	0,1961		0,6795	0,1254	0,1169	0,4852	0,4065	**0,0367**
VAR(2)	0,3678	0,5474		0,2071	0,1243	0,3799	0,3276	0,0782
TVARX(2,1,1)	**0,0221**	**0,0288**	**0,0377**		0,1842	0,1070	0,0914	0,3983
TVAR(2,2)	**0,0092**	**0,0362**	**0,0240**	0,1683		0,0875	0,0922	0,2613
MSI(3)	0,7925	0,1639	0,3311	**0,0183**	**0,0080**		0,7009	**0,0425**
MSIH(3)	0,6520	0,1410	0,2854	**0,0163**	**0,0111**	0,7197		**0,0409**
MSIVARH(3,1)	**0,0307**	0,2259	0,1767	0,9797	0,1821	**0,0281**	**0,0274**	

Panel B: 3-month forecast horizon

	CER	VARX(1,1)	VAR(2)	TVARX(2,1,1)	TVAR(2,2)	MSI(3)	MSIH(3)	MSIVARH(3,1)
CER		0,9848	0,9734	0,4042	0,0677	0,1043	0,5959	0,1120
VARX(1,1)	**0,0195**		0,9779	0,3208	0,0779	0,8955	0,9465	0,5325
VAR(2)	0,4607	0,4035		0,3696	0,0827	0,8937	0,9400	0,5730
TVARX(2,1,1)	**0,0080**	0,2728	0,1442		0,1916	0,3711	0,3781	0,6636
TVAR(2,2)	**0,0079**	0,1786	0,1002	0,5012		0,0590	0,0622	0,1076
MSI(3)	**0,0199**	**0,0108**	0,3692	**0,0048**	**0,0049**		0,5263	0,0587
MSIH(3)	0,8138	**0,0186**	0,4734	**0,0079**	**0,0089**	0,0694		0,1066
MSIVARH(3,1)	0,4518	0,1737	0,7796	0,0667	**0,0157**	0,2942	0,4893	

Panel C: 12-month forecast horizon

	CER	VARX(1,1)	VAR(2)	TVARX(2,1,1)	TVAR(2,2)	MSI(3)	MSIH(3)	MSIVARH(3,1)
CER		0,9315	0,8194	0,3130	0,0578	0,7384	0,7027	**0,0145**
VARX(1,1)	0,1779		0,7996	0,2158	0,0526	0,9265	0,8737	0,0607
VAR(2)	0,4234	0,6907		0,9908	0,9481	0,8203	0,8259	0,6578
TVARX(2,1,1)	**0,0390**	0,0935	0,9813		0,7903	0,3175	0,3504	0,2551
TVAR(2,2)	**0,0304**	0,1521	0,9928	0,9673		0,0577	0,0505	0,1219
MSI(3)	0,4088	0,1751	0,4220	**0,0385**	**0,0297**		0,7303	**0,0140**
MSIH(3)	0,3534	0,2087	0,4381	**0,0481**	**0,0358**	0,3007		**0,0162**
MSIVARH(3,1)	0,3453	0,9048	0,7575	0,6028	0,4905	0,3411	0,3657	

Table 19
Diebold-Mariano test of equal predictive accuracy: Low Volatility factor returns

The table presents *p*-values for Diebold and Mariano's (1995, DM) test comparing forecasts of Low Volatility factor returns. Null hypothesis: no difference in forecast accuracy. *p*-values in bold are below the 5% threshold. The cells below the main diagonal contain *p*-values based on the squared loss function, while the cells below show *p*-values based on the absolute loss function. Each panel focuses on a different forecasting horizon.

Panel A: 1-month forecast horizon

	CER	VARX(1,1)	VAR(2)	TVARX(2,1,1)	TVAR(2,2)	MSI(3)	MSIH(3)	MSIVARH(3,1)
CER		**0,0000**	**0,0000**	**0,0000**	**0,0000**	0,6037	**0,0058**	0,2293
VARX(1,1)	**0,0000**		**0,0074**	0,6719	**0,0004**	**0,0000**	0,7168	**0,0029**
VAR(2)	**0,0000**	0,0701		**0,0049**	0,0957	**0,0000**	0,3812	**0,0002**
TVARX(2,1,1)	**0,0000**	0,2307	**0,0166**		**0,0002**	**0,0000**	0,6051	**0,0128**
TVAR(2,2)	**0,0000**	**0,0010**	**0,0167**	**0,0001**		**0,0000**	0,1740	**0,0000**
MSI(3)	**0,0012**	**0,0000**	**0,0000**	**0,0000**	**0,0000**		**0,0012**	0,1627
MSIH(3)	0,1286	**0,0087**	**0,0013**	**0,0325**	**0,0001**	**0,0009**		**0,0319**
MSIVARH(3,1)	0,3831	**0,0000**	**0,0000**	**0,0000**	**0,0000**	0,2658	**0,0311**	

Panel B: 3-month forecast horizon

	CER	VARX(1,1)	VAR(2)	TVARX(2,1,1)	TVAR(2,2)	MSI(3)	MSIH(3)	MSIVARH(3,1)
CER		**0,0174**	**0,0001**	**0,0115**	**0,0001**	0,2007	0,8902	0,7534
VARX(1,1)	0,1009		**0,0001**	0,3749	**0,0053**	**0,0133**	**0,0164**	**0,0023**
VAR(2)	**0,0001**	**0,0000**		**0,0003**	0,1506	**0,0001**	**0,0001**	**0,0000**
TVARX(2,1,1)	**0,0486**	0,4463	**0,0003**		0,0568	**0,0089**	**0,0063**	**0,0044**
TVAR(2,2)	**0,0001**	**0,0002**	0,5757	**0,0066**		**0,0001**	**0,0001**	**0,0000**
MSI(3)	0,0631	0,0586	**0,0000**	**0,0243**	**0,0001**		0,5647	0,9296
MSIH(3)	0,6517	0,1220	**0,0000**	**0,0449**	**0,0002**	0,1668		0,6895
MSIVARH(3,1)	0,2412	**0,0037**	**0,0000**	**0,0026**	**0,0000**	0,4934	0,1427	

Panel C: 12-month forecast horizon

	CER	VARX(1,1)	VAR(2)	TVARX(2,1,1)	TVAR(2,2)	MSI(3)	MSIH(3)	MSIVARH(3,1)
CER		**0,0408**	**0,0194**	**0,0434**	**0,0059**	**0,0000**	0,3247	0,6876
VARX(1,1)	**0,0479**		0,1147	0,5460	0,2331	**0,0394**	**0,0409**	**0,0183**
VAR(2)	**0,0095**	0,2140		0,1225	0,9198	**0,0190**	**0,0192**	**0,0125**
TVARX(2,1,1)	**0,0406**	0,7943	0,2175		0,2055	**0,0418**	**0,0435**	**0,0196**
TVAR(2,2)	**0,0042**	0,2158	0,6202	0,2007		**0,0057**	**0,0058**	**0,0043**
MSI(3)	**0,0000**	**0,0447**	**0,0089**	**0,0377**	**0,0038**		0,3939	0,7650
MSIH(3)	0,2772	**0,0468**	**0,0092**	**0,0397**	**0,0039**	0,1942		0,7307
MSIVARH(3,1)	0,4497	**0,0067**	**0,0015**	**0,0047**	**0,0009**	0,5303	0,4861	

Figure 1

Empirical distribution of the factor portfolios (full sample)

The top graph is the histogram of the monthly returns of the five factor portfolios over the entire data set (1929:01 – 2012:12). Returns are annualized, continuously compounded. The histogram is plotted against a normal distribution, in red, having mean and volatility equal to the observation. The bottom graph presents the QQ plot on the same data sample (1929:01 – 2012:12). The quantiles of the factors are plotted against the quantiles of the normal distribution, in red. Points above or below the red line represent deviations from the normal distribution.

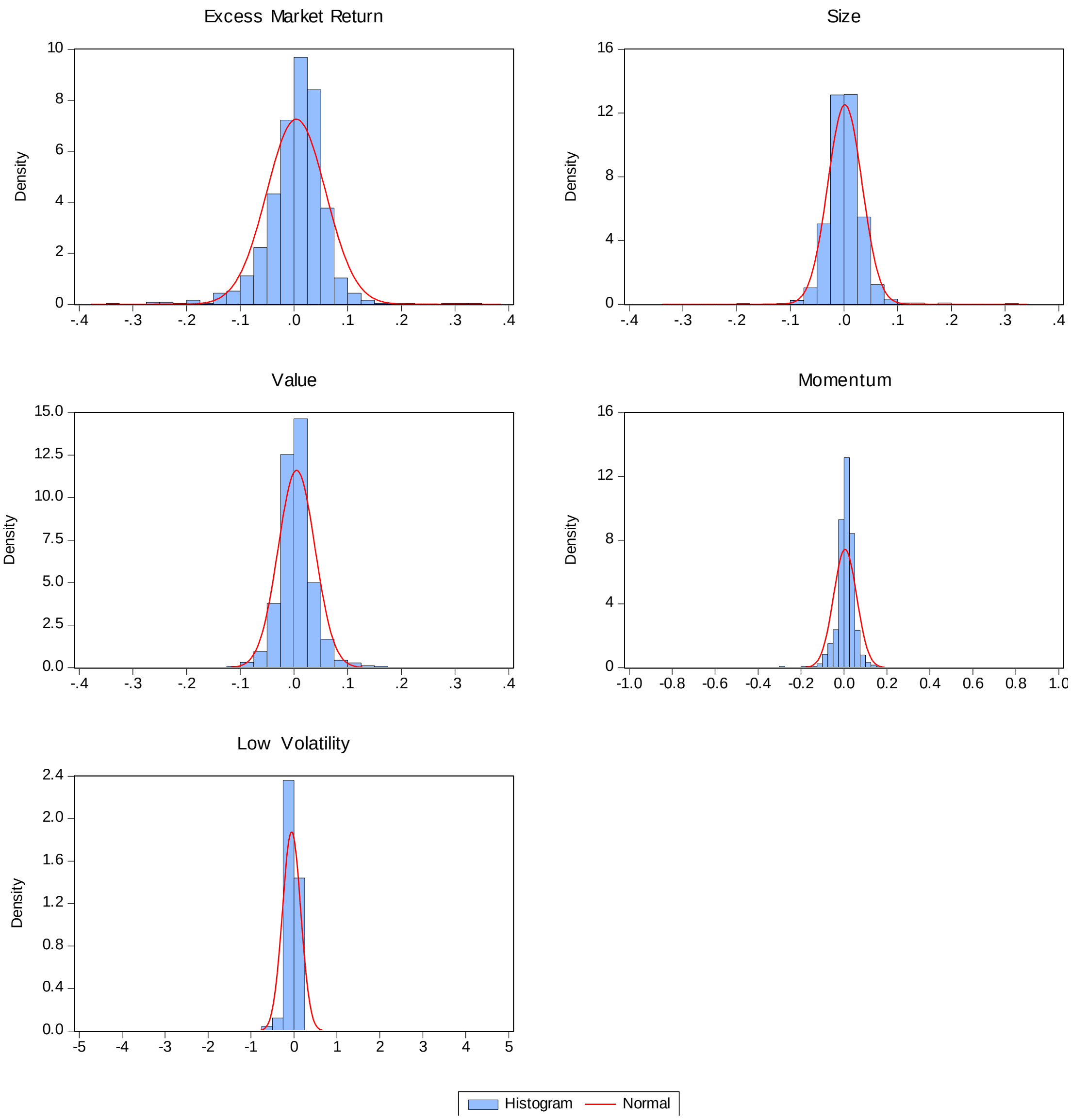

Figure 1 (Continued)
Empirical distribution of the factor portfolios (full sample)

See description above.

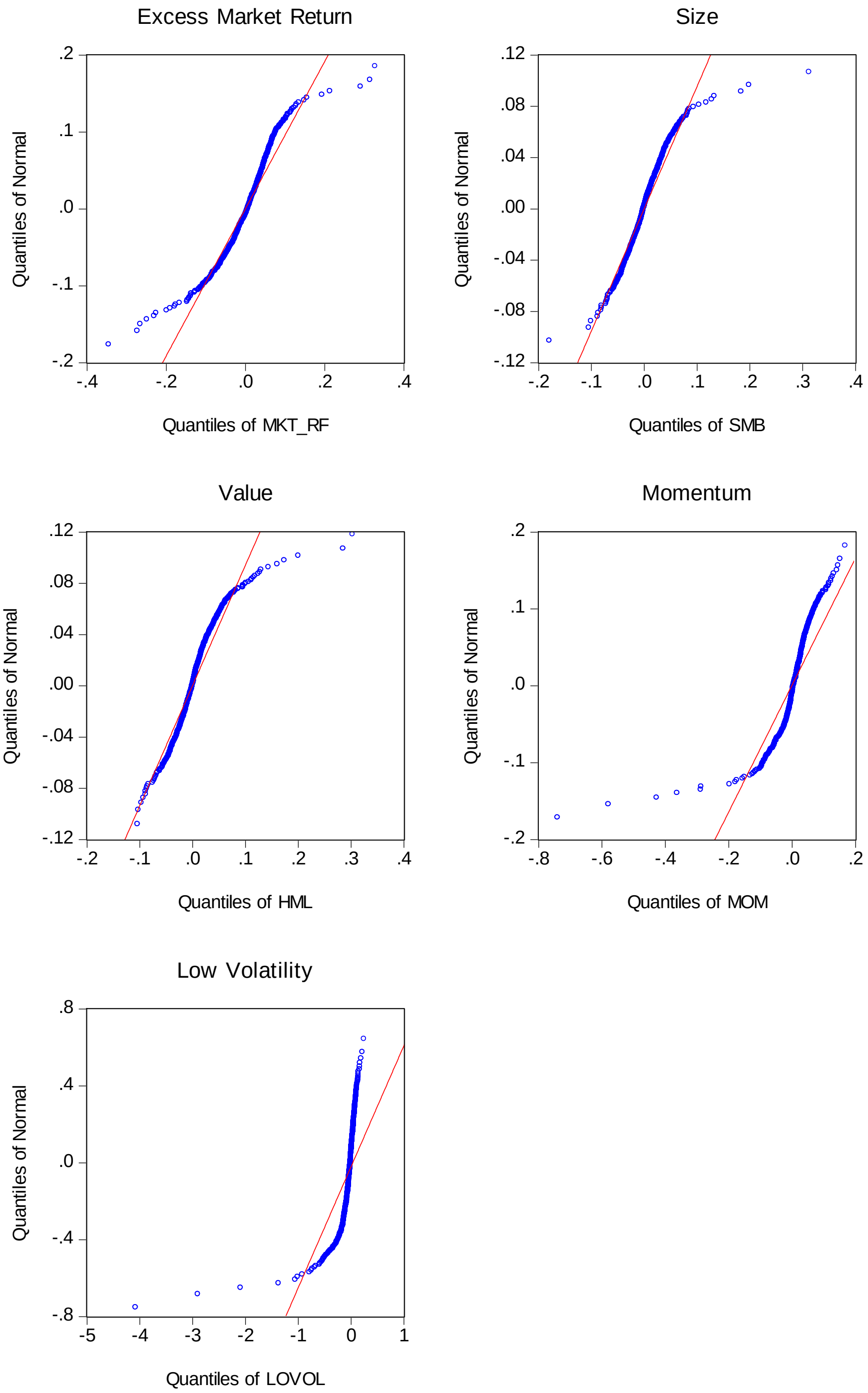

Figure 2

Empirical distribution of the factor portfolios (forecasting window)

Similarly to Figure 1, the top graph is the histogram of the monthly returns of the five factor portfolios over the entire data set (1980:01 – 2012:12). Returns are annualized, continuously compounded. The histogram is plotted against a normal distribution, in red, having mean and volatility equal to the observation. The bottom graph presents the QQ plot on the same data sample (1980:01 – 2012:12). The quantiles of the factors are plotted against the quantiles of the normal distribution, in red. Points above or below the red line represent deviations from the normal distribution.

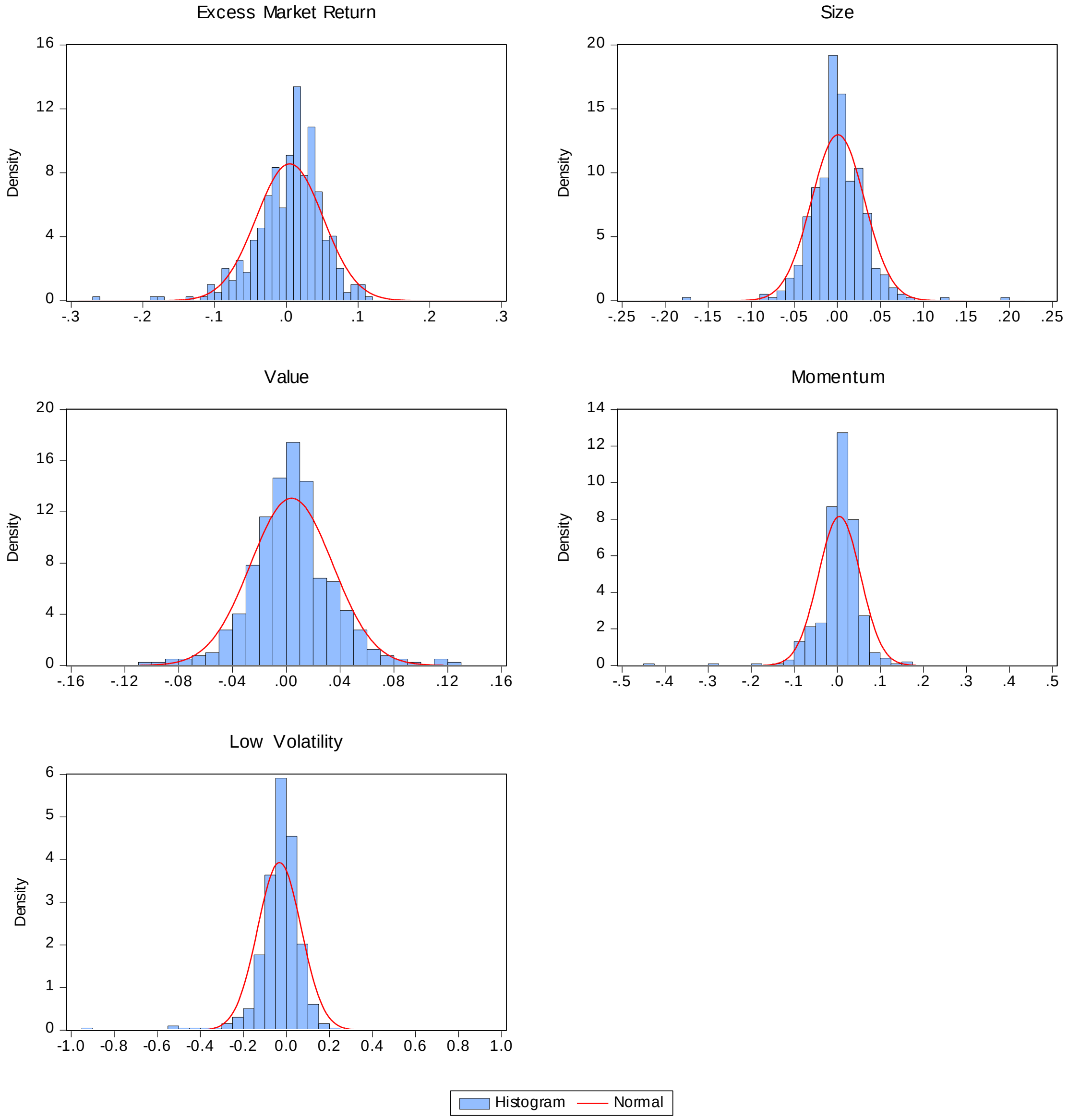

Figure 2 (Continued)
Empirical distribution of the factor portfolios (forecasting window)

See description above.

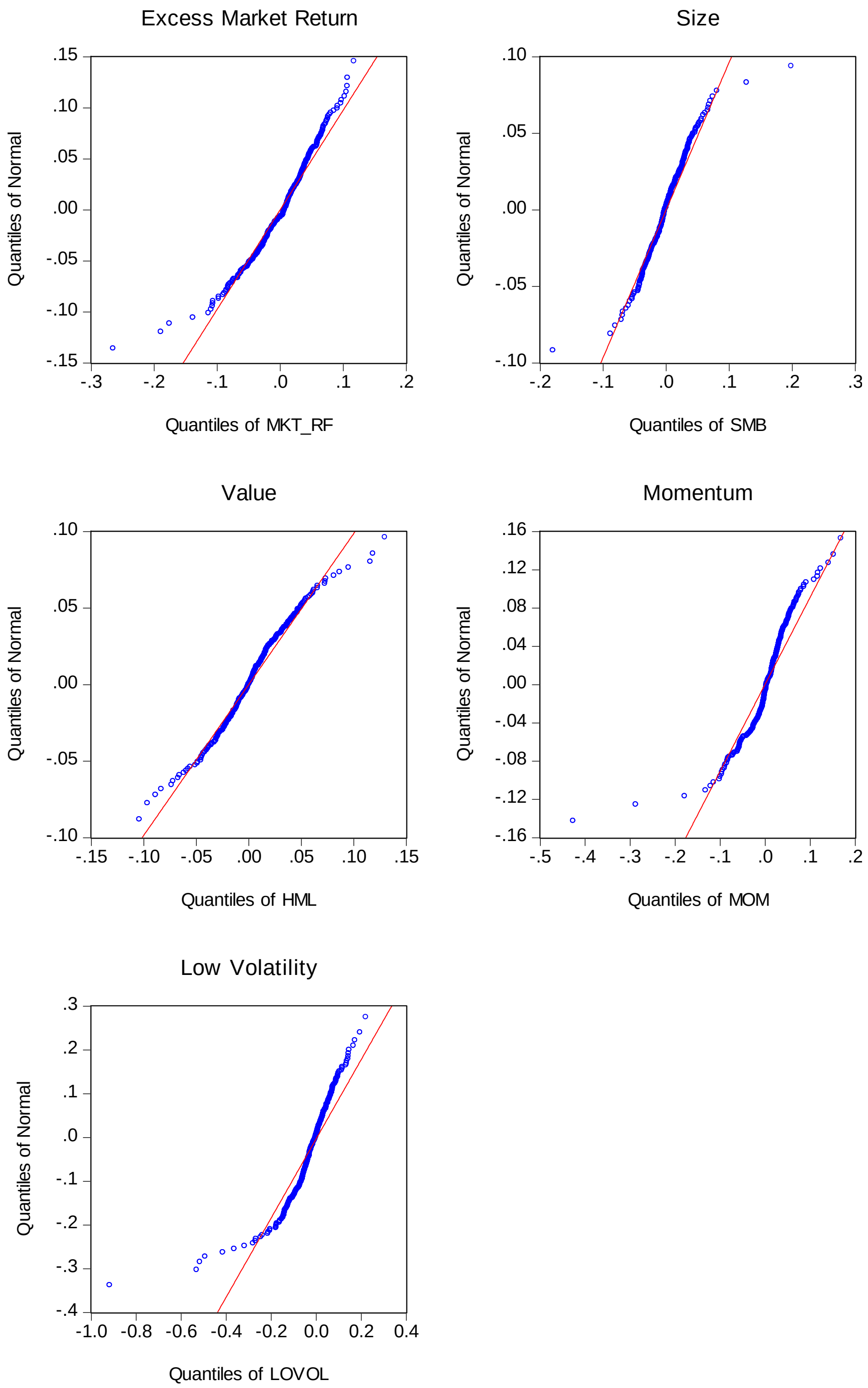

Figure 3
CER model: distribution of the predicted errors

The following two charts show the distribution of the errors predicted by the CER model. Each row covers a different factor portfolio *j*, while each column displays a different forecast horizon *h*. In the top graph, dots represent predicted errors, while the red, horizontal line has 0 as intercept. Optimal forecast should, therefore, be randomly distributed around the red line. The bottom graph, instead, is the histogram of the predicted errors. The distribution in red is Normal, with mean 0 and standard deviation equal to the one of the forecast errors.

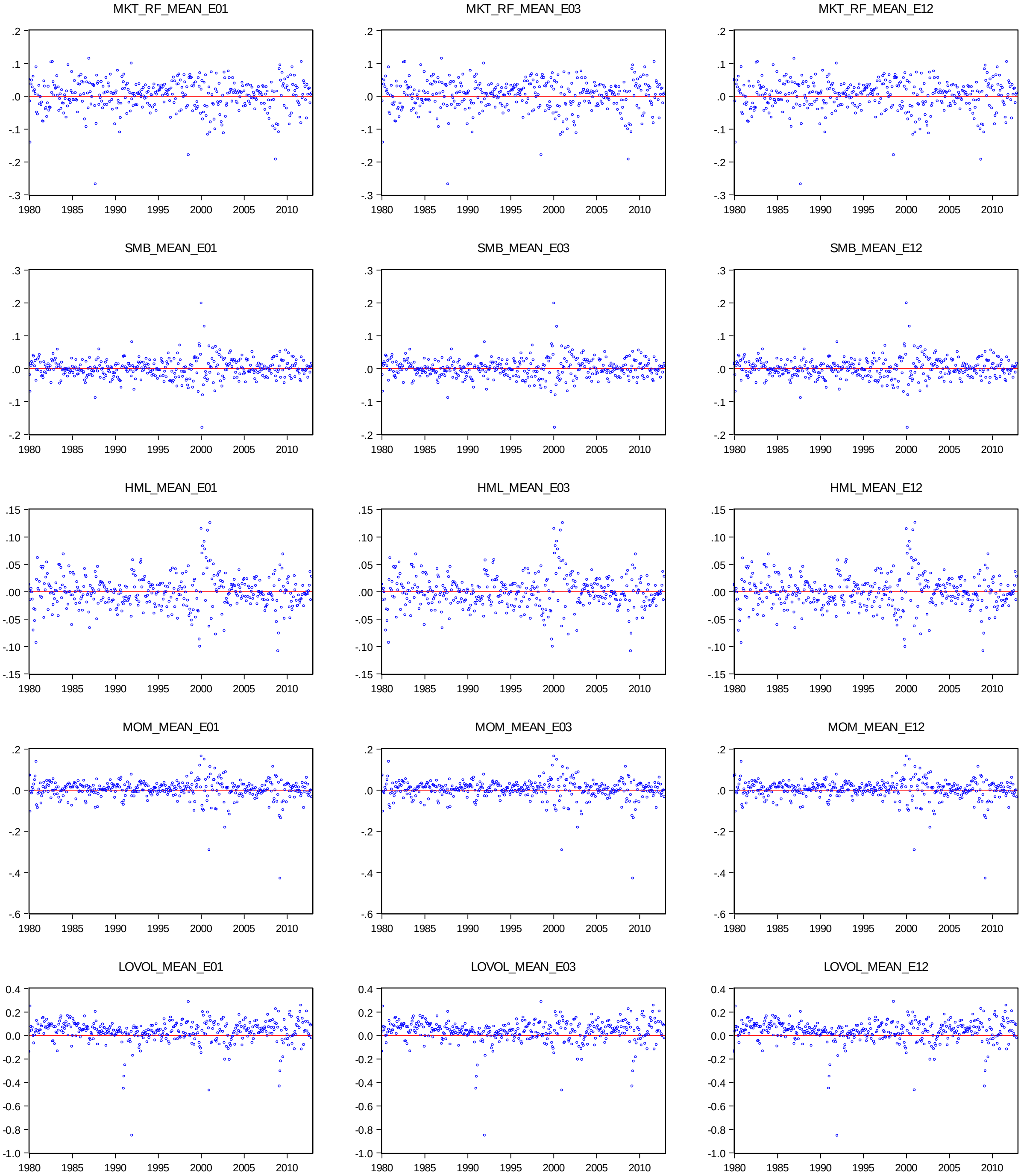

Figure 3 (Continued)

CER model: distribution of the predicted errors

See notes above.

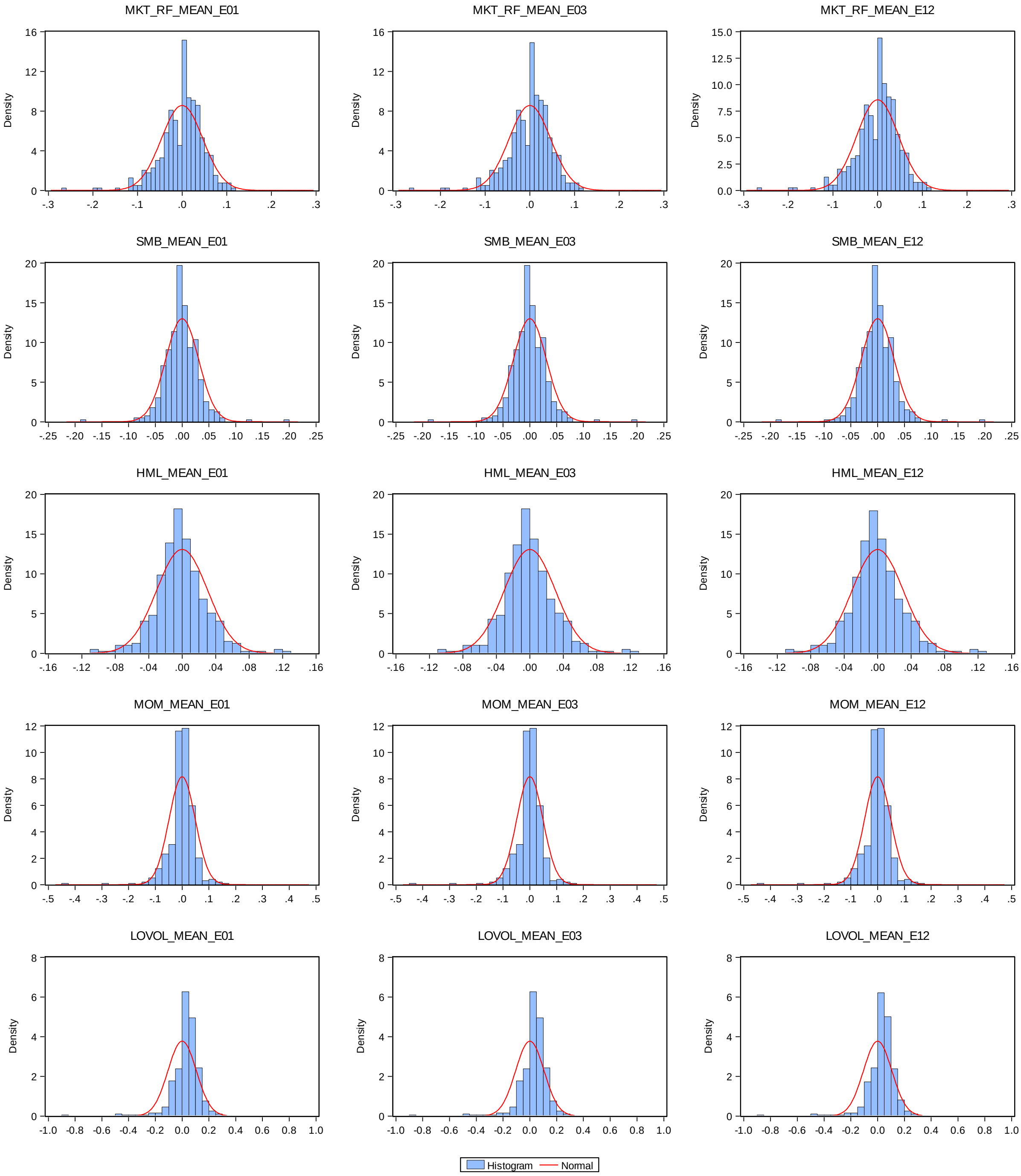

Figure 4
VARX(1,1) model: distribution of the predicted errors

See notes in Figure 3.

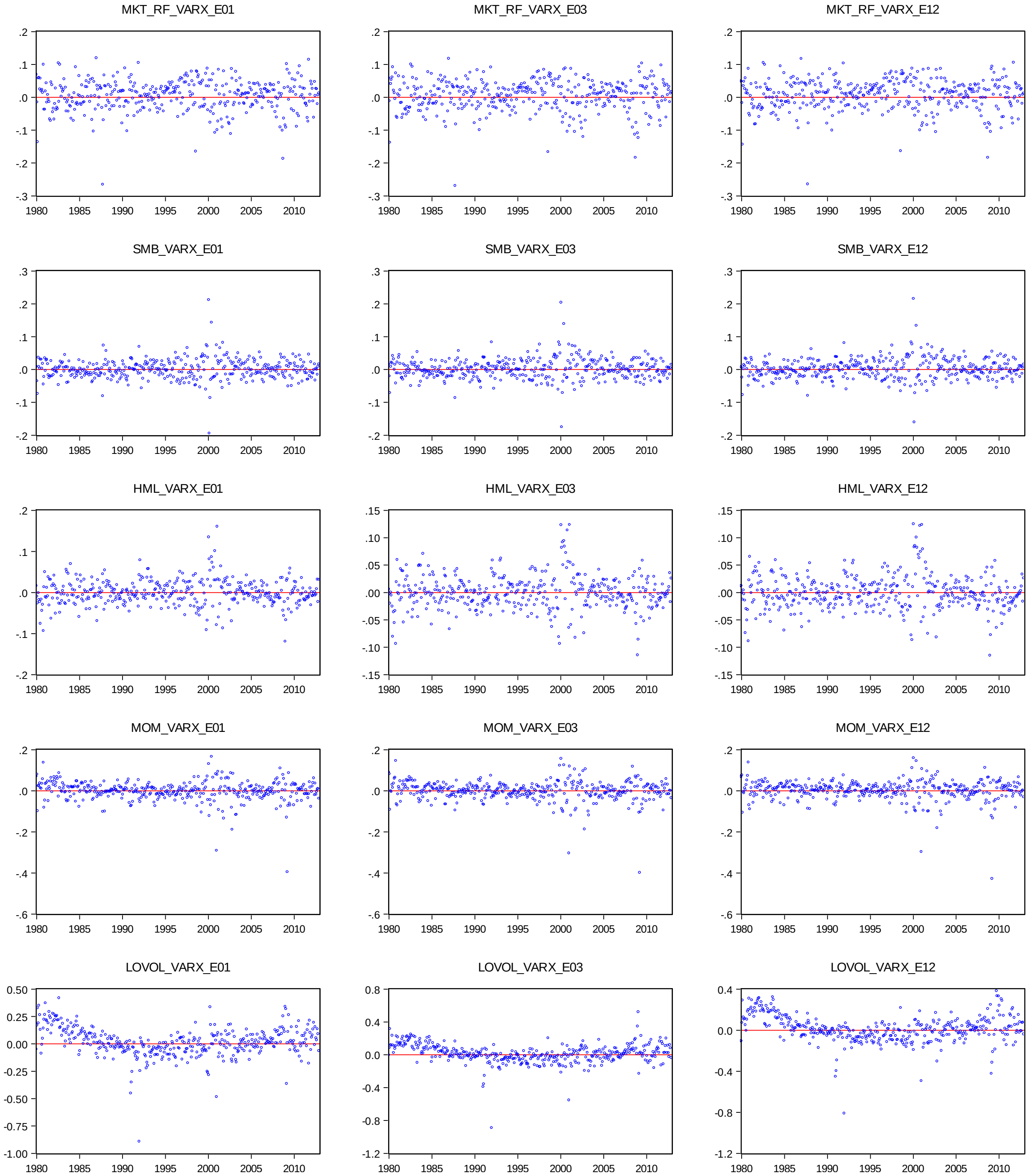

Figure 4 (Continued)
VARX(1,1) model: distribution of the predicted errors

See notes in Figure 3.

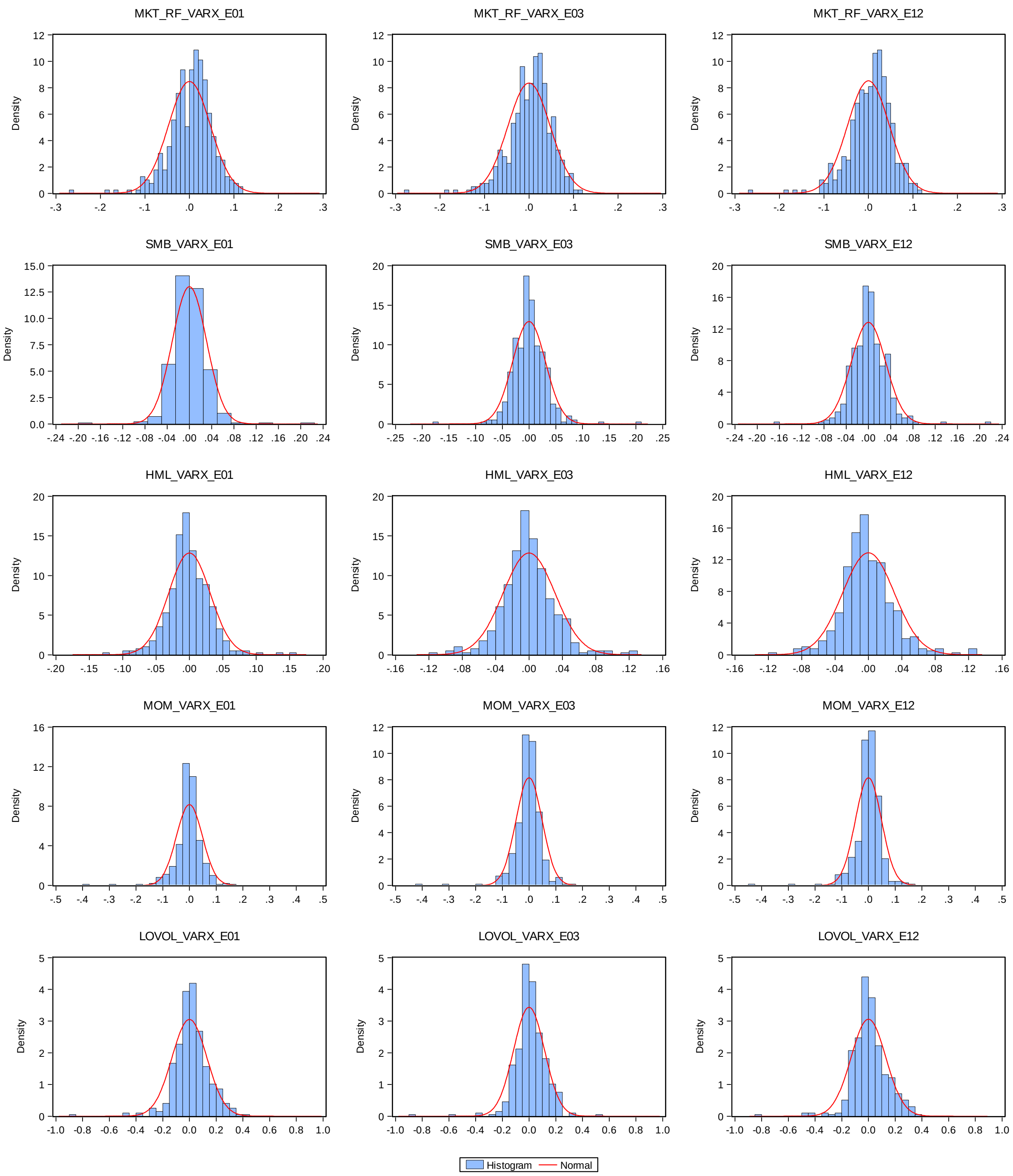

Figure 5
VAR(2) model: distribution of the predicted errors

See notes in Figure 3.

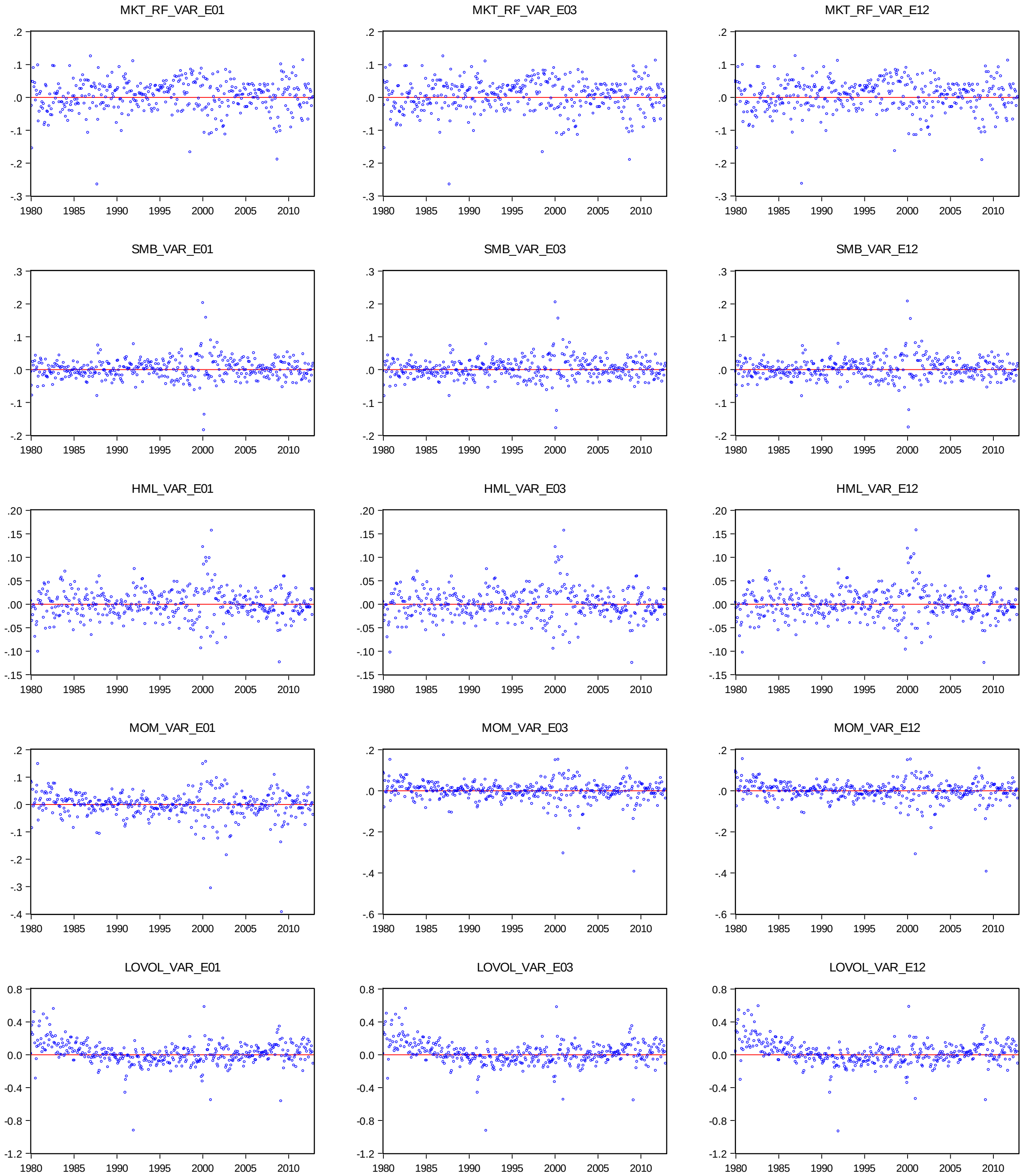

Figure 5 (Continued)

VAR(2) model: distribution of the predicted errors

See notes in Figure 3.

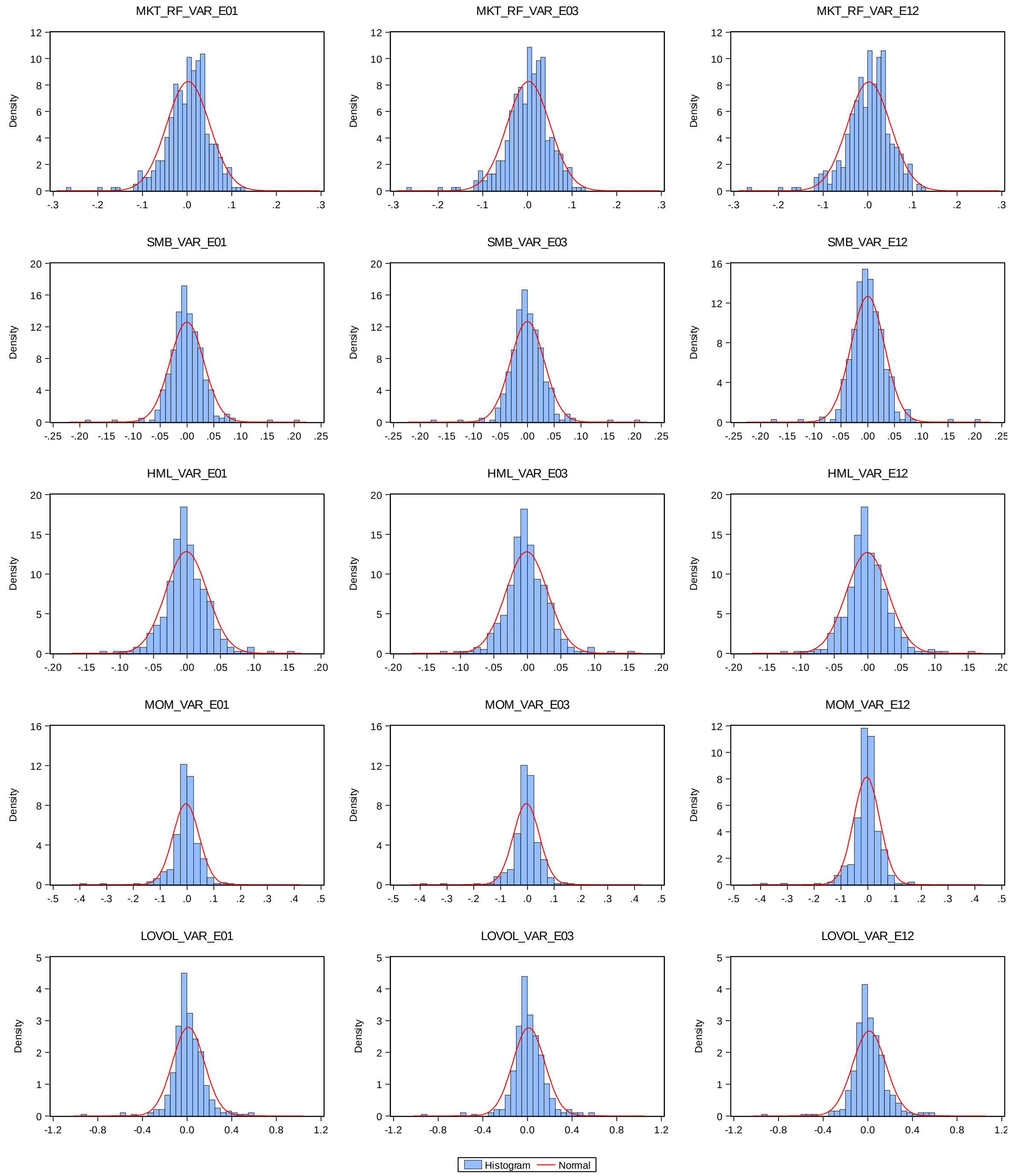

Figure 6

TVARX(2,1,1) model: distribution of the predicted errors

See notes in Figure 3.

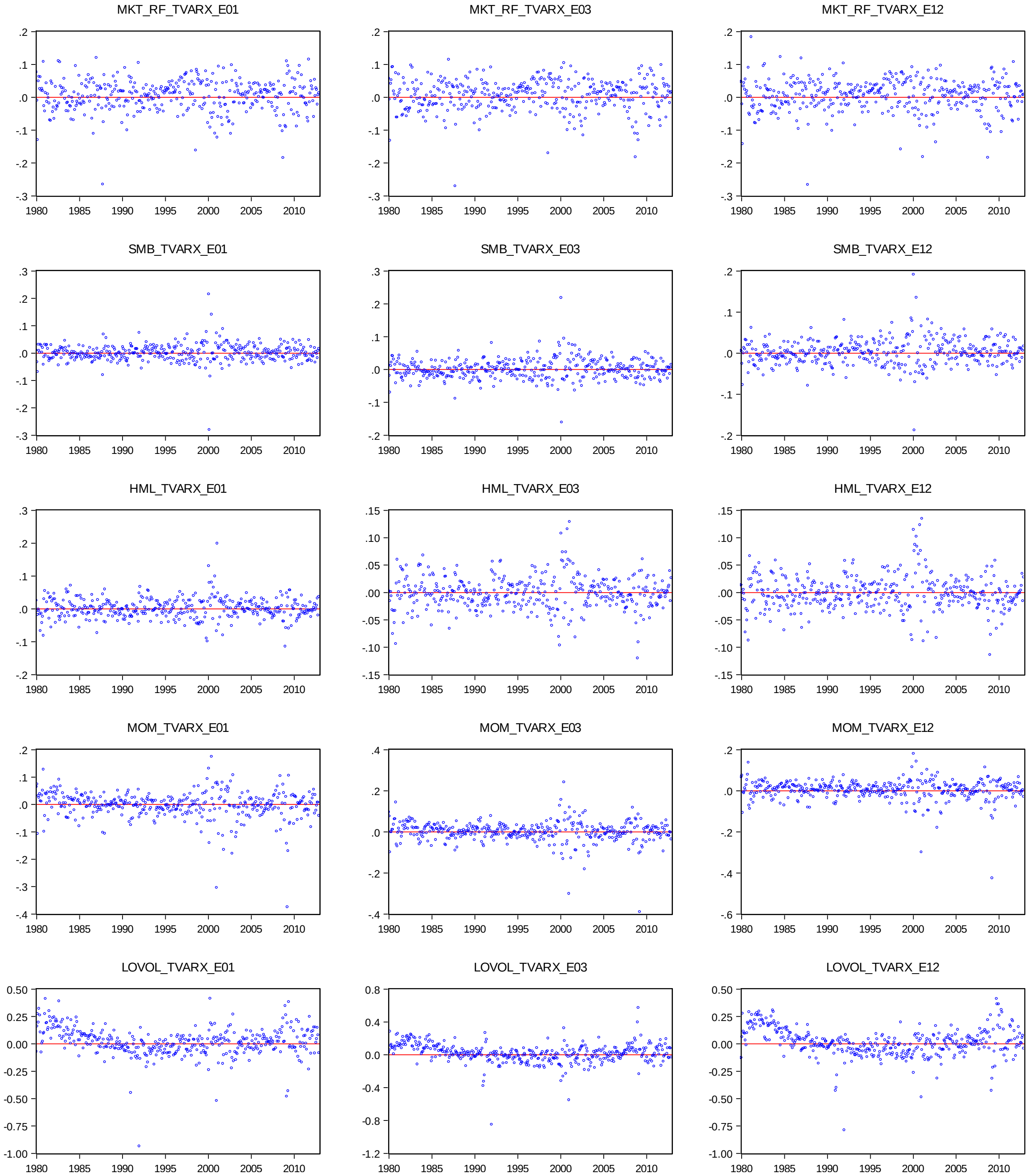

Figure 6 (Continued)

TVARX(2,1,1) model: distribution of the predicted errors

See notes in Figure 3.

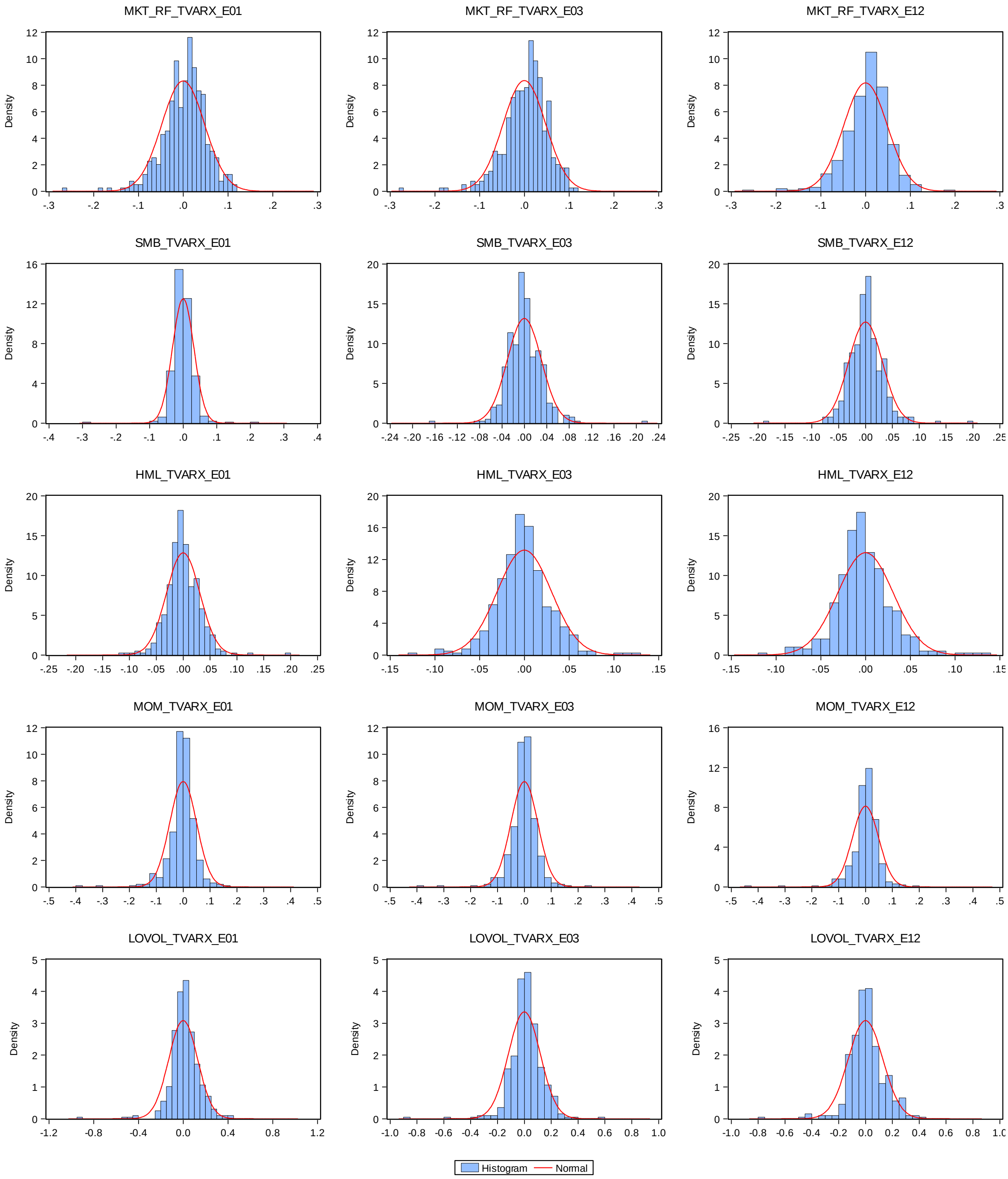

Figure 7
TVAR(2,2) model: distribution of the predicted errors
See notes in Figure 3.

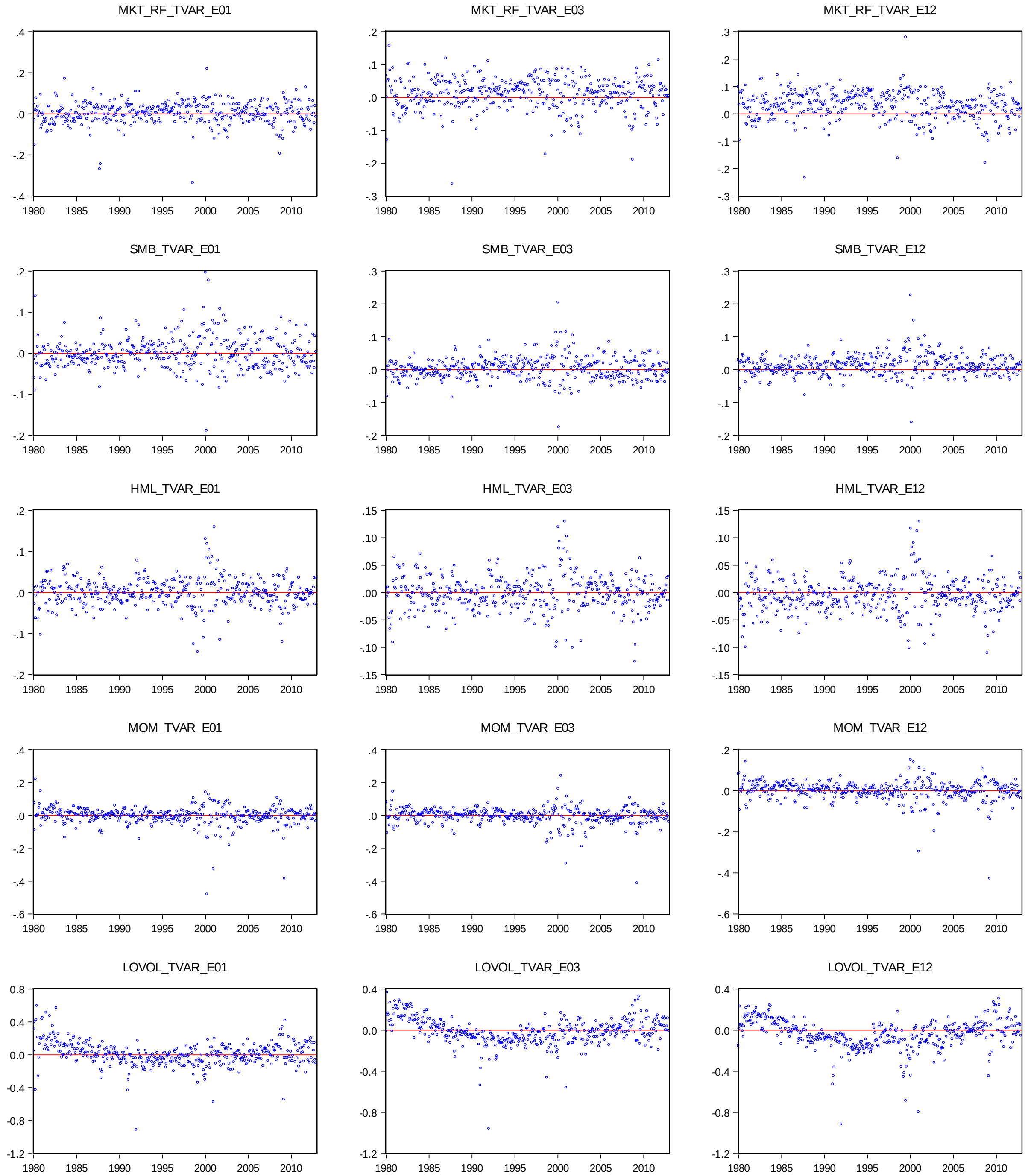

Figure 7 (Continued)

TVAR(2,2) model: distribution of the predicted errors

See notes in Figure 3.

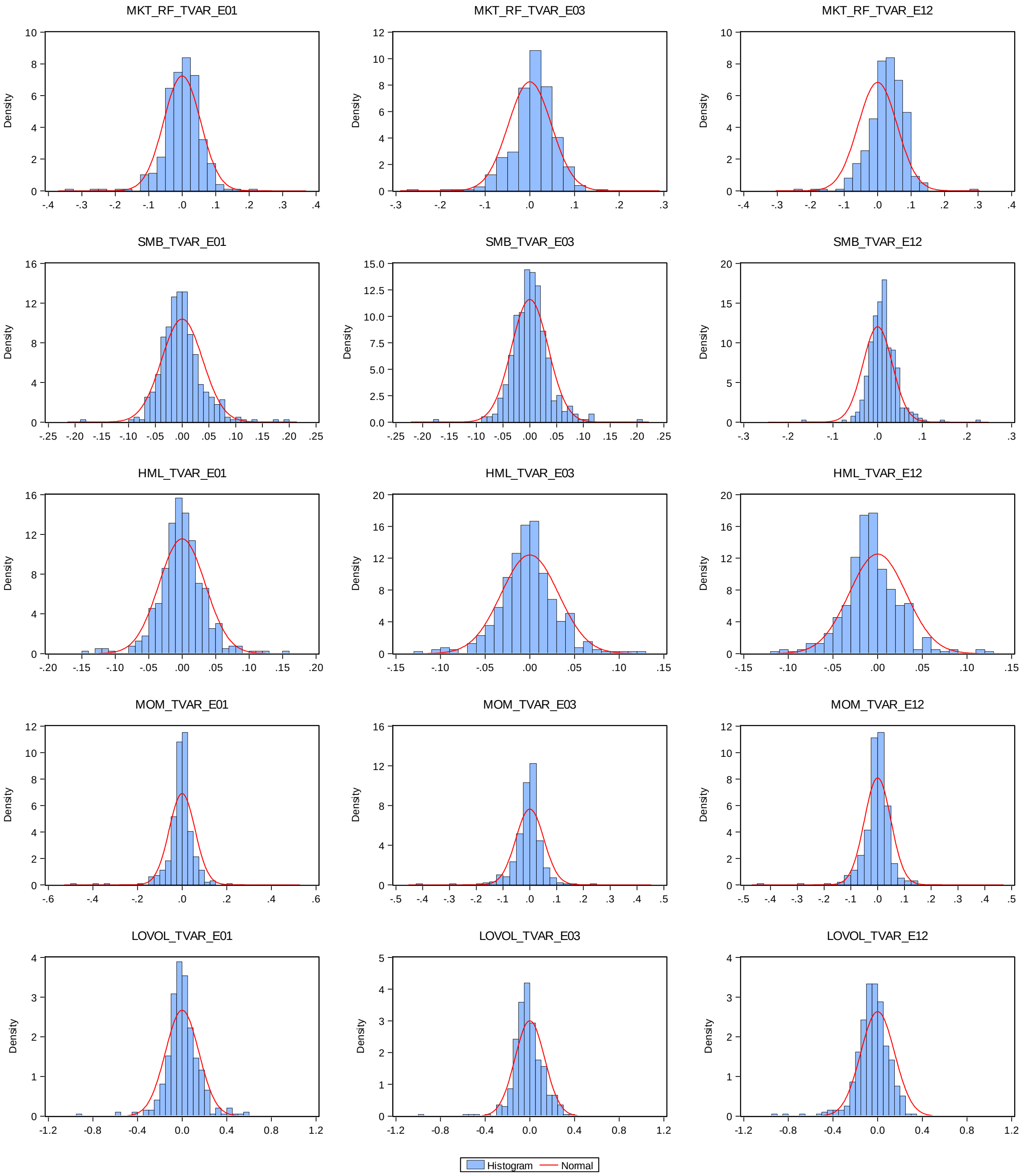

Figure 8

MSI(3) model: distribution of the predicted errors

See notes in Figure 3.

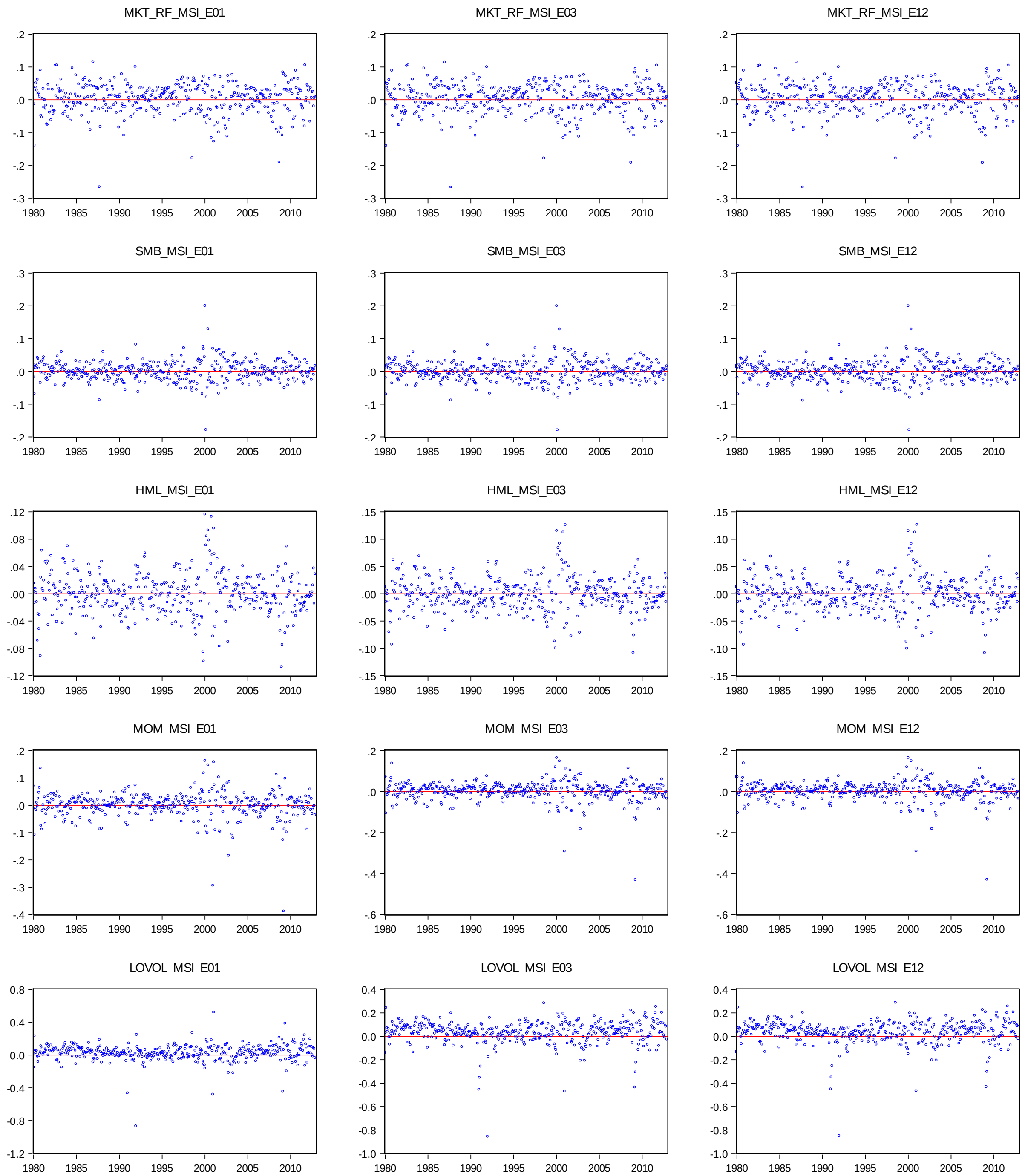

Figure 8 (Continued)

MSI(3) model: distribution of the predicted errors

See notes in Figure 3.

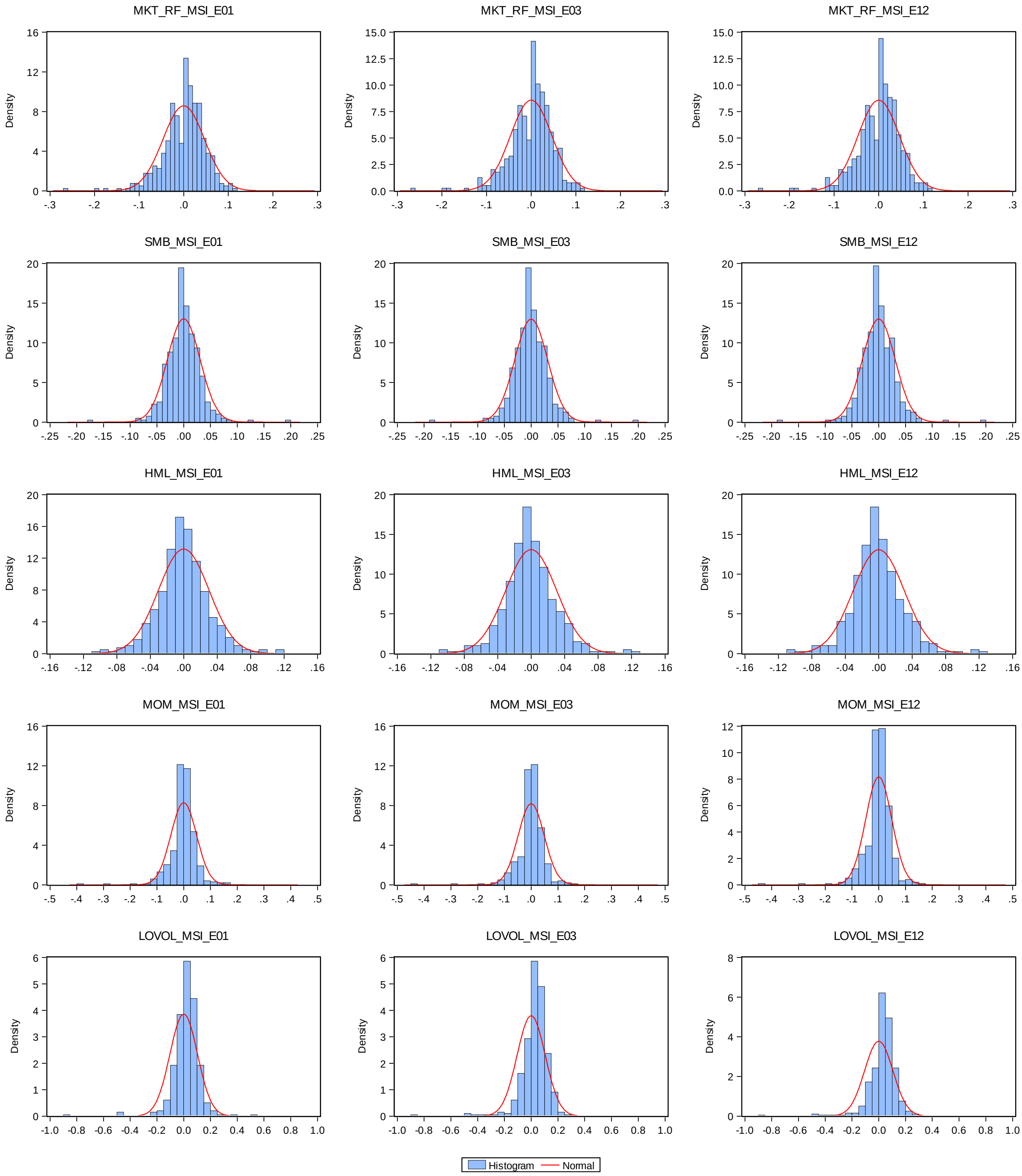

Figure 9
MSIH(3) model: distribution of the predicted errors
See notes in Figure 3.

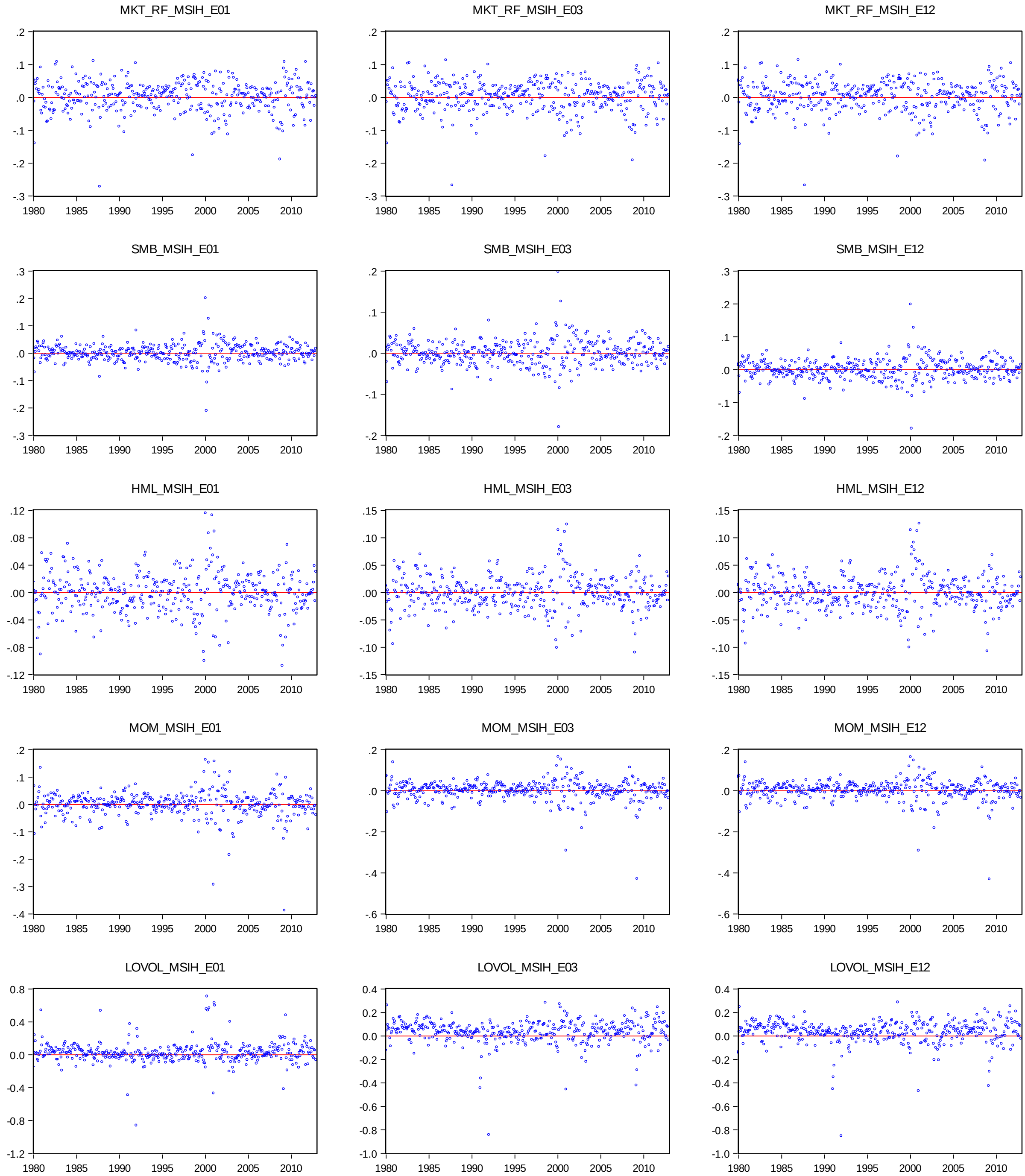

Figure 9 (Continued)

MSIH(3) model: distribution of the predicted errors

See notes in Figure 3.

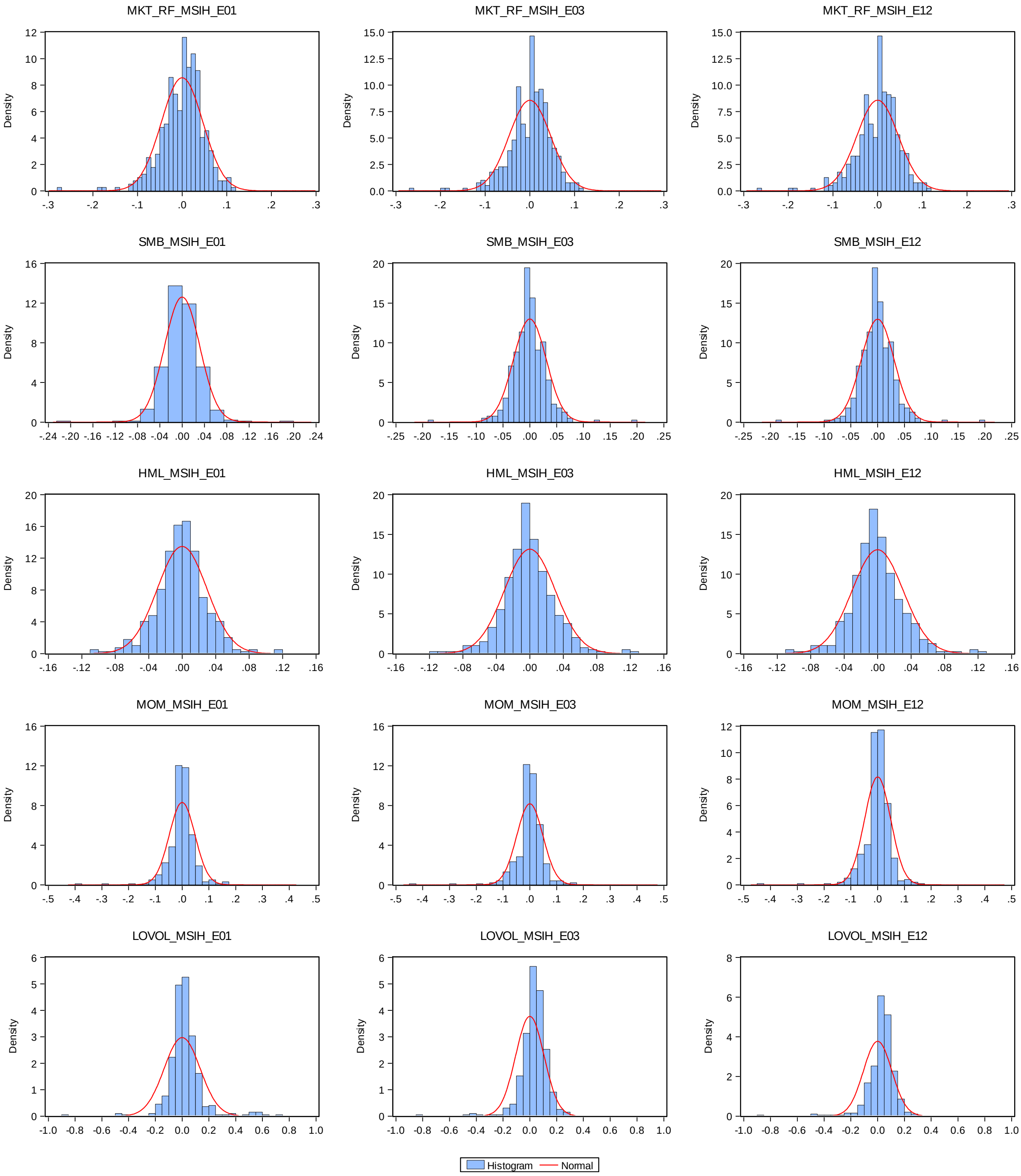

Figure 10
MSIVARH(3,1) model: distribution of the predicted errors

See notes in Figure 3.

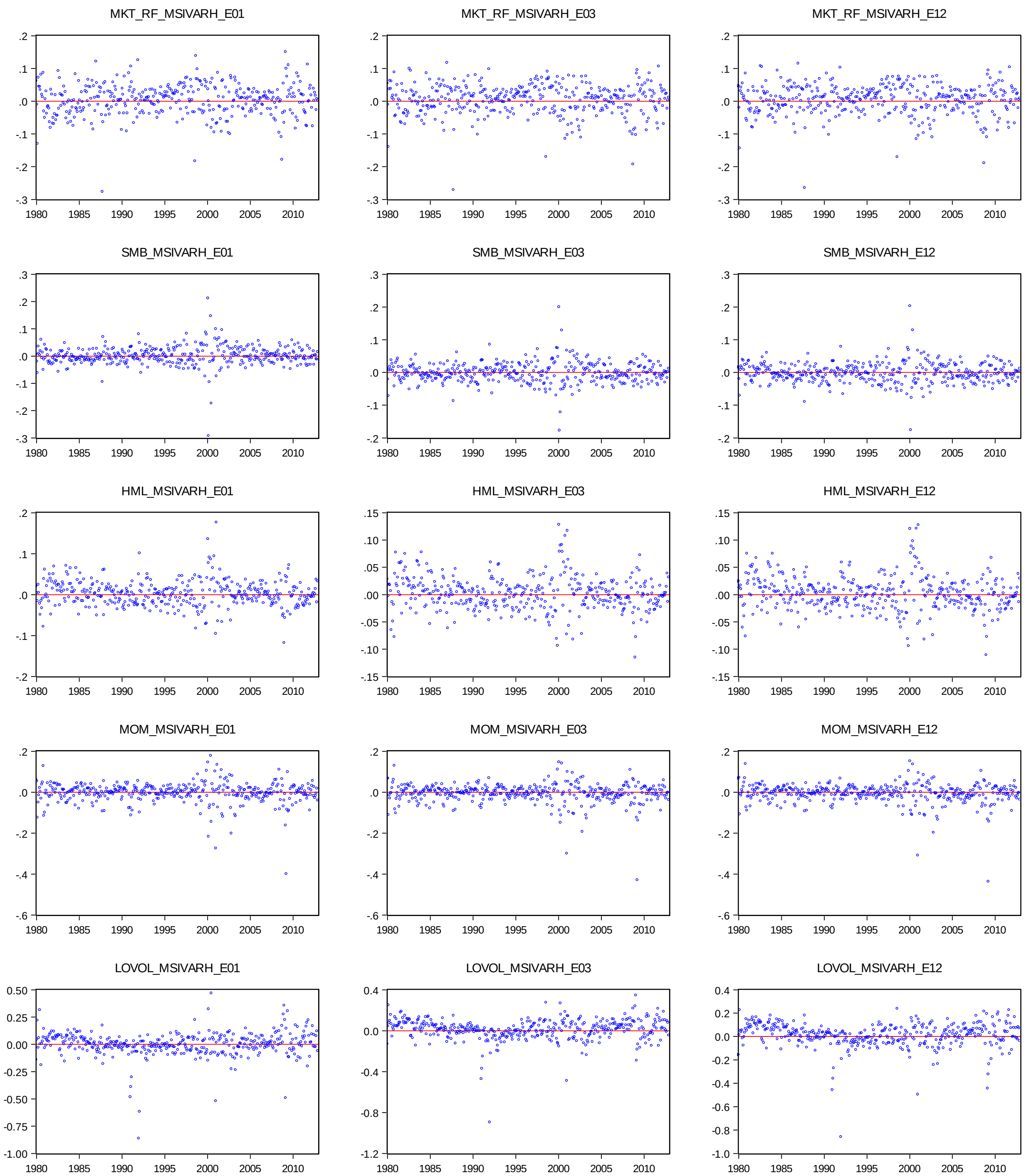

Figure 10 (Continued)

MSIVARH(3,1): distribution of the predicted errors

See notes in Figure 3.