Quantitative Finance

ISSN: 1469-7688 (Print) 1469-7696 (Online) Journal homepage: http://www.tandfonline.com/loi/rquf20

Online learning of time-varying stochastic factor structure by variational sequential Bayesian factor analysis

Hui ‘Fox’ Ling & Christian Franzen

Quantitative Finance, 2017
Vol. 17, No. 8, 1277–1304, http://dx.doi.org/10.1080/14697688.2016.1268708

Online learning of time-varying stochastic factor structure by variational sequential Bayesian factor analysis

HUI 'FOX' LING*† and CHRISTIAN FRANZEN‡

†Allianz Global Investors U.S. LLC, 1633 Broadway, New York, NY 10019, USA
‡Allianz Global Investors GmbH, Bockenheimer Landstr. 42-44, 60323 Frankfurt, Germany

(*Received 9 December 2014; accepted 25 November 2016; published online 17 February 2017*)

Investment tasks include forecasting volatilities and correlations of assets and portfolios. One of the tools widely utilized is stochastic factor analysis on a set of correlated time-series (e.g. asset returns). Published time-series factor models require either sufficiently wide time windows of observed data or numeric solutions by simulations. We developed a 'variational sequential Bayesian factor analysis' (VSBFA) algorithm to make online learning of time-varying stochastic factor structure. The VSBFA is an analytic filter to estimate unknown factor scores, factor loadings and residual variances. The covariance matrix of the time-series predicted by the VSBFA can be decomposed into loadings-based covariance and specific variances, and the former can be expressed by 'explanatory factors' such as systematic components of various financial market indices. We compared the VSBFA with the most practiced factor model relying on wide data windows, the rolling PCA (principal components analysis), by applying them to 9-year daily returns of 200 simulated stocks with the 'true' daily data-generating model completely known, and by using them to forecast volatilities of long-only and long/short global stock portfolios with 25-year monthly returns of more than 800 stocks worldwide. Accuracy of the forecast covariance matrices is measured by a (symmetrized) Kullback–Leibler distance, and accuracy of the forecast portfolio volatilities is measured by bias statistic, log-likelihood, Q-statistic, and portfolio volatility minimization. The factor-based covariance and specific variances predicted by the best VSBFA are significantly more accurate than those by the best rolling PCA.

Keywords: Time-varying factor models of time-series; Online learning of factor model parameters; Variational sequential Bayesian estimates; Variational Bayesian filter to estimate factor model parameters; Forecasting volatilities of portfolios of assets; Predictability of time-varying factor models

JEL Classification: C11, C38

1. Introduction

It is widely recognized that financial assets and portfolios of assets have exposures to a variety of economic, financial and other common factors. These factors are drivers of a large portion of returns and volatilities of the assets and portfolios. Both of the factors (or factor scores) and exposures (or factor loadings) of the assets are time-varying and stochastic. A predictive analysis of the stochastic factor structure is helpful in forecasting future returns and volatilities of the assets and portfolios. Factor analysis (FA) is at the center of investment portfolio management (Axioma 2011, Northfield 2012b, SunGard APT 2012, Ward 2012, numerous other publications). FA is also essential in econometric research (e.g. Stock and Watson 2011, Koopman and van der Wel 2013), image signal processing (Wijnholds *et al.* 2013), EEG (electroencephalogram) analysis (Motta and Ombao 2012), and common spatial pattern analysis (Kang and Choi 2012), to name just a few.

Factors are part of known data inputs in some models, but are among unknown variables to be estimated in others. The unknown loadings are assumed to be constant over time in some FA cases, but must be time-varying in others. A major FA application is to predict the covariance matrix of a set of time-series, but the covariance can also be estimated without a factor model. Therefore, numerous FA

*Corresponding author. Email: fox.ling@allianzgi.com

methods and related approaches in the literature may be grouped into five buckets.

The first group is to take known factors and estimate constant loadings. It includes the ordinary least squares (OLS) regressions, factor-based thresholding estimation of the covariance matrix (Fan et al. 2011), and high-dimensional factor models with high-frequency (HF) data and block-diagonal residual covariance (Fan et al. 2016). The second group is to take known factors and estimate time-varying loadings. It includes the simultaneous graphical dynamic linear model (SGDLM) with common factors as the parental set (Gruber and West 2016), the dynamic dependence network model (DDNM) with a parental set of common factors (Zhao et al. 2015), and variational approximation of sequential Bayesian inference (Ling and Stone 2016).

The third group is to estimate the unknown factors and constant loadings jointly. It includes the principal components analysis (PCA) factor model (SunGard APT 2012), asymptotic PCA (Axioma 2011), PCA augmented with additional Bayesian analysis (Northfield 2012a), the thresholding principal orthogonal complements method (Fan, Liao and Mincheva 2013), maximum likelihood (ML) dynamic factor analysis (Jungbacker and Koopman 2014), expectation-maximization (EM) factor analysis (Dempster, Laird and Rubin 1977), Bayesian factor analysis (Rowe 2003), the ML-EM-Bayesian factor model (Ward 2012), variational Bayesian factor analysis (VBFA) (Ghahramani and Beal 2000), VBFA algorithm (Beal 2003), VBFA approach (Nielsen 2004), VBFA models (Luttinen and Ilin 2009, 2010), multi-step VBFA (Zhao and Yu 2009), variational Bayesian PCA (Bishop 1999), Markov chain Monte Carlo (MCMC) Bayesian PCA (Ding et al. 2011), the dynamic factor model (DFM) (Stock and Watson, 2011), and DFMs (Barhoumi, Darné and Ferrara 2013). The fourth group is to estimate the unknown factors and time-varying loadings jointly. It includes MCMC Bayesian factor analysis (Lopes and Carvalho 2007), Gibbs sampling for DFM (Del Negro and Otrok 2008), a model of dynamic latent factors and time-varying sparse loadings with MCMC solutions (Zhou, Nakajima and West 2014), the VSBFA algorithm (this paper), and all of the approaches in the third group when using moving or rolling data windows.

The fifth group is to estimate the unknown time-varying covariance matrix directly without using common factors. It includes state-space model for symmetric positive-definite matrices (Windle and Carvalho 2014), the SGDLM with parental sets specific to individual time-series (Gruber and West 2016), the DDNM with specific parental sets (Zhao *et al*. 2015), a vast volatility matrix for HF data (Fan *et al.* 2012), and a number of high-dimensional sparse volatility matrices for HF data (Wang and Zou 2014).

When applying any of the factor models of the third group to hundreds or thousands of assets in order to find out the 'current' factor structure, a sufficiently wide time window is required to contain a sufficient history of asset returns. The factors, loadings and variances estimated this way are often more relevant to the average behaviour of the assets inside the time window and, thus, less relevant to the 'current time'. An exponentially decayed weighting may be applied to the historic data in order to make the FA relatively more relevant to the current market structure (e.g. Axioma 2011). Even though making time-varying forecasts with the rolling-window or time-weighted factor analyses are supported by numerous empirical studies, they are not fully anchored on rigorous theoretical foundations. The published time-varying factor models in the fourth group (Lopes and Carvalho 2007, Del Negro and Otrok 2008, Zhou, Nakajima and West 2014) are theoretically well defined. But their numeric solutions by Monte Carlo simulations make them harder to apply to a high-dimensional analysis.

Various methods estimating a vast volatility matrix for HF financial data have been developed in the current decade (Wang and Zou 2014, Fan, Furger and Xiu 2016). The use of HF data is also effective for forecasting time-varying covariance matrices assuming they are continuous in time, because the relatively short forecasting horizons are well matched by the similarly short data windows (Fan, Li and Yu 2012). To deal with microstructure noise, non-synchronization and irregular time spaces in the HF data, many approaches have been suggested (Wang and Zou 2014) which include previous tick, refresh time, generalized sampling time, two time scale realized (co)volatility, multi-scale realized (co)volatility, realized kernel estimation, quasi-ML estimation, a pre-averaging estimator, HY estimator, and sparsity constraints on the volatility matrices themselves. For high-dimensional estimation, concentration inequalities for convergence rates were established (Fan *et al*. 2012), and a more realistic sparsity, the block-diagonal covariance of factor model residuals, was proposed (Fan *et al*. 2016). Limitations of the HF methods are due to the facts that not all assets have HF data, not all HF data are available to every market participant, and data frequencies of relatively illiquid assets may not be high enough for the size of the asset universe.

We have developed a 'variational sequential Bayesian factor analysis (VSBFA)' algorithm to perform online learning of time-varying stochastic factor scores, factor loadings and residual variances. Online learning estimates the parameters or distributions of factors, loadings and variances at time t based on the previous estimates made at $t-1$ plus the observed data newly obtained at time t. In our Bayesian analysis, 'sequential priors' incorporate the previous estimates, likelihood contains the current observations, and 'variational approximating posteriors' give rise to the new estimates. The VSBFA is an analytic Bayesian filter, which essentially assigns adaptive weights to historic data instead of mechanical ones such as equal or exponential weights. Our VSBFA research was encouraged by the published contributions to variational Bayesian inference from Ghahramani and Beal (2000), Beal (2003), Nielsen (2004), Luttinen and Ilin (2009, 2010), Zhao and Yu (2009), Shutin *et al*. (2011), and Kang and Choi (2012).

In the academic literature, factor models of constant loadings are compared or selected by information criteria such as the Akaike information criteria (AICs) and the Bayesian information criteria (BICs) which are log likelihood penalized, based on theories and assumptions, by functions of the number of parameters adjusted with sample size and with constant covariance of residuals (Bai and Ng 2002, Lopes and West 2004). For models of time-varying loadings and covariance estimated by Bayesian filtering, further research is needed to determine effective sample size, 'effective' covariance matrix and, therefore, relevant information criteria. In our simulation case study, the (symmetrized) Kullback–Leibler distance between estimated and 'true' covariance matrices is used to compare the models. In our empirical case study, the bias statistic, log-likelihood, Q-statistic, and portfolio volatility minimization are applied to out-of-sample forecasts in volatilities of long-only and long/short random portfolios in order to compare the models. These case studies demonstrate that 'online factor learning with sequential Bayes and variational approximation' by the VSBFA achieves higher predictive power than the rolling-window FA, makes more accurate predictions for factor-based covariance and specific variances, and applies to high-dimensional data-sets which may prohibit the use of Monte Carlo simulations.

The rest of the paper is organized as follows. Section 2, 'A variational sequential Bayesian analysis of time-varying factor models', introduces the concepts and basic assumptions. Section 3, 'VSBFA algorithm and predicted covariance matrix', summarizes the entire set of statements of the VSBFA algorithm, the covariance matrix of VSBFA estimates, and explanatory factors. Section 4, 'Evaluating VSBFA with data of a known factor model', evaluates the VSBFA with 9-year daily returns of 200 simulated stocks. Section 5, 'A VSBFA model of international stocks', describes a data-set of 25-year monthly returns of more than 800 stocks worldwide, a VSBFA factor model of the stocks, and volatilities of stock portfolios. Section 6, 'Forecasting volatilities of global stock portfolios', compares predictive powers of the VSBFA and rolling PCA by their forecasts in volatilities of random portfolios. Section 7, 'Explanatory effects on portfolio volatilities', illustrates a set of explanatory factors and their effects on portfolio volatilities. Section 8, 'Conclusion', summarizes the VSBFA method. Appendix A, 'Details of VSBFA algorithm', derives in detail the VSBFA formulas. Appendix B, 'Predicted covariance matrix based on VSBFA estimates', derives the covariance matrix predicted by the VSBFA, use of initial value adjustments, relevance levels of the factors, and the explanatory factors.

2. A variational sequential Bayesian analysis of time-varying factor models

Assume a need of time-varying FA on a large number of observed stochastic time-series $r_j(t), j = 1, 2, \ldots, n \gg 1$, such as returns of assets worldwide. Estimated time-varying mean and variance of each time-series based on all information available by time t can be denoted as $\mu_j(t|t)$ and $\sigma_j^2(t|t)$. Denote $n \times 1$ vectors $r_t = [r_1(t), r_2(t), \ldots, r_n(t)]^T$ and $\mu_{t|t} = [\mu_1(t|t), \mu_2(t|t), \ldots, \mu_n(t|t)]^T$, $n \times n$ diagonal matrix of standard deviations $D_{t|t} = diag([\sigma_1(t|t), \sigma_2(t|t), \ldots, \sigma_n(t|t)]^T)$, and $n \times 1$ vector of standardized time-series,

$$y_t = D_{t|t}^{-1}(r_t - \mu_{t|t}), \tag{1}$$

where the superscript T indicates the transposition of a vector or matrix, $diag(v)$ is a diagonal matrix formed by the vector v, and the superscript -1 denotes the inverse of a matrix. A time-varying factor model of the standardized time-series y_t with m common factors, $m \ll n$, is,

$$y_t = X_t f_t + e_t, \tag{2}$$

where f_t is an $m \times 1$ vector of unobservable factor scores, X_t an $n \times m$ matrix of unobservable factor loadings, and e_t an $n \times 1$ vector of unobservable residual errors. In a frequentist analysis, covariance of f_t, the loadings X_t and variance of e_t are unknown non-random parameters to be estimated. In a Bayesian analysis, however, the factors f_t, the covariance of f_t, the loadings X_t, covariance of columns of X_t, the residuals e_t and the variance of e_t are unknown random variables whose joint distribution is to be estimated.

In this paper, following assumptions, to be explained or defined in the Appendices, are applied to our Bayesian analysis. The factors f_t, jth column $X_{j-col,t}$ of the loadings X_t and the residuals e_t are random vectors of multivariate normal distributions given covariance matrices: $f_t \sim N(E(f_t), P_t)$, $X_{j-col,t} \sim N(E(X_{j-col,t}), U_{j,t})$ and $e_t \sim N(0, R_t)$. The $m \times m$ covariance matrix P_t, the $n \times n$ covariance matrix $U_{j,t}$ and the $n \times n$ covariance matrix R_t are diagonal random matrices of 'diagonal inverse-gamma distributions'. Dynamic evolutions of the random elements of P_t, X_t, $\{U_{j,t}\}$ and R_t are of random walks. The random elements of f_t, P_t, X_t, $\{U_{j,t}\}$ and R_t are mutually independent. With these assumptions, equation (2) represents a time-varying factor model of stochastic factors, loadings and variances.

A sequential Bayesian analysis for the joint distribution at time t of the time-varying factor model (2) is given by (A.21) in appendix A as,

$$\begin{aligned} &p(y_t, f_t, P_t, X_t, \{U_{j,t}\}, R_t) \\ &\quad = p(y_t|X_t, f_t, R_t)p(f_t|P_t)p(P_t)p(X_t|\{U_{j,t}\})p(\{U_{j,t}\})p(R_t), \end{aligned} \tag{3}$$

where $p(y_t|X_t, f_t, R_t)$ is the likelihood and $p(f_t|P_t)$, $p(P_t)$, $p(X_t|\{U_{j,t}\})$, $p(\{U_{j,t}\})$ and $p(R_t)$ are the priors whose expected values equal to those of the variables, $f_{t|t-1}$, $P_{t|t-1}$, $X_{t|t-1}$, $\{U_{j,t|t-1}\}$ and $R_{t|t-1}$, predicted at the earlier time $t-1$. The objective of a sequential Bayesian analysis is to estimate the posterior $p(f_t, P_t, X_t, \{U_{j,t}\}, R_t|y_t)$ based on the joint distribution (3).

Since the posterior of the time-varying factor model (2) is analytically intractable even with our simplest realistic assumptions about the likelihood and priors, either a stochastic or an analytic approximation has to be applied. A stochastic approximation is to estimate the posterior by random sampling. It is asymptotically accurate: when the sample size is larger and larger, the estimates can be more and more accurate. An analytic approximation is to estimate the posterior by tractable formulas. Variational approximation (VA) is an analytic one (Bishop 2006, Tzikas, Likas and Galatsanos 2008, Ormerod and Wand 2010, Grimmer 2011). It is to approximate the posterior by a factorized distribution, i.e. a product of independent simpler distributions. The assumption of factorization, or independence between (groups of) variables, is fairly reasonable for many applications in economic, financial and investment analyses (e.g. Ling and Stone 2016). The set of optimal VA distributions are interrelated in nonlinear ways, but they can always be reached by iterations with reasonable starting points (Bishop 2006, Tzikas *et al*. 2008, Ormerod and Wand 2010). If the size of random sampling is large enough relative to the size of factor model, the stochastic approximation methods (Lopes and Carvalho 2007, Del Negro and Otrok 2008, Zhou, Nakajima and West 2014) should, in theory, produce more accurate solutions than any of the VAs.

3. VSBFA algorithm and predicted covariance matrix

Our VSBFA algorithm is to make online learning of the time-varying stochastic factor model (2): estimating the distributions of f_t, P_t, X_t, $\{U_{j,t}\}$ and R_t at time t by the previously estimated expected values $P_{t-1|t-1}$, $X_{t-1|t-1}$, $\{U_{j,t-1|t-1}\}$ and $R_{t-1|t-1}$ plus the new observations y_t. An important application of the factor model (2) in financial analysis is to predict time-varying covariance matrix of a large number of time-series, e.g. daily, weekly or monthly returns of a huge universe of assets. In addition, using a set of real financial indicators to explain the predicted covariance matrix may help investors gain understandable and actionable insights.

3.1. Statements of the VSBFA algorithm

There are three groups of VSBFA statements for each time step t. All of them are explicitly listed here and are easy to be implemented in, for example, MATLAB (www.mathworks.com).

Beginning part of the VSBFA algorithm is initialization. Figure 1 lists the set of initialization formulas, where $diag(A)$ is a diagonal matrix formed by the diagonal elements of matrix A, and $tr(A)$ is the trace of matrix A. The initial values $f_{t|t}^{(0)}$, $P_{t|t}^{(0)}$, $X_{t|t}^{(0)}$, $\{U_{j,t|t}^{(0)}\}$ and $R_{t|t}^{(0)}$ for estimating the distributions of f_t, P_t, X_t, $\{U_{j,t}\}$ and R_t are calculated by (A.37) to (A.41), using previous estimates $P_{t-1|t-1}$, $X_{t-1|t-1}$, $\{U_{j,t-1|t-1}\}$ and $R_{t-1|t-1}$. To prevent the estimated factors and loadings from over- and under-fitting the data, relative levels of the initial variances $\{U_{j,t|t}^{(0)}\}$ and $R_{t|t}^{(0)}$ are adjusted by (B.27), (B.28) and (B.29), in appendix B, to aim at a prescribed 'residual-to-specific variance ratio target h', $0 \ll h < 1$. According to the discussions in appendix section B.2, the over- and under-fitting is prevented by maintaining a balanced 'randomness allocation' between $\{U_{j,t|t}^{(0)}\}$ and $R_{t|t}^{(0)}$. The function of the residual-to-specific variance ratio target h is similar to that of the 'error reduction target g' for the variational sequential Bayesian solution to state space models by Ling and Stone (2016). In addition, if the columns of estimated loadings $X_{t|t}$ need to be nearly orthogonal, the initial loadings columns $\{X_{j-col,t|t}^{(0)}\}$ can be orthogonalized by (B.30) and (B.31). Then, two additional initial values needed later are assigned by (A.42) and (A.43).

Middle part of the VSBFA algorithm contains seven VA iteration statements. Figure 2 lists the set of seven VA iteration equations (A.44) to (A.50), with the superscript (v) as iteration index, $v = 1, 2, \ldots, L$. The parameter T_0 is an 'effective number of data points' supporting the initial variances $P_{t|t}^{(0)}$, $\{U_{j,t|t}^{(0)}\}$ and $R_{t|t}^{(0)}$. For an analysis with daily, weekly or monthly data, $5 \le T_0 \le 20$ can serve as a sensible choice. A reasonable example of the number of VA iterations is $L = 10$.

Ending part of the VSBFA algorithm takes the results of the last VA iteration as the new estimates. Figure 3 lists these estimates: $f_{t|t}$, $P_{t|t}$, $X_{t|t}$, $\{U_{j,t|t}\}$ and $R_{t|t}$ by (A.51) to (A.55). They are estimated expected values of the distributions of f_t, P_t, X_t, $\{U_{j,t}\}$ and R_t. Additional estimates needed by initial values at next time step $t+1$ are calculated by (A.56) to (A.59).

When the next observations y_{t+1} become available, the new estimates of the stochastic factor structure for time $t+1$ can be obtained by applying the VSBFA algorithm (i.e. the statements in figures 1–3) again. Therefore, the VSBFA is a variational Bayesian filter. In asset and investment management practices, ability to add new and delete old assets on demand is a necessity of a factor model. The basic VSBFA algorithm statements listed here assume no additions and deletions of individual time-series. Further research is needed in this aspect.

3.2. Predicted covariance matrix

With the VSBFA estimates $P_{t|t}$, $X_{t|t}$, $\{U_{j,t|t}\}$ and $R_{t|t}$ by equations (b) to (e) in figure 3 for the standardized time-series y_t in (1), as summarized by equations listed in figure 4, factor loadings matrix $X_{t|t}^{(r)}$ and diagonal covariance matrices $\{U_{j,t|t}^{(r)}\}$ and $R_{t|t}^{(r)}$ of the original time-series r_t can be defined by (B.12), (B.13) and (B.15). Then predicted time-varying covariance matrix $C_{t+1|t}$ of the time-series r_t can be expressed by (B.16) or (B.17) to (B.20). The VSBFA-based covariance $C_{t+1|t}$ is a sum of three components. The first one, $C_{t+1|t}^{(X)}$, is loadings-based covariance which fully accounts for predicted covariance values of all pairs of different time-series $r_j(t)$ and $r_k(t)$, $k \ne j$. The sum of the last two, $C_{t+1|t}^{(Spec)} = C_{t+1|t}^{(U)} + C_{t+1|t}^{(R)}$, is specific variance because it is a diagonal matrix contributing only to the predicted variance of individual timer-series $r_j(t)$.

A related topic is importance levels of the common factors whose definitions are listed in figure 4 as well. For the standardized time-series y_t, contribution from the jth stochastic factor alone to the factor model (2), or importance or relevance level of the jth factor, over a given time period $t \in [t_1, t_2]$ is F_{j,t_1,t_2} by (B.32). Total contribution from all m factors, or relevance level of the m-factor model, is $F_{t_1,t_2}^{(m)}$ by (B.33). Incremental relevance level between the model of m-1 factors and that of m factors is $\Delta F_{t_1,t_2}^{(m)}$ by (B.34), which is expected to become smaller as the number of factors m increases. The incremental relevance level $\Delta F_{t_1,t_2}^{(m)}$ can help make a practical decision when to stop introducing an additional common factor.

(a) $f_{t|t}^{(0)} = 0$,

(b) $P_{t|t}^{(0)} = n^{-1}\, tr(\, diag(\, X_{t-1|t-1}\, P_{t-1|t-1}\, X_{t-1|t-1}^{T}\,)^{-1}\, S_{t-1|t-1}^{(X)}\,)\, P_{t-1|t-1}$,

(c) $X_{t|t}^{(0)} = X_{t-1|t-1}$

(d) $U_{j,t|t}^{(0)} = n^{-1}\, tr(\, diag(\, \sum_{j=1}^{m} P_{j,t-1|t-1}\, U_{j,t-1|t-1} + R_{t-1|t-1}\,)^{-1}\, S_{t-1|t-1}^{(Err)}\,)\, U_{j,t-1|t-1}$,

$j = 1, 2, \cdots, m,$

(e) $R_{t|t}^{(0)} = n^{-1}\, tr(\, diag(\, \sum_{j=1}^{m} P_{j,t-1|t-1}\, U_{j,t-1|t-1} + R_{t-1|t-1}\,)^{-1}\, S_{t-1|t-1}^{(Err)}\,)\, R_{t-1|t-1}$,

(f) $W_{t|t}^{(0)} = \sum_{j=1}^{m} P_{j,t|t}^{(0)}\, U_{j,t|t}^{(0)} + R_{t|t}^{(0)}$,

(g) $U_{j,t|t}^{(0)} = \{\, (\, W_{t|t}^{(0)} - R_{t|t}^{(0)}\,)^{-1}\, [\, (1-h)\, W_{t|t}^{(0)}\,]\, \}\, U_{j,t|t}^{(0)}$,

(h) $R_{t|t}^{(0)} = [\, (R_{t|t}^{(0)})^{-1}\, (\, h\, W_{t|t}^{(0)}\,)\,]\, R_{t|t}^{(0)}$,

(i) $Q_{t|t}^{(0)} = P_{t|t}^{(0)}$,

(j) $V_{j,t|t}^{(0)} = U_{j,t|t}^{(0)}$, $j = 1, 2, \cdots, m$.

Figure 1. Initialization formulas of the VSBFA algorithm. Statements (a) to (e) are initial value assignments by (A.37) to (A.41), where $S_{t-1|t-1}^{(X)}$ and $S_{t-1|t-1}^{(Err)}$ are calculated by equations (h) and (i) in figure 3. Statements (f) to (h) rescale $\{U_{j,t|t}^{(0)}\}$ and $R_{t|t}^{(0)}$ by (B.27) to (B.29) given the parameter h ($0 \ll h < 1$). Statements (i) and (j) are for additional initial values by (A.42) and (A.43). Orthogonalization procedure (B.30) and (B.31) can be inserted after the assignment (c).

(a) $Q_{j,t|t}^{(v)} = \{\, (P_{j,t|t}^{(v-1)})^{-1} + tr(\, (R_{t|t}^{(v-1)})^{-1}\, [\, V_{j,t|t}^{(v-1)} + X_{j-col,t|t}^{(v-1)}\, (X_{j-col,t|t}^{(v-1)})^{T}\,]\,)\, \}^{-1}$,

(b) $f_{j,t|t}^{(v)} = Q_{j,t|t}^{(v)}\, \{\, (P_{j,t|t}^{(v-1)})^{-1}\, f_{j,t|t}^{(0)}$

$+ tr(\, (R_{t|t}^{(v-1)})^{-1}\, [\, y_t - \sum_{k\neq j} f_{k,t|t}^{(v-1)} X_{k-col,t|t}^{(v-1)}\,]\, (X_{j-col,t|t}^{(v-1)})^{T}\,)\, \}$,

(c) $P_{j,t|t}^{(v)} = (T_0+1)^{-1}\, [\, T_0\, P_{j,t|t}^{(0)} + Q_{j,t|t}^{(v)} + (\, f_{j,t|t}^{(v)} - f_{j,t|t}^{(0)})^2\,]$,

(d) $V_{j,t|t}^{(v)} = U_{j,t|t}^{(v-1)} - U_{j,t|t}^{(v-1)}\, \{\, U_{j,t|t}^{(v-1)} + R_{t|t}^{(v-1)} / \, [\, Q_{j,t|t}^{(v)} + (f_{j,t|t}^{(v)})^2\,]\, \}^{-1}\, U_{j,t|t}^{(v-1)}$,

(e) $X_{j-col,t|t}^{(v)} = V_{j,t|t}^{(v)}\, \{\, (U_{j,t|t}^{(v-1)})^{-1}\, X_{j-col,t|t}^{(0)} + (R_{t|t}^{(v-1)})^{-1}\, f_{j,t|t}^{(v)}\, [\, y_t - \sum_{k\neq j} f_{k,t|t}^{(v)}\, X_{k-col,t|t}^{(v-1)}\,]\, \}$,

(f) $U_{j,t|t}^{(v)} = (T_0+1)^{-1}\, diag(\, T_0\, U_{j,t|t}^{(0)} + V_{j,t|t}^{(v)}$

$+ (\, X_{j-col,t|t}^{(v)} - X_{j-col,t|t}^{(0)}\,)\, (\, X_{j-col,t|t}^{(v)} - X_{j-col,t|t}^{(0)}\,)^{T}\,)$,

(g) $R_{t|t}^{(v)} = (T_0+1)^{-1}\, diag(\, T_0\, R_{t|t}^{(0)}$

$+ \sum_{j=1}^{m} [\, Q_{j,t|t}^{(v)}\, V_{j,t|t}^{(v)} + Q_{j,t|t}^{(v)}\, X_{j-col,t|t}^{(v)}\, (X_{j-col,t|t}^{(v)})^{T} + (f_{j,t|t}^{(v)})^2\, V_{j,t|t}^{(v)}\,]$

$+ (\, y_t - X_{t|t}^{(v)} f_{t|t}^{(v)}\,)\, (\, y_t - X_{t|t}^{(v)} f_{t|t}^{(v)}\,)^{T}\,)$.

Figure 2. Variational approximation (VA) iteration equations of the VSBFA algorithm. Seven statements (a) to (g) perform the VA iteration by (A.44) to (A.50), where $j = 1, 2, \ldots, m$ and the iteration index $v = 1, 2, \ldots, L$.

3.3. *Explanatory indicators and explanatory factors*

Equations (d) to (g) in figure 4 explain the predicted covariance matrix $C_{t+1|t}$ of the time-series r_t by m stochastic factors in terms of $P_{t|t}$, $X_{t|t}^{(r)}$, $\{U_{j,t|t}^{(r)}\}$ and $R_{t|t}^{(r)}$ which are lack of practical and actionable explanations. Fortunately, the factor-based covariance $C_{t+1|t}$ can be explained by a set of l real financial or market time-series $r_t^{(EI)} = [r_1^{(EI)}(t), r_2^{(EI)}(t), \ldots, r_l^{(EI)}(t)]^T$, here the superscript ($EI$) is for 'time-series serving as explanatory indicators whose systematic components are an alternative set of common factors serving as explanatory factors'. In an analysis on a global market of n stocks, l explanatory indicators and m stochastic

(a) $f_{t|t} = f^{(L)}_{t|t}$,

(b) $P_{t|t} = P^{(L)}_{t|t}$,

(c) $X_{t|t} = X^{(L)}_{t|t}$,

(d) $U_{j,t|t} = U^{(L)}_{j,t|t}$, $j = 1, 2, \cdots, m$,

(e) $R_{t|t} = R^{(L)}_{t|t}$,

(f) $y_{t|t} = X_{t|t} [(X^T_{t|t} X_{t|t})^{-1} (X^T_{t|t} y_t)]$,

(g) $e_{t|t} = y_t - y_{t|t}$,

(h) $S^{(X)}_{t|t} = (T_0 + 1)^{-1} diag(T_0 S^{(X)}_{t-1|t-1} + y_{t|t} y^T_{t|t})$,

(i) $S^{(Err)}_{t|t} = (T_0 + 1)^{-1} diag(T_0 S^{(Err)}_{t-1|t-1} + e_{t|t} e^T_{t|t})$.

Figure 3. Estimation outcomes of the VSBFA algorithm. Statements (a) to (e) take the results of the last VA iteration as the VSBFA estimates by (A.51) to (A.55). Calculations (f) to (i) by (A.56) to (A.59) are needed for the initial values by equations (b), (d) and (e) in figure 1 at next time step $t + 1$.

(a) $X^{(r)}_{t|t} = D_{t|t} X_{t|t}$,

(b) $U^{(r)}_{j,t|t} = D^2_{t|t} U_{j,t|t}$, $j = 1, 2, \cdots, m$,

(c) $R^{(r)}_{t|t} = D^2_{t|t} R_{t|t}$,

(d) $C^{(X)}_{t+1|t} = X^{(r)}_{t|t} P_{t|t} (X^{(r)}_{t|t})^T$,

(e) $C^{(U)}_{t+1|t} = U^{(r)}_{t|t} = \sum_{j=1}^m P_{j,t|t} U^{(r)}_{j,t|t}$,

(f) $C^{(R)}_{t+1|t} = R^{(r)}_{t|t}$,

(g) $C_{t+1|t} = C^{(X)}_{t+1|t} + C^{(U)}_{t+1|t} + C^{(R)}_{t+1|t} = C^{(X)}_{t+1|t} + C^{(Spec)}_{t+1|t}$,

(h) $F_{j,t_1,t_2} = \{ \sum_{t=t_1}^{t_2} P_{j,t|t} [tr(U_{j,t|t}) + (X^T_{j-col,t|t} X_{j-col,t|t})] \}$
$/ \sum_{t=t_1}^{t_2} tr(X_{t|t} P_{t|t} X^T_{t|t} + U_{t|t} + R_{t|t})$,

(i) $F^{(m)}_{t_1,t_2} = [\sum_{t=t_1}^{t_2} tr(X_{t|t} P_{t|t} X^T_{t|t} + U_{t|t})]$
$/ \sum_{t=t_1}^{t_2} tr(X_{t|t} P_{t|t} X^T_{t|t} + U_{t|t} + R_{t|t})$,

(j) $\Delta F^{(m)}_{t_1,t_2} = F^{(m)}_{t_1,t_2} - F^{(m-1)}_{t_1,t_2} > 0$.

Figure 4. Time-varying covariance matrix predicted by the VSBFA estimates. Expressions (a) to (c) are factor model variables estimated by (B.12), (B.13) and (B.15) for the original time-series r_t. Expressions (d) to (g) are components and predicted covariance matrix by (B.18) to (B.20) and (B.17) for the time-series r_t. Expressions (h) and (i) are relevance levels of a single factor by (B.32) and of the factor model by (B.33). Expression (j) is incremental relevance level of the factor models by (B.34).

factors, $n \gg l$ and $n \gg m$, for example, the explanatory indicators may include values of, and differences between, region, country, sector, style and volatility indices and valuation and momentum indicators.

(a) $y^{(EI)}_t = (D^{(EI)}_{t|t})^{-1} (r^{(EI)}_t - \mu^{(EI)}_{t|t})$,

(b) $\begin{pmatrix} y_t \\ y^{(EI)}_t \end{pmatrix} = \begin{pmatrix} X_t \\ X^{(EI)}_t \end{pmatrix} f_t + \begin{pmatrix} e_t \\ e^{(EI)}_t \end{pmatrix}$,

(c) $X^{(EI)(r)}_{t|t} = D^{(EI)}_{t|t} X^{(EI)}_{t|t}$,

(d) $f^{(Expl)}_t = X^{(EI)(r)}_{t|t} f_t = D^{(EI)}_{t|t} X^{(EI)}_{t|t} f_t$,

(e) $P^{(Expl)}_{t|t} = X^{(EI)(r)}_{t|t} P_{t|t} (X^{(EI)(r)}_{t|t})^T$,

(f) $(X^{(EI)(r)}_{t|t})^+ = [(X^{(EI)(r)}_{t|t})^T X^{(EI)(r)}_{t|t}]^{-1} (X^{(EI)(r)}_{t|t})^T$,

(g) $X^{(Expl)}_{t|t} = X^{(r)}_{t|t} (X^{(EI)(r)}_{t|t})^+$,

(h) $C^{(X)}_{t+1|t} = C^{(Expl)}_{t+1|t} = X^{(Expl)}_{t|t} P^{(Expl)}_{t|t} (X^{(Expl)}_{t|t})^T$,

(i) $G^{(l,m)}_{t_1,t_2} = [\sum_{t=t_1}^{t_2} tr(X^{(EI)}_{t|t} P_{t|t} (X^{(EI)}_{t|t})^T)]$
$/ \sum_{t=t_1}^{t_2} tr(X^{(EI)}_{t|t} P_{t|t} (X^{(EI)}_{t|t})^T + U^{(EI)}_{t|t} + R^{(EI)}_{t|t}))$.

Figure 5. Explanatory indicators and explanatory factors. When explanatory indicators $r^{(EI)}_t$ in (a) or (B.35) are part of a joint factor model (b) or (B.36), their loadings $X^{(EI)(r)}_{t|t}$ is (c) or (B.37). Systematic part of the indicators are explanatory factors $f^{(Expl)}_t$ by (d) or (B.42) with non-diagonal covariance matrix $P^{(Expl)}_{t|t}$ by (e) or (B.39). Loadings $X^{(Expl)}_{t|t}$ of the data r_t to the factors $f^{(Expl)}_t$ is (g) or (B.40) with pseudoinverse (f) or (B.38). Therefore, the loadings-based covariance component $C^{(X)}_{t+1|t}$ by equation (d) in figure 4 can be explained with the explanatory factors in terms of (h) or (B.41). Information utilization rate of the explanatory factors is (i) or (B.43).

Figure 5 lists the equations about the explanatory indicators and explanatory factors. With an $l \times m$ factor loadings matrix $X^{(EI)(r)}_{t|t}$ of the indicators $r^{(EI)}_t$ estimated by (B.35) to (B.37), 'explanatory factor scores' $f^{(Expl)}_t$ by (B.42) are systematic components of the indicators $r^{(EI)}_t$. Covariance matrix of $f^{(Expl)}_t$ is $P^{(Expl)}_{t|t}$ by (B.39) which is no longer a diagonal matrix in general. Assuming the loadings matrix $X^{(EI)(r)}_{t|t}$ is of full rank and $l \geq m$, its pseudoinverse is $(X^{(EI)(r)}_{t|t})^+$ by (B.38). Then 'explanatory factor loadings matrix' of the data time-series r_t to the explanatory factors $f^{(Expl)}_t$ is $X^{(Expl)}_{t|t}$ by (B.40). The loadings-based component $C^{(X)}_{t+1|t}$ of the covariance matrix $C_{t+1|t}$ can then be explained by the explanatory factors in terms of (B.41). The explanatory factors $f^{(Expl)}_t$, i.e. the systematic components of the indicators $r^{(EI)}_t$, are indeed an alternative set of common factors for the data time-series r_t. Since only the systematic part of the indicators serves as the explanatory factors, information utilization rate of the l indicators within the m-factor model over the given time period $t \in [t_1, t_2]$ is $G^{(l,m)}_{t_1,t_2}$ by (B.43).

4. Evaluating VSBFA with data of a known factor model

It is generally preferred to evaluate a FA method using a set of simulated data with its true data-generating model completely known.

4.1. True model and data

To make a simulation as relevant to a real financial market as possible, a time-varying factor model based on 9-year daily returns of 200 stocks in Russell Top 200 Index is used as the data-generating model. The 200 stocks are chosen from those included in the index in the years of 2010 to 2014. Daily total returns (in decimal) of these stocks from January 2006 to December 2014, obtained from FactSet (www.factset.com), are ranging from −0.369 to +0.383, have an average of 0.0006, median of 0.0005, standard deviation of 0.0215, skewness of 0.370 and kurtosis of 18.70. A 10-factor rolling PCA is applied to the data to come up with a time-varying factor model, which is used as the 'true' data-generating model of the simulated data.

When the number of time-series $n = 200$ and number of factors $m = 10$, the parameter of the rolling PCA model is its window size $T_{MW} > m$. Denoting the decimal returns of stocks in day t by vector r_t. The time-varying PCA factor model for time t is constructed by the T_{MW} returns r_k within the trailing window $k \in [t - T_{MW} + 1, t]$. The estimated mean vector $\mu_{t|t;T_{MW}}$, diagonal standard deviation matrix $D_{t|t;T_{MW}}$ and standardized return vector $y_{k|t;T_{MW}}$ are

$$\mu_{t|t;T_{MW}} = \frac{1}{T_{MW}} \sum_{k=t-T_{MW}+1}^{t} r_k, \tag{4.a}$$

$$D_{t|t;T_{MW}} = \left\{ \frac{1}{T_{MW} - 1} \sum_{k=t-T_{MW}+1}^{t} diag((r_k - \mu_{t|t;T_{MW}}) \times (r_k - \mu_{t|t;T_{MW}})^T) \right\}^{1/2}, \tag{4.b}$$

$$y_{k|t;T_{MW}} = D^{-1}_{t|t;T_{MW}} (r_k - \mu_{t|t;T_{MW}}). \tag{4.c}$$

With the $n \times T_{MW}$ data matrix, $[y_{t-T_{MW}+1|t;T_{MW}}, y_{t-T_{MW}+2|t;T_{MW}}, \ldots, y_{t|t;T_{MW}}]$, a publically available statistical software package, such as MATLAB and its Toolboxes, can generate the PCA results. Then $y_{k|t;T_{MW}}$ can be expressed by the m-factor PCA model as

$$y_{k|t;T_{MW}} = X_{t|t;T_{MW}} f_{k|t;T_{MW}} + e_{k|t;T_{MW}},$$

where $X_{t|t;T_{MW}}$ is the $n \times m$ loadings matrix for the trailing window, $f_{k|t;T_{MW}}$ the $m \times 1$ factors vector for r_k and $e_{k|t;T_{MW}}$ the $n \times 1$ residuals vector of r_k. The PCA-based $m \times m$ factor covariance and $n \times n$ residual covariance matrices are,

$$P_{t|t;T_{MW}} = \frac{1}{T_{MW} - 1} \sum_{k=t-T_{MW}+1}^{t} f_{k|t;T_{MW}} f^T_{k|t;T_{MW}},$$

$$R_{t|t;T_{MW}} = \frac{1}{T_{MW} - 1} \sum_{k=t-T_{MW}+1}^{t} diag(e_{k|t;T_{MW}} e^T_{k|t;T_{MW}}),$$

where $P_{t|t;T_{MW}}$ is always diagonal. Of the stock returns r_t, the covariance matrix predicted by the rolling PCA factor model of T_{MW} is,

$$C_{t+1|t;T_{MW}} = D_{t|t;T_{MW}} (X_{t|t;T_{MW}} P_{t|t;T_{MW}} X^T_{t|t;T_{MW}} + R_{t|t;T_{MW}}) D_{t|t;T_{MW}}. \tag{5}$$

In our simulation case study, the rolling PCA factor model of $n = 200$, $m = 10$ and $T_{MW} = 65$ is used as the true data-generating model. Then a set of 200 daily time-series, $r^{(s)}_t$, of multivariate normal distribution is generated by this known model as

$$r^{(s)}_t = X^{(s)}_t f^{(s)}_t + e^{(s)}_t, \tag{6}$$

where the superscript (s) indicates 'simulation', the loadings $X^{(s)}_t = D_{t|t;T_{MW}} X_{t|t;T_{MW}}$, the factors $f^{(s)}_t \sim N(0, P_{t|t;T_{MW}})$, the residuals $e^{(s)}_t \sim N(0, R^{(s)}_t)$ and the residual covariance $R^{(s)}_t = D_{t|t;T_{MW}} R_{t|t;T_{MW}} D_{t|t;T_{MW}}$. The simulated 9-year daily returns, $r^{(s)}_t$, are ranging from −0.326 to +0.316, have an average of −0.0004, median of −0.0003, standard deviation of 0.0218, skewness of −0.171 and kurtosis of 12.71. The first 1-year daily data are for model initialization and the remaining 8-year daily data are for the factor model evaluation. The true daily covariance matrix $C^{(s)}_t$ of the simulated time-series $r^{(s)}_t$ by (6) is

$$C^{(s)}_t = X^{(s)}_t P_{t|t;T_{MW}} (X^{(s)}_t)^T + R^{(s)}_t. \tag{7}$$

The estimated daily covariance matrix, $C_{t|t}$, estimated by a factor model for the simulated data $r^{(s)}_t$ is certainly different from the true covariance $C^{(s)}_t$. The difference can be measured with various formulas, such as Euclidean distance, log-Euclidean metric, Frobenius distance, geodesic distance (of Riemannian symmetric space), Kullback–Leibler divergence and Bhattacharyya divergence (Moakher and Batchelor 2006, de Luis-García *et al*. 2012, Moakher 2012). Since the covariance matrices here represent multivariate normal distributions, a distance defined by square root of a symmetrized Kullback–Leibler divergence (de Luis-García *et al*. 2012, Moakher, 2012) between the estimated $C_{t|t}$ and the true $C^{(s)}_t$,

$$d_{KL}(C_{t|t}, C^{(s)}_t) = [tr(C^{-1}_{t|t} C^{(s)}_t + (C^{(s)}_t)^{-1} C_{t|t})/(2n) - 1]^{1/2}, \tag{8}$$

may be regarded as the most relevant measure for us to evaluate the factor models.

4.2. Estimated covariance matrices

It is impossible to recover the true daily covariance $C_t^{(s)}$ by (7) from the simulated data $r_t^{(s)}$ by (6). A numeric lower bound of the distance (8) can be located by a 'perfectly simulated estimate' of covariance matrix. Having a perfect knowledge of the data-generating model (6), a 'sample factor covariance' and a 'sample residual variance' can be obtained 'separately', combined together into a sample covariance matrix, and then time-weighted into a perfectly simulated estimate,

$$C_{t|t}^{(K,w_0)} = (w_0/K)\sum_{k=1}^{K}[X_t^{(s)}f_t^{(k)}(X_t^{(s)}f_t^{(k)})^T + diag(e_t^{(k)}(e_t^{(k)})^T)] + (1-w_0)C_{t-1|t-1}^{(K,w_0)}, \tag{9}$$

where $f_t^{(k)}$ and $e_t^{(k)}$ are the kth sample values of $f_t^{(s)}$ and $e_t^{(s)}$, $K \geq 1$ is the sample size, and $w_0 \leq 1$ is the weight of current sample value. Only when $K = \infty$ and $w_0 = 1$, the estimated $C_{t|t}^{(K,w_0)}$ is the same as the true $C_t^{(s)}$. Table 1 shows 8-year average of daily distances $d_{t;K,w_0} = d_{KL}(C_{t|t}^{(K,w_0)}, C_t^{(s)})$ by (8) between $C_{t|t}^{(K,w_0)}$ and $C_t^{(s)}$. For the typical time-series of single realization at any given time t, sample size $K = 1$ represents the reality. Table 1 indicates that the estimate $C_{t|t}^{(K,w_0)}$ of $K = 1$ and $w_0 = 0.1$ can be regarded as a 'lower bound estimate'. The distance between an estimated $C_{t|t}$ and the true $C_t^{(s)}$ is larger than this lower bound because the components $X_t^{(s)}f_t^{(s)}$ and $e_t^{(s)}$ can never be separated.

Table 2 shows average daily distances for estimates made by factor models with different number of factors, $m = 5, 10, 15$. The left panel tabulates 8-year average of daily distances $d_{t;T_{MW}}^{(PCA)} = d_{KL}(C_{t|t;T_{MW}}^{(PCA)}, C_t^{(s)})$ by (8) for covariance estimates $C_{t|t;T_{MW}}^{(PCA)}$ by (5) with rolling PCA of various window size T_{MW}. The right panel tabulates 8-year average of daily distances $d_{t;h}^{(VSB)} = d_{KL}(C_{t|t;h}^{(VSB)}, C_t^{(s)})$ for covariance estimates $C_{t|t;h}^{(VSB)}$ by equation (g) in figure 4 with the VSBFA of various residual-to-specific variance ratio target h, where the effective number of data points is reasonably chosen as $T_0 = 15$ (for the daily data), the number of VA iterations is conveniently set to $L = 10$, the rolling data standardization (4) is applied with $T_{MW} = 65$, and the initial estimates $P_{0|0}$, $X_{0|0}$, $\{U_{j,0|0}\}$ and $R_{0|0}$ is made by the PCA of $T_{MW} = 65$. When the number of factors is underspecified at modelling $m = 5$ (vs. the true $m = 10$), both the rolling PCA and the VSBFA achieve better estimates evidenced by smaller distances than those of the correctly specified at $m = 10$. The best rolling PCA is of $T_{MW} = 65$ while the best VSBFA is of $h = 0.94$.

Table 3 shows annual averages of the daily lower bound distances, $d_{K,w_0}^{(LB)} = \frac{1}{t_2-t_1+1}\sum_{t=t_1}^{t_2} d_{t;K,w_0}$, of the best rolling PCA distances, $d_{T_{MW}}^{(PCA)} = \frac{1}{t_2-t_1+1}\sum_{t=t_1}^{t_2} d_{t;T_{MW}}^{(PCA)}$, and of the best VSBFA distances, $d_h^{(VSB)} = \frac{1}{t_2-t_1+1}\sum_{t=t_1}^{t_2} d_{t;h}^{(VSB)}$, here $K = 1$, $w_0 = 0.1$, $T_{MW} = 65$ and $h = 0.94$. The average VSBFA distance $d_h^{(VSB)}$ is always smaller than the corresponding average PCA distance $d_{T_{MW}}^{(PCA)}$. With the lower bound distance $d_{K,w_0}^{(LB)}$, a relative difference,

$$q_{VSB-PCA} = (d_h^{(VSB)} - d_{T_{MW}}^{(PCA)})/(d_{T_{MW}}^{(PCA)} - d_{K,w_0}^{(LB)}), \tag{10}$$

can be used to measure how much smaller $d_h^{(VSB)}$ is vs. $d_{T_{MW}}^{(PCA)}$. The last two columns show the value of $q_{VSB\text{-}PCA}$ for $m = 5$ and 10. The average distances achieved by the best VSBFA are more than 10% smaller than those by the best rolling PCA when the number of factors is underspecified at $m = 5$, and more than 30% smaller when correctly specified at $m = 10$.

Table 1. 8-Year average of daily (symmetrized) Kullback–Leibler distances by (8) between the 'true' covariance matrix $C_t^{(s)}$ by (7) of simulated daily returns of 200 stocks (from January 2007 to December 2014) and the 'perfectly simulated estimate' $C_{t|t}^{(K,w_0)}$ by (9) of various sample size K and current weight w_0. The estimated $C_{t|t}^{(K,w_0)} = C_t^{(s)}$ only when $K = \infty$ and $w_0 = 1$. Sample size $K = 1$ represents the reality of time-series with single realization at any given time t. The perfectly simulated estimate of $K = 1$ and $w_0 = 0.1$ is a 'lower bound estimate'.

	Sample size K			
w_0	$K = 1$	$K = 2$	$K = 3$	$K = \infty$
0.02	0.545	0.528	0.520	0.510
0.04	0.455	0.424	0.413	0.393
0.06	0.417	0.373	0.357	0.327
0.08	0.404	0.346	0.325	0.284
0.10	0.403	0.332	0.306	0.252
0.12	0.410	0.326	0.295	0.229
0.14	0.422	0.325	0.289	0.210
0.16	0.437	0.328	0.288	0.194
0.18	0.455	0.334	0.288	0.181
0.20	0.474	0.341	0.291	0.170
1.00	25.989	2.401	1.378	0

Table 2. 8-Year average of daily (symmetrized) Kullback–Leibler distances by (8) between the 'true' covariance matrix $C_t^{(s)}$ by (7) of simulated daily returns of 200 stocks (from January 2007 to December 2014) and covariance matrices estimated by m-factor models with the number of factors underspecified at $m = 5$ and correctly specified at $m = 10$. The left panel is for estimate $C_{t|t;T_{MW}}^{(PCA)}$ by (5) with rolling PCA factor models of various window size T_{MW}. The best rolling PCA model is of $T_{MW} = 65$. The right panel is for estimate $C_{t|t;h}^{(VSB)}$ by equation (g) in figure 4 with the VSBFA factor model of various residual-to-specific variance ratio target h (and $T_0 = 15$ and $L = 10$). The best VSBFA model is of $h = 0.94$.

Rolling PCA				VSBFA			
T_{MW}	$m = 5$	$m = 10$	$m = 15$	h	$m = 5$	$m = 10$	$m = 15$
25	0.737	1.034	1.568	0.86	0.634	0.639	0.647
30	0.696	0.908	1.241	0.87	0.631	0.636	0.643
35	0.673	0.837	1.081	0.88	0.628	0.633	0.640
40	0.661	0.793	0.987	0.89	0.626	0.630	0.638
45	0.654	0.765	0.927	0.90	0.624	0.628	0.635
50	0.653	0.748	0.887	0.91	0.623	0.626	0.633
55	0.654	0.738	0.859	0.92	0.621	0.625	0.632
60	0.657	0.732	0.840	0.93	0.620	0.624	0.631
65	0.662	0.730	0.827	0.94	0.620	0.625	0.631
70	0.668	0.730	0.818	0.95	0.621	0.625	0.632

Table 3. Annual and 8-year average of daily (symmetrized) Kullback–Leibler distances by (8) between the 'true' covariance matrix $C_t^{(s)}$ by (7) of simulated daily returns of 200 stocks and estimated covariance matrices. The first data column is for the 'lower bound estimate' $C_{t|t}^{(LB)}$ by (9) with $K = 1$ and $w_0 = 0.1$. The second data panel is for the best rolling PCA estimate $C_{t|t;T_{MW}}^{(PCA)}$ by (5) of $T_{MW} = 65$. The third panel is for the best VSBFA estimate $C_{t|t;h}^{(VSB)}$ by equation (g) in figure 4 of $h = 0.94$. The last panel shows the relative difference $q_{VSB\text{-}PCA}$ by (10). Relative to the PCA distances, the VSBFA distances are more than 10% smaller at $m = 5$ and more than 30% smaller at $m = 10$.

		Rolling PCA			VSBFA			VSBFA vs. PCA	
Year	Lower bound	$m = 5$	$m = 10$	$m = 15$	$m = 5$	$m = 10$	$m = 15$	$m = 5$ (%)	$m = 10$ (%)
2007	0.399	0.648	0.725	0.829	0.614	0.619	0.627	−13	−32
2008	0.412	0.680	0.768	0.882	0.650	0.660	0.669	−11	−30
2009	0.408	0.715	0.759	0.837	0.642	0.653	0.664	−24	−30
2010	0.393	0.650	0.713	0.805	0.606	0.611	0.616	−17	−32
2011	0.405	0.644	0.718	0.819	0.615	0.623	0.630	−12	−30
2012	0.401	0.660	0.725	0.816	0.616	0.618	0.622	−17	−33
2013	0.402	0.641	0.710	0.804	0.604	0.603	0.607	−16	−35
2014	0.405	0.660	0.725	0.823	0.613	0.610	0.614	−19	−36
'07–'14	0.403	0.662	0.730	0.827	0.620	0.625	0.631	−16	−32

5. A VSBFA model of international stocks

To evaluate performance of the VSBFA factor model in real investment settings, it is applied to a large universe of stocks worldwide. Volatilities of randomly formed stock portfolios are forecasted using the covariance matrix predicted by the factor model. Accuracy of the volatility forecasts can serve as a measure for the predictive power of the FA.

5.1. A VSBFA model of global stock data

There were 1610 stocks in the MSCI World Index as of 31 January 2014. Monthly total returns in USD of the stocks are obtained from FactSet. Total of 807 stocks in 21 countries and 10 GICS sectors were chosen for factor model evaluation because they have a complete 25-year history of monthly returns data from January 1989 to December 2013. The first 60-month period is for the model initialization. The remaining 240-month or 20-year period is for the model evaluation. Table 4 lists the number of stocks chosen in each country and sector. Of the 300 monthly returns (in decimal) of the 807 stocks ranging from −0.835 to +2.597, average and median rates are +0.01200 and +0.0100; standard deviation, skewness and kurtosis are 0.0968, +0.7923 and 15.3943.

In this empirical case study, the VSBFA model parameters are chosen as follows: the effective number of data points is $T_0 = 6$ (for the monthly data), the number of VA iterations is $L = 10$, and the residual-to-specific variance ratio target is $h = 0.86$. The value of h is selected for a higher predictive power as discussed in the next section. For the 807 global stocks, 5- and 20-year VSBFA model relevance levels $F_{t_1,t_2}^{(m)}$ by equation (i) in figure 4 and the incremental relevance levels $\Delta F_{t_1,t_2}^{(m)}$ by equation (j) in figure

Table 4. Country of domicile and GICS sector classification of the 807 stocks chosen from the 1610 stocks in MSCI World Index as of 31 January 2014. All of the 807 stocks have a complete 25-year history of monthly returns in USD from January 1989 to December 2013. The 10 GICS sectors are: Consumer Discretionary (CD), Consumer Staples (CS), Energy (En), Financials (Fin), Health Care (HC), Industrials (Ind), Information Technology (IT), Materials (Mat), Telecommunication Services (Tel) and Utilities (Ut).

	GICS sectors										
Country of domicile	CD	CS	En	Fin	HC	Ind	IT	Mat	Tel	Ut	Total
Australia		1	4	8		3		8		1	25
Austria			1								1
Belgium		2		3	1			1			7
Canada			1					1			2
Denmark				1	1	2					4
Finland				1		2					3
France	8	4	2	10	2	9	2	3	1		41
Germany	5	2		4	2	5		5		2	25
Hong Kong				13		3				3	19
Ireland				1				1			2
Italy	2		1	7		1			1		12
Japan	37	15	2	23	15	57	27	30	1	12	219
Luxembourg	1										1
Netherlands	2	4	1	2		2		1			12
Norway		1						1			2
Singapore	2			7		4					13
Spain				4		2			1	2	9
Sweden	2	1		3		5	1				12
Switzerland	2	2		5	1	5		3			18
UK	12	9	6	15	2	14		5	2		65
US	41	31	27	52	29	47	32	22	4	30	315
Total	114	72	45	159	53	161	62	81	10	50	807

Table 5. 5-Year (60-month) and 20-year (240-month) VSBFA factor model relevance level $F^{(m)}_{t_1,t_2}$ by equation (i) in figure 4 and incremental relevance level $\Delta F^{(m)}_{t_1,t_2}$ by equation (j) in figure 4. The VSBFA parameters are $T_0 = 6$, $L = 10$ and $h = 0.86$. For the 807 stocks chosen from the MSCI World Index, the incremental relevance levels $\Delta F^{(m)}_{t_1,t_2}$ are small for larger models with number of factors $m > 10$. A 10-factor VSBFA model is a reasonable choice for this universe of stocks.

	1994–1998		1999–2003		2004–2008		2009–2013		1994–2013	
VSBFA Model	$F^{(m)}$	$\Delta F^{(m)}$	$F^{(m)}$	$\Delta F^{(m)}$	$F^{(m)}$	$\Delta F^{(m)}$	$F^{(m)}$	$\Delta F^{(m)}$	$F^{(m)}$	$\Delta F^{(m)}$
1-Factor	.338		.342		.389		.515		.395	
2-Factor	.421	.0829	.392	.0495	.424	.0342	.543	.0274	.444	.0490
3-Factor	.446	.0245	.424	.0320	.444	.0202	.578	.0354	.473	.0285
4-Factor	.459	.0136	.447	.0233	.458	.0141	.592	.0141	.489	.0164
5-Factor	.467	.0072	.456	.0090	.475	.0172	.602	.0099	.500	.0107
6-Factor	.472	.0056	.463	.0069	.481	.0060	.607	.0045	.506	.0058
7-Factor	.477	.0049	.467	.0037	.487	.0063	.610	.0032	.510	.0046
8-Factor	.481	.0036	.470	.0032	.492	.0047	.614	.0045	.514	.0040
9-Factor	.484	.0029	.473	.0026	.494	.0014	.616	.0018	.517	.0022
10-Factor	.487	.0034	.477	.0047	.498	.0046	.620	.0033	.521	.0040
11-Factor	.488	.0014	.479	.0015	.502	.0040	.621	.0015	.523	.0021
12-Factor	.492	.0038	.482	.0026	.505	.0028	.626	.0050	.526	.0036
13-Factor	.493	.0012	.483	.0015	.507	.0017	.629	.0026	.528	.0018
14-Factor	.495	.0017	.485	.0024	.508	.0019	.630	.0016	.530	.0019
15-Factor	.498	.0025	.487	.0016	.510	.0019	.633	.0023	.532	.0021
16-Factor	.498	.0005	.488	.0008	.511	.0007	.633	.0007	.533	.0007
17-Factor	.499	.0012	.491	.0032	.512	.0013	.635	.0018	.535	.0019
18-Factor	.500	.0006	.492	.0010	.514	.0013	.636	.0011	.536	.0010
19-Factor	.501	.0011	.494	.0015	.515	.0011	.638	.0016	.537	.0014
20-Factor	.502	.0012	.495	.0015	.515	.0007	.639	.0011	.538	.0011

Table 6. Annual (12-month) relevance levels of individual factors F_{j,t_1,t_2} by equation (h) in figure 4 of the VSBFA factor model of 10 factors. The VSBFA parameters are $T_0 = 6$, $L = 10$ and $h = 0.86$. The last column is annual relevance level $F^{(m)}_{t_1,t_2}$ by equation (i) in figure 4 of the entire VSBFA factor model of 10 factors.

Year	F_1	F_2	F_3	F_4	F_5	F_6	F_7	F_8	F_9	F_{10}	$F^{(10)}$
1994	.232	.102	.046	.038	.015	.019	.009	.009	.006	.012	.488
1995	.200	.114	.033	.049	.013	.014	.006	.006	.005	.010	.450
1996	.167	.103	.030	.053	.013	.014	.007	.006	.009	.009	.412
1997	.232	.130	.045	.031	.010	.016	.011	.009	.006	.009	.499
1998	.238	.150	.040	.041	.018	.008	.011	.012	.004	.009	.532
1999	.233	.097	.028	.049	.029	.008	.008	.012	.005	.010	.479
2000	.187	.084	.050	.056	.019	.013	.006	.009	.005	.016	.445
2001	.196	.077	.055	.071	.015	.015	.010	.009	.005	.015	.468
2002	.253	.099	.049	.045	.012	.013	.006	.009	.004	.009	.500
2003	.267	.103	.041	.052	.013	.010	.004	.013	.003	.007	.512
2004	.248	.099	.034	.038	.016	.010	.004	.014	.003	.009	.475
2005	.234	.079	.028	.024	.035	.015	.012	.013	.003	.011	.455
2006	.222	.056	.035	.034	.044	.012	.012	.009	.003	.009	.437
2007	.167	.075	.038	.046	.039	.007	.016	.011	.004	.014	.416
2008	.378	.065	.048	.022	.034	.013	.008	.010	.002	.011	.590
2009	.502	.059	.029	.020	.022	.007	.004	.010	.001	.006	.662
2010	.461	.063	.027	.025	.015	.009	.005	.009	.001	.007	.623
2011	.426	.096	.032	.033	.014	.008	.003	.008	.002	.006	.628
2012	.406	.075	.035	.040	.018	.009	.003	.006	.002	.004	.599
2013	.275	.072	.051	.072	.012	.011	.008	.006	.003	.005	.517

Table 7. 5-Year and 20-year contributions to the predicted covariance matrix $C_{t+1|t}$ from the three components $C^{(X)}_{t+1|t}$, $C^{(U)}_{t+1|t}$ and $C^{(R)}_{t+1|t}$ by equations (d) to (g) in figure 4 based on the VSBFA factor model of 10 factors. The contributions are measured by 60-month and 240-month averages of contribution ratios $c^{(X)}_{t|t} = tr(C^{(X)}_{t+1|t})/tr(C_{t+1|t})$, $c^{(U)}_{t|t} = tr(C^{(U)}_{t+1|t})/tr(C_{t+1|t})$ and $c^{(R)}_{t|t} = tr(C^{(R)}_{t+1|t})/tr(C_{t+1|t})$. The VSBFA parameters are $T_0 = 6$, $L = 10$ and $h = 0.86$. The 10 VSBFA factors account for roughly half of the variances of time-series r_t.

Time period	$c^{(X)}$ (%)	$c^{(U)}$ (%)	$c^{(R)}$ (%)
1994–1998	38.3	8.0	53.7
1999–2003	39.5	7.9	52.6
2004–2008	41.1	7.6	51.3
2009–2013	54.9	6.0	39.2
1994–2013	43.4	7.4	49.2

4 are shown in table 5. Because $\Delta F^{(m)}_{t_1,t_2}$ are small for $m > 10$, the model of 10 factors is a reasonable choice. The annual (12-month) individual factor relevance levels F_{j,t_1,t_2} by equation (h) in figure 4 of the VSBFA model are tabulated in table 6. The last column is the annual VSBFA model relevance level $F^{(m)}_{t_1,t_2}$. Table 7 shows, for the VSBFA model, 5- and 20-year contributions of the three components, $C^{(X)}_{t+1|t}$, $C^{(U)}_{t+1|t}$ and $C^{(R)}_{t+1|t}$ by equations (d) to (f) in figure 4, to the predicted covariance matrix $C_{t+1|t}$, measured with averages of monthly contribution ratios such as $c^{(X)}_{t|t} = tr(C^{(X)}_{t+1|t})/tr(C_{t+1|t})$.

5.2. *Volatilities of stock portfolios*

Of the 807 global stocks, 1000 random portfolios and 1000 pairs of random portfolios are formed for the factor model evaluation. Each portfolio has a random number of stocks between 50 and 150. All portfolio holdings are assigned random positive weights between minimum and maximum, and summed to 1. Ratio of the maximum to minimum random weight is 10. Denoting the decimal weights and returns of stocks by vectors $w = (w_1, w_2, \cdots, w_S)^T$ and $r_t = [r_1(t), r_2(t), \ldots, r_S(t)]^T$, the decimal return of portfolio w is

$$r^{(w)}_t = w^T r_t = r^T_t w, \tag{11}$$

where the stock weights of the random portfolios are set constant through time for simplicity. The returns of an individual portfolio are referred to as 'returns of a long-only portfolio' because $w_j > 0$. Differential returns between a pair of portfolios are referred to as 'returns of a long/short portfolio', because they are the same as those by long (or owning) one portfolio and short (or borrowing) the other. Therefore, a pair of portfolios is treated together as a 'long/short portfolio' with some stock weights positive, some negative, and all weights sum to 0.

Denoting predicted expectation of the stock returns r_t as $\mu_{t+1|t}$ and predicted covariance matrix of r_t as $C_{t+1|t}$, the predicted expected value $\mu^{(w)}_{t+1|t}$ and variance $(\sigma^{(w)}_{t+1|t})^2$ of the portfolio returns $r^{(w)}_t$ in (11) are

$$\mu^{(w)}_{t+1|t} = w^T_t \mu_{t+1|t} = \mu^T_{t+1|t} w_t, \tag{12.a}$$

$$
\begin{aligned}
(\sigma_{t+1|t}^{(w)})^2 &= E_{t+1|t}((r_{t+1}^{(w)} - \mu_{t+1|t}^{(w)})(r_{t+1}^{(w)} - \mu_{t+1|t}^{(w)})^T) \\
&= w_t^T E_{t+1|t}((r_{t+1} - \mu_{t+1|t})(r_{t+1} - \mu_{t+1|t})^T) w_t \\
&= w_t^T C_{t+1|t} w_t,
\end{aligned} \tag{12.b}
$$

where $C_{t+1|t}$ can be estimated by the VSBFA equation (g) in figure 4, a rolling PCA result (5), or any of other FA methods. With the VSBFA-based covariance $C_{t+1|t}$ by equations (d) to (g) in figure 4, the portfolio variance $(\sigma_{t+1|t}^{(w)})^2$ is a sum of three components:

$$
\begin{aligned}
(\sigma_{t+1|t}^{(w)})^2 &= (\sigma_{t+1|t}^{(w|X)})^2 + (\sigma_{t+1|t}^{(w|U)})^2 + (\sigma_{t+1|t}^{(w|R)})^2 \\
&= (\sigma_{t+1|t}^{(w|X)})^2 + (\sigma_{t+1|t}^{(w|Spec)})^2,
\end{aligned} \tag{13.a}
$$

$$(\sigma_{t+1|t}^{(w|X)})^2 = w_t^T C_{t+1|t}^{(X)} w_t, \tag{13.b}$$

$$(\sigma_{t+1|t}^{(w|U)})^2 = w_t^T C_{t+1|t}^{(U)} w_t, \tag{13.c}$$

$$(\sigma_{t+1|t}^{(w|R)})^2 = w_t^T C_{t+1|t}^{(R)} w_t. \tag{13.d}$$

Table 8 summarizes 5- and 20-year averages of the three components, in contribution ratios such as $v_{t|t}^{(w|X)} = (\sigma_{t+1|t}^{(w|X)})^2/(\sigma_{t+1|t}^{(w)})^2$, for 1000 random long-only and 1000 random long/short portfolios. For the long-only ones, the loadings-based component $(\sigma_{t+1|t}^{(w|X)})^2$ is overwhelmingly dominant because the residuals of stocks are mostly diversified away. For the long/short ones, the specific component $(\sigma_{t+1|t}^{(w|U)})^2 + (\sigma_{t+1|t}^{(w|R)})^2$ is dominant because the effects of common factors are partially cancelled out by the positive and negative weights. Therefore, forecasting the volatilities of long-only portfolios can be used to evaluate the accuracy of estimated factors and loadings, while forecasting the volatilities of long/short portfolios can be used to evaluate the accuracy of estimated specific variances.

6. Forecasting volatilities of global stock portfolios

To evaluate the predictive power of the VSBFA approach relative to that of a high-dimensional FA relying on wide data windows, the VSBFA and the widely practiced rolling PCA factor models are applied to the universe of 807 global stocks. Volatilities of random long-only and random long/short portfolios are forecasted by the two factor models separately and the accuracies of the forecasts are compared.

6.1. Measuring the accuracy of portfolio volatility forecasts

To measure the accuracy of forecasts $\mu_{t+1|t}^{(w)}$ and $(\sigma_{t+1|t}^{(w)})^2$ by (12), a 'z-score squared of the forecasts' defined by

$$(z_{t+1|t}^{(w)})^2 = (r_{t+1}^{(w)} - \mu_{t+1|t}^{(w)})^2/(\sigma_{t+1|t}^{(w)})^2, \tag{14}$$

is handy. According to Litterman and Winkelmann (1998), Patton (2011), Menchero *et al*. (2013) and Fan *et al*. (2016), the accuracy of portfolio volatility forecasts over a given time period $(t+1) \in [t_1, t_2]$ can be measured by bias statistic $BS_{t_1,t_2}^{(w)}$, log-likelihood $LL_{t_1,t_2}^{(w)}$, Q-statistic $QS_{t_1,t_2}^{(w)}$ and volatility minimization $VM_{t_1,t_2}^{(w)}$ defined as

$$BS_{t_1,t_2}^{(w)} = \left[\frac{1}{t_2 - t_1} \sum_{t=t_1-1}^{t_2-1} (z_{t+1|t}^{(w)})^2\right]^{1/2}, \tag{15.a}$$

$$LL_{t_1,t_2}^{(w)} = -\frac{1/2}{t_2 - t_1 + 1} \sum_{t=t_1-1}^{t_2-1} [\ln(2\pi) + (z_{t+1|t}^{(w)})^2 + \ln(\sigma_{t+1|t}^{(w)})^2], \tag{15.b}$$

Table 8. 5-Year and 20-year average contributions to the predicted portfolio variance $(\sigma_{t+1|t}^{(w)})^2$ from the three components $(\sigma_{t+1|t}^{(w|X)})^2$, $(\sigma_{t+1|t}^{(w|U)})^2$ and $(\sigma_{t+1|t}^{(w|R)})^2$ by (13) based on the VSBFA factor model of 10 factors. The contributions are measured by 60-month and 240-month averages of contribution ratios $v_{t|t}^{(w|X)} = (\sigma_{t+1|t}^{(w|X)})^2/(\sigma_{t+1|t}^{(w)})^2$, $v_{t|t}^{(w|U)} = (\sigma_{t+1|t}^{(w|U)})^2/(\sigma_{t+1|t}^{(w)})^2$ and $v_{t|t}^{(w|R)} = (\sigma_{t+1|t}^{(w|R)})^2/(\sigma_{t+1|t}^{(w)})^2$, also averaging over 1000 random long-only portfolios (the left panel) and 1000 random long/short portfolios (the right panel). The VSBFA parameters are $T_0 = 6$, $L = 10$ and $h = 0.86$. The loadings-based component $(\sigma_{t+1|t}^{(w|X)})^2$ overwhelmingly dominates the variances of long-only portfolios, while the specific component $(\sigma_{t+1|t}^{(w|U)})^2 + (\sigma_{t+1|t}^{(w|R)})^2$ dominates the variances of long/short portfolios.

Time period	Long-only portfolios			Long/short portfolios		
	$v^{(w\|X)}$ (%)	$v^{(w\|U)}$ (%)	$v^{(w\|R)}$ (%)	$v^{(w\|X)}$ (%)	$v^{(w\|U)}$ (%)	$v^{(w\|R)}$ (%)
1994–1998	94.8	0.7	4.6	24.6	9.8	65.7
1999–2003	95.2	0.6	4.2	25.8	9.7	64.5
2004–2008	95.7	0.6	3.8	25.1	9.7	65.2
2009–2013	98.0	0.3	1.7	30.3	9.2	60.5
1994–2013	95.9	0.5	3.6	26.4	9.6	64.0

$$QS_{t_1,t_2}^{(w)} = \frac{1}{t_2 - t_1 + 1} \sum_{t=t_1-1}^{t_2-1} [(z_{t+1|t}^{(w)})^2 - \ln(z_{t+1|t}^{(w)})^2], \quad (15.c)$$

$$VM_{t_1,t_2}^{(w)} = sd_{t=t_1-1}^{t_2-1}(r_{t+1}^T w_{t+1|t}^{(QP)}) - sd_{t=t_1-1}^{t_2-1}(r_{t+1}^T w), \quad (15.d)$$

where $sd_{t=t_1}^{t_2}(\cdot)$ denotes sample standard deviation and $w_{t+1|t}^{(QP)}$ is time-varying stock weights of the minimum variance portfolio obtained by quadratic programming (QP) using the predicted covariance matrix $C_{t+1|t}$. A bias statistic $BS_{t_1,t_2}^{(w)} > 1$ or $BS_{t_1,t_2}^{(w)} < 1$ shows an under- or over-prediction of volatility. A higher log-likelihood $LL_{t_1,t_2}^{(w)}$ or a lower Q-statistic $QS_{t_1,t_2}^{(w)}$ indicates more accurate forecasts. A lower value of volatility minimization $VM_{t_1,t_2}^{(w)}$ means better for portfolio optimization. In the quadratic programming, constraint for a long-only portfolio is to set the maximum stock weight at 0.1, and additional constraint for a long/short portfolio is to keep the short-side portfolio unchanged.

6.2. Portfolio volatility forecasts by rolling PCA factor model

With a 10-factor rolling PCA model, substituting the moving average $\mu_{t+1|t;T_{MW}} = \mu_{t|t;T_{MW}}$ by (4.a) and the covariance $C_{t+1|t;T_{MW}}$ by (5) into (12) yields the PCA-based forecasts $\mu_{t+1|t;T_{MW}}^{(w|PCA)}$ and $(\sigma_{t+1|t;T_{MW}}^{(w|PCA)})^2$. Then the z-score squared $(z_{t+1|t;T_{MW}}^{(w|PCA)})^2$, bias statistic $BS_{t_1,t_2;T_{MW}}^{(w|PCA)}$, log-likelihood $LL_{t_1,t_2;T_{MW}}^{(w|PCA)}$, Q-statistic $QS_{t_1,t_2;T_{MW}}^{(w|PCA)}$ and volatility minimization $VM_{t_1,t_2;T_{MW}}^{(w|PCA)}$ of the forecasts are calculated by (14) and (15).

Averaging over 1000 random long-only and 1000 random long/short portfolios, the average bias statistic $BS_{t_1,t_2;T_{MW}}^{(PCA)}$, log-likelihood $LL_{t_1,t_2;T_{MW}}^{(PCA)}$, Q-statistic $QS_{t_1,t_2;T_{MW}}^{(PCA)}$ and volatility minimization $VM_{t_1,t_2;T_{MW}}^{(PCA)}$ are functions of T_{MW}. Table 9 tabulates 20-year values of the four statistics by the rolling PCA models of various T_{MW}. A greater $LL_{t_1,t_2;T_{MW}}^{(PCA)}$, a smaller $QS_{t_1,t_2;T_{MW}}^{(PCA)}$ or a lower $VM_{t_1,t_2;T_{MW}}^{(PCA)}$ indicates more accurate forecasts. Table 9 illustrates that the PCA model of $T_{MW} = 37$ may be regarded as the best, with the performances of $T_{MW} > 37$ trending worse and those of $T_{MW} < 37$ without a clear trend. Table 10 shows 5- and 20-year values of the four statistics by this best rolling PCA of $T_{MW} = 37$.

6.3. Portfolio volatility forecasts by VSBFA factor model

To compare with the best rolling PCA model, the data standardization (1) for the VSBFA models is implemented by the same rolling standardization (4) with $T_{MW} = 37$. The initial estimates $P_{0|0}$, $X_{0|0}$, $\{U_{j,0|0}\}$ and $R_{0|0}$ are calculated by the PCA of $T_{MW} = 37$. With the number of common factors $m = 10$, the effective number of data points $T_0 = 6$ and the number of VA iterations $L = 10$, the parameter of the VSBFA factor model is the residual-to-specific variance ratio target h. Substituting the moving average $\mu_{t+1|t;T_{MW}} = \mu_{t|t;T_{MW}}$ by (4.a) and the covariance $C_{t+1|t;h}$ by equation (g) in figure 4 into (12) yields the VSBFA-based forecasts $\mu_{t+1|t}^{(w|VSB)}$ and $(\sigma_{t+1|t;h}^{(w|VSB)})^2$. Then the z-score squared $(z_{t+1|t;h}^{(w|VSB)})^2$, bias statistic $BS_{t_1,t_2;h}^{(w|VSB)}$, log-likelihood $LL_{t_1,t_2;h}^{(w|VSB)}$, Q-statistic $QS_{t_1,t_2;h}^{(w|VSB)}$ and volatility minimization $VM_{t_1,t_2;h}^{(w|VSB)}$ of the forecasts are calculated by (14) and (15).

Table 9. Average 20-year (240 months from January 1994 to December 2013) bias statistic $BS_{t_1,t_2;T_{MW}}^{(PCA)}$, log-likelihood $LL_{t_1,t_2;T_{MW}}^{(PCA)}$, Q-statistic $QS_{t_1,t_2;T_{MW}}^{(PCA)}$ and volatility minimization $VM_{t_1,t_2;T_{MW}}^{(PCA)}$ of portfolio volatility forecasts based on 10-factor rolling PCA models of various values of T_{MW}, averaging over 1000 random long-only portfolios (the left panel) and 1000 random long/short portfolios (the right panel). The size of moving-window T_{MW} ranges from 30 to 42 months. A bias statistic $BS_{t_1,t_2;T_{MW}}^{(PCA)} < 1$ or $BS_{t_1,t_2;T_{MW}}^{(PCA)} > 1$ indicates an over- or under-prediction of portfolio volatility. A greater $LL_{t_1,t_2;T_{MW}}^{(PCA)}$, a smaller $QS_{t_1,t_2;T_{MW}}^{(PCA)}$ or a lower $VM_{t_1,t_2;T_{MW}}^{(PCA)}$ indicates more accurate portfolio volatility forecasts. The rolling PCA model of $T_{MW} = 37$ can be regarded as the best for forecasting the volatilities of both long-only and long/short portfolios.

	Long-only portfolios				Long/short portfolios			
T_{MW}	$BS^{(PCA)}$	$LL^{(PCA)}$	$QS^{(PCA)}$	$VM^{(PCA)}$ (%)	$BS^{(PCA)}$	$LL^{(PCA)}$	$QS^{(PCA)}$	$VM^{(PCA)}$ (%)
30	1.143	1.6436	2.5879	−4.09	1.060	2.9000	2.4428	−0.56
31	1.141	1.6441	2.5837	−4.11	1.059	2.8998	2.4442	−0.56
32	1.141	1.6423	2.5953	−4.11	1.057	2.8995	2.4473	−0.57
33	1.140	1.6422	2.5945	−4.14	1.055	2.9000	2.4453	−0.57
34	1.137	1.6430	2.6036	−4.17	1.053	2.9004	2.4448	−0.58
35	1.133	1.6460	2.6105	−4.17	1.052	2.9003	2.4457	−0.59
36	1.136	1.6409	2.6207	−4.16	1.050	2.9005	2.4467	−0.59
37	1.132	1.6433	2.6224	−4.18	1.049	2.8999	2.4465	−0.60
38	1.135	1.6380	2.6285	−4.21	1.048	2.8992	2.4485	−0.60
39	1.136	1.6352	2.6351	−4.20	1.047	2.8989	2.4495	−0.60
40	1.136	1.6335	2.6413	−4.24	1.046	2.8981	2.4524	−0.61
41	1.135	1.6318	2.6527	−4.24	1.045	2.8973	2.4543	−0.61
42	1.135	1.6299	2.6722	−4.27	1.044	2.8966	2.4557	−0.61

Table 10. Average 5- and 20-year bias statistic $BS^{(PCA)}_{t_1,t_2;T_{MW}}$, log-likelihood $LL^{(PCA)}_{t_1,t_2;T_{MW}}$, Q-statistic $QS^{(PCA)}_{t_1,t_2;T_{MW}}$ and volatility minimization $VM^{(PCA)}_{t_1,t_2;T_{MW}}$ of portfolio volatility forecasts based on the best 10-factor rolling PCA model of $T_{MW} = 37$, averaging over 1000 random long-only portfolios (the left panel) and 1000 random long/short portfolios (the right panel).

	Long-only portfolios				Long/short portfolios			
Rolling PCA $T_{MW} = 37$	$BS^{(PCA)}$	$LL^{(PCA)}$	$QS^{(PCA)}$	$VM^{(PCA)}$ (%)	$BS^{(PCA)}$	$LL^{(PCA)}$	$QS^{(PCA)}$	$VM^{(PCA)}$ (%)
1994–1998	1.222	1.799	2.616	−2.47	1.115	2.912	2.444	−0.59
1999–2003	0.995	1.698	2.314	−3.14	1.029	2.692	2.364	−0.72
2004–2008	1.350	1.603	3.161	−3.97	1.058	3.039	2.479	−0.41
2009–2013	0.938	1.474	2.398	−6.84	1.004	2.956	2.499	−0.61
1994–2013	1.132	1.643	2.622	−4.18	1.049	2.900	2.447	−0.60

Table 11. Average 20-year (240 months from January 1994 to December 2013) bias statistic $BS^{(VSB)}_{t_1,t_2;h}$, log-likelihood $LL^{(VSB)}_{t_1,t_2;h}$, Q-statistic $QS^{(VSB)}_{t_1,t_2;h}$ and volatility minimization $VM^{(VSB)}_{t_1,t_2;h}$ of portfolio volatility forecasts based on 10-factor VSBFA models of various values of h, averaging over 1000 random long-only portfolios (the left panel) and 1000 random long/short portfolios (the right panel). The residual-to-specific variance ratio target, h, ranges from $h = 0.80$ to $h = 0.95$. Other VSBFA parameters are $T_0 = 6$ and $L = 10$. A bias statistic $BS^{(VSB)}_{t_1,t_2;h} < 1$ or $BS^{(VSB)}_{t_1,t_2;h} > 1$ indicates an over- or under-prediction of portfolio volatility. A greater $LL^{(VSB)}_{t_1,t_2;h}$, a smaller $QS^{(VSB)}_{t_1,t_2;h}$ or a lower $VM^{(VSB)}_{t_1,t_2;h}$ indicates more accurate portfolio volatility forecasts. The VSBFA factor model of $h = 0.86$ can be regarded as the best for forecasting the volatilities of both long-only and long/short portfolios.

	Long-only portfolios				Long/short portfolios			
h	$BS^{(VSB)}$	$LL^{(VSB)}$	$QS^{(VSB)}$	$VM^{(VSB)}$ (%)	$BS^{(VSB)}$	$LL^{(VSB)}$	$QS^{(VSB)}$	$VM^{(VSB)}$ (%)
0.80	0.970	1.7485	2.4121	−4.44	1.030	2.9453	2.3559	−0.69
0.81	0.976	1.7487	2.4116	−4.43	1.029	2.9454	2.3556	−0.69
0.82	0.981	1.7489	2.4112	−4.42	1.028	2.9455	2.3553	−0.69
0.83	0.987	1.7490	2.4110	−4.41	1.027	2.9456	2.3551	−0.69
0.84	0.993	1.7490	2.4110	−4.39	1.025	2.9457	2.3549	−0.69
0.85	1.000	1.7489	2.4113	−4.37	1.024	2.9458	2.3547	−0.69
0.86	1.007	1.7486	2.4118	−4.34	1.022	2.9459	2.3546	−0.69
0.87	1.015	1.7482	2.4126	−4.32	1.020	2.9460	2.3545	−0.69
0.88	1.023	1.7476	2.4138	−4.29	1.018	2.9460	2.3544	−0.69
0.89	1.032	1.7468	2.4154	−4.25	1.016	2.9460	2.3544	−0.69
0.90	1.041	1.7457	2.4176	−4.22	1.014	2.9460	2.3544	−0.69
0.91	1.051	1.7443	2.4203	−4.17	1.012	2.9460	2.3545	−0.68
0.92	1.062	1.7426	2.4239	−4.13	1.010	2.9459	2.3546	−0.68
0.93	1.074	1.7404	2.4283	−4.08	1.007	2.9458	2.3548	−0.67
0.94	1.087	1.7376	2.4339	−4.03	1.004	2.9456	2.3551	−0.67
0.95	1.102	1.7340	2.4409	−3.97	1.001	2.9455	2.3555	−0.66

Table 12. Average 5- and 20-year bias statistic $BS^{(VSB)}_{t_1,t_2;h}$, log-likelihood $LL^{(VSB)}_{t_1,t_2;h}$, Q-statistic $QS^{(VSB)}_{t_1,t_2;h}$ and volatility minimization $VM^{(VSB)}_{t_1,t_2;h}$ of portfolio volatility forecasts based on the best 10-factor VSBFA model of $h = 0.86$, averaging over 1000 random long-only portfolios (the left panel) and 1000 random long/short portfolios (the right panel).

	Long-only portfolios				Long/short portfolios			
VSBFA $h = 0.86$	$BS^{(VSB)}$	$LL^{(VSB)}$	$QS^{(VSB)}$	$VM^{(VSB)}$ (%)	$BS^{(VSB)}$	$LL^{(VSB)}$	$QS^{(VSB)}$	$VM^{(VSB)}$ (%)
1994–1998	1.076	1.903	2.408	−3.17	1.076	2.957	2.354	−0.65
1999–2003	0.961	1.641	2.427	−3.24	0.982	2.711	2.325	−0.81
2004–2008	1.110	1.911	2.545	−3.25	1.076	3.110	2.339	−0.51
2009–2013	0.890	1.539	2.267	−7.45	0.959	3.006	2.400	−0.76
1994–2013	1.007	1.749	2.412	−4.34	1.022	2.946	2.355	−0.69

Averaging over the same 1000 random long-only and 1000 random long/short portfolios, the average bias statistic $BS^{(VSB)}_{t_1,t_2;h}$, log-likelihood $LL^{(VSB)}_{t_1,t_2;h}$, Q-statistic $QS^{(VSB)}_{t_1,t_2;h}$ and volatility minimization $VM^{(VSB)}_{t_1,t_2;h}$ are functions of h. Table 11 tabulates 20-year values of the four statistics by the VSBFA models of various h. A greater $LL^{(VSB)}_{t_1,t_2;h}$, a smaller $QS^{(VSB)}_{t_1,t_2;h}$ or a lower $VM^{(VSB)}_{t_1,t_2;h}$ indicates more accurate forecasts. Table 11 illustrates that the VSBFA model of $h = 0.86$ can be regarded as the best. Table 12 shows 5- and 20-year values of the four statistics by this best VSBFA model of $h = 0.86$.

Table 13. Average rolling 5-year and average 20-year bias statistics $BS^{(VSB)}_{t_1,t_2;h}$ and $BS^{(PCA)}_{t_1,t_2;T_{MW}}$, log-likelihood difference $\Delta LL_{t_1,t_2}$, Q-statistic difference $\Delta QS_{t_1,t_2}$ and volatility minimization difference $\Delta VM_{t_1,t_2}$ of portfolio volatility forecasts by (15.a) and (16) based on the 10-factor VSBFA model of $h = 0.86$ vs. the rolling PCA model of $T_{MW} = 37$, averaging over 1000 random long-only portfolios. A bias statistic $BS < 1$ or $BS > 1$ indicates an over- or under-prediction of portfolio volatility. A positive ΔLL, a negative ΔQS or a negative ΔVM indicates that the VSBFA model has higher predictive power than the rolling PCA model. A small p-value of the difference indicates that the difference is statistically significant. The VSBFA model has significantly higher predictive power than the PCA model in 11 or 14 out of the 16 rolling 5-year periods and in the entire 20-year period.

	Random long-only portfolios							
	Bias statistic		Log-likelihood		Q-statistic		Volatility minim	
VSBFA vs. PCA	$BS^{(VSB)}$	$BS^{(PCA)}$	ΔLL	p-Val	ΔQS	p-Val	ΔVM (%)	p-Val
1994–1998	1.076	1.222	0.104	0.0000	−0.208	0.0000	−0.71	0.0000
1995–1999	1.035	1.211	0.090	0.0000	−0.179	0.0000	−0.66	0.0000
1996–2000	1.025	1.223	0.079	0.0000	−0.158	0.0000	−0.44	0.0000
1997–2001	1.061	1.263	0.054	0.0000	−0.108	0.0000	−0.53	0.0000
1998–2002	1.046	1.218	0.034	0.0000	−0.068	0.0000	−0.52	0.0000
1999–2003	0.961	0.995	−0.057	0.0000	0.113	0.0000	−0.10	0.0000
2000–2004	0.989	0.974	−0.023	0.0000	0.046	0.0000	−0.17	0.0000
2001–2005	0.996	0.939	0.003	0.0000	−0.007	0.0000	−0.34	0.0000
2002–2006	0.952	0.898	0.017	0.0000	−0.034	0.0000	−0.10	0.0000
2003–2007	0.954	0.867	0.026	0.0000	−0.053	0.0000	−0.04	0.0022
2004–2008	1.110	1.350	0.308	0.0000	−0.616	0.0000	0.72	0.0000
2005–2009	1.112	1.476	0.339	0.0000	−0.677	0.0000	0.32	0.0000
2006–2010	1.115	1.494	0.312	0.0000	−0.624	0.0000	0.41	0.0000
2007–2011	1.111	1.473	0.330	0.0000	−0.660	0.0000	0.34	0.0000
2008–2012	1.088	1.438	0.319	0.0000	−0.638	0.0000	0.33	0.0000
2009–2013	0.890	0.938	0.066	0.0000	−0.131	0.0000	−0.60	0.0000
1994–2013	1.007	1.132	0.105	0.0000	−0.211	0.0000	−0.16	0.0000

Table 14. Average rolling 5-year and average 20-year bias statistics $BS^{(VSB)}_{t_1,t_2;h}$ and $BS^{(PCA)}_{t_1,t_2;T_{MW}}$, log-likelihood difference $\Delta LL_{t_1,t_2}$, Q-statistic difference $\Delta QS_{t_1,t_2}$ and volatility minimization difference $\Delta VM_{t_1,t_2}$ of portfolio volatility forecasts by (15.a) and (16) based on the 10-factor VSBFA model of $h = 0.86$ vs. the rolling PCA model of $T_{MW} = 37$, averaging over 1000 random long/short portfolios. A bias statistic $BS < 1$ or $BS > 1$ indicates an over- or under-prediction of portfolio volatility. A positive ΔLL, a negative ΔQS or a negative ΔVM indicates that the VSBFA model has higher predictive power than the rolling PCA model. A small p-value of the difference indicates that the difference is statistically significant. The VSBFA model has significantly higher predictive power than the PCA model in all of the 16 rolling 5-year periods and in the entire 20-year period.

	Random long/short portfolios							
	Bias statistic		Log-likelihood		Q-statistic		Volatility minim	
VSBFA vs. PCA	$BS^{(VSB)}$	$BS^{(PCA)}$	ΔLL	p-Val	ΔQS	p-Val	ΔVM (%)	p-Val
1994–1998	1.076	1.115	0.045	0.0000	−0.090	0.0000	−0.05	0.0000
1995–1999	1.087	1.175	0.047	0.0000	−0.093	0.0000	−0.07	0.0000
1996–2000	1.076	1.215	0.051	0.0000	−0.103	0.0000	−0.07	0.0000
1997–2001	1.054	1.210	0.052	0.0000	−0.104	0.0000	−0.07	0.0000
1998–2002	1.025	1.152	0.047	0.0000	−0.094	0.0000	−0.09	0.0000
1999–2003	0.982	1.029	0.019	0.0000	−0.038	0.0000	−0.09	0.0000
2000–2004	0.970	0.944	0.027	0.0000	−0.055	0.0000	−0.10	0.0000
2001–2005	0.979	0.865	0.033	0.0000	−0.065	0.0000	−0.11	0.0000
2002–2006	1.003	0.865	0.033	0.0000	−0.066	0.0000	−0.11	0.0000
2003–2007	1.019	0.900	0.033	0.0000	−0.066	0.0000	−0.06	0.0000
2004–2008	1.076	1.058	0.070	0.0000	−0.140	0.0000	−0.09	0.0000
2005–2009	1.080	1.187	0.091	0.0000	−0.181	0.0000	−0.13	0.0000
2006–2010	1.029	1.179	0.078	0.0000	−0.157	0.0000	−0.14	0.0000
2007–2011	1.016	1.159	0.087	0.0000	−0.174	0.0000	−0.15	0.0000
2008–2012	1.001	1.128	0.089	0.0000	−0.179	0.0000	−0.16	0.0000
2009–2013	0.959	1.004	0.050	0.0000	−0.100	0.0000	−0.15	0.0000
1994–2013	1.022	1.049	0.046	0.0000	−0.092	0.0000	−0.10	0.0000

6.4. Comparing the VSBFA and rolling PCA factor models

Eyeballing tables 10 and 12, the average $LL^{(VSB)}_{t_1,t_2;h}$ are generally greater than $LL^{(PCA)}_{t_1,t_2;T_{MW}}$, the average $QS^{(VSB)}_{t_1,t_2;h}$ are generally smaller than $QS^{(PCA)}_{t_1,t_2;T_{MW}}$ and the average $VM^{(VSB)}_{t_1,t_2;h}$ are generally lower than $VM^{(PCA)}_{t_1,t_2;T_{MW}}$, i.e. in general the VSBFA of $h = 0.86$ has a higher predictive power than the rolling PCA of $T_{MW} = 37$. Tables 13 and 14 formalize the comparison. For the same random portfolio over the same time period, log-likelihood difference $\Delta LL^{(w)}_{t_1,t_2}$, Q-statistic difference $\Delta QS^{(w)}_{t_1,t_2}$ and volatility minimization difference $\Delta VM^{(w)}_{t_1,t_2}$ between the two factor models are

$$\Delta LL^{(w)}_{t_1,t_2} = LL^{(w|VSB)}_{t_1,t_2;h} - LL^{(w|PCA)}_{t_1,t_2;T_{MW}}, \tag{16.a}$$

$$\Delta QS^{(w)}_{t_1,t_2} = QS^{(w|VSB)}_{t_1,t_2;h} - QS^{(w|PCA)}_{t_1,t_2;T_{MW}}, \tag{16.b}$$

$$\Delta VM^{(w)}_{t_1,t_2} = VM^{(w|VSB)}_{t_1,t_2;h} - VM^{(w|PCA)}_{t_1,t_2;T_{MW}}, \tag{16.c}$$

where $h = 0.86$ and $T_{MW} = 37$. Over 1000 random long-only or 1000 random long/short portfolios, average differences $\Delta LL_{t_1,t_2}$, $\Delta QS_{t_1,t_2}$ and $\Delta VM_{t_1,t_2}$ and p-values of the differences are calculated.

Table 13 shows rolling 5-year and entire 20-year average bias statistics $BS^{(VSB)}_{t_1,t_2;h}$ and $BS^{(PCA)}_{t_1,t_2;T_{MW}}$, average log-likelihood difference $\Delta LL_{t_1,t_2}$, average Q-statistic difference $\Delta QS_{t_1,t_2}$ and average volatility minimization difference $\Delta VM_{t_1,t_2}$, averaging over 1000 random long-only portfolios, as well as the p-values of the differences. Table 14 shows these statistics over 1000 random long/short portfolios. A positive $\Delta LL_{t_1,t_2}$, a negative $\Delta QS_{t_1,t_2}$ or a negative $\Delta VM_{t_1,t_2}$ indicates that the VSBFA factor model makes more accurate forecasts of portfolio volatilities than the rolling PCA model. A p-value smaller than an α level, say $p < \alpha = 0.01$, implies that the difference is statistically significant at $\alpha = 0.01$. Forecasts made by the VSBFA model are more accurate than those by the PCA model in 11 or 14 out of 16 rolling 5-year periods and in the entire 20-year period for the long-only portfolios as demonstrated in Table 13, and more accurate in all of the 16 rolling 5-year periods and in the entire 20-year period for the long/short portfolios as demonstrated in Table 14.

Table 15. 5-Year and 20-year utilization rates, $G^{(l,m)}_{t_1,t_2}$ by equation (i) in figure 5, of 34 explanatory indicators. 'Avg Europe' (N.Am. or Pacific) is a simple average of countries in the Europe (North America or Pacific) region. The first 3 explanatory indicators are simply MSCI indices (long-only portfolios). The remaining are differential indicators (long/short portfolios).

Indicators $r^{(EI)}_t$	1994–1998 (%)	1999–2003(%)	2004–2008(%)	2009–2013(%)	1994–2013(%)
Europe	87.9	77.7	94.3	95.2	89.8
North America	90.0	75.5	93.8	94.0	88.7
Pacific	91.5	84.5	92.1	96.8	92.2
Avg Europe—Eur	45.7	40.4	26.2	35.2	36.3
Avg N.Am.—N.Am.	34.9	37.2	34.8	39.9	36.7
Avg Pacific—Pacific	81.0	73.9	71.4	69.2	73.5
Eur Growth—Eur	34.5	31.5	28.2	26.1	29.8
Eur Value—Eur	33.1	32.3	27.1	25.5	29.1
N.Am. Gro—N.Am.	44.7	26.7	28.2	32.3	31.8
N.Am. Val—N.Am.	46.0	28.1	28.3	33.3	32.8
Pac Growth—Pac	34.5	23.3	23.5	34.3	28.3
Pac Value—Pac	32.9	22.9	22.7	32.7	27.3
Austria—Avg Eur	42.3	23.8	38.6	29.5	32.1
Belgium—Avg Eur	32.1	28.6	17.4	22.7	24.2
Denmark—Avg Eur	31.9	26.8	24.1	26.8	27.4
Finland—Avg Eur	36.0	29.3	35.1	30.5	32.3
France—Avg Eur	39.2	28.8	26.5	26.7	29.7
Germany—Avg Eur	40.9	41.6	35.2	36.3	38.4
Ireland—Avg Eur	31.9	29.1	22.7	34.5	29.5
Italy—Avg Eur	29.4	28.4	20.5	16.0	22.5
Netherland—Av Eur	38.9	36.4	36.6	48.1	40.2
Norway—Avg Eur	35.3	36.0	36.7	52.7	38.9
Portugal—Avg Eur	25.9	20.2	17.6	17.1	20.1
Spain—Avg Eur	32.8	38.2	30.1	31.4	32.8
Sweden—Avg Eur	34.6	30.0	30.4	32.7	32.0
Switzerland—Av Eu	26.4	27.5	20.4	21.4	23.9
U.K.—Avg Eur	48.6	55.3	40.7	41.4	45.6
Canada—Avg N.Am.	35.2	37.4	33.9	39.7	36.6
U.S.A.—Avg N.Am.	35.2	37.4	33.9	39.7	36.6
Australia—Avg Pac	45.8	54.7	48.1	53.1	50.3
Hong Kong—Av Pac	42.8	51.2	44.8	47.7	46.3
Japan—Avg Pac	80.7	74.9	70.4	67.8	72.9
New Zeal.—Av Pac	31.2	35.5	31.9	36.7	33.9
Singapore—Avg Pac	27.4	35.2	37.0	29.1	31.5

7. Explanatory effects on portfolio volatilities

In many investment portfolio analyses, simply decomposing the predicted portfolio variance $(\sigma_{t+1|t}^{(w)})^2$ into the loadings-based and specific components $(\sigma_{t+1|t}^{(w|X)})^2$ and $(\sigma_{t+1|t}^{(w|Spec)})^2$ by (13) does not provide sufficient information. Expressing the loadings-based covariance $C_{t+1|t}^{(X)}$ by equation (h) in figure 5 in terms of explanatory factors $f_t^{(Expl)}$ by equation (d) in figure 5 may help investors meaningfully understand the sources of portfolio risks. Well defined and widely watched MSCI indices of stocks in the MSCI World can be used to construct useful explanatory indicators for the stocks worldwide. Table 15 lists a set of $l = 34$ explanatory indicators formed by, including averages of and differences between, such indices. The 'average North America', for example, is a simple average of the countries inside the North America region. The table also shows 5- and 20-year information utilization rates by equation (i) in figure 5 of the 34 explanatory indicators. Since the MSCI indices are long-only portfolios, the first three simple indicators (e.g. North America) are long-only portfolios while the remaining differential indicators (e.g. Average North America minus North America) are long/short portfolios. The utilization rates of individual indicators are similar to typical levels of the loadings-based contribution ratio exemplified in table 8: high for indicators of long-only portfolios and low for indicators of long/short portfolios.

The loadings-based portfolio variance $(\sigma_{t+1|t}^{(w|X)})^2$ by (13.b) can be expressed in terms of the l explanatory factors, $l \geq m$, by substituting equation (h) in figure 5 into (13.b):

$$
\begin{aligned}
(\sigma_{t+1|t}^{(w|X)})^2 &= w_t^T C_{t+1|t}^{(Expl)} w_t = w_t^T X_{t|t}^{(Expl)} P_{t|t}^{(Expl)} (X_{t|t}^{(Expl)})^T w_t \\
&= (x_{t|t}^{(w)})^T P_{t|t}^{(Expl)} x_{t|t}^{(w)} \\
&= \sum_{j=1}^{l} (x_{t|t}^{(w)})^T (P_{t|t}^{(Expl)})_{j-col} (x_{t|t}^{(w)})_j,
\end{aligned}
$$

where $x_{t|t}^{(w)} = (X_{t|t}^{(Expl)})^T w_t$ is the $l \times 1$ loadings vector of the portfolio to the explanatory factors $f_t^{(Expl)}$. Therefore, contribution from the jth explanatory factor $f_{j,t}^{(Expl)}$ to portfolio variance, or the 'jth explanatory effect on portfolio variance', can be defined as

Table 16. Distributions, in percentiles, of jth explanatory effects, $a_{j,t|t}^{(w)}$ by (17), on variances of 1000 monthly random long-only portfolios (the left panel) and 1000 monthly random long/short portfolios (the right panel) from 1994 to 2013. The explanatory effects are well defined numerically in at least 98% of cases between the percentiles of 1 and 99%.

	Long-only portfolios					Long/short portfolios				
Explanatory factors	Min	1%	Med	99%	Max	Min	1%	Med	99%	Max
Europe	.115	.174	.301	.591	.805	−1.10	−0.32	0.04	1.09	1.87
North America	.054	.114	.266	.440	.578	−0.77	−0.32	0.05	0.80	1.50
Pacific	.049	.148	.319	.538	.619	−0.62	−0.22	0.08	0.61	1.13
Avg Europe—Eur	−.008	−.005	−.001	.002	.007	−0.03	−0.01	0.00	0.03	0.12
Avg N.Am.—N.Am.	−.010	−.002	.004	.016	.030	−0.06	−0.02	0.00	0.06	0.13
Avg Pacific—Pacific	−.058	−.033	−.009	.013	.038	−0.17	−0.06	0.02	0.20	0.34
Eur Growth—Eur	−.016	−.007	−.001	.002	.008	−0.09	−0.02	0.00	0.05	0.15
Eur Value—Eur	−.018	−.008	−.001	.001	.006	−0.10	−0.02	0.00	0.05	0.17
N.Am. Gro—N.Am.	−.011	−.003	.001	.009	.017	−0.08	−0.02	0.00	0.05	0.20
N.Am. Val—N.Am.	−.011	−.003	.002	.010	.019	−0.08	−0.02	0.00	0.06	0.20
Pac Growth—Pac	−.012	−.008	−.002	.000	.003	−0.03	−0.01	0.00	0.04	0.12
Pac Value—Pac	−.011	−.007	−.002	.000	.002	−0.03	−0.01	0.00	0.04	0.11
Austria—Avg Eur	−.011	−.006	−.002	.006	.031	−0.14	−0.04	0.02	0.24	0.54
Belgium—Avg Eur	−.007	−.001	.007	.019	.043	−0.10	−0.02	0.01	0.12	0.25
Denmark—Avg Eur	.002	.010	.026	.057	.099	−0.15	−0.03	0.01	0.16	0.43
Finland—Avg Eur	−.056	−.027	−.003	.026	.064	−0.26	−0.08	0.02	0.39	0.82
France—Avg Eur	−.009	−.003	.000	.008	.021	−0.09	−0.02	0.00	0.09	0.36
Germany—Avg Eur	−.020	−.012	−.002	.007	.034	−0.24	−0.06	0.01	0.25	0.54
Ireland—Avg Eur	−.055	−.019	.024	.066	.120	−0.38	−0.13	0.02	0.27	0.78
Italy—Avg Eur	−.003	−.001	.000	.004	.016	−0.13	−0.02	0.01	0.18	0.54
Netherland—Av Eur	−.010	.009	.026	.057	.092	−0.11	−0.02	0.02	0.17	0.31
Norway—Avg Eur	−.043	−.018	−.001	.017	.076	−0.32	−0.11	0.01	0.39	0.92
Portugal—Avg Eur	−.020	−.008	−.003	.005	.023	−0.14	−0.05	0.00	0.13	0.41
Spain—Avg Eur	−.022	.002	.016	.045	.097	−0.22	−0.06	0.01	0.25	0.62
Sweden—Avg Eur	−.050	−.012	.015	.055	.104	−0.21	−0.09	0.02	0.33	0.69
Switzerland—Av Eu	−.015	−.008	−.003	.000	.014	−0.12	−0.04	0.00	0.12	0.35
U.K.—Avg Eur	−.025	−.017	−.004	.012	.029	−0.12	−0.05	0.01	0.18	0.47
Canada—Avg N.Am.	−.013	−.002	.005	.021	.038	−0.09	−0.02	0.00	0.08	0.17
U.S.A.—Avg N.Am.	−.013	−.002	.005	.021	.038	−0.09	−0.02	0.00	0.08	0.17
Australia—Avg Pac	−.052	−.032	−.003	.028	.061	−0.59	−0.17	0.01	0.30	0.79
Hong Kong—Av Pac	−.039	−.012	−.002	.003	.029	−0.15	−0.07	0.01	0.27	0.91
Japan—Avg Pac	−.122	−.059	−.018	.028	.071	−0.48	−0.16	0.04	0.40	0.66
New Zeal.—Av Pac	−.055	−.037	−.010	.019	.048	−0.21	−0.09	0.01	0.21	0.54
Singapore—Avg Pac	−.013	−.006	.007	.031	.079	−0.14	−0.05	0.01	0.18	0.44

$$a_{j,t|t}^{(w)} = [(x_{t|t}^{(w)})_j (P_{t|t}^{(Expl)})_{j-col}^T x_{t|t}^{(w)}]/(\sigma_{t+1|t}^{(w|X)})^2. \quad (17)$$

When $l \geq m$, the entire loadings-based component $(\sigma_{t+1|t}^{(w|X)})^2$ of predicted portfolio variance can be explained by the explanatory factors $f_t^{(Expl)}$. The predicted portfolio variance $(\sigma_{t+1|t}^{(w)})^2$ can then be explained by a sum of $l+1$ terms as

$$(\sigma_{t+1|t}^{(w)})^2 = \sum_{j=1}^{l} a_{j,t|t}^{(w)} (\sigma_{t+1|t}^{(w|X)})^2 + (\sigma_{t+1|t}^{(w|Spec)})^2.$$

Table 16 shows distributions, in percentiles, of the explanatory effects $a_{j,t|t}^{(w)}$ of the 34 explanatory factors on the predicted variances of 1000 monthly random long-only (in the left panel) and long/short portfolios (in the right panel) for the 20-year period. In 98% of cases between the percentiles of 1 and 99%, the explanatory effects are well defined numerically.

8. Conclusion

We developed a VSBFA algorithm to perform online learning of time-varying stochastic factor model structures. An online learning is to, based on previously estimated factor scores, factor loadings and residual variances, update the factors, loadings and variances by only the current observations. In our VSBFA, previous estimates are introduced by sequential priors, the current observations are incorporated in the likelihood, and the VA is applied to the posterior to get the new estimates. In short, the VSBFA is an analytic Bayesian filter to estimate factors, loadings and variances. The covariance matrix predicted by the VSBFA can be decomposed into the loadings-based covariance and specific variance, and the former can be expressed in terms of explanatory factors which are systematic components of real-world indicators. Both simulated and empirical case studies demonstrate that the VSBFA can estimate and predict more accurately time-varying factor model structure, for both factor-based covariance and specific variance, than the moving-window-based FA, and can be applied to high-dimensional data which are difficult for FA with stochastic approximations. The effects of the explanatory factors on portfolio volatilities are estimated and examined. We expect the VSBFA algorithm to become a useful forecasting solution to asset management tasks in particular, and to many economic, scientific and engineering problems in general.

Acknowledgement

The authors truly appreciate the two anonymous referees for their great and invaluable insights in their critiques, suggestions for additional discussions, indications of further researches, and guidance on academic presentations.

Disclosure statement

No potential conflict of interest was reported by the authors.

References

Axioma, *Axioma Robust Risk Model Handbook*, 2011 (Axioma: New York).

Barhoumi, K., Darné, O. and Ferrara, L., Dynamic factor models: A review of the literature. *OECD J. Bus. Cycle Measure. Anal.*, 2013, **2**, 73–107.

Bai, J. and Ng, S., Determining the number of factors in approximate factor models. *Econometrica*, 2002, **70**(1), 191–221.

Beal, M.J., Variational algorithms for approximate Bayesian inference. A Thesis for the Degree of Doctor of Philosophy, University of London, 2003.

Ben-Israel, A. and Greville, T.N.E., *Generalized Inverses: Theory and Applications*, 2nd ed.., 2003 (Springer-Verlag: New York).

Bishop, C.M., Variational principal components. In *Proceedings Ninth International Conference on Artificial Neural Networks, ICANN '99, Volume 1*, pp. 509–514, 1999 (Institution of Electrical Engineers: Stevenage).

Bishop, C.M., *Pattern Recognition and Machine Learning*, 2006 (Springer Science+Business Media, LLC: New York).

de Luis-García, R., Alberola-López, C. and Westin, C.-F., On the choice of a tensor distance for DTI white matter segmentation. In *New Developments in the Visualization and Processing of Tensor Fields*, edited by D.H. Laidlaw and A. Vilanova, pp. 283–306, 2012 (Springer-Verlag: Berlin).

Del Negro, M. and Otrok, C., Dynamic factor models with time-varying parameters: Measuring changes in international business cycles. Staff Reports, No. 326, Federal Reserve Bank of New York, 2008.

Dempster, A.P., Laird, N.M. and Rubin, D.B., Maximum likelihood from incomplete data via the EM algorithm. *J. R. Stat. Soc. Series B*, 1977, **39**(1), 1–38.

Ding, X., He, L. and Carin, L., Bayesian robust principal component analysis. *IEEE T. Image Process.*, 2011, **20**(12), 3419–3430.

Fan, J., Furger, A. and Xiu, D., Incorporating global industrial classification standard into portfolio allocation: A simple factor-based large covariance matrix estimator with high frequency data. *J. Bus. Econom. Stat.*, 2016, **34**(4), 489–503.

Fan, J., Li, Y. and Yu, K., Vast volatility matrix estimation using high-frequency data for portfolio selection. *J. Amer. Statist. Assoc.*, 2012, **107**(497), 412–428.

Fan, J., Liao, Y. and Mincheva, M., High-dimensional covariance matrix estimation in approximate factor models. *Ann. Stat.*, 2011, **39**(6), 3320–3356.

Fan, J., Liao, Y. and Mincheva, M., Large covariance estimation by thresholding principal orthogonal complements. *J. R. Statist. Soc. B*, 2013, **75**(4), 603–680.

Frishman, F., On the arithmetic means and variances of products and ratios of random variables. In *A Modern Course on Statistical Distributions in Scientific Work*, Vol. 1, edited by G.P. Patil, S. Kotz and J.K. Ord, pp. 401–406, 1975 (Springer: Dordrecht).

Ghahramani, Z. and Beal, M.J., Variational inference for Bayesian mixtures of factor analysers. In *Advances in Neural Information Processing Systems 12*, edited by S.A. Solla, T.K. Leen and K.R. Müller, pp. 449–455, 2000 (MIT Press: Cambridge, MA).

Grimmer, J., An introduction to Bayesian inference via variational approximations. *Polit. Anal.*, 2011, **19**, 32–47.

Gruber, L. and West, M., GPU-accelerated Bayesian learning and forecasting in simultaneous graphical dynamic linear models. *Bayesian Anal.*, 2016, **11**(1), 125–149.

Jungbacker, B. and Koopman, S.J., Likelihood-based dynamic factor analysis for measurement and forecasting. *Econom. J.*, 2014, **17**, 1–21.
Kang, H. and Choi, S., Probabilistic models for common spatial patterns: Parameter-expanded EM and variational Bayes. In *Proceedings of the Twenty-Sixth AAAI Conference on Artificial Intelligence*, pp. 970–976, 2012 (The AAAI Press: Palo Alto, CA).
Koopman, S.J. and van der Wel, M., Forecasting the US term structure of interest rates using a macroeconomic smooth dynamic factor model. *Int. J. Forecasting*, 2013, **29**, 676–694.
Ling, H. and Stone, D.B., Time-varying forecasts by variational approximation of sequential Bayesian inference. *Quant. Finance*, 2016, **16**(1), 43–67. doi:10.1080/14697688.2015.1034759.
Litterman, R. and Winkelmann, K., *Estimating Covariance Matrices*. Risk Management Series, 1998 (Goldman Sachs Group: New York).
Lopes, H.F. and Carvalho, C.M., Factor stochastic volatility with time varying loadings and Markov switching regimes. *J. Stat. Plan. Inference*, 2007, **137**, 3082–3091.
Lopes, H.F. and West, M., Bayesian model assessment in factor analysis. *Stat. Sinica*, 2004, **14**, 41–67.
Luttinen, J. and Ilin, A., Variational Gaussian-process factor analysis for modeling spatio-temporal data. In *Advances in Neural Information Processing Systems 22: 23rd Annual Conference on Neural Information Processing Systems 2009*, pp. 1177–1185, 2009 (Neural Information Processing Systems Foundation: La Jolla, CA).
Luttinen, J. and Ilin, A., Transformations in variational Bayesian factor analysis to speed up learning. *Neurocomputing*, 2010, **73**, 1093–1102.
Menchero, J., Morozov, A. and Pasqua, A., *Predicting risk at short horizons: A case study for the USE4D model*. MSCI Model Insight, MSCI, 2013.
Moakher, M., Divergence measures and means of symmetric positive-definite matrices. In *New Developments in the Visualization and Processing of Tensor Fields*, edited by D.H. Laidlaw and A. Vilanova, pp. 307–321, 2012 (Springer-Verlag: Berlin).
Moakher, M. and Batchelor, P.G., Symmetric positive-definite matrices: From geometry to applications and visualization. In *Visualization and Processing of Tensor Fields*, edited by J. Weickert and H. Hagen, pp. 285–298, 2006 (Springer-Verlag: Berlin).
Motta, G. and Ombao, H., Evolutionary factor analysis of replicated time series. *Biometrics*, 2012, **68**, 825–836.
Muirhead, R.J., *Aspects of Multivariate Statistical Theory*, 2005 (John Wiley: Hoboken, NJ).
Nielsen, F.B., Variational approach to factor analysis and related models. Thesis for the degree Master of Science in Engineering, Technical University of Denmark, 2004.
Northfield, Adaptive near horizon risk model, 2012a. Available online at: www.northinfo.com (accessed 12 February 2014).
Northfield, U.S. short term equity risk model, 2012b. Available online at: www.northinfo.com (accessed 12 February 2014).
Ormerod, J.T. and Wand, M.P., Explaining variational approximations. *Am. Stat.*, 2010, **64**(2), 140–153.
Patton, A.J., Volatility forecast comparison using imperfect volatility proxies. *J. Econom.*, 2011, **160**(1), 246–256.
Rao, C.R., *Linear Statistical Inference and its Applications*, 2nd ed.., 1973 (John Wiley: New York).
Rowe, D.B., *Multivariate Bayesian Statistics: Models for Source Separation and Signal Unmixing*, 2003 (Chapman and Hall/CRC: Boca Raton, FL).
Shutin, D., Buchgraber, T., Kulkarni, S.R. and Poor, H.V., Fast variational sparse Bayesian learning with automatic relevance determination for superimposed signals. *IEEE T. Signal Process.*, 2011, **59**(12), 6257–6261.
Stock, J. and Watson, M., Dynamic factor models. In *The Oxford Handbook of Economic Forecasting*, edited by M.P. Clements and D.F. Hendry, pp. 35–60, 2011 (Oxford University Press: New York).
SunGard APT, APT factor models: Modeling guide. In *Technical Reference, Issue 1.4*, 2012 (SunGard Data Systems: Wayne, PA).
Tzikas, D.G., Likas, A.C. and Galatsanos, N.P., The variational approximation for Bayesian inference. *IEEE Signal Proc. Mag.*, 2008, **25**(6), 131–146.
Wang, Y. and Zou, J., Volatility analysis in high-frequency financial data. *Wiley Interdiscip. Rev. Comput. Stat.*, 2014, **6**(6), 393–404.
Ward, P., *The Barra Europe stochastic factor model (EURS1)*. MSCI Research Notes, MSCI, 2012.
Wijnholds, S., Sardarabadi, A.M. and van der Veen, A.J., Factor analysis as a tool for signal processing. Paper presented at the International Workshop on Biomedical and Astronomical Signal Processing (BASP) Frontiers 2013, Villars-sur-Ollon, Switzerland, 27 January–1 February, 2013.
Windle, J. and Carvalho, C.M., A tractable state-space model for symmetric positive-definite matrices. *Bayesian Anal.*, 2014, **9**(4), 759–792.
Zhao, Z.Y., Xie, M. and West, M., Dynamic dependence networks: Financial time series forecasting & portfolio decisions. Duke University Working Paper, 2015. Available online at: https://stat.duke.edu/sites/default/files/papers/2014-03_0.pdf
Zhao, J.H. and Yu, P.L.H., A note on variational Bayesian factor analysis. *Neural Net.*, 2009, **22**(7), 988–997.
Zhou, X., Nakajima, J. and West, M., Bayesian forecasting and portfolio decisions using dynamic dependent sparse factor models. *Int. J. Forecasting*, 2014, **30**, 963–980.

Appendix A. Details of VSBFA algorithm

This appendix derives and implements the VSBFA algorithm.

A.1. Multivariate normal distribution and product of inverse-gamma distributions

When an $m \times 1$ random vector $x = (x_1, x_2, \ldots, x_m)^T$ is of multivariate normal distribution with $m \times 1$ expected value vector μ and $m \times m$ covariance matrix Σ, $x \sim N(\mu, \Sigma)$, its probability density function is

$$\begin{aligned} p(x;\mu,\Sigma) &= N(x;\mu,\Sigma) \\ &= (2\pi)^{-m/2}|\Sigma|^{-1/2}\exp[(-1/2)(x-\mu)^T\Sigma^{-1}(x-\mu)] \\ &= (2\pi)^{-m/2}|\Sigma|^{-1/2}\exp[(-1/2)tr(\Sigma^{-1}(x-\mu)(x-\mu)^T)], \end{aligned} \tag{A.1}$$

where $|\Sigma|$ is determinant of Σ, and trace $tr(AB) = tr(BA)$. We will denote $(v)_j$, $(A)_{jk}$ and $(A)_{j-col}$ as the jth element of vector v, the (j, k)th element of matrix A and the jth column of A, respectively. Log density of the multivariate normal distribution is

$$\ln N(x;\mu,\Sigma) = (-1/2)[\ln|\Sigma| + tr(\Sigma^{-1}(x-\mu)(x-\mu)^T)] + const.$$

Expected value of a random function $z = z(x)$ with respect to the distribution of x is $E_x(z) = \int z p(x;\mu,\Sigma)dx$, and $E_x(x) = \mu$, $E_x((x-\mu)(x-\mu)^T) = \Sigma$, and $E_x((x-x_0)(x-x_0)^T) =$

$\Sigma + (\mu - x_0)(\mu - x_0)^T$, here x_0 is a non-random vector. If $x_0 = 0$, $E_x(xx^T) = \Sigma + \mu\mu^T$. Quadratic and linear terms of x in the log density are

$$\ln N(x;\mu,\Sigma) = (-1/2)tr(\Sigma^{-1}(x-\mu)(x-\mu)^T) + const$$
$$= (-1/2)tr(\Sigma^{-1}xx^T) + tr(\Sigma^{-1}\mu x^T) + const.$$

This expression indicates that, if the log density of an $m \times 1$ random vector z is

$$\ln p(z) = (-1/2)tr(Czz^T) + tr(bz^T) + const, \tag{A.2}$$

where C is an $m \times m$ symmetric positive-definite parameter matrix and b an $m \times 1$ parameter vector, then z is multivariate Gaussian, $p(z) = N(z; \mu_z, \Sigma_z)$, with its covariance matrix $\Sigma_z = C^{-1}$ and expected value vector $\mu_z = C^{-1}b$. In some cases, the Woodbury matrix identity,

$$(A + BCD)^{-1} = A^{-1} - A^{-1}B(C^{-1} + DA^{-1}B)^{-1}DA^{-1}, \tag{A.3}$$

can simplify the expression of Σ_z.

Assume that a normal distribution ($m = 1$) is conditioned on variance $\sigma^2 = \Sigma$, where σ^2 is also a random variable and its distribution can be refined by Bayes rule and data. It is well known that if the prior of σ^2 is an inverse-gamma distribution (a conjugate prior), the posterior of σ^2 is also inverse-gamma. Density function of an inverse-gamma variable x can be expressed as

$$p(x;\alpha,\beta) = IG(x;\alpha,\beta) = \Gamma(\alpha)^{-1}\beta^{\alpha}x^{-(\alpha+1)}\exp(-\beta/x), \tag{A.4}$$

and its log density is $\ln IG(x;\alpha,\beta) = (-1)[(\alpha+1)\ln x + \beta/x] + const$, where α and β are shape and scale parameters, and $\Gamma(\cdot)$ is a gamma function. Expected value of x^{-1} is

$$E_x(x^{-1}) = \alpha/\beta. \tag{A.5}$$

Assume that an $m \times m$ diagonal matrix X is formed by m mutually independent inverse-gamma variables x_j of (A.4), i.e. $X = diag((x_1, x_2, \ldots, x_m)^T)$ and $p(x_j; \alpha_j, \beta_j) = IG(x_j; \alpha_j, \beta_j)$, $j = 1, 2, \ldots, m$. Density function of X is the product of the independent inverse-gamma distributions, $p(X; \{\alpha_j, \beta_j; j = 1, 2, \ldots, m\}) = \prod_{j=1}^{m} IG(x_j; \alpha_j, \beta_j)$. As discussed by Ling and Stone (2016), when all α_j take the same value $\alpha_j = \alpha$ while all β_j form an $m \times m$ diagonal matrix $B = diag((\beta_1, \beta_2, \ldots, \beta_m)^T)$, the density function of the diagonal random matrix X becomes

$$p(X;\alpha,B) = IG_D(X;\alpha,B)$$
$$= \Gamma(\alpha)^{-m}|B|^{\alpha}|X|^{-(\alpha+1)}\exp[-tr(BX^{-1})], \tag{A.6}$$

and this product of independent inverse-gamma distributions is referred to as 'diagonal inverse-gamma distribution' because both the random variable X and the parameter B are diagonal matrices. Log density of the 'diagonal inverse-gamma distribution' is

$$\ln IG_D(X;\alpha,B) = (-1)[(\alpha+1)\ln|X| + tr(BX^{-1})] + const. \tag{A.7}$$

According to (A.5), expected value of X^{-1} is $E_X(X^{-1}) = \alpha B^{-1}$. The density (A.6) or (A.7) is in a matrix form borrowed from inverse-Wishart distribution, which is the multivariate generalization of inverse-gamma distribution (Muirhead 2005).

A.2. Joint distribution of the time-varying stochastic factor model

In the time-varying factor model (2), assume that (a) the observations y_t, factor scores f_t and columns of factor loadings $\{X_{j-col,t}\}$ are stochastic vectors normally distributed given stochastic covariance matrices R_t, P_t and $\{U_{j,t}\}$, respectively; (b) covariance between two elements of the observed data vector, $(y_t)_j$ and $(y_t)_k$, $k \neq j$, are represented by the expected values $E(X_t)$, $E(f_t)$ and $E(P_t)$; (c) the stochastic covariance matrices P_t, $\{U_{j,t}\}$ and R_t are of 'diagonal invers-gamma distributions'; (d) the dynamic evolutions of the random elements of P_t, X_t, $\{U_{j,t}\}$ and R_t are of random walks; and (e) all random elements of f_t, P_t, X_t, $\{U_{j,t}\}$ and R_t are mutually independent. The multivariate normal distribution has the maximum entropy, i.e. is the most appropriate to be assumed among all possible distributions on R^m, when only the mean μ and covariance Σ are given (Rao 1973). The 'diagonal inverse-gamma distribution' is the conjugate prior for the diagonal stochastic covariance matrix of a multivariate normal distribution. A random walk is one of the simplest stochastic dynamics when lack of additional information.

Joint distribution of the factor model (2) at time t is $p(y_t, f_t, P_t, X_t, \{U_{j,t}\}, R_t)$ which can be built by a Bayesian formulation with likelihood of y_t and priors of all of the random variables. Since the priors are initial guesses about the distributions before seeing the data, expected value of the prior of a variable can be set as the expected value of the same variable predicted at the earlier time $t-1$. Denoting the expected values previously estimated at the time $t-1$ as $P_{t-1|t-1}$, $X_{t-1|t-1}$, $\{U_{j,t-1|t-1}\}$ and $R_{t-1|t-1}$. Based on (B.4) to (B.8) and the random walk assumption, the expected values of the priors are $f_{t|t-1} = 0$, $P_{t|t-1} = P_{t-1|t-1}$, $X_{t|t-1} = X_{t-1|t-1}$, $\{U_{j,t|t-1} = U_{j,t-1|t-1}\}$ and $R_{t|t-1} = R_{t-1|t-1}$. Therefore, these priors are referred to as 'sequential priors' and our FA is referred to as 'sequential Bayesian'.

According to the factor model (2), the assumptions on probability distributions and (A.1), density function of the likelihood of the observations y_t, given f_t, X_t and R_t, is multivariate Gaussian,

$$\begin{aligned} p(y_t|X_t,f_t,R_t) &= N\left(y_t; \sum_{j=1}^{m} X_{j-col,t}f_{j,t}, R_t\right) \\ &= (2\pi)^{-n/2}|R_t|^{-1/2} \\ &\quad \times \exp\left\{(-1/2)\ tr\left(R_t^{-1}(y_t - \sum_{j=1}^{m} X_{j-col,t}f_{j,t}) \times (y_t - \sum_{j=1}^{m} X_{j-col,t}f_{j,t})^T\right)\right\}, \end{aligned} \tag{A.8}$$

where $f_{j,t} = (f_t)_j$. The use of vector summation $\sum_{j=1}^{m} X_{j-col,t}f_{j,t}$ corresponds to the individual priors of $\{X_{j-col,t}\}$ and $\{U_{j,t}\}$. Log density of the likelihood is

$$\begin{aligned} \ln p(y_t|X_t,f_t,R_t) &= \ln N(y_t; \sum_{j=1}^{m} X_{j-col,t}f_{j,t}, R_t) \\ &= (-1/2)\left[\ln|R_t| + tr\left(R_t^{-1}(y_t - \sum_{j=1}^{m} X_{j-col,t}f_{j,t}) \times (y_t - \sum_{j=1}^{m} X_{j-col,t}f_{j,t})^T\right)\right] + const. \end{aligned} \tag{A.9}$$

Density function of the prior of factor scores f_t in (A.8) and (A.9), given the predicted expectation $f_{t|t-1}$ and the diagonal covariance matrix P_t, is multivariate Gaussian or a product of normal distributions,

$$p(f_t|P_t) = N(f_t; f_{t|t-1}, P_t) = \prod_{j=1}^{m} N(f_{j,t}; f_{j,t|t-1}, P_{j,t}) \\ = (2\pi)^{-m/2} \prod_{j=1}^{m} P_{j,t}^{-1/2} \exp[-\frac{1}{2}\sum_{j=1}^{m} P_{j,t}^{-1}(f_{j,t} - f_{j,t|t-1})^2], \tag{A.10}$$

where $P_{j,t} = (P_t)_{jj}$. Log density of the prior of f_t is

$$\ln p(f_t|P_t) = \sum_{j=1}^{m} \ln p(f_{j,t}|P_{j,t}) = \sum_{j=1}^{m} \ln N(f_{j,t}; f_{j,t|t-1}, P_{j,t}) \\ = (-1/2)\sum_{j=1}^{m}[\ln P_{j,t} + P_{j,t}^{-1}(f_{j,t} - f_{j,t|t-1})^2] + const. \tag{A.11}$$

Density function of the prior of diagonal covariance matrix P_t in (A.10) and (A.11) is a 'diagonal inverse-gamma distribution',

$$p(P_t) = IG_D(P_t; \alpha_{t|t-1}, B_{P,t|t-1}) \\ = \Gamma(\alpha_{t|t-1})^{-m}|B_{P,t|t-1}|^{\alpha_{t|t-1}} \\ \times |P_t|^{-(\alpha_{t|t-1}+1)} \exp[-tr(P_t^{-1}B_{P,t|t-1})]. \tag{A.12}$$

The shape $\alpha_{t|t-1}$ and scale $B_{P,t|t-1}$ can be set, given the predicted expectation $P_{t|t-1}$, as

$$\alpha_{t|t-1} = T_0/2, \tag{A.13}$$

$$B_{P,t|t-1} = \alpha_{t|t-1}P_{t|t-1} = (T_0/2)P_{t|t-1},$$

where T_0 can be regarded as an effective number of data points supporting the prior of covariance by comparing the prior shape (A.13) and the posterior shape (A.31). Log density of the prior of P_t is

$$\ln p(P_t) = \ln IG_D(P_t; \alpha_{t|t-1}, B_{P,t|t-1}) \\ = (-1)\sum_{j=1}^{m}[(T_0/2+1)\ln P_{j,t} + P_{j,t}^{-1}(T_0/2)P_{j,t|t-1}] \\ + const. \tag{A.14}$$

Density function of the prior of factor loadings matrix X_t or the collection of its columns $\{X_{j-col,t}\}$ in (A.8) and (A.9), given the predicted expectation $X_{t|t-1}$ and the diagonal covariance matrices $\{U_{j,t}\}$, is a product of multivariate normal distributions,

$$p(X_t|\{U_{j,t}\}) = \prod_{j=1}^{m} N(X_{j-col,t}; X_{j-col,t|t-1}, U_{j,t}) \\ = (2\pi)^{-nm/2}(\prod_{j=1}^{m}|U_{j,t}|^{-1/2}) \exp[-\frac{1}{2}\sum_{j=1}^{m} tr(U_{j,t}^{-1} \\ \times (X_{j-col,t} - X_{j-col,t|t-1})(X_{j-col,t} \\ - X_{j-col,t|t-1})^T)]. \tag{A.15}$$

Log density of the prior of X_t is

$$\ln p(X_t|\{U_{j,t}\}) = \sum_{j=1}^{m} \ln N(X_{j-col,t}; X_{j-col,t|t-1}, U_{j,t}) \\ = (-1/2)\sum_{j=1}^{m}[\ln|U_{j,t}| + tr(U_{j,t}^{-1}(X_{j-col,t} \\ - X_{j-col,t|t-1})(X_{j-col,t} - X_{j-col,t|t-1})^T)] \\ + const. \tag{A.16}$$

Density function of the prior of collection of diagonal covariance matrices $\{U_{j,t}\}$ in (A.15) and (A.16) is a product of 'diagonal inverse-gamma distributions',

$$p(\{U_{j,t}\}) = \prod_{j=1}^{m} IG_D(U_{j,t}; \alpha_{t|t-1}, B_{U,j,t|t-1}) \\ = \Gamma(\alpha_{t|t-1})^{-nm}\prod_{j=1}^{m}\{|B_{U,j,t|t-1}|^{\alpha_{t|t-1}} \\ \times |U_{j,t}|^{-(\alpha_{t|t-1}+1)} \exp[-tr(U_{j,t}^{-1}B_{U,j,t|t-1})]\}. \tag{A.17}$$

The common shape $\alpha_{t|t-1} = T_0/2$ by (A.13) and the scale $\{B_{U,j,t|t-1}\}$ can be set, given the predicted expectation $\{U_{j,t|t-1}\}$, as

$$B_{U,j,t|t-1} = \alpha_{t|t-1}U_{j,t|t-1} = (T_0/2)U_{j,t|t-1}, j = 1, 2, \ldots, m.$$

Log density of the prior of collection $\{U_{j,t}\}$ is

$$\ln p(\{U_{j,t}\}) = \sum_{j=1}^{m} \ln IG_D(U_{j,t}; \alpha_{t|t-1}, B_{U,j,t|t-1}) \\ = (-1)\sum_{j=1}^{m}[(T_0/2+1)\ln|U_{j,t}| \\ + tr(U_{j,t}^{-1}(T_0/2)U_{j,t|t-1})] + const. \tag{A.18}$$

Finally, density function of the prior of diagonal residual error covariance matrix R_t in (A.8) and (A.9) is a 'diagonal inverse-gamma distribution',

$$p(R_t) = IG_D(R_t; \alpha_{t|t-1}, B_{R,t|t-1}) \\ = \Gamma(\alpha_{t|t-1})^{-n}|B_{R,t|t-1}|^{\alpha_{t|t-1}} \\ \times |R_t|^{-(\alpha_{t|t-1}+1)} \exp[-tr(R_t^{-1}B_{R,t|t-1})]. \tag{A.19}$$

Again, the common shape $\alpha_{t|t-1} = T_0/2$ by (A.13) and the scale $B_{R,t|t-1}$ can be set, given the predicted expectation $R_{t|t-1}$, as

$$B_{R,t|t-1} = \alpha_{t|t-1}R_{t|t-1} = (T_0/2)R_{t|t-1},$$

Log density of the prior of R_t is

$$\ln p(R_t) = \ln IG_D(R_t; \alpha_{t|t-1}, B_{R,t|t-1}) \\ = (-1)[(T_0/2+1)\ln|R_t| + tr(R_t^{-1}(T_0/2)R_{t|t-1})] \\ + const. \tag{A.20}$$

The priors (A.10) of f_t and (A.15) of X_t are hierarchical priors, i.e. being expressed in terms of random parameters P_t and $\{U_{j,t}\}$

which have their own priors. With the density functions of the likelihood (A.8) and of all of the sequential priors (A.10), (A.12), (A.15), (A.17) and (A.19), density function of the joint distribution for the time-varying factor model (2) is

$$\begin{aligned}
&p(y_t, f_t, P_t, X_t, \{U_{j,t}\}, R_t) \\
&= p(y_t|X_t, f_t, R_t)p(f_t|P_t)p(P_t)p(X_t|\{U_{j,t}\})p(\{U_{j,t}\})p(R_t) \\
&= N(y_t; \sum_{j=1}^{m} X_{j-col,t}f_{j,t}, R_t) \times N(f_t; f_{t|t-1}, P_t) \\
&\quad \times IG_D(P_t; \alpha_{t|t-1}, B_{P,t|t-1}) \times \prod_{j=1}^{m} N(X_{j-col,t}; X_{j-col,t|t-1}, U_{j,t}) \\
&\quad \times \prod_{j=1}^{m} IG_D(U_{j,t}; \alpha_{t|t-1}, B_{U,j,t|t-1}) \times IG_D(R_t; \alpha_{t|t-1}, B_{R,t|t-1}).
\end{aligned} \tag{A.21}$$

Log density of the joint distribution, by combining (A.9), (A.11), (A.14), (A.16), (A.18) and (A.20), is

$$\begin{aligned}
&\ln p(y_t, f_t, P_t, X_t, \{U_{j,t}\}, R_t) \\
&= \ln N\left(y_t; \sum_{j=1}^{m} X_{j-col,t}f_{j,t}, R_t\right) + \ln N(f_t; f_{t|t-1}, P_t) \\
&\quad + \ln IG_D(P_t; \alpha_{t|t-1}, B_{P,t|t-1}) \\
&\quad + \sum_{j=1}^{m} \ln N(X_{j-col,t}; X_{j-col,t|t-1}, U_{j,t}) \\
&\quad + \sum_{j=1}^{m} \ln IG_D(U_{j,t}; \alpha_{t|t-1}, B_{U,j,t|t-1}) \\
&\quad + \ln IG_D(R_t; \alpha_{t|t-1}, B_{R,t|t-1}) \\
&= (-1/2)[\ln |R_t| \\
&\quad + tr(R_t^{-1}(y_t - \sum_{j=1}^{m} X_{j-col,t}f_{j,t})(y_t - \sum_{j=1}^{m} X_{j-col,t}f_{j,t})^T)] \\
&\quad + (-1/2)\sum_{j=1}^{m}[\ln P_{j,t} + P_{j,t}^{-1}(f_{j,t} - f_{j,t|t-1})^2] \\
&\quad + (-1)\sum_{j=1}^{m}[(T_0/2 + 1)\ln P_{j,t} + P_{j,t}^{-1}(T_0/2)P_{j,t|t-1}] \\
&\quad + (-1/2)\sum_{j=1}^{m}[\ln |U_{j,t}| \\
&\quad + tr(U_{j,t}^{-1}(X_{j-col,t} - X_{j-col,t|t-1})(X_{j-col,t} - X_{j-col,t|t-1})^T)] \\
&\quad + (-1)\sum_{j=1}^{m}[(T_0/2 + 1)\ln |U_{j,t}| + tr(U_{j,t}^{-1}(T_0/2)U_{j,t|t-1})] \\
&\quad + (-1)[(T_0/2 + 1)\ln|R_t| + tr(R_t^{-1}(T_0/2)R_{t|t-1})] + const.
\end{aligned} \tag{A.22}$$

A.3. VA of the posterior

The objective is to estimate the posterior $p(f_t, P_t, X_t, \{U_{j,t}\}, R_t|y_t)$ of the time-varying factor model (2) based on the joint distribution $p(y_t, f_t, P_t, X_t, \{U_{j,t}\}, R_t)$ in (A.21). Even though only the simplest realistic distributions are assumed, the posterior is still too complicated to be analytically tractable. Either a stochastic or an analytic approximation has to be applied. A VA is an analytic one.

The VA is to approximate a posterior by a product of simpler tractable distributions of only one or a group of random variables. Specific to the factor model (2), it is to approximate $p(f_t, P_t, X_t, \{U_{j,t}\}, R_t|y_t)$ by a factorized distribution,

$$\begin{aligned}
q(f_t, P_t, X_t, \{U_{j,t}\}, R_t) &= \prod_{j=1}^{m} q_{f,j}(f_{j,t}) \times \prod_{j=1}^{m} q_{P,j}(P_{j,t}) \\
&\quad \times \prod_{j=1}^{m} q_{X,j}(X_{j-col,t}) \times \prod_{j=1}^{m} q_{U,j}(U_{j,t}) \\
&\quad \times q_R(R_t).
\end{aligned} \tag{A.23}$$

The VA is a free form approximation: there is no need to specify in advance the actual density functions in (A.23). The actual forms of the distributions will be identified in the VA process.

The VA method needs expected values with respect to the approximating distributions. Assume that z is a function of some of the random variables, these expected values can be denoted by 'expectation operators',

$$E_{f,j}z = E_{f,j}(z) = \int z q_{f,j}(f_{j,t}) \mathrm{d}f_{j,t},$$

$$E_{P,j}z = E_{P,j}(z) = \int z q_{P,j}(P_{j,t}) \mathrm{d}P_{j,t},$$

$$E_{X,j}z = E_{X,j}(z) = \int z q_{X,j}(X_{j-col,t}) \mathrm{d}X_{j-col,t},$$

$$E_{U,j}z = E_{U,j}(z) = \int z q_{U,j}(U_{j,t}) \mathrm{d}U_{j,t},$$

$$E_R z = E_R(z) = \int z q_R(R_t) \mathrm{d}R_t,$$

and by the products of the same groups of expectation operators,

$$E_f z = \left(\prod_{j=1}^{m} E_{f,j}\right) z = \int z \prod_{j=1}^{m} q_{f,j}(f_{j,t}) \mathrm{d}f_{j,t},$$

$$E_P z = \left(\prod_{j=1}^{m} E_{P,j}\right) z = \int z \prod_{j=1}^{m} q_{P,j}(P_{j,t}) \mathrm{d}P_{j,t},$$

$$E_X z = \left(\prod_{j=1}^{m} E_{X,j}\right) z = \int z \prod_{j=1}^{m} q_{X,j}(X_{j-col,t}) \mathrm{d}X_{j-col,t},$$

$$E_U z = \left(\prod_{j=1}^{m} E_{U,j}\right) z = \int z \prod_{j=1}^{m} q_{U,j}(U_{j,t}) \mathrm{d}U_{j,t}.$$

According to the VA theory (Bishop 2006, Tzikas, Likas and Galatsanos 2008, Ormerod and Wand 2010, Grimmer 2011), the optimal approximating distributions in (A.23) are the solutions to the following set of VA optimization equations of the log density functions for $j = 1, 2, \ldots, m$:

$$\ln q_{f,j}(f_{j,t}) = \left(\prod_{k \neq j} E_{f,k}\right) E_P E_X E_U E_R \ln p(y_t, f_t, P_t, X_t, \{U_{j,t}\}, R_t) + const, \tag{A.24}$$

$$\ln q_{P,j}(P_{j,t}) = E_f \left(\prod_{k \neq j} E_{P,k}\right) E_X E_U E_R \ln p(y_t, f_t, P_t, X_t, \{U_{j,t}\}, R_t) + const, \tag{A.25}$$

$$\ln q_{X,j}(X_{j-col,t}) = E_f E_P (\prod_{k \neq j} E_{X,k}) E_U E_R \ln p(y_t, f_t, P_t, X_t, \{U_{j,t}\}, R_t) + const, \tag{A.26}$$

$$\ln q_{U,j}(U_{j,t}) = E_f E_P E_X (\prod_{k \neq j} E_{U,k}) E_R \ln p(y_t, f_t, P_t, X_t, \{U_{j,t}\}, R_t) + const, \tag{A.27}$$

$$\ln q_R(R_t) = E_f E_P E_X E_U \ln p(y_t, f_t, P_t, X_t, \{U_{j,t}\}, R_t) + const. \tag{A.28}$$

These optimization equations are interrelated: each optimal approximating distribution is formulated in terms of all other optimal ones. A fortunate characteristic of the VA theory is that an iterative approach with a reasonable set of initial values always converges to a local optimal solution (Bishop 2006, Tzikas *et al.* 2008, Ormerod and Wand 2010).

A.4. Distributions and parameters of the approximating posteriors

Solving for the set of VA optimization equations is conceptually straightforward, though practically tedious. Substituting the log density of the joint distribution (A.22) into the VA equations (A.24) to (A.28) one by one, simplifying the expressions and comparing the simplified forms with the known distributions (A.2) or (A.7), or other equations in appendix A.1, will identify the forms and parameters of the approximating distributions in (A.23). The parameters of the approximating distributions will be subscripted by '$t|t$' in order to differentiate them from the parameters of the sequential priors subscripted by '$t|t-1$'. This process of identifying distributions and parameters leads to the following results.

The approximating distribution $q_{f,j}(f_{j,t})$ is a normal distribution with expected value $f_{j,t|t}$ and variance $Q_{j,t|t}$, i.e. $q_{f,j}(f_{j,t}) = N(f_{j,t}; f_{j,t|t}, Q_{j,t|t})$. The parameter $Q_{j,t|t}$ is for the VA estimation, while the random variable $P_{j,t}$ with its own approximating distribution is the stochastic variance of factor score to be estimated. The approximating distribution $q_{P,j}(P_{j,t})$ is an inverse-gamma distribution with shape $\alpha_{t|t}$ and scale $B_{P,j,t|t}$, i.e. $q_{P,j}(P_{j,t}) = IG(P_{j,t}; \alpha_{t|t}, B_{P,j,t|t})$, and expected value $P^{-1}_{j,t|t} = E_{P,j}(P^{-1}_{j,t}) = \alpha_{t|t} B^{-1}_{P,j,t|t}$. The approximating distribution $q_{X,j}(X_{j-col,t})$ is a multivariate normal distribution with expected value vector $X_{j-col,t|t}$ and covariance matrix $V_{j,t|t}$, i.e. $q_{X,j}(X_{j-col,t}) = N(X_{j-col,t}; X_{j-col,t|t}, V_{j,t|t})$. The parameter $V_{j,t|t}$ is for the VA estimation, while the random variable $U_{j,t}$ with its own approximating distribution is the stochastic covariance matrix of factor loadings to be estimated. The approximating distribution $q_{U,j}(U_{j,t})$ is a 'diagonal inverse-gamma distribution' with shape $\alpha_{t|t}$ and scale $B_{U,j,t|t}$, i.e. $q_{U,j}(U_{j,t}) = IG_D(U_{j,t}; \alpha_{t|t}, B_{U,j,t|t})$, and expected value $U^{-1}_{j,t|t} = E_{U,j}(U^{-1}_{j,t}) = \alpha_{t|t} B^{-1}_{U,j,t|t}$, if the expectation $E_{X,j}((X_{j-col,t} - X_{j-col,t|t-1})(X_{j-col,t} - X_{j-col,t|t-1})^T)$ is diagonal by data or by shrinking the off-diagonal elements to 0s according to the model assumption (the random elements of the stochastic vectors are mutually independent). Finally, the approximating distribution $q_R(R_t)$ is a 'diagonal inverse-gamma distribution' with shape $\alpha_{t|t}$ and scale $B_{R,t|t}$, i.e. $q_R(R_t) = IG_D(R_t; \alpha_{t|t}, B_{R,t|t})$, and expected value $R^{-1}_{t|t} = E_R(R^{-1}_t) = \alpha_{t|t} B^{-1}_{R,t|t}$, if the sum of four terms,

$$\sum_{j=1}^{m} E_{f,j}\left(f_{j,t}^2\right) E_{X,j}\left(X_{j-col,t} X_{j-col,t}^T\right) + \sum_{k \neq j} \sum_{j=1}^{m} E_{f,k}(f_{k,t}) E_{f,j}(f_{j,t}) E_{X,k}(X_{k-col,t}) E_{X,j}(X_{j-col,t})^T + (-2) \sum_{j=1}^{m} E_{f,j}(f_{j,t}) y_t E_{X,j}(X_{j-col,t})^T + y_t y_t^T,$$

is diagonal by data or by shrinking the off-diagonal elements to 0s according to the model assumption.

After identifying the approximating distributions, we are now ready to evaluate all of their parameters. The variance $Q_{j,t|t}$ of the approximating distribution $q_{f,j}(f_{j,t})$ is

$$Q_{j,t|t} = [P^{-1}_{j,t|t} + tr(R^{-1}_{t|t}(V_{j,t|t} + X_{j-col,t|t} X^T_{j-col,t|t}))]^{-1}, \tag{A.29}$$

and the expected value $f_{j,t|t}$ is

$$f_{j,t|t} = Q_{j,t|t} \left\{ P^{-1}_{j,t|t} f_{j,t|t-1} + tr\left(R^{-1}_{t|t} [y_t + (-1) \sum_{k \neq j} f_{k,t|t} X_{k-col,t|t}] X^T_{j-col,t|t} \right) \right\}. \tag{A.30}$$

The shape and scale parameters of the approximating distribution $q_{P,j}(P_{j,t})$ are

$$\alpha_{t|t} = (T_0 + 1)/2, \tag{A.31}$$

and $B_{P,j,t|t} = [T_0 P_{j,t|t-1} + Q_{j,t|t} + (f_{j,t|t} - f_{j,t|t-1})^2]/2$, and the expected value $P_{j,t|t}$ is

$$P_{j,t|t} = \alpha^{-1}_{t|t} B_{P,j,t|t} = [T_0 P_{j,t|t-1} + Q_{j,t|t} + (f_{j,t|t} - f_{j,t|t-1})^2]/(T_0 + 1). \tag{A.32}$$

The covariance $V_{j,t|t}$ of the approximating distribution $q_{X,j}(X_{j-col,t})$ is

$$V_{j,t|t} = [U_{j,t|t}^{-1} + R_{t|t}^{-1}(Q_{j,t|t} + f_{j,t|t}^2)]^{-1} = U_{j,t|t} - U_{j,t|t}[U_{j,t|t} + R_{t|t}/(Q_{j,t|t} + f_{j,t|t}^2)]^{-1}U_{j,t|t}, \tag{A.33}$$

and the expected value $X_{j-col,t|t}$ is

$$X_{j-col,t|t} = V_{j,t|t}\{U_{j,t|t}^{-1}X_{j-col,t|t-1} + R_{t|t}^{-1}f_{j,t|t}[y_t + (-1)\sum_{k\neq j} f_{k,t|t}X_{k-col,t|t}]\}. \tag{A.34}$$

The Woodbury matrix identity (A.3) is applied in (A.33). The shape and scale parameters of the approximating distribution $q_{U,j}(U_{j,t})$ are $\alpha_{t|t} = (T_0 + 1)/2$ by (A.31) and $B_{U,j,t|t} = diag(T_0U_{j,t|t-1} + V_{j,t|t} + (X_{j-col,t|t} - X_{j-col,t|t-1})(X_{j-col,t|t} - X_{j-col,t|t-1})^T)/2$, and the expected value $U_{j,t|t}$ is

$$U_{j,t|t} = \alpha_{t|t}^{-1}B_{U,j,t|t} = (T_0+1)^{-1}diag(T_0U_{j,t|t-1} + V_{j,t|t} + (X_{j-col,t|t} - X_{j-col,t|t-1})(X_{j-col,t|t} - X_{j-col,t|t-1})^T). \tag{A.35}$$

The shape and scale parameters of the approximating distribution $q_R(R_t)$ are $\alpha_{t|t} = (T_0 + 1)/2$ by (A.31) and $B_{R,t|t} = diag(T_0R_{t|t-1} + \sum_{j=1}^m [Q_{j,t|t}V_{j,t|t} + Q_{j,t|t}X_{j-col,t|t}X_{j-col,t|t}^T + f_{j,t|t}^2V_{j,t|t}] + (y_t - X_{t|t}f_{t|t})(y_t - X_{t|t}f_{t|t})^T)/2$, and the expected value $R_{t|t}$ is

$$R_{t|t} = \alpha_{t|t}^{-1}B_{R,t|t} = (T_0+1)^{-1}diag(T_0R_{t|t-1} + \sum_{j=1}^m [Q_{j,t|t}V_{j,t|t} + Q_{j,t|t}X_{j-col,t|t}X_{j-col,t|t}^T + f_{j,t|t}^2V_{j,t|t}] + (y_t - X_{t|t}f_{t|t})(y_t - X_{t|t}f_{t|t})^T). \tag{A.36}$$

A.5. VA iteration to estimate the posterior parameters

The convergence of iterative solutions to the VA optimization equations (A.24) to (A.28) indicates that the expected values $Q_{j,t|t}$, $f_{j,t|t}$, $P_{j,t|t}$, $V_{j,t|t}$, $X_{j-col,t|t}$, $U_{j,t|t}$ and $R_{t|t}$ by (A.29) to (A.36), $j = 1, 2, \ldots, m$, can be estimated iteratively. A reasonable assumption is to use the expected values of the priors discussed in section A.2 as the initial values of the iterative estimates. Therefore, the initial values $f_{t|t}^{(0)}$, $P_{t|t}^{(0)}$, $X_{t|t}^{(0)}$, $\{U_{j,t|t}^{(0)}\}$ and $R_{t|t}^{(0)}$ are predictions $f_{t|t-1}$, $P_{t|t-1}$, $X_{t|t-1}$, $\{U_{j,t|t-1}\}$ and $R_{t|t-1}$ adjusted by estimated variability of data y_{t-1}. Based on (B.4) to (B.8), (B.23) and (B.24),

$$f_{t|t}^{(0)} = 0, \tag{A.37}$$

$$P_{t|t}^{(0)} = n^{-1}tr(diag(X_{t-1|t-1}P_{t-1|t-1}X_{t-1|t-1}^T)^{-1}S_{t-1|t-1}^{(X)}) \times P_{t-1|t-1}, \tag{A.38}$$

$$X_{t|t}^{(0)} = X_{t-1|t-1}, \tag{A.39}$$

$$U_{j,t|t}^{(0)} = n^{-1}tr\left(diag\left(\sum_{j=1}^m P_{j,t-1|t-1}U_{j,t-1|t-1} + R_{t-1|t-1}\right)^{-1} S_{t-1|t-1}^{(Err)}\right) \times U_{j,t-1|t-1},\ j = 1, 2, \ldots, m, \tag{A.40}$$

$$R_{t|t}^{(0)} = n^{-1}tr\left(diag\left(\sum_{j=1}^m P_{j,t-1|t-1}U_{j,t-1|t-1} + R_{t-1|t-1}\right)^{-1} S_{t-1|t-1}^{(Err)}\right) \times R_{t-1|t-1}, \tag{A.41}$$

where $S_{t-1|t-1}^{(X)}$ and $S_{t-1|t-1}^{(Err)}$ by (A.58) and (A.59) are the components of the variability of y_{t-1}. With the relations (A.32) and (A.35), the initial values $Q_{j,t|t}^{(0)}$ and $V_{j,t|t}^{(0)}$ can be set as,

$$Q_{t|t}^{(0)} = P_{t|t}^{(0)}, \tag{A.42}$$

$$V_{j,t|t}^{(0)} = U_{j,t|t}^{(0)},\ j = 1, 2, \ldots, m. \tag{A.43}$$

Denote the iterative estimates by $Q_{j,t|t}^{(v)}$, $f_{j,t|t}^{(v)}$, $P_{j,t|t}^{(v)}$, $V_{j,t|t}^{(v)}$, $X_{j-col,t|t}^{(v)}$, $U_{j,t|t}^{(v)}$ and $R_{t|t}^{(v)}$, respectively, with the superscript (v) as the iteration index, $v = 1, 2, \ldots, L$. Replacing the expected values of the approximating distributions by their iterative estimates in a logical and executable way, the seven expressions (A.29), (A.30), and (A.32) to (A.36) can be translated into the following seven VA iteration statements. In a given VA iteration (v), the factor parameters $Q_{j,t|t}^{(v)} = (Q_{t|t}^{(v)})_{jj}$, $f_{j,t|t}^{(v)} = (f_{t|t}^{(v)})_j$ and $P_{j,t|t}^{(v)} = (P_{t|t}^{(v)})_{jj}$ are updated first for $j = 1, 2, \ldots, m$,

$$Q_{j,t|t}^{(v)} = \{(P_{j,t|t}^{(v-1)})^{-1} + tr((R_{t|t}^{(v-1)})^{-1}[V_{j,t|t}^{(v-1)} + X_{j-col,t|t}^{(v-1)}(X_{j-col,t|t}^{(v-1)})^T])\}^{-1}, \tag{A.44}$$

$$f_{j,t|t}^{(v)} = Q_{j,t|t}^{(v)}\{(P_{j,t|t}^{(v-1)})^{-1}f_{j,t|t}^{(0)} + tr((R_{t|t}^{(v-1)})^{-1}[y_t - \sum_{k\neq j} f_{k,t|t}^{(v-1)}X_{k-col,t|t}^{(v-1)}](X_{j-col,t|t}^{(v-1)})^T)\}, \tag{A.45}$$

$$P_{j,t|t}^{(v)} = (T_0+1)^{-1}[T_0P_{j,t|t}^{(0)} + Q_{j,t|t}^{(v)} + (f_{j,t|t}^{(v)} - f_{j,t|t}^{(0)})^2], \tag{A.46}$$

Then the loading parameters $V_{j,t|t}^{(v)}$, $X_{j-col,t|t}^{(v)}$ and $U_{j,t|t}^{(v)}$ are updated for $j = 1, 2, \ldots, m$,

$$V_{j,t|t}^{(v)} = U_{j,t|t}^{(v-1)} - U_{j,t|t}^{(v-1)}\{U_{j,t|t}^{(v-1)} + R_{t|t}^{(v-1)}/[Q_{j,t|t}^{(v)} + (f_{j,t|t}^{(v)})^2]\}^{-1}U_{j,t|t}^{(v-1)}, \tag{A.47}$$

$$X_{j-col,t|t}^{(v)} = V_{j,t|t}^{(v)}\{(U_{j,t|t}^{(v-1)})^{-1}X_{j-col,t|t}^{(0)} + (R_{t|t}^{(v-1)})^{-1}f_{j,t|t}^{(v)}[y_t - \sum_{k\neq j} f_{k,t|t}^{(v)}X_{k-col,t|t}^{(v-1)}]\}, \tag{A.48}$$

$$U_{j,t|t}^{(v)} = (T_0+1)^{-1}diag(T_0U_{j,t|t}^{(0)} + V_{j,t|t}^{(v)} + (X_{j-col,t|t}^{(v)} - X_{j-col,t|t}^{(0)})(X_{j-col,t|t}^{(v)} - X_{j-col,t|t}^{(0)})^T), \tag{A.49}$$

Finally, the residual parameter $R^{(v)}_{t|t}$ is updated,

$$R^{(v)}_{t|t} = (T_0+1)^{-1} diag(T_0 R^{(0)}_{t|t} + \sum_{j=1}^{m}[Q^{(v)}_{j,t|t}V^{(v)}_{j,t|t} + Q^{(v)}_{j,t|t}X^{(v)}_{j-col,t|t}(X^{(v)}_{j-col,t|t})^T + (f^{(v)}_{j,t|t})^2 V^{(v)}_{j,t|t}] + (y_t - X^{(v)}_{t|t}f^{(v)}_{t|t})(y_t - X^{(v)}_{t|t}f^{(v)}_{t|t})^T). \quad \text{(A.50)}$$

After the VA iterations, the last iterative results when $v = L$ are the new estimated expected values $f_{t|t}$, $P_{t|t}$, $X_{t|t}$, $\{U_{j,t|t}\}$ and $R_{t|t}$ of the approximating distributions:

$$f_{t|t} = f^{(L)}_{t|t}, \quad \text{(A.51)}$$

$$P_{t|t} = P^{(L)}_{t|t}, \quad \text{(A.52)}$$

$$X_{t|t} = X^{(L)}_{t|t}, \quad \text{(A.53)}$$

$$U_{j,t|t} = U^{(L)}_{j,t|t},\ j = 1,2,\ldots,m, \quad \text{(A.54)}$$

$$R_{t|t} = R^{(L)}_{t|t}. \quad \text{(A.55)}$$

Then, the observed data y_t can be divided into two components, $y_t = y_{t|t} + e_{t|t}$, as

$$y_{t|t} = X_{t|t}[(X^T_{t|t}X_{t|t})^{-1}(X^T_{t|t}y_t)], \quad \text{(A.56)}$$

$$e_{t|t} = y_t - y_{t|t}, \quad \text{(A.57)}$$

$$S^{(X)}_{t|t} = (T_0+1)^{-1} diag(T_0 S^{(X)}_{t-1|t-1} + y_{t|t}y^T_{t|t}), \quad \text{(A.58)}$$

$$S^{(Err)}_{t|t} = (T_0+1)^{-1} diag(T_0 S^{(Err)}_{t-1|t-1} + e_{t|t}e^T_{t|t}), \quad \text{(A.59)}$$

where $y_{t|t}$ and $e_{t|t}$ are least-squares conditional mean and residual, and $S^{(X)}_{t|t}$ and $S^{(Err)}_{t|t}$ are estimated mean squares of $y_{t|t}$ and $e_{t|t}$.

Appendix B. Predicted covariance matrix based on VSBFA estimates

The VSBFA solutions can be used to estimate the predicted time-varying covariance matrices of the time-series y_t and r_t. The concepts can also be used to derive the initial value adjustments for the VSBFA, the relevance levels of the stochastic factors, and the explanatory factors.

B.1. Predicted covariance matrix

At the time t, 1-step ahead stochastic factors f_{t+1} and loadings X_{t+1} can be approximately expressed by the VSBFA predictions and unobserved random components as,

$$f_{t+1} \approx f_{t+1|t} + \delta_{t+1}, \quad \text{(B.1)}$$

$$X_{j-col,t+1} \approx X_{j-col,t+1|t} + \xi_{j,t+1},\ j = 1,2,\ldots,m. \quad \text{(B.2)}$$

Substituting (B.1) and (B.2) into (2) and noticing $X_{j-col,t+1|t} = (X_{t+1|t})_{j-col}$, $f_{j,t+1|t} = (f_{t+1|t})_j$ and $\delta_{j,t+1} = (\delta_{t+1})_j$, the standardized time-series at time $t + 1$,

$$y_{t+1} \approx X_{t+1|t}f_{t+1|t} + X_{t+1|t}\delta_{t+1} + \sum_{j=1}^{m}\xi_{j,t+1}f_{j,t+1|t} + \sum_{j=1}^{m}\xi_{j,t+1}\delta_{j,t+1} + e_{t+1},$$

and predicted expectation,

$$\begin{aligned}&E_{t+1|t}(y_{t+1}y^T_{t+1})\\ &= E_{t+1|t}\Bigg(\Bigg[X_{t+1|t}f_{t+1|t} + X_{t+1|t}\delta_{t+1} + \sum_{j=1}^{m}\xi_{j,t+1}f_{j,t+1|t} + \sum_{j=1}^{m}\xi_{j,t+1}\delta_{j,t+1} + e_{t+1}\Bigg]\\ &\times\Bigg[X_{t+1|t}f_{t+1|t} + X_{t+1|t}\delta_{t+1} + \sum_{j=1}^{m}\xi_{j,t+1}f_{j,t+1|t} + \sum_{j=1}^{m}\xi_{j,t+1}\delta_{j,t+1} + e_{t+1}\Bigg]^T\Bigg).\end{aligned}$$

According to the VSBFA assumptions and estimates, the random components are of zero-mean normal distributions, $\delta_{t+1} \approx N(\delta_{t+1}; 0, P_{t+1|t})$, $\xi_{j,t+1} \approx N(\xi_{j,t+1}; 0, U_{j,t+1|t})$ and $e_{t+1} \approx N(e_{t+1}; 0, R_{t+1|t})$, with predicted diagonal covariance matrices $P_{t+1|t}$, $\{U_{j,t+1|t}\}$ and $R_{t+1|t}$. In addition, $E(\delta_{t+1}\xi^T_{j,t+1}) = 0$, $E(\delta_{t+1}e^T_{t+1}) = 0$, $E(\xi_{j,t+1}\xi^T_{k,t+1}) = 0$ and $E(\xi_{j,t+1}e^T_{t+1}) = 0$, here $1 \le j, k \le m$, and $k \ne j$. Therefore, the predicted expectation,

$$E_{t+1|t}(y_{t+1}y^T_{t+1}) = X_{t+1|t}(P_{t+1|t} + f_{t+1|t}f^T_{t+1|t})X^T_{t+1|t} + \sum_{j=1}^{m}(P_{j,t+1|t} + f^2_{j,t+1|t})U_{j,t+1|t} + R_{t+1|t}. \quad \text{(B.3)}$$

This expression is consistent with the formula by Frishman (1975).

For the financial market analyses discussed in this paper where the factors f_t change relatively quickly while the loadings X_t and the covariance matrices P_t, $\{U_{j,t}\}$ and R_t evolve as slower random walks, following predictions are appropriate:

$$f_{t+1|t} = 0, \quad \text{(B.4)}$$

$$P_{t+1|t} = P_{t|t}, \quad \text{(B.5)}$$

$$X_{t+1|t} = X_{t|t}, \quad \text{(B.6)}$$

$$U_{j,t+1|t} = U_{j,t|t},\ j = 1,2,\ldots,m, \quad \text{(B.7)}$$

$$R_{t+1|t} = R_{t|t}. \quad \text{(B.8)}$$

Substituting (B.4) to (B.8) into (B.3), the predicted expectation,

$$E_{t+1|t}(y_{t+1}y_{t+1}^T) = X_{t|t}P_{t|t}X_{t|t}^T + \sum_{j=1}^{m} P_{j,t|t}U_{j,t|t} + R_{t|t} = X_{t|t}P_{t|t}X_{t|t}^T + U_{t|t} + R_{t|t}, \tag{B.9}$$

where $U_{t|t}$ is defined as

$$U_{t|t} = \sum_{j=1}^{m} P_{j,t|t}U_{j,t|t}. \tag{B.10}$$

With the definition (1), the time-varying covariance matrix of the original time-series r_t can be predicted by

$$C_{t+1|t} = E_{t+1|t}((r_{t+1} - \mu_{t+1|t})(r_{t+1} - \mu_{t+1|t})^T) = E_{t+1|t}(D_{t+1|t}y_{t+1}y_{t+1}^T D_{t+1|t}) = D_{t|t}E_{t+1|t}(y_{t+1}y_{t+1}^T)D_{t|t}. \tag{B.11}$$

Define loadings matrix $X_{t|t}^{(r)}$ and diagonal covariance matrices $\{U_{j,t|t}^{(r)}\}$ and $R_{t|t}^{(r)}$ for r_t as

$$X_{t|t}^{(r)} = D_{t|t}X_{t|t}, \tag{B.12}$$

$$U_{j,t|t}^{(r)} = D_{t|t}^2 U_{j,t|t},\ j = 1, 2, \ldots, m, \tag{B.13}$$

$$U_{t|t}^{(r)} = D_{t|t}^2 U_{t|t}, \tag{B.14}$$

$$R_{t|t}^{(r)} = D_{t|t}^2 R_{t|t}. \tag{B.15}$$

Substituting (B.9), (B.10) and (B.12) to (B.15) into (B.11), the predicted covariance matrix $C_{t+1|t}$ of the time-series r_t is

$$C_{t+1|t} = X_{t|t}^{(r)}P_{t|t}(X_{t|t}^{(r)})^T + U_{t|t}^{(r)} + R_{t|t}^{(r)}. \tag{B.16}$$

The VSBFA-based covariance is a sum of three components:

$$C_{t+1|t} = C_{t+1|t}^{(X)} + C_{t+1|t}^{(U)} + C_{t+1|t}^{(R)} = C_{t+1|t}^{(X)} + C_{t+1|t}^{(Spec)}, \tag{B.17}$$

$$C_{t+1|t}^{(X)} = X_{t|t}^{(r)}P_{t|t}(X_{t|t}^{(r)})^T, \tag{B.18}$$

$$C_{t+1|t}^{(U)} = U_{t|t}^{(r)} = \sum_{j=1}^{m} P_{j,t|t}U_{j,t|t}^{(r)}, \tag{B.19}$$

$$C_{t+1|t}^{(R)} = R_{t|t}^{(r)}. \tag{B.20}$$

The first one, $C_{t+1|t}^{(X)}$, is loadings-based covariance for all pairs of different time-series, while the sum of the last two, $C_{t+1|t}^{(Spec)} = C_{t+1|t}^{(U)} + C_{t+1|t}^{(R)}$, is specific variance for individual timer-series.

B.2. Initial value alignments and adjustments

Based on (B.9), predicted expectation $E_{t|t-1}(y_t y_t^T)$ is

$$E_{t|t-1}(y_t y_t^T) = X_{t-1|t-1}P_{t-1|t-1}X_{t-1|t-1}^T + \sum_{j=1}^{m} P_{j,t-1|t-1}U_{j,t-1|t-1} + R_{t-1|t-1}.$$

It suggests that the previous estimates $P_{t-1|t-1}$, $X_{t-1|t-1}$, $\{U_{j,t-1|t-1}\}$ and $R_{t-1|t-1}$ can serve as initial values $P_{t|t}^{(0)}$, $X_{t|t}^{(0)}$, $\{U_{j,t|t}^{(0)}\}$ and $R_{t|t}^{(0)}$ for the estimates at the time t, and their characteristics can be summarized by two matrices: a loadings-based symmetric matrix for the correlations between elements of the data vector y_{t-1},

$$C_C(P_{t|t}^{(0)}, X_{t|t}^{(0)}) = X_{t|t}^{(0)}P_{t|t}^{(0)}(X_{t|t}^{(0)})^T, \tag{B.21}$$

and a diagonal matrix for the specific variances of the elements,

$$C_D(P_{t|t}^{(0)}, \{U_{j,t|t}^{(0)}\}, R_{t|t}^{(0)}) = \sum_{j=1}^{m} P_{j,t|t}^{(0)}U_{j,t|t}^{(0)} + R_{t|t}^{(0)}. \tag{B.22}$$

When the actual variability of the data y_{t-1} is estimated as $S_{t-1|t-1}^{(X)}$ plus $S_{t-1|t-1}^{(Err)}$ by (A.58) and (A.59), both $C_C(P_{t|t}^{(0)}, X_{t|t}^{(0)})$ and $S_{t-1|t-1}^{(X)}$ describe the loadings-based covariance, while both $C_D(P_{t|t}^{(0)}, \{U_{j,t|t}^{(0)}\}, R_{t|t}^{(0)})$ and $S_{t-1|t-1}^{(Err)}$ describe the specific variances. Therefore, to make the initial values aligned with the observed data, an 'average loadings-based variance bias ratio',

$$c_{t|t-1}^{(0)} = n^{-1}tr(diag(X_{t-1|t-1}P_{t-1|t-1}X_{t-1|t-1}^T)^{-1}S_{t-1|t-1}^{(X)}), \tag{B.23}$$

can be used to rescale $P_{t-1|t-1}$ into $P_t^{(0)}$ by (A.38), while an 'average specific variance bias ratio',

$$d_{t|t-1}^{(0)} = n^{-1}tr(diag(\sum_{j=1}^{m} P_{j,t-1|t-1}U_{j,t-1|t-1} + R_{t-1|t-1})^{-1}S_{t-1|t-1}^{(Err)}), \tag{B.24}$$

can be used to rescale $\{U_{j,t-1|t-1}\}$ and $R_{t-1|t-1}$ into $\{U_{j,t|t}^{(0)}\}$ and $R_{t|t}^{(0)}$ by (A.40) and (A.41).

Relative levels of variances $\{U_{j,t}\}$ and R_t have real impact on estimating the loadings $X_{t|t}$ by (A.48). If the residual error variances (the diagonal of R_t) are too small or too large relative to the variances of loadings (the diagonals of $\{U_{j,t}\}$), the estimated factors $f_{t|t}$ and loadings $X_{t|t}$ will over- or under-fit the observed data y_t and thus lead to a poor predictability. To achieve a higher predictive power, desired relative levels of variances $\{U_{j,t}\}$ and R_t need to be targeted by adjusting the relative levels of their initial values $\{U_{j,t|t}^{(0)}\}$ and $R_{t|t}^{(0)}$. The structure of specific variances (B.22) can be used to rescale $\{U_{j,t|t}^{(0)}\}$ and $R_{t|t}^{(0)}$ with a prescribed 'residual-to-specific variance ratio target h', $0 \ll h < 1$, while keeping the specific variances unchanged. The rescaling coefficient matrices $u_{t|t-1}^{(0)}$ and $r_{t|t-1}^{(0)}$, for $U_{j,t|t}^{(0)}$ and $R_{t|t}^{(0)}$, respectively, are to make the adjusted residual-to-specific variance ratio equal to h,

$$(r_{t|t-1}^{(0)}R_{t|t}^{(0)})C_D(P_{t|t}^{(0)}, \{u_{t|t-1}^{(0)}U_{j,t|t}^{(0)}\}, r_{t|t-1}^{(0)}R_{t|t}^{(0)})^{-1} = hI_n, \tag{B.25}$$

and to keep the adjusted specific variances unchanged,

$$C_D(P^{(0)}_{t|t}, \{u^{(0)}_{t|t-1} U^{(0)}_{j,t|t}\}, r^{(0)}_{t|t-1} R^{(0)}_{t|t}) = W^{(0)}_{t|t} = C_D(P^{(0)}_{t|t}, \{U^{(0)}_{j,t|t}\}, R^{(0)}_{t|t}). \tag{B.26}$$

The solutions to (B.25) and (B.26) are

$$u^{(0)}_{t|t-1} = (W^{(0)}_{t|t} - R^{(0)}_{t|t})^{-1}[(1-h)W^{(0)}_{t|t}],$$

$$r^{(0)}_{t|t-1} = (R^{(0)}_{t|t})^{-1}(hW^{(0)}_{t|t}).$$

Therefore, after the initial value assignments (A.40) and (A.41), $\{U^{(0)}_{j,t|t}\}$ and $R^{(0)}_{t|t}$ need to be rescaled by

$$W^{(0)}_{t|t} = \sum_{j=1}^{m} P^{(0)}_{j,t|t} U^{(0)}_{j,t|t} + R^{(0)}_{t|t}, \tag{B.27}$$

$$U^{(0)}_{j,t|t} = \{(W^{(0)}_{t|t} - R^{(0)}_{t|t})^{-1}[(1-h)W^{(0)}_{t|t}]\} U^{(0)}_{j,t|t}, \tag{B.28}$$

$$R^{(0)}_{t|t} = [(R^{(0)}_{t|t})^{-1}(hW^{(0)}_{t|t})] R^{(0)}_{t|t}. \tag{B.29}$$

In addition, the columns of initial loadings $X^{(0)}_{j-col,t|t}$ can be orthogonalized in order to keep the columns of estimated loadings nearly orthogonal. For the jth column, $j \geq 2$, denoting $n \times (j-1)$ matrix $X^{(0)}_{(1:j-1)col,t|t} = (X^{(0)}_{t|t})_{1-col:(j-1)-col}$, we have,

$$b^{(0)}_{j,t|t} = [(X^{(0)}_{(1:j-1)col,t|t})^T X^{(0)}_{(1:j-1)col,t|t}]^{-1} (X^{(0)}_{(1:j-1)col,t|t})^T X^{(0)}_{j-col,t|t}, \tag{B.30}$$

$$X^{(0)}_{j-col,t|t} = X^{(0)}_{j-col,t|t} - X^{(0)}_{(1:j-1)col,t|t} b^{(0)}_{j,t|t}, \tag{B.31}$$

where $b^{(0)}_{j,t|t}$ is the OLS solution to express the jth column $X^{(0)}_{j-col,t|t}$ by the columns of $X^{(0)}_{(1:j-1)col,t|t}$, and the matrix $[(X^{(0)}_{(1:j-1)col,t|t})^T X^{(0)}_{(1:j-1)col,t|t}]$ is diagonal.

B.3. Relevance levels

Based on (B.9) and (B.10), contribution from the jth stochastic factor alone to the factor model (2) over a given time period $t \in [t_1, t_2]$,

$$F_{j,t_1,t_2} = \left\{ \sum_{t=t_1}^{t_2} P_{j,t|t}\left[tr(U_{j,t|t}) + (X^T_{j-col,t|t} X_{j-col,t|t})\right] \right\} \Big/ \sum_{t=t_1}^{t_2} tr(X_{t|t} P_{t|t} X^T_{t|t} + U_{t|t} + R_{t|t}), \tag{B.32}$$

can be used to measure the degree of importance, or relevance level, of the jth factor. Total contribution from all m factors,

$$F^{(m)}_{t_1,t_2} = \left[\sum_{t=t_1}^{t_2} tr(X_{t|t} P_{t|t} X^T_{t|t} + U_{t|t})\right] \Big/ \sum_{t=t_1}^{t_2} tr(X_{t|t} P_{t|t} X^T_{t|t} + U_{t|t} + R_{t|t}), \tag{B.33}$$

can be used to measure the relevance level of the m-factor model. Incremental relevance level between the model of m-1 factors and that of m factors,

$$\Delta F^{(m)}_{t_1,t_2} = F^{(m)}_{t_1,t_2} - F^{(m-1)}_{t_1,t_2} > 0, \tag{B.34}$$

is expected to become smaller as m increases.

B.4. Explanatory indicators and explanatory factors

The formula (B.16) or (B.17) explains the predicted covariance matrix $C_{t+1|t}$ by m stochastic factors. The factor-based covariance $C_{t+1|t}$ can also be explained by a set of l real financial or market time-series $r^{(EI)}_j(t)$, $j = 1, 2, \ldots, l$, serving as explanatory indicators (EI). If $\mu^{(EI)}_j(t|t)$ and $(\sigma^{(EI)}_j(t|t))^2$ denote estimated time-varying mean and variance of $r^{(EI)}_j(t)$, vectors $r^{(EI)}_t = [r^{(EI)}_1(t), r^{(EI)}_2(t), \ldots, r^{(EI)}_l(t)]^T$ and $\mu^{(EI)}_{t|t} = [\mu^{(EI)}_1(t|t), \mu^{(EI)}_2(t|t), \ldots, \mu^{(EI)}_l(t|t)]^T$, and diagonal matrix of standard deviations $D^{(EI)}_{t|t} = diag([\sigma^{(EI)}_1(t|t), \sigma^{(EI)}_2(t|t), \ldots, \sigma^{(EI)}_l(t|t)]^T)$, then standardized explanatory indicators are

$$y^{(EI)}_t = (D^{(EI)}_{t|t})^{-1}(r^{(EI)}_t - \mu^{(EI)}_{t|t}). \tag{B.35}$$

Assume that the standardized time-series y_t of (1) and indictors $y^{(EI)}_t$ of (B.35) are modelled jointly by a time-varying factor model,

$$\begin{pmatrix} y_t \\ y^{(EI)}_t \end{pmatrix} = \begin{pmatrix} X_t \\ X^{(EI)}_t \end{pmatrix} f_t + \begin{pmatrix} e_t \\ e^{(EI)}_t \end{pmatrix}. \tag{B.36}$$

In estimating the VSBFA parameters of the joint factor model (B.36), the indicators $y^{(EI)}_t$ are not involved in estimating the factors f_t and P_t, and not involved in calculating the rescaling coefficients for the initial value adjustments.

According to (B.35) and (B.36), the $l \times m$ loadings matrix of the indicators $r^{(EI)}_t$ is

$$X^{(EI)(r)}_{t|t} = D^{(EI)}_{t|t} X^{(EI)}_{t|t}. \tag{B.37}$$

When $X^{(EI)(r)}_{t|t}$ is of full rank and $l \geq m$, its Moore–Penrose pseudoinverse $(X^{(EI)(r)}_{t|t})^+$ is an $m \times l$ matrix (Ben-Israel and Greville 2003),

$$(X^{(EI)(r)}_{t|t})^+ = X^{-(EI)(r)}_{t|t} = [(X^{(EI)(r)}_{t|t})^T X^{(EI)(r)}_{t|t}]^{-1} (X^{(EI)(r)}_{t|t})^T, \tag{B.38}$$

where the pseudoinverse is also denoted as $X^{-(EI)(r)}_{t|t}$ for simplicity. For the data time-series r_t, define 'explanatory factor score covariance' $P^{(Expl)}_{t|t}$ and 'explanatory factor loadings' $X^{(Expl)}_{t|t}$ as,

$$P_{t|t}^{(Expl)} = X_{t|t}^{(EI)(r)} P_{t|t} (X_{t|t}^{(EI)(r)})^T, \quad \text{(B.39)}$$

$$X_{t|t}^{(Expl)} = X_{t|t}^{(r)} X_{t|t}^{-(EI)(r)}. \quad \text{(B.40)}$$

Substituting the relation $I_m = X_{t|t}^{-(EI)(r)} X_{t|t}^{(EI)(r)} = (X_{t|t}^{(EI)(r)})^T (X_{t|t}^{-(EI)(r)})^T$ and (B.37) to (B.40) into (B.18), the loadings-based covariance $C_{t+1|t}^{(X)}$ can be explained as,

$$C_{t+1|t}^{(X)} = C_{t+1|t}^{(Expl)} = X_{t|t}^{(Expl)} P_{t|t}^{(Expl)} (X_{t|t}^{(Expl)})^T. \quad \text{(B.41)}$$

A comparison between (B.18) and (B.41) demonstrates that the systematic components,

$$f_t^{(Expl)} = X_{t|t}^{(EI)(r)} f_t = D_{t|t}^{(EI)} X_{t|t}^{(EI)} f_t, \quad \text{(B.42)}$$

of the indicators $r_t^{(EI)}$ can serve as an alternative set of common factors for the data time-series r_t. Therefore, $f_t^{(Expl)}$ by (B.42) are 'explanatory factor scores', $P_{t|t}^{(Expl)}$ by (B.39) is covariance of $f_t^{(Expl)}$, and $X_{t|t}^{(Expl)}$ by (B.40) is loadings matrix of r_t to $f_t^{(Expl)}$.

Since only the systematic part of the indicators $r_t^{(EI)}$ serve as explanatory factors $f_t^{(Expl)}$, information utilization rate of the indicators needs to be calculated. Similar to the relevance level $F_{t_1,t_2}^{(m)}$ by (B.33), noticing that the jth column of loadings $X_{j-col,t}^{(EI)} \sim N(X_{j-col,t|t}^{(EI)}, U_{j,t|t}^{(EI)})$ and the residuals $e_t^{(EI)} \sim N(0, R_{t|t}^{(EI)})$, and denoting $U_{t|t}^{(EI)} = \sum_{j=1}^{m} P_{j,t|t} U_{j,t|t}^{(EI)}$, the utilization rate of the l indicators $y_t^{(EI)}$ within the model of m stochastic factors over the time period $t \in [t_1, t_2]$ can be defined as

$$G_{t_1,t_2}^{(l,m)} = \left[\sum_{t=t_1}^{t_2} tr(X_{t|t}^{(EI)} P_{t|t} (X_{t|t}^{(EI)})^T) \right] / \sum_{t=t_1}^{t_2} tr(X_{t|t}^{(EI)} P_{t|t} (X_{t|t}^{(EI)})^T + U_{t|t}^{(EI)} + R_{t|t}^{(EI)})). \quad \text{(B.43)}$$

The utilization rate of a single indicator $(y_t^{(EI)})_j$ can be calculated by the same formula with the number $l = 1$.