Stock return predictability: A factor-augmented predictive regression system with shrinkage method

Saburo Ohno\textsuperscript{a} and Tomohiro Ando\textsuperscript{b}

\textsuperscript{a}Amundi Japan, Ltd., Chiyoda-ku, Tokyo, Japan; \textsuperscript{b}Graduate School of Business, Keio University Yokohama, Kanagawa, Japan

ABSTRACT

To predict stock market behaviors, we use a factor-augmented predictive regression with shrinkage to incorporate the information available across literally thousands of financial and economic variables. The system is constructed in terms of both expected returns and the tails of the return distribution. We develop the variable selection consistency and asymptotic normality of the estimator. To select the regularization parameter, we employ the prediction error, with the aim of predicting the behavior of the stock market. Through analysis of the Tokyo Stock Exchange, we find that a large number of variables provide useful information for predicting stock market behaviors.

KEYWORDS

Common factor; high-dimensional predictors; model selection; quantiles; regularization

JEL CLASSIFICATION

C13; G12

1. Introduction

The finance literature has long identified sources of comovement in stock returns over time, and the correct model specification for the prediction of asset returns is of central importance to investment practitioners. Some common sources of stock return comovements include the book-to-market ratio (Kothari and Shanken, 1997), the dividend payout ratio (Lamont, 1998), the dividend–price ratio (Fama and French, 1988), expected business conditions (Campbell and Diebold, 2009), interest rate spreads (Campbell, 1987), labor income (Santos and Veronesi, 2006), lagged industry portfolio returns (Hong et al., 2007), nominal interest rates (Ang and Bekaert, 2007), and oil prices (Driesprong et al., 2008), to name but a few. There is also some evidence that we can use a variety of economic variables to predict stock returns. However, evidence of stock return predictability remains controversial in the literature. Moreover, even if a particular asset-pricing model performs well with in-sample forecasting, it may not necessarily have the same accuracy for out-of-sample forecasts. Indeed, Goyal and Welch (2008) concluded that the predictive ability of a variety of popular economic variables does not hold up well in out-of-sample forecasting.

Since the recent subprime mortgage crisis, followed by the collapse of Lehman Brothers and the European sovereign debt crisis, investment practitioners have come to realize that stock prices are highly variable during macroeconomic shocks. As a result, investors now consider macroeconomic information to be an important factor in their decision making. Furthermore, the idea that quantitative models should reflect macroeconomic information in their construction for predicting asset returns has become one of the standards in asset management. Because the stock market reflects the economic outlook, it may be possible to improve the power of a predictive system by incorporating macroeconomic variables (Brown et al., 2009). Chen et al. (1986), for example, argued that innovations in macroeconomic fundamentals could explain expected stock returns, given that economic fundamentals reflect current and future consumption at various horizons. As one development, subsequent studies have confirmed that the book-to-market value and size factors are associated with economic fundamentals (Hahn and Lee, 2006; Petkova, 2006; Vassalou and Xing, 2004).

CONTACT Saburo Ohno\textsuperscript{a} saburo.ohno@jp.amundi.com Amundi Japan, Ltd., Chiyoda-ku, Tokyo, Japan.

© 2018 Taylor & Francis Group, LLC
In practice, the channel through which macroeconomic conditions affect stock returns is expressed by the fundamental equation

\[ P_t = E \left[ \sum_{k=1}^{\infty} \frac{C_{t+k}}{1 + d_{t+k}} \right], \]

where \( P_t \) is a security price at time \( t \), \( C_{t+k} \) is the future cash flow paid at time \( t + k \), and \( d_{t+k} \) is the discount factor, consisting of the risk-free interest rate and the risk premium. The expectation is based on investor expectations about future economic activities. According to this equation, macroeconomic conditions can affect stock prices through cash flows and the discount factor. If the economic outlook is indeed reflected in the stock market, market information, including turnover, liquidity, and the implied cost of equity capital, would then be helpful in stock return predictions. Therefore, awareness of these issues in practice is valuable for prediction of the future behavior of stock indices or prices, including expected returns, early confirmation of market turning points, and changes in market trends. Obviously, we can also employ such information as trading signals for asset allocations. As this viewpoint becomes more widely held, and effective investment styles change on a market-by-market basis, it is now becoming common for investors to consider many types of broader environmental information in their decision making.

Motivated by this practical need, we address the question of which types of information concerning the dynamic behavior of the stock market and the real economy are important for capturing the future characteristics of the stock market. In the analysis, we use a factor-augmented predictive regression system (see Ando and Tsay, 2011, 2014; Bai, 2006; Bernanke et al., 2005; Ludvigson and Ng, 2007; Stock and Watson, 2002a,b) to evaluate the usefulness of the dynamic behavior of the stock market and the real economy in predicting the behavior of the Japanese stock market. Investigation of the Japanese market in particular can assist in assessing the impact of such dynamic behavior on other stock markets, given that Japan is the world’s third-largest economy. Furthermore, the Japanese stock market is a rare example among developed economies of a market where the momentum strategy does not work well, whereas the value strategy works very well (Asness, 2011; Fama and French, 2012).

We construct a predictive system for predicting the conditional mean and the 5% and 95% quantile points. This is in stark contrast with previous studies, which have addressed only the prediction of expected stock returns. Of course, while the conditional mean is the central parameter in most investor decision-making processes, including asset allocation, portfolio optimization, market timing, and so on, the conditional 5% and 95% quantile points also represent important information, especially for those wishing to control tail or downside risks. It is also easy to understand that the 5% quantile points link directly to the concept of “value at risk.” Alternatively, the conditional 95% quantile points provide information on the upside potential, and value at risk involving shorting strategies. This is because if the prediction objective is an index for which derivatives are available, many practitioners use these derivatives to control for the exposure of their position to the market, mainly by taking a short position. For this purpose, the 95% quantile becomes “the value at risk of 5%” in addition to the 5% quantile. Finally, movements or significant shifts in the 5% and 95% quantile points are expected to reflect information on the distance to turning points in the market trend. This information helps practitioners to determine their investment stance, choosing between being, say, a trend follower or a contrarian.

A crucial issue in factor-augmented predictive regression modeling is the selection of the combination of predictors from among the many candidate or potential predictors. We note here that some traditional model selection criteria, such as the Akaike and Bayesian information criteria, make some distributional assumptions for stock returns, even though the true distribution is unknown. In addition, other problems with model selection criteria arise when there is a large set of predictors with no natural ordering, such that the enumeration of a very large number of predictive regressions is necessary. For instance, even with just 20 potential predictor variables, the number of candidate models exceeds a million. Although there has been rapid progress in information technology recently, evaluating such a large number of models is impractical.
We deal with these issues using the regularization method. This approach allows us to select a handful of predictors to be included in the predictive regression system, while the remaining variables are discarded. The idea of using the regularization method has become increasingly popular, with recent work in this area by Caner (2009), Fan and Li (2001), Fan and Lv (2008), Fan and Peng (2004), Tibshirani (1996), Zhang (2010), Zou (2006), Zou and Hastie (2005), and Zou et al. (2007), among many others. Together with the use of the regularization method, our study focuses on the question of what type of information is important for prediction. As discussed later, we select the regularization parameter using the prediction error.

Our study makes the following contributions. First, we construct a predictive system for predicting the 5% and 95% quantile points, as well as the conditional mean, whereas previous studies focused only on the prediction of expected stock returns. Second, we develop the variable selection consistency and asymptotic normality of our estimator. Third, we introduce a number of variables that represent the dynamic behavior of the stock market and the real economy in predicting the characteristics of the stock market. We also construct a set of new factors based on liquidity and the implied cost of equity capital. Fourth, we find that a wide variety of types of information concerning the dynamic behavior of the stock market and the real economy contribute to out-of-sample stock return forecasting. However, most of these contribute little to the prediction of the 5% and 95% quantile points. Finally, we obtain more solid out-of-sample evidence for 12-months-ahead forecasting than for one-month-ahead forecasting. This implies that there is a long-term relationship between a number of the variables representing the dynamic behavior of the stock market and the real economy as well as the future stock market, and that the short-term relationships are much weaker.

The remainder of the article is organized as follows. Sections 2 and 3 review the method of factor-augmented predictive regression with shrinkage. Section 4 describes the data and specifies the model structure. Section 5 presents and analyzes the empirical results. We evaluate the approach using the mean squared forecasting errors (MSFE), mean absolute forecasting errors (MAFE), and correct sign prediction (SIGN). Section 6 concludes the article.

2. Predictive regression system

We focus on stock return predictability, and for this purpose, we review several key methods. Suppose that our aim is to predict the future rate of returns for equity portfolios \( y_{t+h} \) using some predictive regression system, where \( h \) denotes the forecasting horizon. Our predictive system has a factor-augmented predictive regression structure as follows:

\[
y_{t+h} = \alpha'f_t + \beta'w_t + e_{t+h} = \theta'z_t + e_{t+h}, \quad t = 1, \ldots, T,
\]

where \( h \) is the lead time between information availability and the dependent variable, \( \theta = (\alpha', \beta')' \), \( z_t = (f_t', w_t')' \) are \( p \)-dimensional vectors, and \( \alpha \) and \( \beta \) are \( r \)- and \( (p - r) \)-dimensional coefficient parameters, respectively. There are two types of predictor: \( r \)-dimensional unobservable factors \( f_t = (f_{1t}, \ldots, f_{rt})' \), which represent the market dynamics, and \( w_t \), which is a set of prepared predictors. In total, there is a set of \( p \) predictors \( f_t, w_t \).

We assume that the \( N \)-dimensional stock returns \( x_t = (x_{1t}, \ldots, x_{Nt})' \) are derived from the set of \( r \)-factors \( f_t' \):

\[
x_t = \Lambda f_t + \epsilon_t, \quad t = 1, \ldots, T,
\]

where \( \epsilon_t = (\epsilon_{1t}, \ldots, \epsilon_{Nt})' \) is an \( N \)-dimensional random noise vector with mean \( 0 \) and a variance satisfying a condition described later. The \( N \times r \) matrix \( \Lambda_r = (\lambda_1, \ldots, \lambda_N)' \) contains the factor loadings. Instead of \( f_t \), we observe a panel of data \( X = (x_1, \ldots, x_T)' \) that contains information about \( f_t \). Then, (1) and (2) constitute the diffusion index forecasting model in Stock and Watson (2002a).

In this article, we also consider a quantile regression model with factor-augmented predictors. Quantile regression (Koenker and Bassett, 1978), a comprehensive methodology for estimating models of conditional quantile functions, is widely used in the economics literature. By complementing the focus
of classical linear regression on the conditional mean, quantile regression allows us to estimate the effects of covariates, not only on the center of a distribution but also on the upper and lower tails. In contrast to the factor-augmented regression model in (1), in which a conditional expectation of the response variable is the focus, quantile regression attempts to estimate the \( \tau \) th conditional quantile of \( y_{t+h} \) given \( f_t \) and \( w_t \):

\[
q_{\tau}(y_{t+h}|f_t, w_t; \alpha(\tau), \beta(\tau)) = \alpha(\tau)' f_t + \beta(\tau)' w_t = \theta(\tau)' z_t,
\]

where \( \alpha(\tau) \) and \( \beta(\tau) \) are the vectors of coefficients that depend on the quantile \( \tau \).

Section 3.2 in this article provides details of the assumptions imposed on the factor-augmented regression models in (1) and (3). Explanations of the factor method are available in the literature (Bai, 2001; Forni and Reichlin, 1998; Forni et al., 2000, 2001, 2004; Koop and Potter, 2004; Stock and Watson, 2003; Bai and Ng, 2002, 2008; Bernanke and Boivin, 2003; Connor and Korajczyk, 1986; Forni and Lippi, 2002a, b). In the following section, we provide the inference procedure.

3. Methods

3.1. Inference with shrinkage methods

The inference procedure consists of two stages. In the first stage, we estimate common factors from the panel of data using the principal components method. In the second stage, we apply the estimated factors, as described below.

The full set of observations \( X \) can be expressed in matrix form \( X = FA' + E \), where \( X = (x_1, \ldots, x_r)' \), \( F = (f_1, \ldots, f_r)' \), and \( E = (e_1, \ldots, e_r)' \). To estimate the factor model in (2), we use the method of asymptotic principal components (Connor and Korajczyk, 1986). Given a value of the dimension of \( r \), the estimates \( \hat{F} \) or \( \hat{\Lambda} \) of \( F \) and \( \Lambda \) can be obtained by minimizing \( \text{tr} \left\{ (X - FA')' (X - FA') \right\} / NT \), where the \( T \times r \) matrix \( F \) and the \( N \times r \) matrix \( \Lambda \) are subject to the normalization condition \( F'F/T = I_r \), where \( I_r \) is the \( r \times r \) identity matrix. Although the number of factors is unknown, it can be determined by some model evaluation criteria, including those explained in Bai and Ng (2002).

The asymptotic principal component estimate of \( \hat{F} \) is \( \sqrt{T} \) times the eigenvectors corresponding to the \( r \) largest eigenvalues of \( XX' / (NT) \). Substituting the estimated factors into the factor-augmented regression models in (1) and (3) yields

\[
y_{t+h} = \alpha' \hat{f}_t + \beta' w_t + e_{t+h} = \theta' \hat{z}_t + e_{t+h}
\]

and

\[
q_{\tau}(y_{t+h}|\hat{f}_t, w_t; \alpha(\tau), \beta(\tau)) = \alpha(\tau)' \hat{f}_t + \beta(\tau)' w_t = \theta(\tau)' \hat{z}_t,
\]

where \( \hat{z}_t = (\hat{f}_t, w_t)' \). The unknown parameters \( \theta = (\alpha', \beta')' \) in the factor-augmented regression model (4) are estimated by a shrinkage method, obtained by minimizing

\[
\frac{1}{T} \sum_{t=1}^T (y_{t+h} - \alpha' \hat{f}_t - \beta' w_t)^2 + p(\theta),
\]

where \( p(\theta) \) is a function of the coefficients indexed by the regularization parameter \( \kappa \) that controls the trade-off between the loss function and the penalty. In a similar manner, the unknown parameters \( \theta(\tau) = (\alpha(\tau)', \beta(\tau)')' \) in (5) are estimated by minimizing the objective function

\[
\frac{1}{T} \sum_{t=1}^T \rho_{\tau}(y_{t+h} - \alpha(\tau)' \hat{f}_t - \beta(\tau)' w_t) + p(\theta(\tau)),
\]

with \( \rho_{\tau}(u) = u(\tau - I(u < 0)) \).
We use the shrinkage method to estimate coefficients for irrelevant predictors to be exactly equal to zero. This operation is equivalent to selection of the relevant predictors. Fan and Li (2001) have shown that the Lasso penalty produces biases because of its linear increase of the penalty on regression coefficients. To remedy this bias issue, several procedures are proposed, including the Smoothly Clipped Absolute Deviation (SCAD) penalty (Forni and Lippi, 2001) and the adaptive Lasso penalty (Zou, 2006). The adaptive Lasso requires data-adaptive weights for the penalty term. To avoid this difficulty, we use the SCAD penalty for $p(\cdot)$, which is formally given as $p(\theta) = \sum_{k=1}^p p_{k,2}(\theta_k)$ with

$$p_{k,2}(\theta_k) = \begin{cases} \kappa |\theta_k| & (|\theta_k| \leq \kappa) \\ \gamma \kappa |\theta_k| - 0.5(\theta_k^2 + \kappa^2) & (\kappa < |\theta_k| \leq \gamma \kappa) \\ 0.5 \kappa^2 (\gamma + 1) & (\gamma \kappa < |\theta_k|) \end{cases}$$

for $\kappa > 0$ and $\gamma > 2$. This penalty first applies the same rate of penalization as the Lasso and then reduces the rate to zero as it moves further away from zero. Therefore, this penalty can avoid the bias issue of Lasso. For the quantile regression with shrinkage methods, see, for example, Koenker and Xiao (2006), Wu and Liu (2009), and Belloni and Chernozhukov (2011). An estimation algorithm for quantile regression with the SCAD penalty is described in Wu and Liu (2009).

If we take an extremely large value of the regularization parameter $\kappa$, almost all elements of $\theta$ will be estimated as zero. In such a case, we could exclude an important predictor. Conversely, too small a regularization parameter could include a number of unrelated predictors because almost all elements of $\theta$ will not disappear at zero. Therefore, we need to balance these options and determine a proper size for the regularization parameter $\kappa$. To select a penalty size, we use the model selection procedure described in Section 3.3.

### 3.2. Some asymptotic results

In this section, we investigate some asymptotic properties of the parameter estimates. We first clarify the assumptions imposed on the factor-augmented regression model in (1) for predicting the conditional mean. The following assumptions are adopted.

**Assumption A.** Assumption A in Bai and Ng (2008) holds.

**Assumption B.** The noise term $e_{t+h}$ is independently and identically distributed (i.i.d.) and has a zero mean; that is, $E[e_{t+h}|I_t] = 0$, where $I_t$ denotes all information available up to time $t$. Also, it is independent of $(f_t), (\lambda_t), e_t$.

**Assumption C.** $(e_{t+h}, f_t, w_t, e_t)$ is strictly stationary and ergodic.

**Assumption D.** Let $\theta^0 = (\theta_{10}^0, \theta_{20}^0)'$ be the true parameter value of the factor-augmented regression model in (1), and $\theta = (\theta_1', \theta_2')'$ be the corresponding parameter values. Without loss of generality, we assume that $\theta_{20} = 0$. Let $h(y_{t+h}, \theta^0, f_t, w_t)$ denote the first derivative of the loss function with respect to $\theta_1$ (which corresponds to the true predictor) such that the observation $(y_{t+h}, f_t, w_t)$ is yielded. Then, $T^{-1/2} \sum_{t=1}^T h(y_{t+h}, \theta^0, f_t, w_t) \to N(0, \Sigma_{zz,e})$ in distribution with $\Sigma_{zz,e}$ being a positive definite matrix.

**Assumption D’.** Let $\theta^0(\tau) = (\theta_{10}(\tau)', \theta_{20}(\tau)')'$ be the true parameter value of the conditional quantile system in (3), $\theta(\tau) = (\theta_1(\tau)', \theta_2(\tau)')'$ be the corresponding parameter values, and let $\theta_{20}(\tau) = 0$. Let $h(y_{t+h}, \theta^0, f_t, w_t)$ denote the first derivative of the loss function with respect to $\theta_1$ (which corresponds to the true predictor) such that the observation $(y_{t+h}, f_t, w_t)$ is yielded. Then, $T^{-1/2} \sum_{t=1}^T h(y_{t+h}, \theta^0, f_t, w_t) \to N(0, \Sigma_{zz,e}(\tau))$ in distribution with $\Sigma_{zz,e}(\tau)$ being a positive definite matrix.
Assumption E. The penalty \( p(\theta) \) satisfies \( \max \{ p^\prime(\theta) : \theta \neq 0 \} \rightarrow 0 \), and \( \lim_{T \rightarrow \infty} \lim_{\theta \rightarrow 0^+} p'(\theta)/\kappa > 0 \). The same assumption is imposed on the penalty \( p(\theta(\tau)) \) on the factor-augmented conditional quantile system in (3).

Assumption A concerns the factor model, and are identical to those made in Bai and Ng (2008). It assumes the existence of \( r \) common factors and allows for heteroscedasticity and weak time-series and cross-section dependence in the idiosyncratic component. See also Assumption C in Bai and Ng (2006). Assumption B implies that \( \varepsilon_{t+h} \) is conditionally unpredictable. Assumption C assumes that the data are strictly stationary and ergodic, and is identical to R(ii) of Cheng and Hansen (2013). This assumption simplifies the asymptotic theory. Assumptions D and D’ are similar to M1(iii) in Bai and Ng (2008), and is the central limit theorem. Assumption E is needed for the variable selection consistency, and the use of the SCAD penalty implicitly imposed this assumption on the penalty term \( p(\cdot) \). See both Theorem 1 and Lemma 1 in Forni and Lippi (2001).

We establish the asymptotic normality of the estimated parameter for the model (1). In the next theorem, it is shown that the method described in Section 3.1 can identify the set of true explanatory variables. Recall that \( \theta^0 = (\theta_{10}', \theta_{20}')' \) is the true parameter value and \( \hat{\theta} = (\hat{\theta}_1, \hat{\theta}_2)' \) is the corresponding parameter estimate. In addition to the asymptotic normality of \( \hat{\theta}_1 \), we show that the estimator must possess the sparsity property, \( \hat{\theta}_2 = 0 \). All proofs are provided in the Appendix.

**Theorem 1.** Suppose that Assumptions A–D hold and that the regularization parameter satisfies \( \kappa \rightarrow 0 \) and \( \sqrt{T} \kappa \rightarrow \infty \) as \( T \rightarrow \infty \). Furthermore, if \( \sqrt{T}/N \rightarrow 0 \), then the following variable selection consistency holds:

\[
P(\hat{\theta}_2 = 0) \rightarrow 1, \quad N, T \rightarrow \infty.
\]

Moreover, \( \sqrt{T}(\hat{\theta}_1 - \theta_{10}) \) is asymptotically normal with mean \( \theta \) and variance–covariance matrix \( \Sigma \)

\[
\sqrt{T}(\hat{\theta}_1 - \theta_{10}) \rightarrow_d N(\theta, \Sigma),
\]

with the asymptotic covariance matrix given by

\[
\Sigma = \Phi_0^{-1} \Sigma_{zz}^{-1} \Sigma_{zz,c} \Sigma_{zz}^{-1} \Phi_0^{-1},
\]

where \( \Phi_0 = \text{diag}(V_0^{-1} Q \Sigma_A, I) \) is a block diagonal matrix, and \( V_0 \) is the probability limit of \( \hat{V} \), with \( \hat{V} \) being the \( r_0 \times r_0 \) diagonal matrix consisting of the \( r_0 \) eigenvalues of \( XX'/(TN) \) such that the corresponding \( r_0 \) factors are true predictors in the predictive regression system. Based on the set of \( r_0 \) factors, \( Q \) is defined as the probability limit of \( \hat{V}F/T \), \( \Sigma_A \) is the probability limit of \( N^{-1} \Lambda' \Lambda \), \( \Sigma_{zz} \) is the probability limit of \( \Sigma_{zz} = T^{-1} \sum_{t=1}^T \hat{z}_t \hat{z}_t' + T^{-1} \Sigma_\kappa \) with \( \Sigma_\kappa = \text{diag}(\hat{p}_{k,q}(\theta_{110}), \ldots, \hat{p}_{k,q}(\theta_{10^q})) \), and \( q \) is the dimension of \( \theta_{10} \).

Similar results are also obtained for the estimated quantile forecasting model. The result is summarized in the following theorem.

**Theorem 2.** Suppose that Assumptions A–C and D’ hold and that the regularization parameter satisfies \( \kappa \rightarrow 0 \) and \( \sqrt{T} \kappa \rightarrow \infty \) as \( T \rightarrow \infty \). Furthermore, if \( T^{5/8}/N \rightarrow 0 \), then the following variable selection consistency holds:

\[
P(\hat{\theta}_2(\tau) = 0) \rightarrow 1, \quad N, T \rightarrow \infty.
\]

Moreover, \( \sqrt{T}(\hat{\theta}_1(\tau) - \theta_{10}(\tau)) \) is asymptotically normal with mean \( \theta \) and variance–covariance matrix \( \Sigma(\tau) \)

\[
\sqrt{T}(\hat{\theta}_1(\tau) - \theta_{10}(\tau)) \rightarrow_d N(\theta, \Sigma(\tau)),
\]

where \( \Sigma(\tau) = \Phi_0^{-1} \Sigma_{zz}^{-1} \Sigma_{zz,c} \Sigma_{zz}^{-1} \Phi_0^{-1} \), and \( \Phi_0 = \text{diag}(V_0^{-1} Q \Sigma_A, I) \) is a block diagonal matrix, and \( V_0 \) is the probability limit of \( \hat{V} \), with \( \hat{V} \) being the \( r_0 \times r_0 \) diagonal matrix consisting of the \( r_0 \) eigenvalues of \( XX'/(TN) \) such that the corresponding \( r_0 \) factors are true predictors in the predictive regression system. Based on the set of \( r_0 \) factors, \( Q \) is defined as the probability limit of \( \hat{V}F/T \), \( \Sigma_A \) is the probability limit of \( N^{-1} \Lambda' \Lambda \), \( \Sigma_{zz} \) is the probability limit of \( \Sigma_{zz} = T^{-1} \sum_{t=1}^T \hat{z}_t \hat{z}_t' + T^{-1} \Sigma_\kappa \) with \( \Sigma_\kappa = \text{diag}(\hat{p}_{k,q}(\theta_{110}), \ldots, \hat{p}_{k,q}(\theta_{10^q})) \), and \( q \) is the dimension of \( \theta_{10} \).
with the asymptotic covariance matrix given by

\[ \Sigma(\tau) = J(\tau)^{-1}I(\tau)J(\tau)^{-1}, \]

where \( I(\tau) = \frac{1}{T}(\tau - 1)(HZ)^\prime(HZ), \) \( J(\tau) = \frac{1}{T}(HZ)^\prime M(HZ) + \tau \Sigma_k \) with \( \Sigma_k = \text{diag}(p_k^\prime(\theta_{10}), \ldots, p_q^\prime(\theta_{10})) \), and \( q \) is the dimension of \( \theta_{10} \). \( Z = (z_1, \ldots, z_T)^\prime \) consists of the true regressors, \( M = \text{diag}(g(\xi_1(\tau)), \ldots, g(\xi_n(\tau))) \) is the \( T \)-dimensional diagonal matrix in which the \( t \)th element of \( M \) is the value of the \( t \)th quantile of the density function \( g(\xi_i(\tau)) \), with \( \xi_i(\tau) = G^{-1}(\tau|z_t) \) and \( P(y_t \leq y|z_t) = G(y|z_t) \).

**Remark 1.** In this article, the SCAD penalty is used for the shrinkage. We point out that the oracle property in Theorems 1 and 2 can be obtained even if we use the adaptive Lasso for the penalty. However, in this case, we would require the adaptive weight that satisfies the regularity conditions.

**Remark 2.** We extracted the factor structure through one panel \( X \). There may be a situation where there are several groups of variables; for example, a panel of exchange rates, a panel of leading economic indicators, and so on. We can extract their factor structures through analyzing each of the panels separately. In such a case, we impose Assumptions A on each of the panels and still obtain the same oracle property as in Theorems 1 and 2.

**Remark 3.** In Step 1, the factors are estimated from the full panel \( X \), and the factors are expressed as the linear combinations of the variables in the original panel. As pointed out by a referee, if all variables are included as potential predictors, then there is no more gain from adding factors estimated from the same predictors (see also De Mola et al., 2008). However, we point out that the total number of predictors \( z_t \) (compared with the number of observations \( T \)) should not be large when using the shrinkage procedure. Fan and Lv (2008) reported that in general, it is challenging to estimate the sparse parameter vector \( \theta \) accurately when the number of potential predictors \( z_t \) is much larger than \( T \). Owing to the dimensionality, the maximum spurious correlation between an unrelated predictor and the response can be large (see Fan and Lv, 2008). In our application, the panel \( X \) consists of an individual firm’s stock returns series, noting that we analyze roughly 1,700 firms in total. Thus, the direct use of all variables in the panel might cause the problem of dimensionality. Our empirical result in Section 5.4 also shows the usefulness of the factor-augmented predictive regression system in terms of prediction accuracy.

**Remark 4.** Our theorems consider the factor estimation error, and hence, the development of these results are not straightforward. Obviously, Theorem 1 reduces to the result of Fan and Li (2001), and Theorem 2 reduces to the result of Wu and Liu (2009). We point out that the number of potential predictors \( p \) should be fixed. If it diverges, we need more strict assumptions on the model; in particular, the number of potential predictors \( p \) should be strictly restricted (see, for example, Fan and Peng, 2004). Several existing papers have developed theories for shrinkage estimation using estimated factors as regressors, including Lu and Su (2013) and Cheng et al. (2013). Ando and Bai (2013a,b) developed the variable selection consistency and the asymptotic normality of panel data models with unobserved factor errors. Caner and Han (2014) proposed a group bridge estimator to select the correct number of factors in approximate factor models.

### 3.3. Regularization parameter selection

A natural performance measure for fitted factor-augmented regression models such as (4) is the predictive mean squared error (PMSE). Let \( \hat{y}_{t+h} = \hat{\alpha}' \hat{f}_t + \hat{\beta}' \hat{w}_t \) be the predicted value under the
given value of $\kappa$, and the parameter estimates $\hat{\alpha}$ and $\hat{\beta}$ are obtained by using the past $T$ observations \{$y_{t+h-1}, y_{t+h-2}, \ldots, y_{t+h-T}\}. In other words, $\hat{y}_{t+h}$ is constructed based on the historical information, up to time $t - 1$. The PMSE is then
\[
\text{PMSE}(\kappa) = \sum_{j=1}^{n} (y_{t+h+j} - \hat{y}_{t+h+j})^2,
\]
where $n$ is the number of prediction points. We can select the value of the regularization parameter in (6) by minimizing the PMSE. This PMSE is a measure investigated by Inoue and Kilian (2006). If the dataset involves cross-sectional data, then the PMSE reduces to the well-known cross-validation score that evaluates the regression models. The following theorem justifies the use of the PMSE.

**Theorem 3.** Suppose that Assumptions A–D and $T^{1/2}/N \to 0$ hold. Then, the PMSE is an unbiased estimate of the mean squared error $L(\hat{\mu}) = (\mu - \hat{\mu})'(\mu - \hat{\mu})/n$, between the true expectation $\mu = (\mu_{t+h+1}, \ldots, \mu_{t+h+n})'$ and its estimate $\hat{\mu} = (\hat{\mu}_{t+h+1}, \ldots, \hat{\mu}_{t+h+n})'$, plus a constant term that does not depend on the model specification.

Therefore, we can expect that the proposed approach asymptotically minimizes the expected squared error. Once we select the value of the regularization parameter $\kappa$, we construct the predicted mean at time $t + h + n + 1$ as $\hat{y}_{t+h+n+1} = \hat{\alpha} \hat{f}_{t+h+n+1} + \hat{\beta} w_{t+n+1}$. Cheng and Hansen (2013) proved similar results for the averaging estimator under a weaker assumption, in the sense that the error term $e_{t+h}$ (h > 1) may be serially correlated. In their Theorem 2, it is shown that the cross-validated error is an asymptotically unbiased estimate of the in-sample squared error from the leave-h-out estimator, plus an irrelevant term. See also Theorem 2 of Hansen (2010), where the cross-validation for the multistep forecasting is investigated.

In a similar manner, we can select the value of the regularization parameter in (7) by minimizing the prediction error (PE):
\[
\text{PE}(\kappa) = \sum_{j=1}^{n} \rho_{t}(y_{t+h+j} - \hat{y}_{t+h+j}(\tau)),
\]
where $\hat{y}_{t+h+j} = \hat{\alpha}(\tau)\hat{f}_{t+j} + \hat{\beta}(\tau)'w_{t+j}$ is the predicted value. Using the same argument, we construct the predicted quantile at time $t + h + n + 1$ as $\hat{y}_{t+h+n+1}(\tau) = \hat{\alpha}(\tau)\hat{f}_{t+n+1} + \hat{\beta}(\tau)'w_{t+n+1}$.

**Remark 5.** This out-of-sample procedure is a natural predictive measure and is most common in practice (Inoue and Kilian, 2001). In the context of standard regression, the abovementioned criterion corresponds to the cross-validation (see Forni and Lippi, 2001). However, Theorem 3 does not establish that the selected regularization parameter is asymptotically efficient. Although a rigorous development of the asymptotic optimality is interesting, it is beyond the scope of this article. Nevertheless, the asymptotic unbiasedness of the PMSE is shown in Theorem 3.

### 4. Data and empirical model specifications

We obtain Tokyo Price Index (TOPIX) returns for different maturities. These serve as the dependent variable, $y_{t+h}$, in the factor-augmented regression models in (1) and (3). To capture the dynamics of the stock market, we use returns for individual stocks listed on the Tokyo Stock Exchange (TSE) for $x_t$ in (2). In addition, we collect $w_t$ candidates for potential predictors that might affect the drift term of TOPIX returns, or that bring in new information from outside Japanese stock markets. Thus, some indexes are selected that indicate current market conditions, changes of the oil price, the foreign exchange rate, and the stock indexes of foreign countries. More details of the model description are as follows.
4.1. Daily models

The first model predicts the conditional mean and the 5% and 95% quantiles for short-term TOPIX returns, including daily returns. The lead times between information availability and the dependent variable, $h$, are set to one, two, three, five, 10, and 20 days. We construct the factor-augmented regression models (1) and (3) and develop one-day-ahead predictive regression systems. We use daily returns for individual stocks listed on the TSE for $x_t$ in (2). For the $w_t$ candidates, we specify the variables in Table 1.

It is well known that daily Japanese stock market returns and those for the U.S. stock market on the previous day often move in the same direction. Indeed, many Japanese market participants check the news about the U.S. stock market's movement in newspapers and morning news shows on television. Therefore, it would be useful to include the S&P 500 index ($W_6$). In the U.S. stock market, the Chicago

| Variable | Description |
|----------|-------------|
| $W_1$ | Call rate (overnight) |
| $W_2$ | Interest rate spread |
| $W_3$ | TOPIX return (1 day) |
| $W_4$ | TOPIX return (5 days) |
| $W_5$ | TOPIX return (20 days) |
| $W_6$ | S&P 500 return (1 day) |
| $W_7$ | Nikkei volatility index |
| $W_8$ | Nikkei volatility index (difference 1 day) |
| $W_9$ | Nikkei volatility index (difference 5 days) |
| $W_{10}$ | Nikkei volatility index (difference 20 days) |
| $W_{11}$ | YTM of long term Japanese government bond (difference 1 day) |
| $W_{12}$ | YTM of long term Japanese government bond (difference 5 days) |
| $W_{13}$ | Crude oil return (1 day) |
| $W_{14}$ | Crude oil return (5 days) |
| $W_{15}$ | Crude oil return (20 days) |
| $W_{16}$ | Crude oil return (120 days) |
| $W_{17}$ | Japanese Yen/US dollar exchange rate (1 day) |
| $W_{18}$ | Japanese Yen/US dollar exchange rate (5 days) |
| $W_{19}$ | Japanese Yen/US dollar exchange rate (20 days) |
| $W_{20}$ | Japanese Yen/US dollar exchange rate (120 days) |
| $W_{21}$ | Japanese Yen/Euro exchange rate (1 day) |
| $W_{22}$ | Japanese Yen/Euro exchange rate (5 days) |
| $W_{23}$ | Japanese Yen/Euro exchange rate (20 days) |
| $W_{24}$ | Japanese Yen/Euro exchange rate (120 days) |
| $W_{25}$ | Fama-French 3 factor — market factor (1 days) |
| $W_{26}$ | Fama-French 3 factor — small minus big factor (1 days) |
| $W_{27}$ | Fama-French 3 factor — high minus low factor (1 days) |
| $W_{28}$ | Fama-French 3 factor — market factor (5 days) |
| $W_{29}$ | Fama-French 3 factor — small minus big factor (5 days) |
| $W_{30}$ | Fama-French 3 factor — high minus low factor (5 days) |
| $W_{31}$ | Fama-French 3 factor — market factor (20 days) |
| $W_{32}$ | Fama-French 3 factor — small minus big factor (20 days) |
| $W_{33}$ | Fama-French 3 factor — high minus low factor (20 days) |
| $W_{34}$ | Modified Amihud's illiquidity measure (weighted average) |
| $W_{35}$ | Modified Amihud's small minus big factor |
| $W_{36}$ | Modified Amihud's high minus low factor |
| $W_{37}$ | Modified Amihud's illiquidity measure (weighted average; spot) |
| $W_{38}$ | Modified Amihud's small minus big factor (spot) |
| $W_{39}$ | Modified Amihud's high minus low factor (spot) |
| $W_{40}$ | TSE trading volume/TSE market value |
| $W_{41}$ | TSE yields (weighted average) |
| $W_{42}$ | TSE number of rising minus drops over total number (%) |
| $W_{43}$ | TSE number of priced stocks/TSE number of listed stocks (%) |
| $W_{44}$ | Number of factors identified in the factor analysis on the returns of listed stocks |
| $W_{45}$ | Cumulative eigenvalues up to identified number of factors divided by total sum of eigenvalues (%) |
| $W_{46}$ | First eigenvalues/Total sum of eigenvalues (%) |
Board Options Exchange (CBOE) volatility index is a key measure of market expectations of near-term volatility conveyed by S&P 500 stock index option prices. Thus, we can consider it to be the premier global barometer of investor sentiment and market volatility: it is also known as the “fear index” among practitioners, who consider that it provides information about the degree of possibility of a market crash. In Japan, the Nikkei volatility index \((W_7 - W_{10})\) is a proxy variable for investor sentiment. The interest rate spread \((W_2)\) is one possible proxy variable for the economic outlook. Crude oil prices \((W_{13} - W_{16})\) act as a proxy variable for commodity prices and are a major cost factor for various economic activities in Japan. Currency movements directly affect the earnings of Japanese companies, so we also include the Japanese yen/U.S. dollar exchange rate \((W_{17} - W_{20})\). In addition, this exchange rate may cause changes in monetary policy that give rise to changes in consumer behavior and capital flows to the stock market. To consider trade with Europe, we include the Japanese yen/euro exchange rate \((W_{21} - W_{24})\). The publication cycle for typical macroeconomic variables such as gross domestic product (GDP) growth is monthly or quarterly, and thus these variables become nearly constant in the estimation period. Therefore, the Daily Models include only macroeconomic variables that are updated daily, including the yield spread, the exchange rate, and crude oil price changes.

The three-factor model in Fama and French (1992, 1993), one of the best-known asset-pricing models, argues that \(\beta\), book-to-market, and size factors play an important role in explaining the cross section of expected stock returns. We employed the ordinal types of Fama and French's three factors by varying the term of the return calculation \((W_{25} - W_{33})\).

In this analysis, we introduce a new set of investment-style related factors. The first three factors relate to liquidity and are similar to the illiquid (ILLIQ) measure in Amihud (2002). To calculate the daily changes in stock prices, we adopt the \(\sigma_{ij}\) measure of Garman and Klass (1980), which also uses price information on open, high, and low stocks, rather than just the absolute value of daily returns. As a measure of liquidity, we specify the average value for the past 25 business days, which we hereafter refer to as the modified ILLIQ (MILLIQ).

The set of variables \(W_{34} - W_{39}\) is calculated in much the same way as the three well-known Fama and French (1993) factors. At the end of August each year, we sort all firms listed on the First Section of the TSE by market capitalization and include them in two portfolios, small (S) and big (B). Next, we independently sort the same set of stocks into three portfolios according to their book-to-market value. These portfolios are denoted \(H_1\) (highest), \(H_2\) (medium), and \(H_3\) (lowest). The cutoff points are the 30% and 70% quantiles. We then form six portfolios \((S/H_1, S/H_2, S/H_3, B/H_1, B/H_2, B/H_3)\), comprising combinations of the size and book-to-market measures. Then, we calculate the market capitalization-weighted average of the liquidity measure MILLIQ for the six portfolios for each day over the year following the formation of the portfolio. We repeat this procedure each year and obtain a set of market cap-weighted average MILLIQ series over the various data periods. The \(W_{34}\) variable (market level of MILLIQ) is then the market cap-weighted average of the liquidity measure for all firms listed on the First Section of the TSE. The \(W_{35}\) variable (MILLIQ small-minus-big factor) is the simple average liquidity measure for the three small portfolios minus that for the three big portfolios:

\[
\text{MILLIQ small minus big factor} = \frac{1}{3}(S/H_1 + S/H_2 + S/H_3) - \frac{1}{3}(B/H_1 + B/H_2 + B/H_3).
\]

We calculate the \(W_{36}\) variable (MILLIQ high-minus-low factor) in the same manner:

\[
\text{MILLIQ high minus low factor} = \frac{1}{2}(S/H_1 + B/H_1) - \frac{1}{2}(B/H_3 + S/H_3).
\]

The set of variables \(W_{37} - W_{39}\) are calculated in the same manner as above but using the one-day value of the liquidity measure; i.e., the value before averaging.

The reasons for using these variables are as follows. First, liquidity measures, such as Amihud’s ILLIQ, reflect the liquidity premium and the gap between supply and demand. Second, the Japanese stock
market clearly shows a value effect. We also observe a size effect, although reversal of the estimated sign for the size effect sometimes occurs (Jagannathan et al., 1998; Kubota and Takehara, 2003). Therefore, there is a possibility that the market players may differ across the six groups. Importantly, stock markets do not move as a mass: some segments move first, and other segments follow them, to eliminate any arbitrage opportunities made available. Therefore, by calculating the liquidity variable based on our groupings, we may be able to grasp the potential magnitude of any such distortion in the market.

The TSE trading turnover ratio ($W_{40}$) measures trading activity, with $W_{42}$ calculated as the ratio of the number of differences between rising and falling stocks to the total number of listed stocks measuring the direction of the market. The ratio of the number of traded stocks to the total number of listed stocks ($W_{43}$) also measures trading activity. The TSE yield ($W_{41}$) is a component of the rate of return on equity investment.

In addition to these variables, we include variables $W_{44}$–$W_{46}$ relating to stock market states. We use the approach in Bai and Ng (2002) to identify the number of common factors in this large panel of data $X$. In this article, we adopted their $PC_1$ criterion. Then, the number of factors identified is used to calculate the corresponding cumulative eigenvalue, divided by the total sum of eigenvalues, and the first eigenvalue, divided by the total sum of eigenvalues. Kritzman et al. (2010) introduced similar variables as a measure of implied systemic risk, called the absorption ratio, which equals the fraction of the total variance of a set of asset returns, explained by a fixed number of eigenvectors. They argued that this variable captures the extent to which markets are unified (or tightly coupled). Indeed, the fewer the number of effective factors, the greater is the tendency for the stock market to be more volatile.

Note that Bai and Ng (2002) proved that their $PC$ and $IC$ criteria asymptotically select the true number of factors. Thus, the use of $PC$ criteria and $IC$ criteria are asymptotically identical for constructing $W_{44}$–$W_{46}$. Also, the set of common factors $\hat{f}_1$–$\hat{f}_r$ is prepared based on the asymptotic principal components. Here $\hat{r}$ is the selected number of factors by Bai and Ng’s (2002) $PC_1$ criterion. However, note that the set of predictors of the factor augmented regression models, given in (4) and (5), is selected by using $PMSE(\kappa)$ and $PE(\kappa)$ defined in Section 3.3.

The prediction period is from December 30, 1999 to October 31, 2012. In this interval, full sample estimation is available for all models in this article. Although this period is shorter than that for the Monthly Models, described below, we chose this period because the Japanese yen/euro exchange rate has only been available since 1999. To check the robustness of the results that we obtained, we also separated the forecasting period into two parts: before and after September 15, 2008, the day of the “Lehman shock.”

### 4.2. Monthly models

The second model considers the following specifications for (1) and (3). TOPIX returns are used for $y_{t+h}$. The lead times between information availability and the dependent variable, $h$, are set to one month, two, three, six, and 12 months. Therefore, this model uses longer forecasts. To capture the dynamics of the stock market, we use monthly returns for individual stocks listed on the First Section of the TSE for $x_t$ in (2).

Table 2 provides the set of variables to be used for $w_t$. In addition to the new investment factors based on liquidity, we introduce three factors related to the market-implied cost of equity capital. For Daily Models, we introduced a set of three illiquidity factors. By using the market-implied cost of equity capital, we introduce $W_{40}$–$W_{44}$. In calculating the implied cost of equity, we use the formula in Frankel and Lee (1998) so that the values computed in the evaluation formula and market capitalization match each other at the time of evaluation. We specify the two-period-ahead expected value, according to estimates by stock market analysts, as recorded in a database provided by Toyo Keizai

$$V_t = BV_t + \frac{ROE_{t+1} - r_e}{1 + r_e} BV_t + \frac{ROE_{t+2} - r_e}{(1 + r_e)r_e} BV_{t+1},$$

where $BV_t$, $ROE_t$, and $r_e$ denote the book value, the return on equity, and the implied cost of equity capital, respectively. The price evaluation formula shows that the implied cost of equity capital is the
**Table 2.** A set of variables for $w_t$ in monthly model. In addition to these variables, monthly model incorporates the set of $r$-dimensional factors $f_t$, TSE, Tokyo Stock Exchange.

| $W_1$ | Short term interest rate (CD 3 month) |
| $W_2$ | Interest rate spread |
| $W_3$ | TOPIX return (1 month) |
| $W_4$ | TOPIX return (3 months) |
| $W_5$ | TOPIX return (6 months) |
| $W_6$ | TOPIX return (12 months) |
| $W_7$ | S&P 500 return (1 month) |
| $W_8$ | S&P 500 return (3 month) |
| $W_9$ | S&P 500 return (6 month) |
| $W_{10}$ | S&P 500 return (12 month) |
| $W_{11}$ | Nikkei volatility index |
| $W_{12}$ | Nikkei volatility index (difference 1 month) |
| $W_{13}$ | Nikkei volatility index (difference 3 months) |
| $W_{14}$ | Nikkei volatility index (difference 6 months) |
| $W_{15}$ | Nikkei volatility index (difference 12 months) |
| $W_{16}$ | YTM of long term Japanese government bond (difference 1 month) |
| $W_{17}$ | YTM of long term Japanese government bond (difference 3 months) |
| $W_{18}$ | YTM of long term Japanese government bond (difference 6 months) |
| $W_{19}$ | YTM of long term Japanese government bond (difference 12 months) |
| $W_{20}$ | Crude oil return (1 month) |
| $W_{21}$ | Crude oil return (3 months) |
| $W_{22}$ | Crude oil return (6 months) |
| $W_{23}$ | Crude oil return (12 months) |
| $W_{24}$ | Japanese Yen/US dollar exchange rate (1 month) |
| $W_{25}$ | Japanese Yen/US dollar exchange rate (3 months) |
| $W_{26}$ | Japanese Yen/US dollar exchange rate (6 months) |
| $W_{27}$ | Japanese Yen/US dollar exchange rate (12 months) |
| $W_{28}$ | Fama-French 3 factor — market factor (1 month) |
| $W_{29}$ | Fama-French 3 factor — SMB factor (1 month) |
| $W_{30}$ | Fama-French 3 factor — HML factor (1 month) |
| $W_{31}$ | Fama-French 3 factor — market factor (3 months) |
| $W_{32}$ | Fama-French 3 factor — SMB factor (3 months) |
| $W_{33}$ | Fama-French 3 factor — HML factor (3 days) |
| $W_{34}$ | Fama-French 3 factor — market factor (6 months) |
| $W_{35}$ | Fama-French 3 factor — SMB factor (6 months) |
| $W_{36}$ | Fama-French 3 factor — HML factor (6 months) |
| $W_{37}$ | Fama-French 3 factor — market factor (12 months) |
| $W_{38}$ | Fama-French 3 factor — SMB factor (12 months) |
| $W_{39}$ | Fama-French 3 factor — HML factor (12 months) |
| $W_{40}$ | Implied cost of capital (weighted average) |
| $W_{41}$ | Implied cost of capital (small minus big) |
| $W_{42}$ | Implied cost of capital (high minus low) |
| $W_{43}$ | Implied cost of capital (weighted average; difference 1 month) |
| $W_{44}$ | Implied cost of capital (small minus big; difference 1 month) |
| $W_{45}$ | Implied cost of capital (high minus low; difference 1 month) |
| $W_{46}$ | Implied cost of capital (weighted average; difference 3 months) |
| $W_{47}$ | Implied cost of capital (small minus big; difference 3 months) |
| $W_{48}$ | Implied cost of capital (high minus low; difference 3 months) |
| $W_{49}$ | Implied cost of capital (weighted average; difference 6 months) |
| $W_{50}$ | Implied cost of capital (small minus big; difference 6 months) |
| $W_{51}$ | Implied cost of capital (high minus low; difference 6 months) |
| $W_{52}$ | Implied cost of capital (weighted average; difference 12 months) |
| $W_{53}$ | Implied cost of capital (small minus big; difference 12 months) |
| $W_{54}$ | Implied cost of capital (high minus low; difference 12 months) |
| $W_{55}$ | Modified Amihud’s illiquidity measure (weighted average) |
| $W_{56}$ | Modified Amihud’s small minus big factor |
| $W_{57}$ | Modified Amihud’s high minus low factor |
| $W_{58}$ | Modified Amihud’s illiquidity measure (weighted average; difference 1 month) |
| $W_{59}$ | Modified Amihud’s small minus big factor (difference 1 month) |
| $W_{60}$ | Modified Amihud’s high minus low factor (difference 1 month) |
| $W_{61}$ | Modified Amihud’s illiquidity measure (weighted average; difference 3 months) |
| $W_{62}$ | Modified Amihud’s small minus big factor (difference 3 months) |

(Continued)
Table 2. Continued

| \( W_{63} \) | Modified Amihud's high minus low factor (difference 3 months) |
| \( W_{64} \) | Modified Amihud's high minus low factor (difference 6 months) |
| \( W_{65} \) | Modified Amihud's small minus big factor (difference 6 months) |
| \( W_{66} \) | Modified Amihud's high minus low factor (difference 6 months) |
| \( W_{67} \) | Modified Amihud's illiquidity measure (weighted average; difference 12 months) |
| \( W_{68} \) | Modified Amihud's small minus big factor (difference 12 months) |
| \( W_{69} \) | Modified Amihud's high minus low factor (difference 12 months) |
| \( W_{70} \) | TSE trading volume (rate of change 1 month) |
| \( W_{71} \) | TSE trading volume (rate of change 3 month) |
| \( W_{72} \) | TSE trading volume (rate of change 6 month) |
| \( W_{73} \) | TSE trading volume (rate of change 12 month) |
| \( W_{74} \) | TSE trading volume/TSE market value |
| \( W_{75} \) | TSE yields (weighted average) |
| \( W_{76} \) | TSE number of rising minus drop/TSE number of stocks listed |
| \( W_{77} \) | TSE number of priced stocks/TSE number of stocks listed |
| \( W_{78} \) | Composit Index (Leading Index): |
| \( W_{79} \) | Difference of CI-Leading (1m) |
| \( W_{80} \) | Difference of CI-Leading (3m) |
| \( W_{81} \) | Difference of CI-Leading (6m) |
| \( W_{82} \) | Difference of CI-Leading (12m) |
| \( W_{83} \) | Composit Index (Coincident Index): |
| \( W_{84} \) | Difference of CI-Coincident (1m) |
| \( W_{85} \) | Difference of CI-Coincident (3m) |
| \( W_{86} \) | Difference of CI-Coincident (6m) |
| \( W_{87} \) | Difference of CI-Coincident (12m) |
| \( W_{88} \) | Composit Index (Lagging Index): |
| \( W_{89} \) | Difference of CI-Lagging (1m) |
| \( W_{90} \) | Difference of CI-Lagging (3m) |
| \( W_{91} \) | Difference of CI-Lagging (6m) |
| \( W_{92} \) | Difference of CI-Lagging (12m) |
| \( W_{93} \) | Diffusion Index (Leading Index): |
| \( W_{94} \) | Cumulated of DI-Leading (3m) |
| \( W_{95} \) | Cumulated of DI-Leading (6m) |
| \( W_{96} \) | Cumulated of DI-Leading (12m) |
| \( W_{97} \) | Diffusion Index (Coincident Index): |
| \( W_{98} \) | Cumulated of DI-Coincident (3m) |
| \( W_{99} \) | Cumulated of DI-Coincident (6m) |
| \( W_{100} \) | Cumulated of DI-Coincident (12m) |
| \( W_{101} \) | Diffusion Index (Lagging Index): |
| \( W_{102} \) | Cumulated of DI-Lagging (3m) |
| \( W_{103} \) | Cumulated of DI-Lagging (6m) |
| \( W_{104} \) | Cumulated of DI-Lagging (12m) |
| \( W_{105} \) | Number of factors identified |
| \( W_{106} \) | Cumulative eigenvalues/Total sum of eigenvalues |
| \( W_{107} \) | First eigenvalue/Total sum of eigenvalues |

The motivation for using variables \( W_1 - W_{39}, W_{55} - W_{77}, \) and \( W_{105} - W_{107} \) is as described for the Daily Models. For the Monthly Models, unlike the Daily Models, we can use macroeconomic variables as \( W_t \) candidates. Variables \( W_{78} - W_{104} \) are “Indexes of Business Conditions,” published by the Cabinet Office of the Government of Japan, and their movements. We regard these factors as business cycle variables that play an important role in forecasting stock returns. Lettau and Ludvigson (2001) argued that stock returns are forecastable by business cycle variables. Similar to the Daily models, the set of common factors is prepared based on the asymptotic principal components together with Bai and Ng’s (2002) \( PC_1 \) criterion.

The prediction period for this model is from December 1995 to October 2012. As for the Daily Models, the full sample estimation is available for all models in this interval. Again, we separated the discount rate, extracted from the market price. In other words, this is direct information concerning pricing.
foregoing period into two parts: from December 1995 to August 2008, and from September 2008 to October 2012.

5. Empirical findings

5.1. Forecasting results

In our modeling framework, the factor-augmented regression models (1) and (3) are determined based on the past $n$ observations with the prediction errors (PMSE and PE in Section 3.2). To obtain the prediction errors in each time period, $T$ past observations are used.

We set $n = 60$ and $T = 60$ in this study. Following Fan and Li’s (2001) suggestion, we set $\gamma$ in the SCAD penalty as $\gamma = 3.7$ to reduce the computational burden. The Daily Models start from December 28, 1999 for the full-size estimation periods. From a practical viewpoint, setting the models estimation periods to be equal to 60, or as a multiple of 60, appears to be reasonable because, in the case of the Daily Models, 60 days is approximately equal to a quarter, the shortest interval of analysis of business activity in a fiscal year. For the Monthly Models, 60 months is equal to five years, the basic unit of time from the strategic point of view for businesses. Moreover, many practitioners doubt the idea that the structure of stock price dynamism is stable, so they tend to set shorter estimation periods, with keeping sample size can be thought as enough. However, because there is no theoretical background that supports the optimality of setting $n = 60$, we also discuss the effect of using different estimation periods later.

The regularization parameter takes values of $\kappa = \{10, 1, 0.1, 0.01, 0.001, 0.0001\}$. Given an optimized value of $\kappa$ (i.e., the optimized set of predictors $w_t$ and $f_t$ and the corresponding model parameters), the expected return in the next period $t + h$, $\hat{y}_{t+h}$, and the 5% and 95% quantiles, $\hat{y}_{t+h}(5\%)$ and $\hat{y}_{t+h}(95\%)$, are calculated as

$$\hat{y}_{t+h} = \hat{\alpha} f_t + \hat{\beta} w_t,$$

$$\hat{y}_{t+h}(5\%) = \hat{\alpha}(0.05)' f_t + \hat{\beta}(0.05)' w_t,$$

$$\hat{y}_{t+h}(95\%) = \hat{\alpha}(0.95)' f_t + \hat{\beta}(0.95)' w_t.$$

Rolling the sample forward by one period, we repeat this procedure to the last period. Thus, all empirical tests in this study are based exactly on the outputs from out-of-sample predictions.

We computed the MSFE and the MAPE associated with our sector forecasts, given by

$$\text{MSFE} = \frac{1}{t_{\max} - t_{\min} + 1} \sum_{t = t_{\min}}^{t_{\max}} (y_{t+h} - \hat{y}_{t+h})^2,$$

$$\text{MAPE} = \frac{1}{t_{\max} - t_{\min} + 1} \sum_{t = t_{\min}}^{t_{\max}} |y_{t+h} - \hat{y}_{t+h}|,$$

where $t_{\min}$ and $t_{\max}$ are the forecast start and end points, respectively. We also computed the correct sign (SIGN) prediction measure:

$$\text{SIGN} = \frac{1}{t_{\max} - t_{\min} + 1} \sum_{t = t_{\min}}^{t_{\max}} \delta(y_{t+h}, \hat{y}_{t+h}),$$

where $\delta(y_{t+h}, \hat{y}_{t+h}) = 1$ if the signs for $y_{t+h}$ and $\hat{y}_{t+h}$ are the same, and $\delta(y_{t+h}, \hat{y}_{t+h}) = 0$ otherwise.

Furthermore, we evaluated the statistical relationships between the observed values $y_{t+h} = \mu_{t+h} + \epsilon_{t+h}$ and the predicted values $\hat{y}_{t+h}$. Here, $\mu_{t+h}$ is the true conditional mean, and $\epsilon_{t+h}$ is the idiosyncratic shock. We regressed $y_{t+h}$ on $\hat{y}_{t+h}$ using the simple linear regression $y_{t+h} = \alpha + \beta \hat{y}_{t+h} + \epsilon_{t+h}$. Under Assumptions A–E and $\sqrt{T}/N \to 0$, $\hat{y}_{t+h}$ is a consistent estimator of $\mu_{t+h}$, and $\max_{t=1,...,T} \| \hat{f}_t - Hf_t \| = o_p(1)$ (see Lemma 2 in the Appendix.). Thus, in the asymptotic sense, we can regard the idiosyncratic shock $\epsilon_{t+h}$ as independent of $\hat{y}_{t+h}$, and the estimated $\beta$ is equivalent to the beta coefficient estimated
from the simple linear regression as if the true factor is used in the predictive regression system. We calculated the estimated $\beta$, its $t$-value and $p$-value, and the adjusted $R^2$ score. The adjusted $R^2$ measures how well $y_{t+h}$ predicts $y_{t+h}$. A higher adjusted $R^2$ indicates a better forecasting result. Gathering out-of-sample evidence is important in avoiding the problem of data snooping Lo and MacKinlay (1990).

Table 3 presents MSFE, MAFE, and SIGN values for our forecasts, as well as estimated $\beta$ values and the corresponding $t$-value, $p$-value, and adjusted $R^2$. As shown in Table 3, the result for Daily Model ($h = 1$) represents a good prediction result. We also tested the hypothesis $\beta = 1$ vs. $\beta \neq 1$ and found that the Daily Model with $h = 1$ did not reject the hypothesis $\beta = 1$ at the 5% significance level. The results of the test are not included in Table 3.

Table 4 reports the list of the top-10 variables in the Daily Model for the conditional mean prediction ($h = 1$) in the following order: the number, the average times selected, and the longest period for which the variable was continuously selected. The extracted factors are also included as the top-10 variables. To understand the meanings of the factors, we calculated the correlation between the extracted factors

| Panel A | $h$ | MSE | MAE | SIGN | $\hat{\beta}$ | $t$-value | $p$-value | Adjusted $R^2$ |
|---------|-----|-----|-----|------|---------------|-----------|-----------|----------------|
| Daily Models | 1 day | 0.0002 | 0.0096 | 0.5951 | 0.91000 | 21.7431 | 0.0000 | 0.1301 |
| | 2 days | 0.0004 | 0.0152 | 0.5684 | 0.3227 | 7.5725 | 0.0000 | 0.0175 |
| | 3 days | 0.0008 | 0.0202 | 0.5225 | 0.1297 | 4.1973 | 0.0000 | 0.0052 |
| | 5 days | 0.0022 | 0.0324 | 0.5358 | 0.0681 | 4.6296 | 0.0000 | 0.0064 |
| | 10 days | 0.0053 | 0.0515 | 0.5282 | 0.0570 | 4.9619 | 0.0000 | 0.0074 |
| | 20 days | 0.0101 | 0.0698 | 0.5447 | 0.0742 | 6.5898 | 0.0000 | 0.0133 |
| Monthly Models | 1 month | 0.0030 | 0.0436 | 0.5074 | -0.1608 | -0.7736 | 0.4401 | -0.0020 |
| | 2 month | 0.0116 | 0.0826 | 0.4975 | -0.0621 | -0.7736 | 0.4401 | -0.0020 |
| | 3 month | 0.0279 | 0.1229 | 0.5222 | -0.0511 | -0.8971 | 0.3707 | -0.0010 |
| | 6 month | 0.1053 | 0.2608 | 0.4236 | -0.1305 | -2.9441 | 0.0036 | 0.0366 |
| | 12 month | 0.1710 | 0.3499 | 0.3218 | -0.3173 | -5.1537 | 0.0000 | 0.1128 |

| Panel B | $h$ | MSE | MAE | SIGN | $\hat{\beta}$ | $t$-value | $p$-value | Adjusted $R^2$ |
|---------|-----|-----|-----|------|---------------|-----------|-----------|----------------|
| Daily Models | 1 day | 0.0002 | 0.0094 | 0.5760 | 1.0273 | 14.9687 | 0.0000 | 0.0943 |
| | 2 days | 0.0004 | 0.0147 | 0.5578 | 0.3546 | 5.8924 | 0.0000 | 0.0155 |
| | 3 days | 0.0006 | 0.0195 | 0.5173 | 0.1884 | 4.8936 | 0.0000 | 0.0106 |
| | 5 days | 0.0018 | 0.0304 | 0.5359 | 0.0737 | 3.9136 | 0.0001 | 0.0066 |
| | 10 days | 0.0049 | 0.0492 | 0.5373 | 0.0415 | 2.9435 | 0.0033 | 0.0036 |
| | 20 days | 0.0074 | 0.0619 | 0.5280 | 0.0596 | 3.6164 | 0.0003 | 0.0056 |
| Monthly Models | 1 month | 0.0027 | 0.0421 | 0.4837 | -0.4385 | -1.4616 | 0.1454 | 0.0074 |
| | 2 month | 0.0106 | 0.0803 | 0.5098 | -0.0237 | -0.2486 | 0.8040 | -0.0062 |
| | 3 month | 0.0217 | 0.1127 | 0.5098 | -0.0801 | -0.9535 | 0.3415 | -0.0006 |
| | 6 month | 0.1054 | 0.2693 | 0.3856 | -0.1951 | -3.6562 | 0.0003 | 0.0752 |
| | 12 month | 0.1895 | 0.3694 | 0.2961 | -0.3408 | -5.1069 | 0.0000 | 0.1424 |

| Panel C | $h$ | MSE | MAE | SIGN | $\hat{\beta}$ | $t$-value | $p$-value | Adjusted $R^2$ |
|---------|-----|-----|-----|------|---------------|-----------|-----------|----------------|
| Daily Models | 1 day | 0.0002 | 0.0100 | 0.6354 | 0.8530 | 14.9266 | 0.0000 | 0.1799 |
| | 2 days | 0.0006 | 0.0161 | 0.5909 | 0.2994 | 4.6382 | 0.0000 | 0.0199 |
| | 3 days | 0.0010 | 0.0217 | 0.5336 | 0.0522 | 0.9899 | 0.3223 | -0.0000 |
| | 5 days | 0.0030 | 0.0368 | 0.5356 | 0.0603 | 2.4660 | 0.0137 | 0.0050 |
| | 10 days | 0.0063 | 0.0562 | 0.5089 | 0.0782 | 3.8322 | 0.0001 | 0.0134 |
| Monthly Models | 1 month | 0.0040 | 0.0479 | 0.5000 | -0.0031 | -0.0171 | 0.9863 | -0.0208 |
| | 2 month | 0.0146 | 0.0896 | 0.4600 | -0.1575 | -0.9801 | 0.3282 | -0.0008 |
| | 3 month | 0.0470 | 0.1539 | 0.5600 | -0.0187 | -0.2126 | 0.8318 | -0.0199 |
| | 6 month | 0.1051 | 0.2350 | 0.5400 | 0.0320 | 0.3891 | 0.6976 | -0.0176 |
| | 12 month | 0.1148 | 0.2907 | 0.4000 | -0.3012 | -1.9588 | 0.0515 | 0.0547 |
Table 4. List of the top-10 variables in daily model \((h = 1)\), conditional mean prediction, in order of the number and average times selected and the longest period for which the variable was continuously selected. Prediction periods: Panel A: from December 30, 1999 to October 31, 2012. Panel B: from December 30, 1999 to September 12, 2008. Panel C: from September 16, 2008 to October 31, 2012.

### Panel A

| Rank | Variable | Selected times | Average terms | Longest terms |
|------|----------|----------------|---------------|---------------|
|      |          | times          | days          | days          |
| 1    | \(W_6\)  | 2544           | \(W_6\) 43.64 | \(W_6\) 388   |
| 2    | \(W_{17}\)| 613            | \(W_{17}\) 9.01 | \(W_{17}\) 72  |
| 3    | \(W_{41}\)| 510            | \(W_{13}\) 8.45 | \(W_{13}\) 71  |
| 4    | \(f_3\)  | 397            | \(W_{11}\) 7.55 | \(W_{27}\) 65  |
| 5    | \(W_{13}\)| 361            | \(W_9\) 6.97   | \(W_7\) 57    |
| 6    | \(W_7\)  | 311            | \(W_{57}\) 6.89 | \(W_1\) 50    |
| 7    | \(f_6\)  | 295            | \(W_1\) 6.61   | \(W_{43}\) 49  |
| 8    | \(f_4\)  | 284            | \(W_{41}\) 6.0 | \(W_{38}\) 49  |
| 9    | \(f_9\)  | 272            | \(W_7\) 5.98   | \(W_{33}\) 42  |
| 10   |          |                |               |               |

### Panel B

| Rank | Variable | Selected times | Average terms | Longest terms |
|------|----------|----------------|---------------|---------------|
|      |          | times          | days          | days          |
| 1    | \(W_6\)  | 1683           | \(W_6\) 37.95 | \(W_6\) 294   |
| 2    | \(W_{17}\)| 389            | \(W_{13}\) 10.0 | \(W_{17}\) 71  |
| 3    | \(f_3\)  | 290            | \(W_{57}\) 9.91 | \(W_{13}\) 67  |
| 4    | \(W_{13}\)| 280            | \(W_7\) 8.36   | \(W_{11}\) 58  |
| 5    | \(W_{41}\)| 272            | \(W_{17}\) 8.10 | \(W_7\) 57    |
| 6    | \(W_7\)  | 234            | \(W_9\) 6.88   | \(W_{43}\) 49  |
| 7    | \(W_1\)  | 196            | \(W_{58}\) 6.75 | \(W_{38}\) 49  |
| 8    | \(f_5\)  | 188            | \(W_5\) 6.60   | \(W_{33}\) 42  |
| 9    | \(W_2\)  | 174            | \(W_4\) 6.11   | \(W_5\) 41    |
| 10   | \(W_9\)  | 172            | \(W_2\) 6.0    | \(W_1\) 39    |

### Panel C

| Rank | Variable | Selected times | Average terms | Longest terms |
|------|----------|----------------|---------------|---------------|
|      |          | times          | days          | days          |
| 1    | \(W_6\)  | 861            | \(W_6\) 45.21 | \(W_6\) 337   |
| 2    | \(W_{41}\)| 238            | \(W_1\) 12.07 | \(W_{17}\) 72  |
| 3    | \(W_{17}\)| 224            | \(W_{17}\) 11.20 | \(W_{11}\) 71  |
| 4    | \(W_1\)  | 181            | \(W_{11}\) 10.25 | \(W_{27}\) 65  |
| 5    | \(f_4\)  | 145            | \(W_{27}\) 9.60 | \(W_1\) 50    |
| 6    | \(W_{27}\)| 144            | \(W_{41}\) 7.68 | \(W_9\) 37    |
| 7    | \(W_{11}\)| 123            | \(W_9\) 7.14   | \(W_{41}\) 34  |
| 8    | \(f_1\)  | 113            | \(W_{37}\) 6.67 | \(W_{18}\) 28  |
| 9    | \(f_9\)  | 107            | \(W_{34}\) 5.92 | \(W_{34}\) 27  |
| 10   | \(f_6\)  | 107            | \(W_{43}\) 5.86 | \(W_{42}\) 22  |

and the variables in Table 1. We found that the correlation is time varying, and thus the interpretation of the extracted factors also depends on the time points.

In Table 4, we found that the main explanatory variable for this model is \(W_6\), the S&P 500 index return at time \(t\). For the same calendar day, the closing time of the US stock market is about half a day after the closing time of the Japanese stock market, so investors cannot use the information to make their investment position until the next market opening. The behavior of some of the other variables is comparable to this case. However, this does not indicate that the result for Daily Model \((h = 1)\) are
invalid. Indeed, although the result is difficult to apply to daily base trading strategies, the model still has high power to predict prices in the future. To consider applicability to investment strategies, we also calculated MSFE, MAPE, and so on, using returns from the opening price at time \( t + 1 \) to the closing price at time \( t + h \) as observed values \( y_{t+h} \) instead of the returns from the closing price at time \( t \) to the closing price at time \( t + h \) to eliminate this half-day front-running effect. (Details are summarized in the Appendix.) Compared with Table 3, the results indicate that while the prediction power of Daily Model \((h = 1)\) has become weaker, the result still indicates that a statistical relationship exists between the observed values \( y_{t+1} \) and the predicted values \( \hat{y}_{t+1} \). Overall, almost all of the Daily Models have significant \( p \)-values, although the predicting results are not as strong as the Daily Model \((h = 1)\). We also found that the adjusted \( R^2 \) calculated for “after the Lehman shock” are higher than those values calculated for “before the Lehman shock.” From this evidence, we can conclude that the results that we obtained are robust.

Unfortunately, the monthly return forecasting results were not as robust. First, if the prediction term, \( h \), is shorter than six months, none of the Monthly Models has a significant \( p \)-value, and even if \( h \) is extended to six or 12 months, the \( p \)-values calculated for the “after Lehman” period are insignificant. Although the \( p \)-values and adjusted \( R^2 \) indicate good prediction results for \( h = 6 \) and \( h = 12 \) for the “before Lehman” period, the signs of \( \beta \) for those models are negative. Indeed, the adjusted \( R^2 \) of \( h = 12 \) is high enough if we use \( \hat{y}_{t+h} \) as a predictor, but overall, the resulting outputs from the prediction model are very poor.

Table 5 shows the proportion of observed variables greater than the 95% quantile point and less than the 5% quantile point. As shown in Table 5, only the results for the Daily Models \( h = 1 \) and \( h = 2 \) are close to 5%. Tables 6 and 7 provide a summary of the selected predictors. We can see that the predictors for quantile point forecasting are weaker than those for the conditional mean. The number of times that the top 10 predictors are selected is much smaller than in the case of the conditional mean forecasting model.

We then checked how the number of predictors (the dimension of \( w_t \) and \( f_t \)) varies over time to forecast the conditional mean, and the 5% and 95% quantile points for the TOPIX. Figure 1 shows the total number of predictors for the Daily Model \((h = 1)\). It is clear that the optimal number of predictors is time varying. Moreover, it is evident that a variety of information types relating to the dynamic behavior of the stock market and the real economy contribute to the out-of-sample forecasting of stock returns. Figures (b) and (c) in Fig. 1 summarize the predictor selection results to forecast the conditional mean, and the 5% and 95% quantile points. In contrast to the results illustrated in Fig. (a) in Fig. 1, most of this information does not contribute to the 5% and 95% quantile points. We obtained similar results for the remaining daily and monthly forecasts. This indicates that the information that is useful for constructing the conditional mean and tail points for TOPIX returns may be different.

### 5.2. Role of macroeconomic factors

Table 4 indicates that the suitable variables for the conditional mean prediction of the Daily Model \((h = 1)\) should contain various types of information. For example, this includes the global base “price change” variables such as the S&P 500 index return, the yen/dollar currency, and oil price changes, and the “price dynamics” variables such as Principal Component Analysis (PCA) factors, Fama and French’s style factors (Small-Minus-Big (SMB) and High-Minus-Low (HML)), and the variables of market conditions such as the dividend yield, the volatility index, and the style-related illiquidity factors that we introduced in this article.

To check the role of macroeconomic factors and the new set of style-related illiquidity factors, we undertook prediction with the following two additional models (hereafter, Model A and Model B), which are distinguished from each other and the basic model, the Daily Model \((h = 1)\), by the setting of the candidates for predictors. Candidates for predictors for Model A are the call rate, \( W_1 \), and the set of PCA factors identified by Bai and Ng’s (2002) \( PC_1 \) criterion. For Model B, candidates are the same as for
Table 5. Proportion of realized TOPIX returns that exceeded predicted 5% and 95% quantiles. Prediction periods: Panel A: from December 30, 1999 to October 31, 2012 for daily models; from December, 1995 to October, 2012 for monthly models. Panel B: from December 30, 1999 to September 12, 2008, for daily models; from December, 1995 to August, 2008 for monthly models. Panel C: from September 16, 2008 to October 31, 2012 for daily models; from September, 2008 to October, 2012 for monthly models.

| Panel | $h$ | under the 5% quantile | over the 95% quantile |
|-------|-----|------------------------|-----------------------|
| Daily Model | 1 day | 6.12% | 6.31% |
| | 2 days | 7.32% | 6.69% |
| | 3 days | 7.92% | 7.51% |
| | 5 days | 9.22% | 9.09% |
| | 10 days | 14.13% | 12.39% |
| | 20 days | 21.39% | 24.59% |
| Monthly Model | 1 month | 7.88% | 6.90% |
| | 2 month | 7.39% | 7.39% |
| | 3 month | 8.87% | 8.37% |
| | 6 month | 14.78% | 12.32% |
| | 12 month | 18.04% | 11.34% |
| Panel B | $h$ | under the 5% quantile | over the 95% quantile |
| Daily Model | 1 day | 6.16% | 6.39% |
| | 2 days | 7.42% | 6.78% |
| | 3 days | 8.16% | 7.18% |
| | 5 days | 9.19% | 9.38% |
| | 10 days | 14.09% | 11.38% |
| | 20 days | 23.04% | 21.64% |
| Monthly Model | 1 month | 8.50% | 5.88% |
| | 2 month | 7.84% | 6.54% |
| | 3 month | 8.50% | 7.19% |
| | 6 month | 16.34% | 9.80% |
| | 12 month | 22.92% | 9.72% |
| Panel C | $h$ | under the 5% quantile | over the 95% quantile |
| Daily Model | 1 day | 6.03% | 6.13% |
| | 2 days | 7.12% | 6.52% |
| | 3 days | 7.41% | 8.20% |
| | 5 days | 9.29% | 8.50% |
| | 10 days | 14.23% | 14.53% |
| | 20 days | 17.89% | 30.83% |
| Monthly Model | 1 month | 6.0% | 10.0% |
| | 2 month | 6.0% | 10.0% |
| | 3 month | 10.0% | 12.0% |
| | 6 month | 10.0% | 20.0% |
| | 12 month | 4.0% | 16.0% |

the Daily Models, with the exception of the style-related illiquidity factors, $W_{34}$–$W_{39}$. We calculated the prediction accuracy measures, including MSFE, MAFE, and so on.

Table 8 summarizes the results. The results of Model A, the model with PCA factors, are reasonable but not as strong as those for Model B and the Base Model. The PCA factors, a type of endogenous variable of stock market dynamics, do not have enough power to predict future prices. Adding “price change” variables such as the S&P 500 index return, yen/dollar currency, oil price changes, and some well-known variables of market conditions to the potential predictors of Model A results in Model B. The expansion of potential predictors brings significant predictive power to the model. Finally, adding style-related illiquidity factors improves the forecast accuracy. As shown in Table 8, the $p$-values for the test with the null hypothesis, $\beta = 1$, are more than 1%, and the Base Model is the only model that maintained this $p$-value level even for the “after Lehman shock” period. This evidence indicates the importance of role of macroeconomic factors and market condition variables to predict future prices of TOPIX. Moreover, variables of “price dynamics,” the PCA factors, tend to increase in predictive importance in situations following a large shock such as the Lehman shock.

Now, we compare the variable selection results in Tables 4 with the prediction term $h$ (See also Tables 2–4 in the Appendix). When the prediction term $h$ is extended from 1 to 2 and 20, the PCA
Table 6. List of the top-10 variables in the daily model, 5% quantile prediction ($h = 1$), in order of the number and average times selected and the longest period for which the variable was continuously selected. Prediction periods: Panel A: from December 30, 1999 to October 31, 2012. Panel B: from December 30, 1999 to September 12, 2008. Panel C: from September 16, 2008 to October 31, 2012.

Panel A

| Rank | Variable | Selected times | Average terms | Longest terms |
|------|----------|----------------|---------------|---------------|
|      |          |                 | Variable times | Variable days | Variable days |
| 1    | $W_{42}$ | 1054           | $W_3$         | 21.0          | $W_{16}$ 72  |
| 2    | $W_{16}$ | 452            | $W_6$         | 13.88         | $W_{42}$ 65  |
| 3    | $W_{15}$ | 215            | $W_{13}$      | 1.0           | $W_6$ 47    |
| 4    | $W_{10}$ | 130            | $W_{42}$      | 10.65         | $W_{13}$ 46  |
| 5    | $W_{14}$ | 125            | $W_{16}$      | 8.53          | $W_8$ 42    |
| 6    | $W_6$   | 111            | $W_8$         | 8.17          | $W_{14}$ 38  |
| 7    | $W_5$   | 108            | $W_{10}$      | 7.65          | $W_{10}$ 35  |
| 8    | $W_{13}$| 88             | $W_7$         | 6.75          | $W_7$ 27    |
| 9    | $W_{24}$| 82             | $W_{31}$      | 6.60          | $W_{15}$ 21  |
| 10   | $W_{31}$| 66             | $W_{14}$      | 6.58          | $W_5$ 21    |

Panel B

| Rank | Variable | Selected times | Average terms | Longest terms |
|------|----------|----------------|---------------|---------------|
|      |          |                 | Variable times | Variable days | Variable days |
| 1    | $W_{42}$ | 709            | $W_3$         | 21.0          | $W_{42}$ 44  |
| 2    | $W_{16}$ | 236            | $W_6$         | 11.0          | $W_{16}$ 33  |
| 3    | $W_{15}$ | 152            | $W_{42}$      | 10.58         | $W_6$ 24    |
| 4    | $W_{14}$ | 51             | $W_{13}$      | 8.0           | $W_{15}$ 21  |
| 5    | $W_{13}$| 40             | $W_{31}$      | 7.67          | $W_3$ 21    |
| 6    | $W_6$   | 33             | $W_{16}$      | 6.94          | $W_{13}$ 21  |
| 7    | $W_{24}$| 26             | $W_7$         | 6.0           | $W_{31}$ 17  |
| 8    | $W_7$   | 24             | $W_{14}$      | 5.67          | $W_{24}$ 15  |
| 9    | $W_{31}$| 23             | $W_{24}$      | 5.20          | $W_{14}$ 14  |
| 10   | $W_3$   | 21             | $W_{15}$      | 4.75          | $W_{29}$ 12  |

Panel C

| Rank | Variable | Selected times | Average terms | Longest terms |
|------|----------|----------------|---------------|---------------|
|      |          |                 | Variable times | Variable days | Variable days |
| 1    | $W_{42}$ | 345            | $W_{13}$      | 16.0          | $W_{16}$ 72  |
| 2    | $W_{16}$ | 216            | $W_6$         | 15.60         | $W_{42}$ 47  |
| 3    | $W_{10}$ | 122            | $W_{20}$      | 13.50         | $W_6$ 47    |
| 4    | $W_7$   | 84             | $W_{42}$      | 9.84          | $W_{13}$ 46  |
| 5    | $W_6$   | 78             | $W_{16}$      | 9.26          | $W_8$ 42    |
| 6    | $W_{14}$| 74             | $W_8$         | 9.20          | $W_{14}$ 38  |
| 7    | $W_{15}$| 63             | $W_9$         | 9.17          | $W_{10}$ 35  |
| 8    | $W_{24}$| 56             | $W_{10}$      | 8.13          | $W_7$ 27    |
| 9    | $W_9$   | 55             | $W_{14}$      | 7.40          | $W_{23}$ 17  |
| 10   | $W_{13}$| 48             | $W_7$         | 7.0           | $W_{20}$ 17  |

factors disappear. Moreover, while the change in the commodities price and the foreign exchange rate fell in rank in terms of their predictive power, Japanese domestic market-related variables, such as the yield spread, the interest rate change, the dividend yield, and TOPIX’s own past returns, increased in rank. This indicates that in quite a short period, even a single day, common global information impacts upon the TOPIX return. However, for short-term predictions, market participators tend to consider the condition of the Japanese stock market and their own past behavior. As shown by the list of variables included in the Monthly Model ($h = 12$), style factors using the implied cost of capital and the illiquidity measure are included as model variables at a high frequency. As shown earlier, the three-factor Fama–French model has predictive power on a daily basis; that is, for short-term predictions. Although we created our style factors based on a procedure similar to that developed by Fama and French (1993), our results indicate that our proposed factors potentially include information on the characteristics of
Table 7. List of the top-10 variables in daily model, 95% quantile prediction \((h = 1)\) in order of the number and average times selected and the longest period the variable was continuously selected. Prediction periods; Panel A: from December 30, 1999 to October 31, 2012. Panel B: from December 30, 1999 to September 12, 2008. Panel C: from September 16, 2008 to October 31, 2012.

Panel A

| Rank | Variable | Selected times | Average terms | Longest terms |
|------|----------|----------------|---------------|---------------|
|      |          | times          | days          | days          |
| 1    | \(W_{42}\) | 803            | \(W_{17}\) 10.50 | \(W_{42}\) 55 |
| 2    | \(W_{16}\) | 359            | \(W_{16}\) 7.64  | \(W_{15}\) 42 |
| 3    | \(W_{15}\) | 255            | \(W_{42}\) 7.16  | \(W_{6}\) 42 |
| 4    | \(W_{14}\) | 207            | \(W_{6}\) 6.08   | \(W_{16}\) 42 |
| 5    | \(W_{10}\) | 142            | \(W_{14}\) 5.45  | \(W_{10}\) 40 |
| 6    | \(W_{5}\)  | 127            | \(W_{5}\) 5.08   | \(W_{14}\) 38 |
| 7    | \(W_{6}\)  | 116            | \(W_{23}\) 4.90  | \(W_{6}\) 35 |
| 8    | \(W_{13}\) | 83             | \(W_{7}\) 4.87   | \(W_{5}\) 28 |
| 9    | \(W_{8}\)  | 79             | \(W_{9}\) 4.83   | \(W_{17}\) 26 |
| 10   | \(W_{7}\)  | 73             | \(W_{15}\) 4.72  | \(W_{7}\) 26 |

Panel B

| Rank | Variable | Selected times | Average terms | Longest terms |
|------|----------|----------------|---------------|---------------|
|      |          | times          | days          | days          |
| 1    | \(W_{42}\) | 568            | \(W_{17}\) 10.50 | \(W_{42}\) 55 |
| 2    | \(W_{16}\) | 286            | \(W_{6}\) 7.78   | \(W_{15}\) 42 |
| 3    | \(W_{15}\) | 194            | \(W_{16}\) 7.33  | \(W_{6}\) 42 |
| 4    | \(W_{14}\) | 188            | \(W_{42}\) 6.91  | \(W_{16}\) 42 |
| 5    | \(W_{5}\)  | 89             | \(W_{14}\) 5.33  | \(W_{14}\) 38 |
| 6    | \(W_{6}\)  | 70             | \(W_{23}\) 5.85  | \(W_{5}\) 28 |
| 7    | \(W_{10}\) | 67             | \(W_{5}\) 4.94   | \(W_{17}\) 26 |
| 8    | \(W_{13}\) | 64             | \(W_{15}\) 4.59  | \(W_{23}\) 19 |
| 9    | \(W_{9}\)  | 61             | \(W_{29}\) 4.50  | \(W_{31}\) 18 |
| 10   | \(W_{4}\)  | 53             | \(W_{19}\) 4.33  | \(W_{29}\) 17 |

Panel C

| Rank | Variable | Selected times | Average terms | Longest terms |
|------|----------|----------------|---------------|---------------|
|      |          | times          | days          | days          |
| 1    | \(W_{42}\) | 235            | \(W_{9}\) 11.0  | \(W_{10}\) 40 |
| 2    | \(W_{10}\) | 75             | \(W_{16}\) 9.13 | \(W_{42}\) 35 |
| 3    | \(W_{16}\) | 73             | \(W_{28}\) 9.04 | \(W_{6}\) 35 |
| 4    | \(W_{15}\) | 61             | \(W_{8}\) 9.0    | \(W_{7}\) 26 |
| 5    | \(W_{9}\)  | 55             | \(W_{42}\) 7.83 | \(W_{16}\) 20 |
| 6    | \(W_{8}\)  | 38             | \(W_{7}\) 7.60  | \(W_{15}\) 16 |
| 7    | \(W_{7}\)  | 38             | \(W_{10}\) 7.50 | \(W_{13}\) 14 |
| 8    | \(W_{14}\) | 19             | \(W_{5}\) 5.43  | \(W_{28}\) 14 |
| 9    | \(W_{13}\) | 19             | \(W_{15}\) 5.08 | \(W_{14}\) 13 |
| 10   | \(W_{28}\) | 18             | \(W_{14}\) 4.75 | \(W_{5}\) 11 |

the market pricing system from time to time. That is, they also capture the degree of market distortion by considering value and size measures, and extract information on future trends in the Japanese stock market.

We investigated the time series of the estimated \(\beta\) coefficients for the different prediction period. We found that, in almost all cases, the number of selected factors includes the same variables (MILLIQ or ICC) and style (Market, SMB, or HML) but that the calculation term varies. Therefore, we investigated the sum of the estimated \(\beta\) coefficients with the same name but different terms. The time series plots in Fig. 2 are for the MILLIQ-based style factors in the Daily Models, \(h = 1\). As shown in the figure, the signs of the estimated coefficients are closely clustered. We obtained similar results for the different prediction period (see also Figs. 1–3 in the Appendix).
We obtained similar results for the remaining coefficients. In particular, only a few of the estimated coefficients for the series $w_t$ displayed the same sign over the entire period. However, each plus or minus sign tended to persist for a certain period rather than alternating frequently. This means that the role of each variable in the prediction of future stock price changes depends on differences in the economic environment, business confidence, and the business cycle. In addition, we observed higher predictive power for the 12-months-ahead prediction models than for those models with relatively short-term prediction periods, such as for the one, two, three, and six-months-ahead periods. This implies that
Table 8. Summary of the Mean Squared Error (MSE), the Mean Absolute Error (MAE), and the Correct Sign Prediction Measure (SIGN) for one day-prior-predictions. We regress $y_{t+1}$ on $\hat{y}_{t+1}$ using a simple linear regression: $y_{t+1} = \alpha + \beta \hat{y}_{t+1} + \varepsilon_{t+1}$. The table also includes the estimated $\beta$ coefficients, t-value, p-value, and adjusted $R^2$ scores. The t-value and p-value are not for the test of $H_0: \beta = 0$, but for the test with $H_0: \beta = 1$. Model A: Daily models that include the short-term interest rate and PCA factors derived from TSE individual stock returns only. Model B: Daily models that exclude liquidity variables, $W_{34} - W_{39}$, from Table 1 as potential candidates. Base model: daily model $(h = 1)$ with the full set of potential predictors in Table 1. Therefore, the results are the same as the values appearing in Table 3. Prediction periods: Panel A: from December 30, 1999 to October 31, 2012 for daily models; from December, 1995 to October, 2012 for monthly models. Panel B: from December 30, 1999 to September 12, 2008, for daily models; from December, 1995 to August, 2008 for monthly models. Panel C: from September 16, 2008 to October 31, 2012 for daily models; from September, 2008 to October, 2012 for monthly models.

### Panel A

| Model | $h$ | MSE | MAE | SIGN | $\hat{\beta}$ | t-value | p-value | Adjusted $R^2$ |
|------|-----|-----|-----|------|--------------|---------|---------|---------------|
| A    | 0.0002 | 0.0103 | 0.5152 | 0.3097 | 5.7085 | 0.0000 | 0.0018 |
| B    | 0.0002 | 0.0096 | 0.6014 | 0.8658 | 3.3251 | 0.0099 | 0.1271 |
| Base | 0.0002 | 0.0096 | 0.5951 | 0.9100 | 2.1504 | 0.0316 | 0.1301 |

### Panel B

| Model | $h$ | MSE | MAE | SIGN | $\hat{\beta}$ | t-value | p-value | Adjusted $R^2$ |
|------|-----|-----|-----|------|--------------|---------|---------|---------------|
| A    | 0.0002 | 0.0099 | 0.5117 | 0.1502 | 5.6064 | 0.0000 | 0.0000 |
| B    | 0.0002 | 0.0094 | 0.5835 | 0.9811 | 0.2852 | 0.7755 | 0.0921 |
| Base | 0.0002 | 0.0094 | 0.5760 | 1.0273 | 0.3972 | 0.6912 | 0.0943 |

### Panel C

| Model | $h$ | MSE | MAE | SIGN | $\hat{\beta}$ | t-value | p-value | Adjusted $R^2$ |
|------|-----|-----|-----|------|--------------|---------|---------|---------------|
| A    | 0.0003 | 0.0112 | 0.5227 | 0.4965 | 2.4403 | 0.0148 | 0.0047 |
| B    | 0.0002 | 0.010 | 0.6393 | 0.8104 | 3.4436 | 0.0006 | 0.1759 |
| Base | 0.0002 | 0.010 | 0.6354 | 0.8530 | 2.5723 | 0.0102 | 0.1799 |

the set of variables $w_t$ used in this analysis can be used for predicting whether the stock market direction will continue for a certain period rather than just predicting the effect of shocks that could occur in the very near future.

### 5.3. Rolling effects

Figure 3 plots the time series of the estimated $\beta$ coefficients of $W_6$, the S&P 500 index return, for one-day-ahead prediction models. We found that even though this is the most powerful predictor for the one-day-ahead prediction model, the estimated value is time varying. Therefore, we ask whether the evidence supports the practitioners’ commonsense understanding that the structure of stock price dynamism is not stable. To test this, we compared the results of three types of one-day-ahead prediction models, which are basically very similar to the Daily Model $(h = 1)$ but which vary in terms of the size of the estimation period. We have already seen the result from setting $n = 60$ in this study. Adding to this, we set $n = 120$ and attempted to set $n$ equal to all available periods; i.e., covering all past data. We have daily data used for estimating model parameters from June 30, 1999, so the full period model starts from $n = 125$, and the number of samples can be increased by one, every business day. Table 9 presents MSFE, MAFE, and SIGN values for our forecasts, as well as estimated $\beta$ values and the corresponding t-value, p-value, and adjusted $R^2$. The t-value and p-value are values for testing whether $\hat{\beta} = 1$. Given the excessively time-consuming nature of the full-period prediction, we stopped our calculations before we reached the end of our data. Therefore, MSFE, MAFE, and so on are calculated with samples from December 30, 1999 to January 21, 2008. The highest adjusted $R^2$ is the value of the $n = 60$ in our basic model. Comparing the previously obtained results, the adjusted $R^2$ is more practical. Thus, this confirms the practitioners’ experience.
5.4. The role of factor extraction

In our estimation procedure, the set of factors are extracted from the $T \times N$-dimensional large panels. As noted by a referee, the factors are linear combinations of the variables in the original panel. If all panel variables $x_t$ are included as potential predictors, then there is no gain from adding factors estimated from
Table 9. Summary of the Mean Squared Error (MSE), Mean Absolute Error (MAE), and Correct Sign Prediction Measure (SIGN) for 1 day-prior-prediction. We regress $y_{t+1}$ on $\hat{y}_{t+1}$ using a simple linear regression: $y_{t+1} = \alpha + \beta \hat{y}_{t+1} + \epsilon_{t+1}$. The table also includes the estimated $\beta$ Coefficients, t-value, p-value, and adjusted $R^2$ scores. t-value and p-value are not of test with H0: $\beta = 0$, but of test with H0: $\beta = 1$. Full periods; estimation at time $t$ is done with all data that is available till the time $t$. (120) and (60) are the result with setting the size of estimation period to $T = n = 120$ and $T = n = 60$, respectively. Prediction periods are from December 30, 1999 to January 21, 2008.

| Model     | $h$ | MSE   | MAE   | SIGN  | $\hat{\beta}$ | t-value | P-value | Adjusted $R^2$ |
|-----------|-----|-------|-------|-------|----------------|---------|---------|----------------|
| Full Periods | 0.0001 | 0.0091 | 0.5689 | 0.9942 | 0.9213 | 0.9628 | 0.0783 |
| 120       | 0.0001 | 0.0090 | 0.5957 | 0.9363 | 0.8884 | 0.9493 | 0.1117 |
| 60        | 0.0001 | 0.0088 | 0.6058 | 1.0002 | 0.8495 | 0.9260 | 0.1502 |

We therefore compared the predictive performance of this procedure with the proposed method that employs the factor approach. Setting $h = 1$ in the Daily Models, we evaluated the shrinkage without factor extraction by calculating the mean squared error (MSE), the mean absolute error (MAE), and the SIGN. The prediction period is from December 30, 1999 to October 31, 2012, and the corresponding MSE, MAE, and SIGN are 0.0002, 0.0098, and 0.5947, respectively. These numbers were then compared with those reported in Table 3. We found that the proposed method is better than the shrinkage regression using all $x_t$ as predictors.

We then regressed $y_{t+h}$ on $\hat{y}_{t+h}$ using a simple linear regression: $y_{t+h} = \alpha + \beta \hat{y}_{t+h} + \epsilon_{t+h}$. The estimated coefficient $\hat{\beta} = 0.7694$ rejects the hypothesis $\beta = 1$, whereas the $\hat{\beta}$ from our procedure did not reject the hypothesis. We also found that the adjusted $R^2$ of the shrinkage without factor extraction is 0.070, which is about half of the result, 0.1301, in Table 3. In summary, we believe that the factor extraction provides a useful tool for forecasting.

5.5. Modeling under the presence of structural breaks

In Table 3, we found that the prediction results of each model differ for the “before Lehman bankruptcy” period and the “after Lehman bankruptcy” period. If a structural break occurred in the estimation period, it is natural to consider that the latest information should have more weight in the model parameter estimation. Let $(a_1, \ldots, a_T)'$ be the weight vector that reflects the importance of information. Then unknown parameters $\theta = (\alpha', \beta')'$ in the factor-augmented regression model (4) are estimated by
a shrinkage method, obtained by minimizing

\[
\frac{1}{T} \sum_{t=1}^{T} a_t (y_{t+h} - \alpha^T \hat{f}_t - \beta^T w_t)^2 + p_{k,y}(\theta),
\]

where \( p_{k,y}(\theta) \) is the SCAD penalty, used in Section 3.1. Let \( t_0 \) denote the date of the Lehman bankruptcy. Then, we set

\[
a_t = \begin{cases} 
\frac{1}{NC} & (t < t_0) \\
\phi/NC & (t_0 \leq t)
\end{cases},
\]

where \( NC = \sum_{t=1}^{T} a_t \) is the normalizing constant so that \( \sum_{t=1}^{T} a_t = 1 \), and \( \phi (\cdot > 1) \) puts more weight on the information observed for the Lehman bankruptcy. We point out that various functional forms of \( a_t \) can be considered if the purpose of \( a_t \) is to place more weight on the information observed for the Lehman bankruptcy. However, the exploring the best functional form of \( a_t \) is out of scope in this analysis. In our analysis, we prepared three levels of \( \phi = 1 + 10^k \) \((k = 0, 1, 2)\). The values of the regularization parameter \( \kappa \) and the weight parameter \( \phi \) are selected by minimizing the PMSE score obtained in Section 3.3.

Setting \( h = 20 \), we computed the prediction results using the above approach, and then using the method that did not take into account the structural break. The prediction periods are from September 16, 2008 to September 30, 2010. We found that the consideration of a “structural break” contributed to increasing the adjusted \( R^2 \). The improved adjusted \( R^2 \) (0.0637) is higher than that from the standard method (0.0192).

Recently, several forecasting procedures that take structural breaks into account have been proposed, including Pesaran et al. (2006) and Pesaran and Timmermann (2007). Although a rigorous theoretical investigation of the above methods is interesting, it is beyond the scope of this paper. However, we would like to investigate this problem in a future study.

6. Conclusions

In this article, we addressed the importance of macroeconomic variables in the Japanese stock market in the context of factor-augmented predictive regression. After the recent subprime crisis, followed by the collapse of Lehman Brothers and the European sovereign debt crisis, investment practitioners have realized that stock prices are highly variable during macroeconomic shocks. As awareness of these issues increases in practice, the ability to predict the future behavior of stock markets becomes even more highly valuable.

Motivated by practical views on the stock market, we addressed the question of whether macroeconomic variables play an important role in capturing the characteristics of the Japanese stock market. Using a factor-augmented predictive regression system, we constructed quantitative models for predicting the conditional mean and the 5% and 95% quantile points for the TOPIX. As theoretical contributions, we developed the variable selection consistency and asymptotic normality of our estimators. The results show that macroeconomic variables play an important role in predicting the characteristics of the stock market in Japan, after taking into consideration the various asset-pricing factors. We also checked the robustness of the predictability by regression of the empirical return on the predicted return, which revealed a significant relationship between these variables.

Appendix

A1: Lemma

The results most relevant for our present analysis are summarized in the following lemma. These are well summarized and explained in Bai and Ng (2008). We let \( \| A \| = [tr(A^T A)]^{1/2} \) be the usual norm of a matrix \( A \), where \( tr \) denotes the trace of a square matrix.
Lemma 1 (Bai and Ng, 2002). Suppose that Assumptions A hold. Then
\[
\|\hat{F} - F_0H\|^2 / T = O_p(1 / \min(N, T)),
\]
where \(F_0\) is the true factor, \(H = \hat{V}^{-1}(\hat{F}'F/T)(\Lambda'\Lambda/N)\), with \(\hat{V}\) being the \(r \times r\) diagonal matrix consisting of the \(r\) largest eigenvalues of \(X'X/(TN)\).

Proof of Lemma 1. See Bai and Ng (2002) and also Bai (2003).

Lemma 2 (Bai and Ng, 2008). Suppose that Assumptions A hold. If \(\sqrt{T}/N \to 0\), we have
\[
\max_{t=1,\ldots,T} \|\hat{f}_t - Hf_t\| = O_p(\alpha_T/T) + O_p(1/N^{1/2}) + \max_{t=1,\ldots,T} \left\{ \frac{1}{N} \sum_{i=1}^{N} \lambda_i \varepsilon_{it} \right\},
\]
where \(\max_{t=1,\ldots,T} \|f_t\| = O_p(\alpha_T)\).

Proof of Lemma 2. See the proof of Lemma 2 of Bai and Ng (2008).

As in Bai and Ng (2008), \(\max_{t=1,\ldots,T} \|\hat{f}_t - Hf_t\| = o_p(1)\) if \(T^{1/4}/N \to 0\). Lemma 2 allows us to establish the consistency of our estimators \(\hat{\theta}\). This is because our sample loss function converges uniformly to the population loss function. In other words, we require the uniform convergence of \(\hat{f}_t\) to the space spanned by \(f_t\). More details are discussed in Bai and Ng (2008).

A2: Proof of Theorem 1

We first prove the consistency of \(\hat{\theta} = (\hat{\alpha}', \hat{\beta}')\). Note that \(\hat{\theta}\) satisfies \(\hat{\theta} = \arg\min_{\theta} \sum_{t=1}^{T} (y_{t+h} - \alpha'\hat{f}_t - \beta'w_t)^2 + Tp_{k,y}(\theta)\). Using the consistency of \(\hat{F}\), the difference between \(\sum_{t=1}^{T} (y_{t+h} - \alpha'\hat{f}_t - \beta'w_t)^2\) and \(\sum_{t=1}^{T} (y_{t+h} - \alpha'f_t - \beta'w_t)^2\) can be evaluated as
\[
\frac{1}{T} \sum_{t=1}^{T} \left\{ (y_{t+h} - \alpha'\hat{f}_t - \beta'w_t)^2 - (y_{t+h} - \alpha'f_t - \beta'w_t)^2 \right\}
\leq \frac{1}{T} \sum_{t=1}^{T} \left\{ \|y_{t+h} - \alpha'f_t - \beta'w_t\| \cdot \|\alpha\| \cdot \|\hat{f}_t - f_t\| \right\}
\leq \left[ \frac{1}{T} \sum_{t=1}^{T} (y_{t+h} - \alpha'f_t - \beta'w_t)^2 \right]^{1/2} \left[ \frac{1}{T} \sum_{t=1}^{T} \|\alpha\|^2 \cdot \|\hat{f}_t - f_t\|^2 \right]^{1/2}
= O_p(1) \cdot O_p\left( \frac{1}{\min\{T^{1/2}, N^{1/2}\}} \right) + O_p(1) \cdot O_p\left( \frac{1}{\min\{T, N\}} \right)
= O_p\left( \frac{1}{\min\{T^{1/2}, N^{1/2}\}} \right).
\]
Thus, \(\tilde{\theta}\) and \(\hat{\theta}\), defined as
\[
\hat{\theta} = \arg\min_{\theta} \left\{ \sum_{t=1}^{T} (y_{t+h} - \alpha'f_t - \beta'w_t)^2 + Tp_{k,y}(\theta) \right\},
\]
are asymptotically equivalent. Thus, we obtain
\[ \| \hat{\theta} - \bar{\theta} \| = O_p(1/\min\{T^{1/2}, N^{1/2}\}). \]

Hereafter, we denote \( \left[ \sum_{t=1}^T (y_{t+h} - \alpha' f_t - \beta' w_t)^2 + Tp_{k,y}(\theta) \right] \) as \( R(\theta) \).

We now show that \( \hat{\theta} \) is consistent. Let \( \eta_T = T^{-1/2} + d_T \), with \( d_T = \max\{p'_{k,T,y}(\theta_0^0); \theta_0^0 \neq 0\} \), where \( \theta_0^0 = (\alpha_0', \beta_0')' \) are the true parameter values. Under \( \max\{p'_{k,T,y}(\theta_0^0); \theta_0^0 \neq 0\} \to 0 \), we now show that there exists a local minimizer \( \tilde{\theta} \) of \( R(\theta) \) such that \( \| \hat{\theta} - \tilde{\theta} \| = O_p(T^{-1/2} + d_T) \). As in Forni and Lippi (2001), it is enough to show that for any given \( e > 0 \), there exists a large constant \( C \) such that
\[ P \left[ \min_{\| u \|=C} R(\theta^0 + \eta_T u) > R(\theta^0) \right] \geq 1 - e, \]

because with a probability of at least \( 1 - e \), there exists a local minimum in the ball \( \{ \theta^0 + \eta_T u; \| u \| \leq C \} \), which implies that there exists a local minimizer such that \( \| \hat{\theta} - \theta^0 \| = O_p(\eta_T) \). Here, \( u = (u_\alpha', u_\beta')' \).

Using \( p_{k,T,y}(0) = 0 \), we have
\[ R(\theta^0 + \eta_T u) - R(\theta^0) \geq \sum_{t=1}^T (y_{t+h} - (\alpha_0 + \eta_T u_\alpha)' f_t - (\beta_0 + \eta_T u_\beta)' w_t)^2 + T \sum_{k=1}^q \sum_{k=1}^q p_{k,y}(\theta_k^0 + \eta_T u_k) \]
\[ - \sum_{t=1}^T (y_{t+h} - \alpha_0' f_t - \beta_0' w_t)^2 - T \sum_{k=1}^q p_{k,y}(\theta_k^0), \]
where \( q \) is the number of nonzero components of \( \theta^0 \). By using the Taylor expansion, we have
\[ R(\theta^0 + \eta_T u) - R(\theta^0) \geq -2\eta_T \sum_{t=1}^T (y_{t+h} - \alpha_0' f_t - \beta_0' w_t)f_t' u_\alpha - 2\eta_T \sum_{t=1}^T (y_{t+h} - \alpha_0' f_t - \beta_0' w_t)w_t' u_\beta \]
\[ + \eta_T^2 \sum_{t=1}^T u'_t (f_t' , w_t') (f_t' , w_t') u_T \{ 1 + o_p(1) \} \]
\[ + \sum_{k=1}^q \left[ T \eta_T p'_{k,T,y}(\theta_k^0) \text{sgn}(\theta_k^0) u_k + T \eta_T^2 p'_{k,T,y}(\theta_k^0) u_k^2 \{ 1 + o_p(1) \} \right]. \]

The first and second terms are in the order of \( O_p(\sqrt{T} \eta_T) = O_p(T^{1/2}) \). By choosing a sufficiently large constant, \( C \), the third term dominates the first and second terms uniformly in \( \| u \| = C \). The final term is bounded by \( \sqrt{q} T \eta_T d_T \| u \| + T \eta_T^2 \max\{p'_{k,T,y}(\theta_k^0); \theta_k^0 \neq 0\} \| u \|^2 \). This is also dominated by the third term. In summary, by choosing a sufficiently large constant, \( C, (9) \) holds. Thus, there exists a local minimizer \( \tilde{\theta} \) of \( R(\theta) \) such that \( \| \hat{\theta} - \theta^0 \| = O_p(T^{-1/2} + \eta_T) = o_p(1) \). By choosing a proper \( \kappa_T \), we can obtain a consistent estimator \( \tilde{\theta} \). Thus, we have \( \| \hat{\theta} - \theta^0 \| \leq \| \hat{\theta} - \tilde{\theta} \| + \| \tilde{\theta} - \theta^0 \| = O_p(1/\min\{T^{1/2}, N^{1/2}\}) + O_p(1/\sqrt{T}) = o_p(1) \). The consistency of \( \tilde{\theta} \) is obtained. Moreover, under \( T^{1/2}/N \to 0 \), Theorem 1 of Bai and Ng (2008) provides the convergence rate of \( \tilde{\theta} \); that is,
\[ \| \hat{\theta} - \theta^0 \| = O_p(1/T^{1/2}). \]

Next, we prove the variable selection consistency \( P(\hat{\theta}_{20} = 0) \to 1 \) as \( N, T \to \infty \). It is enough to show that \( P(\hat{\theta}_{20} = 0) \to 1 \) as \( N, T \to \infty \). This part is almost identical to the proof of Forni and Lippi (2001).
It is sufficient to show that with probability tending to one as \( N, T \to \infty \), for some small \( \delta_{N,T} = C/\sqrt{T} \) with a constant \( C \), and for each element of \( \theta_2 = (\theta_{21}, \ldots, \theta_{2p-q}) \), we have

\[
\frac{\partial}{\partial \theta_{2k}} \left[ \sum_{t=1}^{T} (y_{t+h} - \alpha' f_t - \beta' w_t)^2 + T \rho_{k,y} (\theta_{2k}) \right] > 0 \quad (0 < \theta_{2k} < \delta_{N,T}),
\]

\[
\frac{\partial}{\partial \theta_{2k}} \left[ \sum_{t=1}^{T} (y_{t+h} - \alpha' f_t - \beta' w_t)^2 + T \rho_{k,y} (\theta_{2k}) \right] < 0 \quad (-\delta_{N,T} < \theta_{2k} < 0),
\]

for \( k = 1, \ldots, p-q \). We let \( z_{t2} \) denote the \( p-q \) dimensional vector of \( z_t = (f'_t, w'_t) \) that corresponds to the zero coefficients \( \theta_2 \). By Taylor’s expansion and the proof of Bai and Ng (2008), we have

\[
\frac{\partial}{\partial \theta_2} \left[ \frac{1}{T} \sum_{t=1}^{T} (y_{t+h} - \alpha' f_t - \beta' w_t)^2 + \rho_{k,y} (\theta_2) \right]
= \frac{\partial}{\partial \theta_2} \left[ \frac{1}{T} \sum_{t=1}^{T} (y_{t+h} - \alpha_0 f_t - \beta_0 w_t)^2 \right] - 2 \frac{\partial}{\partial \theta_2} \left[ \frac{1}{T} \sum_{t=1}^{T} (y_{t+h} - \alpha' f_t - \beta' w_t) z'_t (\theta - \theta^0) \right]
+ \frac{\partial}{\partial \theta_2} \rho_{k,y} (\theta_2),
\]

where \( \theta^* \) lies between \( \theta \) and \( \theta^0 \). The sum of the first and third terms is dominated by the second term because \( \| \theta - \theta^0 \| = O_p(1/T^{1/2}) \) and \( \partial \left[ \frac{1}{T} \sum_{t=1}^{T} (y_{t+h} - \alpha' f_t - \beta' w_t)^2 \right] / \partial \theta_2 = O_p(1/T^{1/2}) \), we have

\[
\frac{\partial}{\partial \theta_{2k}} \left[ \sum_{t=1}^{T} (y_{t+h} - \alpha' f_t - \beta' w_t)^2 + T \rho_{k,y} (\theta_{2k}) \right]
= T \cdot \kappa \left[ \frac{1}{\kappa} \rho_{k,y} (\theta_{2k}) \text{sign}(\theta_{2k}) + O_p \left( 1/(\sqrt{T} \cdot \kappa) \right) \right].
\]

Now, note that \( \liminf_{T \to \infty} \liminf_{\theta \to 0^+} \frac{1}{T} \rho_{k,y} (\theta) > 0 \) and \( 1/(\sqrt{T} \kappa) \to 0 \). The sign of the derivative is completely determined by that of \( \theta_t \), and this result implies the sign claim. Thus, we can consider the set of true predictors when we develop the asymptotic normality. Finally, the asymptotic normality directly follows from Theorem 1 of Bai and Ng (2006). This completes the proof.

**A3: Proof of Theorem 2**

Under Assumptions A–D, E and using Lemmas 1 and 2, we prove Theorem 2. We first evaluate

\[
B = T^{-1} \sum_{t=1}^{T} \left[ \rho_\tau (y_t - \hat{z}_t' \theta(\tau)) - \rho_\tau (y_t - z_t' \theta(\tau)) \right] \quad \text{where} \quad z_t = (f'_t, w'_t), \quad \hat{z}_t = (\hat{f}'_t, \hat{w}'_t) \quad \text{and} \quad \theta(\tau) = (\alpha(\tau)', \beta(\tau)')\).

We have

\[
B = T^{-1} \sum_{t=1}^{T} \left[ \rho_\tau (y_t - \hat{z}_t' \theta(\tau)) - \rho_\tau (y_t - z_t' \theta(\tau)) \right]
= T^{-1} \sum_{t=1}^{T} \left[ \rho_\tau (y_t - z_t' \theta(\tau) - (\hat{z}_t - z_t)' \theta(\tau)) - \rho_\tau (y_t - z_t' \theta(\tau)) \right]
\]
\[
\frac{-1}{T} \sum_{t=1}^{T} (\hat{z}_t - z_t)\psi_\tau (y_t - z_t^i(\tau)) + \frac{1}{T} \sum_{t=1}^{T} \int_{0}^{I(y_t - z_t^i(\tau) \leq s) - I(y_t - z_t^i(\tau) \leq 0)} ds
\]

\[
= B_1 + B_2.
\]

From the second line to the third line, a Knight's identity is used:

\[
\rho_\tau (u - \nu) - \rho_\tau (u) = -\nu \psi_\tau (u) + \int_{0}^{\nu} (I(u \leq s) - I(u \leq 0)) ds,
\]

with \( \psi_\tau (u) = \tau - I(u \leq 0) \).

The two terms can be evaluated as follows. The first term \( B_1 \) is

\[
B_1 \leq \frac{1}{T} \left\| \sum_{t=1}^{T} ||\hat{z}_t - z_t|| \cdot \psi_\tau (y_t - z_t^i(\tau)) \right\| \|\theta(\tau)\|
\]

\[
\leq \left( \frac{1}{T} \sum_{t=1}^{T} ||\hat{z}_t - z_t||^2 \right)^{1/2} \left( \frac{1}{T} \sum_{t=1}^{T} \psi_\tau (y_t - z_t^i(\tau))^2 \right)^{1/2} \|\theta(\tau)\|
\]

\[
= O_p \left( \frac{1}{\min \{N^{1/2}, T^{1/2} \}} \right).
\]

Next, we evaluate the term \( B_2 \)

\[
\frac{1}{T} \sum_{t=1}^{T} \int_{0}^{I(y_t - (z_t^i(\tau) \leq s) - I(y_t - z_t^i(\tau) \leq 0))} ds
\]

\[
= \frac{1}{T} \sum_{t=1}^{T} \int_{0}^{\sqrt{N}(\hat{z}_t - z_t)\psi_\tau (y_t - z_t^i(\tau))} (G(\xi_t + s) - G(\xi_t)) ds
\]

\[
= \frac{1}{T\sqrt{N}} \sum_{t=1}^{T} \int_{0}^{\sqrt{N}(\hat{z}_t - z_t)\psi_\tau (y_t - z_t^i(\tau))} (G(\xi_t + w/\sqrt{N}) - G(\xi_t)) dw
\]

\[
= \frac{1}{TN} \sum_{t=1}^{T} \sqrt{N}(\hat{z}_t - z_t)\psi_\tau (y_t - z_t^i(\tau)) (G(\xi_t + w/\sqrt{N}) - G(\xi_t)) dw
\]

\[
= \frac{1}{TN} \sum_{t=1}^{T} \sqrt{N}(\hat{z}_t - z_t)\psi_\tau (y_t - z_t^i(\tau)) g(\xi_t) dw + o_p(T^{-1})
\]

\[
= \frac{1}{2T} \sum_{t=1}^{T} g(\xi_t)(\hat{z}_t - z_t)\psi_\tau (y_t - z_t^i(\tau)) (\hat{z}_t - z_t) + o_p(T^{-1}),
\]

with \( \xi_t(\tau) = G^{-1}(\tau | z_t) \), and \( P(y_t \leq y | z_t) = G(y | z_t) \) is defined by the true quantile structure of \( y_t \). Thus,

\[
B_2 \leq \frac{1}{2} \|\theta(\tau)\|^2 \cdot \left( \frac{1}{T} \sum_{t=1}^{T} g(\xi_t)^2 \right)^{1/2} \cdot \left( \frac{1}{T} \sum_{t=1}^{T} ||\hat{z}_t - z_t||^2 \right)^{1/2}
\]

\[
= O_p \left( \frac{1}{\min \{N^{1/2}, T^{1/2} \}} \right). 
\]
These investigations imply that $B = O_p(1/ \min\{N^{1/2}, T^{1/2}\})$. Thus, $\hat{\theta}$ and $\tilde{\theta}$, defined as

$$
\tilde{\theta} = \arg\min_\theta \sum_{t=1}^{T} \rho_t(y_{t+h} - \alpha' f_t - \beta' w_t)^2 + p_{\lambda,y}(\theta),
$$

are asymptotically equivalent, $\|\hat{\theta} - \tilde{\theta}\| = O_p(1/ \min\{N^{1/2}, T^{1/2}\})$. Combining the variable selection consistency and the asymptotic normality of $\tilde{\theta}$ following Wu and Liu (2009), the proof is complete.

**A4: Proof of Theorem 3**

Here, we show that the MSFE is close to the expected in-sample squared error. From the definition of the PMSE, we have

$$
PMSE = \frac{1}{n} \sum_{j=1}^{n} (y_{t+h+j} - \hat{y}_{t+h+j})^2
$$

$$
= \frac{1}{n} (\mu + \epsilon - \hat{\mu})'(\mu + \epsilon - \hat{\mu})
$$

$$
= \frac{1}{n} L(\hat{\mu}) + \frac{1}{n} \epsilon' \epsilon + \frac{2}{n} \epsilon' (\mu - \hat{\mu}),
$$

where $L(\hat{\mu}) = (\mu - \hat{\mu})'(\mu - \hat{\mu})$, $\hat{y}_{t+h} = \hat{\alpha}' f_t + \hat{\beta}' w_t$ is the predicted value, $\mu = (\mu_{t+h+1}, ..., \mu_{t+h+n})'$, and $\hat{\mu} = (\hat{\mu}_{t+h+1}, ..., \hat{\mu}_{t+h+n})'$. From Assumption B, $E[\epsilon' (\mu - \hat{\mu})/n] = 0$. Also, $\epsilon' \epsilon/n$ is not relevant to the model specification, and $\frac{2}{n} \epsilon' (\mu - \hat{\mu})$ becomes the remaining term. Therefore, the minimization of the PMSE is equivalent to the minimization of $L(\hat{\mu})$. This completes the proof.

**Acknowledgment**

The authors would like to thank three anonymous referees for many excellent suggestions which have substantially improved the article. We are also grateful to them for pointing out several references. We would also like to thank Ruey Tsay for helpful advice and constructive comments.

**Funding**

This research was partially supported by a grant from the Seimeikai Foundation, Japan.

**References**

Amihud, Y. (2002). Illiquidity and stock returns: Cross-section and time-series effects. *Journal of Financial Markets* 5:31–56.

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

Ando, T., Bai, J. (2013a). Asset pricing with general multifactor structure. *Journal of Financial Econometrics*, forthcoming.

Ando, T., Bai, J. (2013b). Panel data models with grouped factor structure under unknown group membership, Working Paper. Columbia university.

Ando, T., Tsay, R. (2011). Quantile regression models with factor-augmented predictors and information criteria. *Econometrics Journal* 14:1–24.

Ando, T., Tsay, R. (2014). Predictive approach for selection of diffusion index models. *Econometric Reviews* 33:68–99.

Asness, C. (2011). Momentum in Japan: The exception that proves the rule. *Journal of Portfolio Management* 37:67–75.

Bai, J. (2003). Inferential theory for factor models of large dimensions. *Econometrica* 71:135–172.

Bai, J., Ng, S. (2002). Determining the number of factors in approximate factor models. *Econometrica* 70:191–221.

Bai, J., Ng, S. (2006). Confidence intervals for diffusion index forecasts and inference with factor-augmented regressions. *Econometrica* 74:1133–1150.

Bai, J., Ng, S. (2008). Extremum estimation when the predictors are estimated from large panels. *Annals of Economics and Finance* 9:201–222.
Koop, G., Potter, S. (2004). Forecasting in dynamic factor models using Bayesian model averaging. *Econometrics Journal* 7:550–565.

Kothari, S. P., Shanken, J. (1997). Book-to-market, dividend yield, and expected market returns: A time-series analysis. *Journal of Financial Economics* 44:169–203.

Kritzman, M., Li, Y., Page, S., Rigobon, R. (2010). Principal components as a measure of systemic risk. Working Paper, MIT.

Kubota, K., Takehara, H. (2003). Financial sector risk and the stock returns: Evidence from Tokyo Stock Exchange firms. *Asia-Pacific Financial Markets* 10:1–28.

Lamont, O. (1998). Earnings and expected returns. *Journal of Finance* 53:1563–1587.

Lettau, M., Ludvigson, S. (2001). Consumption, aggregate wealth, and expected stock returns. *Journal of Finance* 56:815–849.

Lo, A., MacKinlay, C. (1990). Data snooping biases in tests of financial asset pricing models. *Review of Financial Studies* 3:431–467.

Lu, X., Su, L. (2013). Shrinkage estimation of dynamic panel data models with interactive fixed effects, Working Paper, Singapore Management University.

Ludvigson, S. C., Ng, S. (2007). The empirical risk-return relation: A factor analysis approach. *Journal of Financial Economics* 83:171–222.

Pesaran, M. H., Timmermann, A. (2007). Selection of estimation window in the presence of breaks. *Journal of Econometrics* 137:134–161.

Pesaran, M. H., Pettenuzzo, D., Timmermann, A. (2006). Forecasting time series subject to multiple structural breaks. *Review of Economic Studies* 73:1057–1084.

Petkova, R. (2006). Do the Fama-and-French factors proxy for innovations in state variables? *Journal of Finance* 61:581–612.

Santos, T., Veronesi, P. (2006). Labor income and predictable stock returns. *Review of Financial Studies* 19:1–44.

Stock, J. H., Watson, M. W. (2002a). Forecasting using principal components from a large number of predictors. *Journal of the American Statistical Association* 97:1167–1179.

Stock, J. H., Watson, M. W. (2002b). Macroeconomic forecasting using diffusion indexes. *Journal of Business & Economic Statistics* 20:147–162.

Tibshirani, R. (1996). Regression shrinkage and selection via the lasso. *Journal of the Royal Statistical Society* B58:267–288.

Vassalou, M., Xing, Y. (2004). Default risk in equity returns. *Journal of Finance* 59:831–868.

Wu, Y., Liu, Y. (2009). Variable selection in quantile regression. *Statistica Sinica* 19:801–817.

Zhang, C.-H. (2010). Nearly unbiased variable selection under minimax concave penalty. *Annals of Statistics* 38:894–942.

Zou, H. (2006). The adaptive lasso and its oracle properties. *Journal of the American Statistical Association* 101:1418–1429.

Zou, H., Hastie, T. (2005). Regularization and variable selection via the Elastic Net. *Journal of the Royal Statistical Society* B67:301–320.

Zou, H., Hastie, T., Tibshirani, R. (2007). On the degrees of freedom of the lasso. *Annals of Statistics* 35:2173–2192.