UNIVERSITY OF CAMBRIDGE
Faculty of Economics

JANEWAY INSTITUTE CAMBRIDGE

CAMBRIDGE WORKING PAPERS IN ECONOMICS
JANEWAY INSTITUTE WORKING PAPERS

A Structural Dynamic Factor Model for Daily Global Stock Market Returns

Oliver B. Linton	Haihan Tang	Jianbin Wu
University of Cambridge	Fudan University	Nanjing University

Abstract

Most stock markets are open for 6-8 hours per trading day. The Asian, European and American stock markets are separated in time by time-zone differences. We propose a statistical dynamic factor model for a large number of daily returns across multiple time zones. Our model has a common global factor as well as continental factors. Under a mild fixed-signs assumption, our model is identified and has a structural interpretation. We propose several estimators of the model: the maximum likelihood estimator-one day (MLE-one day), the quasi-maximum likelihood estimator (QMLE), an improved estimator from QMLE (QMLE-md), the QMLEres (similar to MLE-one day), and a Bayesian estimator (Gibbs sampling). We establish consistency, the rates of convergence and the asymptotic distributions of the QMLE and the QMLE-md. We next provide a heuristic procedure for conducting inference for the MLE-one day and the QMLE-res. Monte Carlo simulations reveal that the MLE-one day, the QMLE-res and the QMLE-md work well. We then apply our model to two real data sets: (1) equity portfolio returns from Japan, Europe and the US; (2) MSCI equity indices of 41 developed and emerging markets. Some new insights about linkages among different markets are drawn.

Reference Details

2237	Cambridge Working Papers in Economics
2022/14	Janeway Institute Working Paper Series
Published	15 June 2022
Key Words	Daily Global Stock Market Returns, Time-Zone Differences, Structural Dynamic Factor Model, Quasi Maximum Likelihood, Minimum Distance, Expectation Maximization Algorithm
JEL-codes	C55, C58, G15
Websites	www.econ.cam.ac.uk/cwpe www.janeway.econ.cam.ac.uk/working-papers

A Structural Dynamic Factor Model for Daily Global Stock Market Returns*

Oliver B. Linton† University of Cambridge
Haihan Tang‡ Fudan University
Jianbin Wu§ Nanjing University

June 8, 2022

Abstract

Most stock markets are open for 6-8 hours per trading day. The Asian, European and American stock markets are separated in time by time-zone differences. We propose a statistical dynamic factor model for a large number of daily returns across multiple time zones. Our model has a common global factor as well as continental factors. Under a mild fixed-signs assumption, our model is identified and has a structural interpretation. We propose several estimators of the model: the maximum likelihood estimator-one day (MLE-one day), the quasi-maximum likelihood estimator (QMLE), an improved estimator from QMLE (QMLE-md), the QMLE-res (similar to MLE-one day), and a Bayesian estimator (Gibbs sampling). We establish consistency, the rates of convergence and the asymptotic distributions of the QMLE and the QMLE-md. We next provide a heuristic procedure for conducting inference for the MLE-one day and the QMLE-res. Monte Carlo simulations reveal that the MLE-one day, the QMLE-res and the QMLE-md work well. We then apply our model to two real data sets: (1) equity portfolio returns from Japan, Europe and the US; (2) MSCI equity indices of 41 developed and emerging markets. Some new insights about linkages among different markets are drawn.

Keywords: Daily Global Stock Market Returns; Time-Zone Differences; Structural Dynamic Factor Model; Quasi Maximum Likelihood; Minimum Distance; Expectation Maximization Algorithm.

JEL classification C55; C58; G15.

1 Introduction

The world's stock markets are separated in time by substantial time-zone differences, to the extent that for example the US and Chinese markets do no overlap. Nevertheless, it is a common belief that they are becoming more and more connected through international

*We thank Matteo Barigozzi, Liang Jiang, Chen Wang, Fa Wang for helpful discussions. Any remaining errors are our own.

†Faculty of Economics, Austin Robinson Building, Sidgwick Avenue, Cambridge, CB3 9DD. Email: obl20@cam.ac.uk. Thanks to the ERC for financial support.

‡Fanhai International School of Finance, Fudan University. 220 Handan Road, Yangpu District, Shanghai, 200433, China. Email: hhtang@fudan.edu.cn.

§School of Economics, Nanjing Universtiy, Nanjing, 210093, China. Email: wujianbin@nju.edu.cn.

trade and cross border investments. Linkages between different markets were particularly evident during stressful times like the financial crisis in 2008 and the COVID-19 outbreak in 2020. The last three decades have witnessed a heightening interest in measuring and modelling such linkages, whether dubbed as the stock market integration, international return spillovers, cross-market contagions etc. Gagnon and Karolyi (2006) and Sharma and Seth (2012) have carefully reviewed the literature and categorized these studies according to methodologies, data sets and findings.

All the existing studies either examined a small number of entities such as a few market indices, or ignored the time-zone differences whenever using daily data. The use of daily closing prices while ignoring the time-zone differences causes the so-called stale-price problem (Martens and Poon (2001), Connor, Goldberg, and Korajczyk (2010, p.42-44)). The standard approach is to include the lead and lagged covariances to correct for the stale pricing effect. This approach does not allow one to identify the source of variation or its relative impacts.

When studying a large number of entities, it is common practice to aggregate information. We provide a framework to model the correlations of daily stock returns in different markets across multiple time zones. The machinery will be a statistical dynamic factor model (first proposed by Forni, Hallin, Lippi, and Reichlin (2000)), which enables us to work with a large number of stocks. To make the framework tractable, we make the following modelling assumption: All the markets belong to one of three continents: Asia (A), Europe (E) and America (U). Within a calendar day, the Asian markets close first, followed by the European and then American markets. We suppose that *observed* logarithmic 24-hr returns (determined as the close-to-close returns) in each continent follow a dynamic factor model with both global and continental factors. This model reflects a situation in which new global information represented by the global factor affects all three continents simultaneously, but is only revealed in the observed returns in the three continents sequentially as their markets open in turn and trade on the new information (Koch and Koch (1991, p.235)). New information represented by a continental factor accumulated since the last closure of that continent's markets will also have an impact on the upcoming observed logarithmic 24-hr returns of those markets.

The approach of having global and continental factors, in some respects, resembles the GVAR modeling approach (Pesaran, Schuermann, and Weiner (2004)), which was developed to model world low-frequency macroeconomic series. Here we do not have so many relevant variables beyond the prices themselves and hence our model is in terms of unobserved factors. Our approach is also closely related to the nowcasting framework (Giannone, Reichlin, and Small (2008), Banbura, Giannone, Modugno, and Reichlin (2013), Aruoba, Diebold, and Scotti (2009) etc). In the nowcasting literature, researchers use factor models to extract the information contained in the data at higher frequencies than the target variable in order to forecast the target variable. Here, if we make *additional* assumptions on the data generating processes of the *unobserved* logarithmic 24-hr returns, we could also obtain their corresponding nowcasts; this is the similarity. The difference is that, as we shall point out in Section 2, model (2.3) is identified under a mild fixed-signs assumption (Assumption 2.2) and hence has a structural interpretation, whereas in the nowcasting literature, identification of factor models is usually not addressed, and factor models are mere dimension-reducing tools with no structural interpretations. In some sense, our model belongs to the class of structural dynamic factor models (Stock and Watson (2016)).

On the theoretical side, research about estimation of large factor models via the likelihood approach has matured over the last decade. The likelihood approach enjoys several advantages such as efficiency compared to the principal components method (Banbura et al. (2013, p.204)). Doz, Giannone, and Reichlin (2012) established an *average* rate of convergence of the estimated factors using a quasi-maximum likelihood estimator (QMLE) via the Kalman smoother. However, there is a rotation matrix attached to the estimated factors as the authors did not address identification of factor models. Also they did not derive consistency for the estimated factor loadings, or the limiting distributions of any estimate.

Bai and Li (2012) took a different approach to study large exact factor models. They treated factors as fixed parameters instead of random vectors. One nice thing about this approach is that the theoretical results obtained hold for any dynamic pattern of factors. Bai and Li (2012) obtained consistency, the rates of convergence and the limiting distributions of the maximum likelihood estimators (MLE) of the factor loadings, idiosyncratic variances, and sample covariance matrix of factors. In fact, Bai and Li (2012) called their estimators the QMLE instead of the MLE. We decided to re-label them as the MLE since we shall reserve the phrase QMLE for another purpose to be made specific shortly. Factors are then estimated via a generalised least squares (GLS) method. Bai and Li (2016) generalised the results of Bai and Li (2012) to large approximate factor models.

In practice, instead of maximising a likelihood and finding the MLE, people usually use the EM algorithm together with the Kalman smoother to estimate the model (Watson and Engle (1983), Quah and Sargent (1992), Doz et al. (2012), Bai and Li (2012), Bai and Li (2016) etc). Since the EM algorithm runs only for a finite number of iterations, strictly speaking the estimate obtained by the EM algorithm is only an approximation to the MLE. However, in a breakthrough study Barigozzi and Luciani (2022) showed that the estimate obtained by the EM algorithm converges to the MLE sufficiently fast, so it has the same asymptotic distribution as the MLE.

We propose several estimators of our model (2.3): the MLE-one day, the QMLE-res, the QMLE, the QMLE-md, and the Bayesian. The MLE-one day estimator is the usual MLE estimator of our model. The QMLE-res estimator is the MLE estimator of the two-day representation of our model while maintaining the working independence hypothesis (see Section 2.2); in this article we shall refer a likelihood-based estimator obtained under the working independence hypothesis as the QMLE rather than the MLE. The QMLE estimator differs from the QMLE-res in the sense that only a specific subset of restrictions implied by our model is imposed. Since the QMLE is inefficient, we propose an improved estimator, the QMLE-md, which uses the QMLE in the first step and incorporates additional finite number of restrictions implied by our model via the minimum distance method in the second step. Last, the Bayesian estimator uses the Gibbs sampling to estimate the model (2.3). However, the Gibbs sampling is computationally intensive and feasible only for a small number of entities.

The large sample results of the aforementioned studies are not directly applicable to our model because the proofs of these results are identification-scheme dependent. In particular, Bai and Li (2012), Bai and Li (2016) established their results under five popular identification schemes, none of which is consistent with our model. In order to have an identification scheme consistent with our model and at the same time utilise the theories of Bai and Li (2012), we could only impose *some*, not all, of the restrictions implied by our model to derive the first-order conditions (FOC) of the log-likelihood. It took us a considerable amount of work to derive the corresponding large sample results of our QMLE estimator. That is, consistency, the rate of convergence and the asymptotic distribution of the QMLE are established. Then the asymptotic distribution of the QMLE-md could be derived as well. The large sample theories of the MLE-one day and the QMLE-res are beyond the scope of this article, and we leave them for future research. Nevertheless, we provide a heuristic procedure to approximate the standard errors of the MLE-one day and the QMLE-res.

We then conduct some Monte Carlo simulations to evaluate the MLE-one day, the QMLE-res and the QMLE-md in terms of the root mean square errors, average of the standard errors across the Monte Carlo samples, and the coverage probability of the constructed confidence interval. Indeed, these three estimators perform well.

Last, we apply our model to two real data sets. The first data set consists of equity portfolio returns from Japan, Europe and the US; that is, one market per continent. Our methodology quantifies how much the global factor loaded on the returns during a particular fraction of a calendar day, as well as the relative importance of the global and continental factors. We also uncover some interesting time-series patterns. The second data set is MSCI equity indices of the 41 developed and emerging markets. Taking the Asian-Pacific continent as an example, we find that Mainland China and Hong Kong have particularly high loadings on the global factor during the US trading time. Japan has high loadings on the continental factor but small idiosyncratic variances, while other Asian-Pacific markets have statistically insignificant loadings on the continental factor.

We contribute to methodology by providing a new modelling framework for daily global stock market returns. Our framework could easily handle a large number of stocks and at the same time take into account the time-zone differences. Under a mild fixed-signs assumption, our model is identified and has a structural interpretation. We also contribute to theory by deriving the asymptotic results of the QMLE and the QMLE-md. The machinery is based on the theoretical results of Bai and Li (2012), but we demonstrate how one could obtain their results for almost *any* identified dynamic factor model. This is an important contribution as many dynamic factor models, like ours, are motivated by different economic theories and might not be compatible with the five identification schemes of Bai and Li (2012). We last contribute to the applied literature by proposing several practically usable estimators and validate their performances in the Monte Carlo simulations. When applying our model to two real data sets, we draw some new insights about linkages among different stock markets.

The rest of the article is structured as follows. In Section 2 we explain our model and discuss identification while in Section 3 we introduce our estimators. Section 4 presents the large sample theories of the QMLE and the QMLE-md. Section 6 conducts the Monte Carlo simulations to assess those advocated estimators, and Section 7 presents two empirical applications of our model. Section 8 concludes. Major proofs and explanations are to be found in Appendix; secondary materials are put in Supplementary Material (SM in what follows).

1.1 Notation

Let $\mathbb{R}^n$ and $\mathbb{Z}^+$ denote the n-dimensional Euclidean space and set of non-negative integers, respectively. For $x \in \mathbb{R}^n$, let $\|x\|_2 := \sqrt{\sum_{i=1}^n x_i^2}$ and $\|x\|_\infty := \max_{1\leq i\leq n} |x_i|$ denote the Euclidean (ℓ_2) and element-wise maximum (ℓ_∞) norms, respectively. Let A be an $m \times n$ matrix. Let $\operatorname{vec} A$ denote the vector obtained by stacking columns of A one underneath the other. Let unvec denote the reverse operation of vec.

The *commutation matrix* $K_{m,n}$ is an $mn \times mn$ *orthogonal* matrix which translates

$\operatorname{vec} A$ to $\operatorname{vec}(A^\intercal)$, i.e., $\operatorname{vec}(A^\intercal) = K_{m,n} \operatorname{vec}(A)$. If A is a symmetric $n \times n$ matrix, its $n(n-1)/2$ supradiagonal elements are redundant in the sense that they can be deduced from symmetry. If we eliminate these redundant elements from $\operatorname{vec} A$, we obtain a new $n(n+1)/2 \times 1$ vector, denoted $\operatorname{vech} A$. They are related by the full-column-rank, $n^2 \times n(n+1)/2$ *duplication matrix* D_n: $\operatorname{vec} A = D_n \operatorname{vech} A$. Conversely, $\operatorname{vech} A = D_n^+ \operatorname{vec} A$, where D_n^+ is $n(n+1)/2 \times n^2$ and the Moore-Penrose generalized inverse of D_n.

Given a vector v, $\operatorname{diag}(v)$ creates a diagonal matrix whose diagonal elements are elements of v. We use $p(\cdot)$ to denote the (asymptotic) probability density function. $\lfloor x \rfloor$ denotes the greatest integer strictly less than $x \in \mathbb{R}$ and $\lceil x \rceil$ denotes the smallest integer greater than or equal to $x \in \mathbb{R}$. Landau (order) notation in this article, unless otherwise stated, should be interpreted in the sense that $N, T \to \infty$ jointly, where N and T are the cross-sectional and temporal dimensions, respectively. We use C or C with number subscripts to denote *absolute* positive constants (i.e., constants independent of anything which is a function of N and/or T); identities of such Cs might change from one place to another.

2 The Model

Our model is based on the closing prices of the stock markets on the three continents, which occur at different calendar times. We suppose that the closing times are ordered as follows:

$$\begin{array}{cccccccc} & A & E & U & A & E & U & A & \cdots \\ t = & 1 & 2 & 3 & 4 & 5 & 6 & 7 & \cdots \end{array}$$

Note that the unit of t is not a day, but a fraction of a day. This framework could be applied to three markets only (i.e., one market in each continent), or to the case where some continent contains several markets. Let $p_{i,t}^c$ denote the logarithmic closing price of stock i in continent c at time t for $c = A, E, U$. We shall assume that there are no weekends as is the prevalent assumption in single market analysis. Nevertheless, we allow that $p_{i,t}^c$ is not observed in two scenarios:

(i) Missing because of non-synchronized trading. That is, t might not correspond to the closing times of continent c. For example, we do not observe $p_{i,3}^E$ for any stock i in the European continent.

(ii) Missing because of some specific reasons. The reasons could be continent-specific (e.g., Chinese New Year, Christmas), market-specific (e.g., national holidays), or stock-specific (e.g., general meetings of shareholders).

We shall rule out scenario (ii) for the time being and address it to some extent in SM B.1. We next define our model.

Define the logarithmic 24-hr return $y_{i,t}^c := p_{i,t}^c - p_{i,t-3}^c$ for $c = A, E, U$, and $T_A := \{1, 4, 7, \ldots, T-2\}$, $T_E := \{2, 5, 8, \ldots, T-1\}$ and $T_U := \{3, 6, 9, \ldots, T\}$, where T is a multiple of 3. We assume that the *observed* logarithmic 24-hr returns follow the dynamic system: For $c = A, E, U$,

$$y_{i,t}^c = z_i^c(L) f_{g,t} + \tilde{z}_i^c(L) f_{c,t} + e_{i,t}^c, \quad i = 1, \ldots, N_c, \quad t \in T_c \tag{2.1}$$

where $f_{g,t}$ and $f_{c,t}$ are the scalar unobserved global and continental factors, respectively, while

$$z_i^c(L)f_{g,t} := \sum_{j\in\mathbb{Z}^+} z_{i,j}^c f_{g,t-j}, \qquad \tilde{z}_i^c(L)f_{c,t} := \sum_{j\in\mathbb{Z}^+} \tilde{z}_{i,j}^c f_{c,t-3j}$$

for $c = A, E, U$, where L is the lag operator $Lx_t = x_{t-1}$. The model is dynamic in the sense that

$$\begin{aligned} \phi_g(L)f_{g,t+1} &= \eta_{g,t}, \qquad & t &= 0, 1, \ldots, T-1 \\ \phi_c(L)f_{c,t+1} &= \eta_{c,t}, \qquad & t+1 &\in T_c \end{aligned}$$

where

$$\phi_g(L)f_{g,t+1} := \sum_{j\in\mathbb{Z}^+} \phi_{g,j} f_{g,t+1-j}, \qquad \phi_c(L)f_{c,t+1} := \sum_{j\in\mathbb{Z}^+} \phi_{c,j} f_{g,t+1-3j}$$

for $c = A, E, U$. Note that at every t we only observe the logarithmic 24-hr returns for one continent. The lag polynomials acting on $f_{c,t}$ are autoregressive in terms of periods in T_c as we cannot extract $f_{c,t}$ for $t \notin T_c$ without additional assumptions.

Although some of the aforementioned studies allow dynamics of factors, say, factors following an AR(1) process, strictly speaking those factor models are not dynamic factor models in the sense that the lagged factors are not allowed to enter the equation relating factors to the observed series (see Bai and Wang (2015)). Exceptions are Forni et al. (2000) and Barigozzi and Luciani (2022).

2.1 A Particular Form of (2.1) and Its State Space Form

For the rest of the article, we shall focus on a specific form of (2.1) for simplicity: For $c = A, E, U$, $z_i^c(L)f_{g,t} = \sum_{j=0}^2 z_{i,j}^c f_{g,t-j}$, $\tilde{z}_i^c(L)f_{c,t} = \tilde{z}_{i,0}^c f_{c,t}$, $\phi_g(L)f_{g,t+1} = f_{g,t+1} - \phi f_{g,t}$ ($|\phi| < 1$), and $\phi_c(L)f_{c,t+1} = f_{c,t+1}$, so that (2.1) becomes

$$\begin{aligned} y_{i,t}^c &= \sum_{j=0}^2 z_{i,j}^c f_{g,t-j} + \tilde{z}_{i,0}^c f_{c,t} + e_{i,t}^c, \quad i = 1, \ldots, N_c, \quad t \in T_c \\ f_{g,t+1} &= \phi f_{g,t} + \eta_{g,t}, \qquad |\phi| < 1, \qquad t = 0, 1, \ldots, T-1 \\ f_{c,t+1} &= \eta_{c,t} \qquad t+1 \in T_c \end{aligned} \tag{2.2}$$

Model (2.2) is natural for our framework in the sense that all the new information represented by the global and continental factors accumulated since the last closure of continent c will have an impact on the upcoming observed logarithmic 24-hr returns of continent c. For the case of several markets in some continent, the presence of ϕ in (2.2) allows one to capture the effect caused by the fact that some markets in the same continent might have different closing times. However, for the case of one market per continent, the efficient market hypothesis (along with a time invariant risk premium) predicts that $\phi = 0$.

Stacking all the stocks in continent c, we have

$$\boldsymbol{y}_t^c = Z^c \boldsymbol{\alpha}_t + \boldsymbol{e}_t^c, \qquad t \in T_c,$$

where

$$\boldsymbol{y}_t^c := \begin{bmatrix} y_{1,t}^c \\ \vdots \\ y_{N_c,t}^c \end{bmatrix}, \quad \boldsymbol{e}_t^c := \begin{bmatrix} e_{1,t}^c \\ \vdots \\ e_{N_c,t}^c \end{bmatrix} \quad \boldsymbol{\alpha}_t := \begin{bmatrix} f_{g,t} \\ f_{g,t-1} \\ f_{g,t-2} \\ f_{C,t} \end{bmatrix}$$

$$Z^c := \begin{bmatrix} z_{1,0}^c & z_{1,1}^c & z_{1,2}^c & z_{1,3}^c \\ \vdots & \vdots & \vdots & \vdots \\ z_{N_c,0}^c & z_{N_c,1}^c & z_{N_c,2}^c & z_{N_c,3}^c \end{bmatrix} =: \begin{bmatrix} \boldsymbol{z}_0^c & \boldsymbol{z}_1^c & \boldsymbol{z}_2^c & \boldsymbol{z}_3^c \end{bmatrix},$$

where $f_{C,t} := f_{c,t}$ if $t \in T_c$, $z_{i,3}^c := \tilde{z}_{i,0}^c$, and $f_{g,t}, f_{C,t} := 0$ for $t \leq 0$. We compress $f_{A,t}, f_{E,t}, f_{U,t}$ into a "single" continental factor $f_{C,t}$ for the purpose of reducing the number of state variables. Note that $\boldsymbol{y}_t^c, \boldsymbol{e}_t^c$ are $N_c \times 1$ vectors, $\boldsymbol{\alpha}_t$ is 4×1, and Z^c is $N_c \times 4$. We make the following assumptions:

Assumption 2.1. *(i) The idiosyncratic components are i.i.d. across time: $\{\boldsymbol{e}_t^c\}_{t \in T_c} \overset{i.i.d.}{\sim} N(0, \Sigma_c)$, where $\Sigma_c := \operatorname{diag}(\sigma_{c,1}^2, \ldots, \sigma_{c,N_c}^2)$ for $c = A, E, U$. Moreover, $e_{i,t}^A, e_{i,t}^E, e_{i,t}^U$ are mutually independent for all possible i and t. Moreover, $\mathbb{E}[(e_{i,t}^c)^4] \leq C$ for all i, t and c.*

(ii) Assume that $\boldsymbol{\eta}_t := (\eta_{g,t}, \eta_{C,t})^\intercal \overset{i.i.d.}{\sim} N(0, I_2)$ for $t = 0, 1, \ldots, T-1$, where $\eta_{C,t} := \eta_{c,t}$ if $t \in T_c$. Moreover, $\{\boldsymbol{\eta}_t\}_{t=1}^{T-1}$ are independent of $\{\boldsymbol{e}_t^c\}_{t \in T_c}$ for $c = A, E, U$.

Assumption 2.1(i) is the same as Assumption B of Bai and Li (2012). We make the assumption of diagonality of Σ_c for simplicity as our model is already quite involved so we refrain from complicating the model unnecessarily. Assumption 2.1(ii) is a white-noise assumption on the innovations of the factors.

We now cast model (2.2) in the state space form

$$\boldsymbol{y}_t = Z_t \boldsymbol{\alpha}_t + \boldsymbol{\varepsilon}_t, \qquad \boldsymbol{\varepsilon}_t \sim N(0, \Sigma_t), \qquad t = 1, \ldots, T, \tag{2.3}$$

$$\boldsymbol{\alpha}_{t+1} = \begin{bmatrix} \phi & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{bmatrix} \boldsymbol{\alpha}_t + \begin{bmatrix} 1 & 0 \\ 0 & 0 \\ 0 & 0 \\ 0 & 1 \end{bmatrix} \boldsymbol{\eta}_t =: \mathcal{T} \boldsymbol{\alpha}_t + R \boldsymbol{\eta}_t, \quad t = 0, 1, \ldots, T-1,$$

where $\boldsymbol{y}_t = \boldsymbol{y}_t^c, Z_t = Z^c, \boldsymbol{\varepsilon}_t = \boldsymbol{e}_t^c, \Sigma_t = \Sigma_c$ if $t \in T_c$, for $c = A, E, U$. This is a non-standard dynamic factor model. The non-standard features are: (1) The factor loading matrix Z_t is switching among three states $\{Z^A, Z^E, Z^U\}$. (2) The column dimensions of $\boldsymbol{y}_t, Z_t, \boldsymbol{\varepsilon}_t$ are switching among $\{N_A, N_E, N_U\}$. (3) The covariance matrix of $\boldsymbol{\varepsilon}_t$ is switching among $\{\Sigma_A, \Sigma_E, \Sigma_U\}$.

In general for a *static* factor model, say, $\boldsymbol{y}_t = Z\boldsymbol{\alpha}_t + \boldsymbol{\varepsilon}_t$, further identification restrictions are needed in order to separately identify Z and $\boldsymbol{\alpha}_t$ from the term $Z\boldsymbol{\alpha}_t$. In particular, $Z\boldsymbol{\alpha}_t = \mathring{Z}\mathring{\boldsymbol{\alpha}}_t$ for any 4×4 invertible matrix C such that $\mathring{Z} := ZC^{-1}$ and $\mathring{\boldsymbol{\alpha}}_t := C\boldsymbol{\alpha}_t$; we need 4^2 identification restrictions so that the only admissible C is an identity matrix. A classical reference on this issue would be Anderson and Rubin (1956). These restrictions have been ubiquitous in the literature (e.g., Bai and Li (2012), Bai and Li (2016)). One exception is Bai and Wang (2015); Bai and Wang (2015) pointed out that by relying on the dynamic equation of $f_{g,t}$, such as (2.2), one could use far less identification restrictions to identify the model. In our case, we shall only make the following mild fixed-signs assumption to identify the model.

Assumption 2.2. *Estimators of* $\boldsymbol{z}_0^A, \boldsymbol{z}_1^A, \boldsymbol{z}_2^A, \boldsymbol{z}_3^A, \boldsymbol{z}_3^E, \boldsymbol{z}_3^U$ *have the same column signs as* $\boldsymbol{z}_0^A, \boldsymbol{z}_1^A, \boldsymbol{z}_2^A, \boldsymbol{z}_3^A, \boldsymbol{z}_3^E, \boldsymbol{z}_3^U$.

Bai and Li (2012) have made similar assumption as an implicit part of their identification schemes (IC2, IC3 and IC5) (see Bai and Li (2012, p.445, p.463)).

Lemma 2.1. *The parameters of the dynamic factor model (2.3) are identified under Assumption 2.2.*

Our model (2.3) has a structural interpretation under Assumption 2.2, because one could not freely insert a rotation matrix between Z_t and $\boldsymbol{\alpha}_t$. In other words, under Assumption 2.2 we are not estimating the rotations of Z_t or $\boldsymbol{\alpha}_t$; we are estimating the *true* Z_t and $\boldsymbol{\alpha}_t$ of the data generating process. This is a novel feature of our model.

2.2 The Two-Day Representation

The representation of the model in state space form is of course not unique. For simplicity, assume $N := N_A = N_E = N_U$ hereafter. We re-write our model (2.3) in the following two-day representation:

$$\underbrace{\mathring{\boldsymbol{y}}_t}_{6N\times 1} = \underbrace{\Lambda}_{6N\times 14} \underbrace{\boldsymbol{f}_t}_{14\times 1} + \underbrace{\boldsymbol{e}_t}_{6N\times 1} \tag{2.4}$$

for $t = 1, 2, \ldots, T/6 =: T_f$, where we define $\ell := 6(t-1)+1$,

$$\mathring{\boldsymbol{y}}_t := \begin{bmatrix} \boldsymbol{y}_\ell^A \\ \boldsymbol{y}_{\ell+1}^E \\ \boldsymbol{y}_{\ell+2}^U \\ \boldsymbol{y}_{\ell+3}^A \\ \boldsymbol{y}_{\ell+4}^E \\ \boldsymbol{y}_{\ell+5}^U \end{bmatrix} \qquad \boldsymbol{e}_t := \begin{bmatrix} \boldsymbol{e}_\ell^A \\ \boldsymbol{e}_{\ell+1}^E \\ \boldsymbol{e}_{\ell+2}^U \\ \boldsymbol{e}_{\ell+3}^A \\ \boldsymbol{e}_{\ell+4}^E \\ \boldsymbol{e}_{\ell+5}^U \end{bmatrix} \qquad \boldsymbol{f}_t := \begin{bmatrix} f_{g,\ell+5} \\ f_{g,\ell+4} \\ f_{g,\ell+3} \\ f_{g,\ell+2} \\ f_{g,\ell+1} \\ f_{g,\ell} \\ f_{g,\ell-1} \\ f_{g,\ell-2} \\ f_{C,\ell+5} \\ f_{C,\ell+4} \\ f_{C,\ell+3} \\ f_{C,\ell+2} \\ f_{C,\ell+1} \\ f_{C,\ell} \end{bmatrix}$$

$$\Lambda := \begin{bmatrix} \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \boldsymbol{z}_0^A & \boldsymbol{z}_1^A & \boldsymbol{z}_2^A & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \boldsymbol{z}_3^A \\ \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \boldsymbol{z}_0^E & \boldsymbol{z}_1^E & \boldsymbol{z}_2^E & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \boldsymbol{z}_3^E & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \mathbf{0} & \boldsymbol{z}_0^U & \boldsymbol{z}_1^U & \boldsymbol{z}_2^U & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \boldsymbol{z}_3^U & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \boldsymbol{z}_0^A & \boldsymbol{z}_1^A & \boldsymbol{z}_2^A & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \boldsymbol{z}_3^A & \mathbf{0} & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \boldsymbol{z}_0^E & \boldsymbol{z}_1^E & \boldsymbol{z}_2^E & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \boldsymbol{z}_3^E & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} \\ \boldsymbol{z}_0^U & \boldsymbol{z}_1^U & \boldsymbol{z}_2^U & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \boldsymbol{z}_3^U & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} \end{bmatrix}, \tag{2.5}$$

with $\boldsymbol{z}_k^c$ being $N \times 1$ for $k = 0, 1, 2, 3$ and $c = A, E, U$, while

$$\begin{aligned}
&\Sigma_{ee} := \mathbb{E}[\boldsymbol{e}_t \boldsymbol{e}_t^\intercal] \\
&= \operatorname{diag}(\sigma_{A,1}^2, \ldots, \sigma_{A,N}^2, \sigma_{E,1}^2, \ldots, \sigma_{E,N}^2, \sigma_{U,1}^2, \ldots, \sigma_{U,N}^2, \sigma_{A,1}^2, \ldots, \sigma_{A,N}^2, \sigma_{E,1}^2, \ldots, \sigma_{E,N}^2, \sigma_{U,1}^2, \ldots, \sigma_{U,N}^2) \\
&=: \operatorname{diag}(\sigma_{1,1}^2, \ldots, \sigma_{1,N}^2, \sigma_{2,1}^2, \ldots, \sigma_{2,N}^2, \sigma_{3,1}^2, \ldots, \sigma_{3,N}^2, \sigma_{4,1}^2, \ldots, \sigma_{4,N}^2, \sigma_{5,1}^2, \ldots, \sigma_{5,N}^2, \sigma_{6,1}^2, \ldots, \sigma_{6,N}^2).
\end{aligned}$$

Note that Λ consists of six row blocks of dimension $N \times 14$. Let $\boldsymbol{\lambda}_{k,j}^\intercal$ denote the jth row of the kth row block of Λ. In other words, $\boldsymbol{\lambda}_{1,j}^\intercal$ refers to the factor loadings for the jth Asian stock in day one, while $\boldsymbol{\lambda}_{5,j}^\intercal$ refers to the factor loadings for the jth European stock in day two.

The reason for doing so is that we could rely on the information contained in the covariance matrix $M := \mathbb{E}[\boldsymbol{f}_t \boldsymbol{f}_t^\intercal]$ to estimate ϕ, while treating $\{\boldsymbol{f}_t\}$ as i.i.d. across t when setting up the likelihood. Then we are able to use the theoretical results of Bai and Li (2012) to establish the large-sample theories of the QMLE estimator of this representation. Treating $\{\boldsymbol{f}_t\}$ as i.i.d. when setting up the likelihood, albeit incorrectly, will not destroy consistency or the asymptotic normality of the QMLE. This is the idea of *working* independence (Pan and Connett (2002)).

3 Estimation

In this section, we shall outline several different estimation methods applicable for our model (2.3). The estimators start from different formulations of the state space model and impose different subsets of the available parameter restrictions. The reason we consider these different estimators is because of the difficulty of deriving the distribution theory in some cases and the different numerical performance we have uncovered. The different estimators also provide an understanding of the theoretical issues we face.

MLE-one day. This is the MLE estimator of the likelihood function $\{\boldsymbol{y}_t\}_{t=1}^T$, where $\boldsymbol{y}_t = Z_t \boldsymbol{\alpha}_t + \boldsymbol{\varepsilon}_t$. Estimation is done by the EM algorithm (to be explained in Section 3.1).

MLE-two day. This is the MLE estimator of the likelihood function $\{\mathring{\boldsymbol{y}}_t\}_{t=1}^{T_f}$, where $\mathring{\boldsymbol{y}}_t = \Lambda \boldsymbol{f}_t + \boldsymbol{e}_t$. All the restrictions implied by Λ and M, and implied by autocorrelation between $\boldsymbol{f}_t$ and $\boldsymbol{f}_{t-1}$ will be taken into account. This is equivalent to the MLE-one day. This estimator is not used in this article, but will help readers understand the relationships among various estimators. Estimation is done by the EM algorithm.

QMLE-res. This is the QMLE estimator of the likelihood of $\{\mathring{\boldsymbol{y}}_t\}_{t=1}^{T_f}$, where $\mathring{\boldsymbol{y}}_t = \Lambda \boldsymbol{f}_t + \boldsymbol{e}_t$. All the restrictions implied by Λ and M are taken into account. However, autocorrelation between $\boldsymbol{f}_t$ and $\boldsymbol{f}_{t-1}$ is ignored, and $\{\boldsymbol{f}_t\}_{t=1}^{T_f}$ are assumed as i.i.d over t when setting up the likelihood. Estimation is done by the EM algorithm (to be explained in Section 3.4).

QMLE. This is the QMLE estimator of the likelihood of $\{\mathring{\boldsymbol{y}}_t\}_{t=1}^{T_f}$, where $\mathring{\boldsymbol{y}}_t = \Lambda \boldsymbol{f}_t + \boldsymbol{e}_t$. A specific set of 14^2 restrictions implied by Λ and M is employed. Autocorrelation between $\boldsymbol{f}_t$ and $\boldsymbol{f}_{t-1}$ is ignored, and $\{\boldsymbol{f}_t\}_{t=1}^{T_f}$ are assumed as i.i.d over t when setting up the likelihood. This estimator is explained in Section 3.2. Estimation is done by the EM algorithm (see Appendix A.13).

Bai and Li (2012)'s QMLE. These are the QMLE estimators of the likelihood of $\{\mathring{\boldsymbol{y}}_t\}_{t=1}^{T_f}$, where $\mathring{\boldsymbol{y}}_t = \Lambda \boldsymbol{f}_t + \boldsymbol{e}_t$ with five specific sets of 14^2 restrictions consistent with the five identification schemes of Bai and Li (2012). Autocorrelation between $\boldsymbol{f}_t$ and $\boldsymbol{f}_{t-1}$ is ignored, and $\{\boldsymbol{f}_t\}_{t=1}^{T_f}$ are assumed i.i.d over t when setting up the likelihood. Unfortunately, our model is not consistent with any one of the five identification schemes.

QMLE-md. Use the QMLE as the first-step estimator and incorporate additional *finite* number of restrictions implied by our model to obtain an improved estimator via the minimum distance method. This estimator is explained in Section 3.3.

Bayesian. Use the Gibbs sampling to estimate $\boldsymbol{y}_t = Z_t\boldsymbol{\alpha}_t + \boldsymbol{\varepsilon}_t$. This estimator is explained in Appendix A.2. This estimator is computationally intensive and feasible only for not so large N.

3.1 MLE-one day

Define $\boldsymbol{\theta} := \{\phi, Z^c, \Sigma_c, c = A, E, U\}$, $\Xi := (\boldsymbol{\alpha}_1, \ldots, \boldsymbol{\alpha}_T)^\intercal$ and $Y_{1:T} := \{\boldsymbol{y}_1^\intercal, \ldots, \boldsymbol{y}_T^\intercal\}^\intercal$. The log-likelihoods of Ξ and $Y_{1:T}|\Xi$ are

$$\ell(\Xi;\boldsymbol{\theta}) = -T\log(2\pi) - \frac{1}{2}\sum_{t=0}^{T-1}\left[(f_{g,t+1} - \phi f_{g,t})^2 + f_{C,t+1}^2\right] \tag{3.1}$$

$$\ell(Y_{1:T}|\Xi;\boldsymbol{\theta}) = -\frac{TN}{2}\log(2\pi) - \frac{1}{2}\sum_{t=1}^{T}\log|\Sigma_t| - \frac{1}{2}\sum_{t=1}^{T}(\boldsymbol{y}_t - Z_t\boldsymbol{\alpha}_t)^\intercal\Sigma_t^{-1}(\boldsymbol{y}_t - Z_t\boldsymbol{\alpha}_t).$$

The *complete* log-likelihood function of model (2.3), i.e., based on an observed state vector, is hence (omitting constant)

$$\begin{aligned}
&\ell(\Xi, Y_{1:T};\boldsymbol{\theta}) = \ell(Y_{1:T}|\Xi;\boldsymbol{\theta}) + \ell(\Xi;\boldsymbol{\theta}) \\
&= -\frac{1}{2}\sum_{t=1}^{T}\left(\log|\Sigma_t| + \boldsymbol{\varepsilon}_t^\intercal\Sigma_t^{-1}\boldsymbol{\varepsilon}_t\right) - \frac{1}{2}\sum_{t=1}^{T}\eta_{g,t-1}^2 =: -\frac{1}{2}\sum_{t=1}^{T}\left(\ell_{1,t} + \ell_{2,t}\right)
\end{aligned}$$

where $\ell_{1,t} := \log|\Sigma_t| + \operatorname{tr}\left(\boldsymbol{\varepsilon}_t\boldsymbol{\varepsilon}_t^\intercal\Sigma_t^{-1}\right)$ and $\ell_{2,t} := \eta_{g,t-1}^2$.

The EM algorithm consists of an E-step and an M-step.[1] In the E-step, we evaluate a conditional expectation of the complete log-likelihood function given the observed data, while in the M-step we maximize it with respect to parameters. To give the starting values of parameters in the EM algorithm, we first use the MLE to estimate a restricted version of model (2.3). Take the Asian continent as an example. All the elements of $\boldsymbol{z}_0^A, \boldsymbol{z}_1^A, \boldsymbol{z}_2^A$ are set to one scalar, all the elements of $\boldsymbol{z}_3^A$ are set to one scalar, and all the diagonal elements of Σ_A are set to one scalar. This will give reasonably good starting values. We do not use the Principal Component (PC) estimator as the starting values because in finite samples the PC estimator will not ensure that $\hat{\boldsymbol{\alpha}}_{t,1}^{PC} = \hat{\boldsymbol{\alpha}}_{t+1,2}^{PC}$ for all t, where $\hat{\boldsymbol{\alpha}}_t^{PC}$ is the PC estimator of $\boldsymbol{\alpha}_t$.

[1] Motivation of the EM algorithm is reviewed in SM B.2

3.1.1 E-Step

Let $\tilde{\mathbb{E}}$ denote the expectation with respect to the conditional density $p(\Xi|Y_{1:T};\tilde{\boldsymbol{\theta}}^{(i)})$ at $\tilde{\boldsymbol{\theta}}^{(i)}$, where $\tilde{\boldsymbol{\theta}}^{(i)}$ is the estimate of $\boldsymbol{\theta}$ from the ith iteration of the EM algorithm. Taking such expectation on both sides of the preceding display, we hence have

$$\tilde{\mathbb{E}}\left[\ell(\Xi, Y_{1:T};\boldsymbol{\theta})\right] = \text{constant} - \frac{1}{2}\left(\tilde{\mathbb{E}}\sum_{t=1}^{T}\ell_{1,t} + \tilde{\mathbb{E}}\sum_{t=1}^{T}\ell_{2,t}\right).$$

This is the so-called "E" step of the EM algorithm. $\tilde{\mathbb{E}}[\cdot]$ could be computed using the Kalman smoother (KS; see Section A.3 for formulas of the KS).

3.1.2 M-Step

The M step involves maximising $\tilde{\mathbb{E}}\left[\ell(\Xi, Y_{1:T};\boldsymbol{\theta})\right]$ with respect to $\boldsymbol{\theta}$. This is usually done by computing

$$\frac{\partial\tilde{\mathbb{E}}\left[\ell(\Xi, Y_{1:T};\boldsymbol{\theta})\right]}{\partial\boldsymbol{\theta}}$$

and setting the preceding display to zero to obtain the estimate $\tilde{\boldsymbol{\theta}}^{(i+1)}$ of $\boldsymbol{\theta}$ for the $(i+1)$th iteration of the EM algorithm.

M-Step of Z_t and Σ_t We now find values of Z_t and Σ_t to minimize $\tilde{\mathbb{E}}\sum_{t=1}^{T}\ell_{1,t}$. Recall that $Z_t = Z^c, \Sigma_t = \Sigma_c$ if $t \in T_c$ for $c = A, E, U$. Without loss of generality, we shall use the Asian continent to illustrate the procedure. We now find values of Z^A and Σ_A to minimise $\tilde{\mathbb{E}}\sum_{t\in T_A}\ell_{1,t}$. Since $\boldsymbol{\varepsilon}_t = \boldsymbol{y}_t - Z_t\boldsymbol{\alpha}_t$ and $\boldsymbol{\eta}_t = R^\intercal(\boldsymbol{\alpha}_{t+1} - \mathcal{T}\boldsymbol{\alpha}_t)$ (see (2.3)), we have

$$\sum_{t\in T_A}\ell_{1,t} = \frac{T}{3}\log|\Sigma_A| + \sum_{t\in T_A}\operatorname{tr}\left(\left[\boldsymbol{y}_t\boldsymbol{y}_t^\intercal - 2Z^A\boldsymbol{\alpha}_t\boldsymbol{y}_t^\intercal + Z^A\boldsymbol{\alpha}_t\boldsymbol{\alpha}_t^\intercal Z^{A\intercal}\right]\Sigma_A^{-1}\right)$$

and hence

$$\tilde{\mathbb{E}}\sum_{t\in T_A}\ell_{1,t} = \frac{T}{3}\log|\Sigma_A| + \sum_{t\in T_A}\operatorname{tr}\left(\left[\boldsymbol{y}_t\boldsymbol{y}_t^\intercal - 2Z^A\tilde{\mathbb{E}}[\boldsymbol{\alpha}_t]\boldsymbol{y}_t^\intercal + Z^A\tilde{\mathbb{E}}[\boldsymbol{\alpha}_t\boldsymbol{\alpha}_t^\intercal]Z^{A\intercal}\right]\Sigma_A^{-1}\right) \quad (3.2)$$

We now consider Z^A, and take differential of (3.2) with respect to Z^A:

$$\begin{aligned}\mathrm{d}\tilde{\mathbb{E}}\sum_{t\in T_A}\ell_{1,t} &= \sum_{t\in T_A}\operatorname{tr}\left(\left[-2\mathrm{d}Z^A\tilde{\mathbb{E}}[\boldsymbol{\alpha}_t]\boldsymbol{y}_t^\intercal + 2\mathrm{d}Z^A\tilde{\mathbb{E}}[\boldsymbol{\alpha}_t\boldsymbol{\alpha}_t^\intercal]Z^{A\intercal}\right]\Sigma_A^{-1}\right)\\ &= -2\sum_{t\in T_A}\operatorname{tr}\left(\mathrm{d}Z^A\left[\tilde{\mathbb{E}}[\boldsymbol{\alpha}_t]\boldsymbol{y}_t^\intercal - \tilde{\mathbb{E}}[\boldsymbol{\alpha}_t\boldsymbol{\alpha}_t^\intercal]Z^{A\intercal}\right]\Sigma_A^{-1}\right),\end{aligned}$$

whence we have

$$\tilde{Z}^A = \sum_{t\in T_A}\tilde{\mathbb{E}}[\boldsymbol{y}_t\boldsymbol{\alpha}_t^\intercal]\left(\sum_{t\in T_A}\tilde{\mathbb{E}}[\boldsymbol{\alpha}_t\boldsymbol{\alpha}_t^\intercal]\right)^{-1}. \quad (3.3)$$

We next consider Σ_A. Define

$$C_A := \sum_{t\in T_A}\left[\boldsymbol{y}_t\boldsymbol{y}_t^{\intercal} - 2Z^A\tilde{\mathbb{E}}[\boldsymbol{\alpha}_t]\boldsymbol{y}_t^{\intercal} + Z^A\tilde{\mathbb{E}}[\boldsymbol{\alpha}_t\boldsymbol{\alpha}_t^{\intercal}]Z^{A\intercal}\right]$$
$$\tilde{C}_A := \sum_{t\in T_A}\left[\boldsymbol{y}_t\boldsymbol{y}_t^{\intercal} - 2\tilde{Z}^A\tilde{\mathbb{E}}[\boldsymbol{\alpha}_t]\boldsymbol{y}_t^{\intercal} + \tilde{Z}^A\tilde{\mathbb{E}}[\boldsymbol{\alpha}_t\boldsymbol{\alpha}_t^{\intercal}]\tilde{Z}^{A\intercal}\right]$$

Then (3.2) can be written as

$$\tilde{\mathbb{E}}\sum_{t\in T_A}\ell_{1,t} = \frac{T}{3}\log|\Sigma_A| + \mathrm{tr}(C_A\Sigma_A^{-1})$$

Take the differential of $\tilde{\mathbb{E}}\sum_{t\in T_A}\ell_{1,t}$ with respect to Σ_A

$$\mathrm{d}\tilde{\mathbb{E}}\sum_{t\in T_A}\ell_{1,t} = \frac{T}{3}\,\mathrm{tr}(\Sigma_A^{-1}\mathrm{d}\Sigma_A) - \mathrm{tr}(\Sigma_A^{-1}C_A\Sigma_A^{-1}\mathrm{d}\Sigma_A) = \mathrm{tr}\left(\Sigma_A^{-1}\left[\frac{T}{3}\Sigma_A - C_A\right]\Sigma_A^{-1}\mathrm{d}\Sigma_A\right)$$

whence we have, recognising the diagonality of Σ_A,

$$\frac{\partial\tilde{\mathbb{E}}\sum_{t\in T_A}\ell_{1,t}}{\partial\Sigma_A} = \Sigma_A^{-1}\left[\frac{T}{3}\Sigma_A - C_A\right]\Sigma_A^{-1}\circ I_{N_A}$$

where $\circ$ denotes the Hadamard product. The first-order condition of Σ_A is

$$\tilde{\Sigma}_A = \frac{3}{T}(\tilde{C}_A\circ I_N).$$

M-Step of ϕ We now find value of ϕ to minimize $\tilde{\mathbb{E}}\sum_{t=1}^{T}\ell_{2,t}$. We have

$$\tilde{\mathbb{E}}\sum_{t=1}^{T}\ell_{2,t} = \sum_{t=1}^{T}\tilde{\mathbb{E}}\eta_{g,t-1}^2 = \sum_{t=1}^{T}\left(\tilde{\mathbb{E}}[f_{g,t}^2] - 2\phi\tilde{\mathbb{E}}[f_{g,t}f_{g,t-1}] + \phi^2\tilde{\mathbb{E}}[f_{g,t-1}^2]\right),$$

whence the first order condition of ϕ gives

$$\tilde{\phi} = \left(\sum_{t=1}^{T}\tilde{\mathbb{E}}[f_{g,t-1}^2]\right)^{-1}\sum_{t=1}^{T}\tilde{\mathbb{E}}[f_{g,t}f_{g,t-1}]. \tag{3.4}$$

Remark 3.1. *As mentioned before, our model is perfectly geared for the scenario of missing observations due to non-synchronized trading (scenario (i)). In SM B.1, we discuss how to alter the EM algorithm if we include the scenario of missing observations due to continent-specific reasons such as continent-wide public holidays (e.g., Chinese New Year). That is, both scenario (i) and a specific form of scenario (ii) are present in the data. We do not consider other forms of scenario (ii) - missing observations due to market-specific, stock-specific reasons - in this article.*

3.2 QMLE

We work with the two-day representation (2.4). We define the following quantities:

$$S_{yy} := \frac{1}{T_f}\sum_{t=1}^{T_f} \mathring{\boldsymbol{y}}_t \mathring{\boldsymbol{y}}_t^\intercal, \qquad \Sigma_{yy} := \Lambda M \Lambda^\intercal + \Sigma_{ee},$$

where $M := \mathbb{E}[\boldsymbol{f}_t \boldsymbol{f}_t^\intercal]$ and $T_f := T/6$. It can be seen that the 14×14 matrix M and the 8×8 matrix Φ are as follows:

$$M = \begin{bmatrix} \Phi & 0 \\ 0 & I_6 \end{bmatrix}, \qquad \Phi := \frac{1}{1-\phi^2}\begin{bmatrix} 1 & \phi & \phi^2 & \phi^3 & \phi^4 & \phi^5 & \phi^6 & \phi^7 \\ \phi & 1 & \phi & \phi^2 & \phi^3 & \phi^4 & \phi^5 & \phi^6 \\ \phi^2 & \phi & 1 & \phi & \phi^2 & \phi^3 & \phi^4 & \phi^5 \\ \phi^3 & \phi^2 & \phi & 1 & \phi & \phi^2 & \phi^3 & \phi^4 \\ \phi^4 & \phi^3 & \phi^2 & \phi & 1 & \phi & \phi^2 & \phi^3 \\ \phi^5 & \phi^4 & \phi^3 & \phi^2 & \phi & 1 & \phi & \phi^2 \\ \phi^6 & \phi^5 & \phi^4 & \phi^3 & \phi^2 & \phi & 1 & \phi \\ \phi^7 & \phi^6 & \phi^5 & \phi^4 & \phi^3 & \phi^2 & \phi & 1 \end{bmatrix}. \tag{3.5}$$

Given the assumption of $|\phi| < 1$ in (2.2), we have $M = O(1)$ and $M^{-1} = O(1)$.

Treating $\{\boldsymbol{f}_t\}_{t=1}^{T_f}$ as i.i.d over t, we can write down the log-likelihood of $\{\mathring{\boldsymbol{y}}_t\}_{t=1}^{T_f}$ (scaled by $1/(NT_f)$):

$$\frac{1}{NT_f}\ell\left(\{\mathring{\boldsymbol{y}}_t\}_{t=1}^{T_f}; \boldsymbol{\theta}\right) = -3\log(2\pi) - \frac{1}{2N}\log|\Sigma_{yy}| - \frac{1}{2N}\operatorname{tr}(S_{yy}\Sigma_{yy}^{-1}). \tag{3.6}$$

We shall only utilise the information that M is symmetric, positive definite and that Σ_{ee} is diagonal to derive the generic first-order conditions (FOCs).[2] In Appendix A.4, we derive such FOCs and identify the QMLE estimators $\hat{\Lambda}, \hat{M}, \hat{\Sigma}_{ee}$ after imposing 14^2 identification restrictions. The QMLE estimators $\hat{\Lambda}, \hat{M}, \hat{\Sigma}_{ee}$ satisfy the equations

$$\begin{aligned} \hat{\Lambda}^\intercal \hat{\Sigma}_{yy}^{-1}(S_{yy} - \hat{\Sigma}_{yy}) &= 0 \\ \operatorname{diag}(\hat{\Sigma}_{yy}^{-1}) &= \operatorname{diag}(\hat{\Sigma}_{yy}^{-1} S_{yy} \hat{\Sigma}_{yy}^{-1}), \end{aligned} \tag{3.7}$$

where $\hat{\Sigma}_{yy} := \hat{\Lambda}\hat{M}\hat{\Lambda}^\intercal + \hat{\Sigma}_{ee}$. Display (3.7) is the same as (2.7) and (2.8) of Bai and Li (2012). Bai and Li (2012) considered five identification schemes, none of which is consistent with Λ and M defined in (2.5) and (3.5), respectively. Actually Λ and M imply more than 14^2 restrictions, but in order to have a solution for the generic FOCs, we could only impose 14^2 restrictions on $\hat{\Lambda}$ and $\hat{M}$. We call the resulting estimators the QMLE rather than the MLE because $\{\boldsymbol{f}_t\}_{t=1}^{T_f}$ are assumed i.i.d over t when setting up the likelihood. How to select these 14^2 restrictions from those implied by Λ and M are crucial because we cannot afford imposing a restriction which is not instrumental for establishing large-sample theories later. We painstakingly explain our procedure in the proofs of Proposition 4.1 and Theorems 4.1, 4.3. Our procedure is quite ingenious and does not exist in the proofs of Bai and Li (2012).

[2]The word *generic* means that specific forms of Λ given by (2.5) and of M given by (3.5) are not utilised to derive the FOCs.

3.3 QMLE-md

The QMLE estimator we defined above only used 14^2 restrictions and there are additional restrictions implied by our model ((2.3), (2.2)). In this subsection, we propose an improved estimator (the QMLE-md) by including *some* of these additional restrictions via the minimum distance method. Recall that $\boldsymbol{\lambda}_{k,j}^{\intercal}$ denote the jth row of the kth row block of Λ, so we can also use $\{\hat{\boldsymbol{\lambda}}_{k,j}, \hat{\sigma}_{k,j}, \hat{M} : k = 1, \ldots, 6, j = 1, \ldots, N\}$ to denote the QMLE estimator. Suppose that we take a *finite* number c_1 of elements of the QMLE to form a column vector $\hat{\boldsymbol{h}}$. Note that the QMLE $\hat{\boldsymbol{h}}$ is estimating $h(\boldsymbol{\theta}_m)$, where $\boldsymbol{\theta}_m \subset \boldsymbol{\theta}$ is of finite-dimension c_2 ($c_2 < c_1$) and $h(\cdot) : \mathbb{R}^{c_2} \to \mathbb{R}^{c_1}$.

Example 3.1. *As an illustration, one could take*

$$\begin{aligned}
\underbrace{\hat{\boldsymbol{h}}}_{59\times 1} &= (\hat{\boldsymbol{\lambda}}_{1,2}^{\intercal}, \hat{\boldsymbol{\lambda}}_{4,2}^{\intercal}, \hat{\boldsymbol{\lambda}}_{3,5}^{\intercal}, \hat{\boldsymbol{\lambda}}_{6,5}^{\intercal}, \hat{M}_{1,1}, \hat{M}_{2,1}, \hat{\sigma}_{1,5}^2)^{\intercal}, \\
\underbrace{\boldsymbol{\theta}_m}_{10\times 1} &= (z_{2,0}^A, z_{2,1}^A, z_{2,2}^A, z_{2,3}^A, z_{5,0}^U, z_{5,1}^U, z_{5,2}^U, z_{5,3}^U, \phi, \sigma_{1,5}^2)^{\intercal}, \\
\underbrace{h(\boldsymbol{\theta}_m)}_{59\times 1} &= (\boldsymbol{\lambda}_{1,2}^{\intercal}, \boldsymbol{\lambda}_{4,2}^{\intercal}, \boldsymbol{\lambda}_{3,5}^{\intercal}, \boldsymbol{\lambda}_{6,5}^{\intercal}, M_{1,1}, M_{2,1}, \sigma_{1,5}^2)^{\intercal}
\end{aligned}$$

The expression of the 59×10 *derivative matrix* $\partial h(\boldsymbol{\theta}_m)/\partial \boldsymbol{\theta}_m$ *is given in SM B.3.*

Let W denote a $c_1 \times c_1$ symmetric, positive definite weighting matrix and define the minimum distance estimator

$$\check{\boldsymbol{\theta}}_m := \arg \min_{\boldsymbol{b} \in \mathbb{R}^{c_2}} \left[\hat{\boldsymbol{h}} - h(\boldsymbol{b})\right]^{\intercal} W \left[\hat{\boldsymbol{h}} - h(\boldsymbol{b})\right]. \tag{3.8}$$

3.4 QMLE-res

In this subsection, we propose to estimate the two-day representation of our model via the EM algorithm. We shall incorporate *all* the restrictions implied by Λ and M defined in (2.5) and (3.5), respectively, but assume the working independence hypothesis: $\{\boldsymbol{f}_t\}_{t=1}^{T_f}$ are assumed as i.i.d. when setting up the likelihood. We call this the QMLE-res estimator.

The log-likelihoods of $\{\mathring{\boldsymbol{y}}_t\}_{t=1}^{T_f} | \{\boldsymbol{f}_t\}_{t=1}^{T_f}$ and $\{\boldsymbol{f}_t\}_{t=1}^{T_f}$ are, respectively,

$$\ell\left(\{\mathring{\boldsymbol{y}}_t\}_{t=1}^{T_f} | \{\boldsymbol{f}_t\}_{t=1}^{T_f}; \boldsymbol{\theta}\right) = -\frac{T_f(6N)}{2} \log(2\pi) - \frac{T_f}{2} \log|\Sigma_{ee}| - \frac{1}{2} \sum_{t=1}^{T} (\mathring{\boldsymbol{y}}_t - \Lambda \boldsymbol{f}_t)^{\intercal} \Sigma_{ee}^{-1} (\mathring{\boldsymbol{y}}_t - \Lambda \boldsymbol{f}_t)$$

$$\ell\left(\{\boldsymbol{f}_t\}_{t=1}^{T_f}; \boldsymbol{\theta}\right) = -\frac{T_f 14}{2} \log(2\pi) - \frac{T_f}{2} \log|M| - \frac{1}{2} \sum_{t=1}^{T_f} \boldsymbol{f}_t^{\intercal} M^{-1} \boldsymbol{f}_t.$$

The complete log-likelihood function of the two-day representation of our model is hence (omitting constant)

$$\begin{aligned}
&\ell\left(\{\mathring{\boldsymbol{y}}_t\}_{t=1}^{T_f}, \{\boldsymbol{f}_t\}_{t=1}^{T_f}; \boldsymbol{\theta}\right) = \ell\left(\{\mathring{\boldsymbol{y}}_t\}_{t=1}^{T_f} | \{\boldsymbol{f}_t\}_{t=1}^{T_f}; \boldsymbol{\theta}\right) + \ell\left(\{\boldsymbol{f}_t\}_{t=1}^{T_f}; \boldsymbol{\theta}\right) \\
&= -\frac{T_f}{2} \log|\Sigma_{ee}| - \frac{1}{2} \sum_{t=1}^{T} (\mathring{\boldsymbol{y}}_t - \Lambda \boldsymbol{f}_t)^{\intercal} \Sigma_{ee}^{-1} (\mathring{\boldsymbol{y}}_t - \Lambda \boldsymbol{f}_t) - \frac{T_f}{2} \log|M| - \frac{1}{2} \sum_{t=1}^{T_f} \boldsymbol{f}_t^{\intercal} M^{-1} \boldsymbol{f}_t \\
&=: -\frac{1}{2} \left(\sum_{t=1}^{T_f} \vec{\ell}_{1,t} + \sum_{t=1}^{T_f} \vec{\ell}_{2,t} \right)
\end{aligned} \tag{3.9}$$

where

$$\sum_{t=1}^{T_f} \vec{\ell}_{1,t} := T_f \log|\Sigma_{ee}| + \sum_{t=1}^{T} \mathrm{tr}\left[(\mathring{\boldsymbol{y}}_t - \Lambda \boldsymbol{f}_t)(\mathring{\boldsymbol{y}}_t - \Lambda \boldsymbol{f}_t)^\intercal \Sigma_{ee}^{-1}\right]$$
$$\sum_{t=1}^{T_f} \vec{\ell}_{2,t} := T_f \log|M| + \sum_{t=1}^{T_f} \mathrm{tr}\left[\boldsymbol{f}_t \boldsymbol{f}_t^\intercal M^{-1}\right].$$

Let $\vec{\mathbb{E}}$ denote the expectation with respect to the conditional density $p\left(\{\boldsymbol{f}_t\}_{t=1}^{T_f} | \{\mathring{\boldsymbol{y}}_t\}_{t=1}^{T_f}; \vec{\boldsymbol{\theta}}^{(i)}\right)$ at $\vec{\boldsymbol{\theta}}^{(i)}$, where $\vec{\boldsymbol{\theta}}^{(i)}$ is the estimate of $\boldsymbol{\theta}$ from the ith iteration of the EM algorithm. Taking such an expectation on both sides of (3.9), we hence have

$$\vec{\mathbb{E}}\left[\ell\left(\{\mathring{\boldsymbol{y}}_t\}_{t=1}^{T_f}, \{\boldsymbol{f}_t\}_{t=1}^{T_f}; \boldsymbol{\theta}\right)\right] = -\frac{1}{2}\left(\vec{\mathbb{E}}\sum_{t=1}^{T_f} \vec{\ell}_{1,t} + \vec{\mathbb{E}}\sum_{t=1}^{T_f} \vec{\ell}_{2,t}\right).$$

This is the E-step. We now find values of Z^c to minimise $\vec{\mathbb{E}}\sum_{t=1}^{T_f} \vec{\ell}_{1,t}$. We will illustrate the procedure using the first row of Z^A, denoted $(Z^A)_1$. Define the 4×14 selection matrices L_1 and L_4, so that $\boldsymbol{\lambda}_{1,1}^\intercal = (Z^A)_1 L_1$ and $\boldsymbol{\lambda}_{4,1}^\intercal = (Z^A)_1 L_4$. One can show that

$$\begin{aligned}\vec{\mathbb{E}}\sum_{t=1}^{T_f} \vec{\ell}_{1,t} \propto & \sum_{t=1}^{T_f}\left[-2(Z^A)_1 L_1 \vec{\mathbb{E}}[\boldsymbol{f}_t \mathring{\boldsymbol{y}}_{1,1}]/\sigma_{1,1}^2 + (Z^A)_1 L_1 \vec{\mathbb{E}}[\boldsymbol{f}_t \boldsymbol{f}_t^\intercal] L_1^\intercal (Z^A)_1^\intercal/\sigma_{1,1}^2\right] \\ & + \sum_{t=1}^{T_f}\left[-2(Z^A)_1 L_4 \vec{\mathbb{E}}[\boldsymbol{f}_t \mathring{\boldsymbol{y}}_{1,3N+1}]/\sigma_{1,1}^2 + (Z^A)_1 L_4 \vec{\mathbb{E}}[\boldsymbol{f}_t \boldsymbol{f}_t^\intercal] L_4^\intercal (Z^A)_1^\intercal/\sigma_{1,1}^2\right].\end{aligned}$$

Taking the differential with respect to $(Z^A)_1$, recognising the derivative and setting that to zero, we have

$$(\vec{Z}^A)_1^\intercal = \left[\sum_{t=1}^{T_f}\left(L_1 \vec{\mathbb{E}}[\boldsymbol{f}_t \boldsymbol{f}_t^\intercal] L_1^\intercal + L_4 \vec{\mathbb{E}}[\boldsymbol{f}_t \boldsymbol{f}_t^\intercal] L_4^\intercal\right)\right]^{-1}\left[\sum_{t=1}^{T_f}\left(L_1 \vec{\mathbb{E}}[\boldsymbol{f}_t \mathring{\boldsymbol{y}}_{1,1}] + L_4 \vec{\mathbb{E}}[\boldsymbol{f}_t \mathring{\boldsymbol{y}}_{1,3N+1}]\right)\right].$$

In a similar way, we can obtain the QMLE-res for other factor loadings.

We now find values of Σ_{ee} to minimise $\vec{\mathbb{E}}\sum_{t=1}^{T_f} \vec{\ell}_{1,t}$. We can show that

$$\vec{\mathbb{E}}\sum_{t=1}^{T_f} \vec{\ell}_{1,t} \propto T_f \log|\Sigma_{ee}| + \mathrm{tr}\left[C_e \Sigma_{ee}^{-1}\right] = T_f \sum_{k=1}^{3N} 2\log\sigma_k^2 + \sum_{k=1}^{3N} \frac{C_{e,k,k} + C_{e,3N+k,3N+k}}{\sigma_k^2}$$

where $C_e := \sum_{t=1}^{T_f}\left(\mathring{\boldsymbol{y}}_t \mathring{\boldsymbol{y}}_t^\intercal - 2\Lambda \vec{\mathbb{E}}[\boldsymbol{f}_t \mathring{\boldsymbol{y}}_t^\intercal] + \Lambda \vec{\mathbb{E}}[\boldsymbol{f}_t \boldsymbol{f}_t^\intercal] \Lambda^\intercal\right)$, and the single-index σ_k^2 is defined as $\sigma_k^2 := \sigma^2_{\lceil \frac{k}{N} \rceil, k - \lfloor \frac{k}{N} \rfloor N}$.[3] Taking the derivative with respect to σ_k^2 and setting that to zero, we have

$$\vec{\sigma}_k^2 = \frac{1}{T_f}\frac{C_{e,k,k} + C_{e,3N+k,3N+k}}{2}.$$

We provide the formulas for $\vec{\mathbb{E}}[\boldsymbol{f}_t \boldsymbol{f}_t^\intercal]$ and $\vec{\mathbb{E}}[\boldsymbol{f}_t \mathring{\boldsymbol{y}}_t^\intercal]$ in Appendix A.12.

Next, we find values of ϕ to minimise $\vec{\mathbb{E}}\sum_{t=1}^{T_f} \vec{\ell}_{2,t}$. It is difficult to derive the analytical solution for $\vec{\phi}$ so we will obtain $\vec{\phi}$ in a numerical way.

[3] Note that $k \mapsto (\lceil \frac{k}{N} \rceil, k - \lfloor \frac{k}{N} \rfloor N)$ is a bijection from $\{1, ..., 6N\}$ to $\{1, ..., 6\} \times \{1, ..., N\}$.

4 Large Sample Theories

4.1 QMLE

We now present the large sample theories of the QMLE. The idea of the proof is based on that of Bai and Li (2012), but is considerably more involved because our identification scheme is non-standard. Hence, we provide a recipe for obtaining results similar to those of Bai and Li (2012) for almost *any* identified dynamic factor model. This does have a practical importance because many dynamic factor models, like ours, are coming from different economic theories and might not conform to the five identification schemes of Bai and Li (2012).

Recall that we could use $\{\hat{\boldsymbol{\lambda}}_{k,j}, \hat{\sigma}_{k,j}, \hat{M} : k = 1, \ldots, 6, j = 1, \ldots, N\}$ to denote the QMLE. We make the following assumption.

Assumption 4.1. *(i) The factor loadings $\{\boldsymbol{\lambda}_{k,j}\}$ satisfy $\|\boldsymbol{\lambda}_{k,j}\|_2 \leq C$ for all k and j.*

(ii) Assume $C^{-1} \leq \sigma^2_{k,j} \leq C$ for all k and j. Also $\hat{\sigma}^2_{k,j}$ is restricted to a compact set $[C^{-1}, C]$ for all k and j.

(iii) $\hat{M}$ is restricted to be in a set consisting of all positive definite matrices with all the elements bounded in the interval $[C^{-1}, C]$.

(iv) Suppose that $Q := \lim_{N\to\infty} \frac{1}{N}\Lambda^\intercal \Sigma_{ee}^{-1}\Lambda$ is a positive definite matrix.

Assumption 4.1 is standard in the literature of factor models and has been taken from the assumptions of Bai and Li (2012).

Proposition 4.1. *Suppose that Assumptions 2.1, 4.1 hold. When $N, T_f \to \infty$, with the identification condition outlined in the proof of this proposition (i.e., 14^2 particular restrictions imposed on $\hat{\Lambda}$ and $\hat{M}$), and the requirement that $\hat{\Lambda}$ and Λ have the same column signs, we have*

$$\hat{\boldsymbol{\lambda}}_{k,j} - \boldsymbol{\lambda}_{k,j} = o_p(1) \tag{4.1}$$

$$\frac{1}{6N}\sum_{k=1}^{6}\sum_{j=1}^{N}(\hat{\sigma}^2_{k,j} - \sigma^2_{k,j})^2 = o_p(1) \tag{4.2}$$

$$\hat{M} - M = o_p(1). \tag{4.3}$$

for $k = 1, \ldots, 6, j = 1, \ldots, N$.

Display (4.1) and (4.3) establish consistency for the individual loading estimator $\hat{\boldsymbol{\lambda}}_{k,j}$ and $\hat{M}$, respectively, while display (4.2) establishes some average consistency for $\{\hat{\sigma}^2_{k,j}\}$.

Theorem 4.1. *Under the assumptions of Proposition 4.1, we have*

$$\|\hat{\boldsymbol{\lambda}}_{k,j} - \boldsymbol{\lambda}_{k,j}\|_2^2 = O_p(T_f^{-1}) \tag{4.4}$$

$$\frac{1}{6N}\sum_{k=1}^{6}\sum_{j=1}^{N}(\hat{\sigma}^2_{k,j} - \sigma^2_{k,j})^2 = O_p(T_f^{-1}) \tag{4.5}$$

$$\hat{M} - M = O_p(T_f^{-1/2}). \tag{4.6}$$

for $k = 1, \ldots, 6, j = 1, \ldots, N$.

Theorem 4.1 resembles Theorem 5.1 of Bai and Li (2012) and establishes the rate of convergence for the QMLE. The only difference is while Bai and Li (2012) only established an average rate of convergence for $\{\hat{\boldsymbol{\lambda}}_{k,j}\}$, we managed to establish a rate of convergence for the individual loading estimator $\hat{\boldsymbol{\lambda}}_{k,j}$.

Theorem 4.2. *Under the assumptions of Proposition 4.1, we have, for $k = 1, \ldots, 6, j = 1, \ldots, N$,*

(i)

$$\hat{\sigma}_{k,j}^2 - \sigma_{k,j}^2 = \frac{1}{T_f} \sum_{t=1}^{T_f} \left(e_{(k-1)N+j,t}^2 - \sigma_{k,j}^2 \right) + o_p(T_f^{-1/2}),$$

where $e_{(k-1)N+j,t}$ is the $[(k-1)N + j]$th element of $\boldsymbol{e}_t$.

(ii) As $N, T_f \to \infty$,

$$\sqrt{T_f}(\hat{\sigma}_{k,j}^2 - \sigma_{k,j}^2) \xrightarrow{d} N(0, 2\sigma_{k,j}^4).$$

Theorem 4.2(i) gives the asymptotic representation of $\hat{\sigma}_{k,j}^2$. Theorem 4.2(ii) is the same as Theorem 5.4 of Bai and Li (2012) and establishes the asymptotic distribution of $\hat{\sigma}_{k,j}^2$.

Theorem 4.3. *Suppose that the assumptions of Proposition 4.1 hold.*

(i) For $k = 1, \ldots, 6,\ j = 1, \ldots, N$, we have

$$\sqrt{T_f}(\hat{\boldsymbol{\lambda}}_{k,j} - \boldsymbol{\lambda}_{k,j}) = (\boldsymbol{\lambda}_{k,j}^{\intercal} \otimes I_{14}) \Gamma \frac{1}{\sqrt{T_f}} \sum_{t=1}^{T_f} \left(\boldsymbol{e}_t^{\dagger} \otimes \boldsymbol{f}_t \right) + M^{-1} \frac{1}{\sqrt{T_f}} \sum_{t=1}^{T_f} \boldsymbol{f}_t e_{(k-1)N+j,t} + o_p(1),$$

where $\boldsymbol{e}_t^{\dagger}$ is a 24×1 vector consisting of $e_{(p-1)N+q,t}$ for $p = 1, \ldots, 6$ and $q = 1, \ldots, 4$, and Γ is a 196×336 matrix, whose elements are known (but complicated) linear functions of elements of (inverted) submatrices of Λ and M, satisfying

$$\operatorname{vec} A = \Gamma \times \frac{1}{T_f} \sum_{t=1}^{T_f} \left(\boldsymbol{e}_t^{\dagger} \otimes \boldsymbol{f}_t \right) + o_p(T_f^{-1/2}), \qquad A := (\hat{\Lambda} - \Lambda)^{\intercal} \hat{\Sigma}_{ee}^{-1} \hat{\Lambda} (\hat{\Lambda}^{\intercal} \hat{\Sigma}_{ee}^{-1} \hat{\Lambda})^{-1}.$$

(ii) As $N, T_f \to \infty$, for $k = 1, \ldots, 6,\ j = 5, \ldots, N$, we have

$$\sqrt{T_f}(\hat{\boldsymbol{\lambda}}_{k,j} - \boldsymbol{\lambda}_{k,j}) \xrightarrow{d} N \left(\mathbf{0}, (\boldsymbol{\lambda}_{k,j}^{\intercal} \otimes I_{14}) \Gamma (\Sigma_{ee}^{\dagger} \otimes M) \Gamma^{\intercal} (\boldsymbol{\lambda}_{k,j} \otimes I_{14}) + M^{-1} \sigma_{k,j}^2 \right),$$

and for $k = 1, \ldots, 6,\ j = 1, \ldots, 4$, we have

$$\sqrt{T_f}(\hat{\boldsymbol{\lambda}}_{k,j} - \boldsymbol{\lambda}_{k,j}) \xrightarrow{d} N \left(\mathbf{0}, (\boldsymbol{\lambda}_{k,j}^{\intercal} \otimes I_{14}) \Gamma (\Sigma_{ee}^{\dagger} \otimes M) \Gamma^{\intercal} (\boldsymbol{\lambda}_{k,j} \otimes I_{14}) + M^{-1} \sigma_{k,j}^2 + cov_{k,j} + cov_{k,j}^{\intercal} \right),$$

where $\Sigma_{ee}^{\dagger}$ is a 24×24 diagonal matrix whose $[4(p-1) + q]$th diagonal element is $\sigma_{p,q}^2$ for $p = 1, \ldots, 6$ and $q = 1, \ldots, 4$. The 14×14 matrix $cov_{k,j}$ is defined as

$$cov_{k,j} := (\boldsymbol{\lambda}_{k,j}^{\intercal} \otimes I_{14}) \Gamma \left[\boldsymbol{\iota}_{k,j} \otimes I_{14} \right] \sigma_{k,j}^2,$$

where $\boldsymbol{\iota}_{k,j}$ is a 24×1 zero vector with its $[4(k-1) + j]$th element replaced by one.

Theorem 4.3 presents the asymptotic representation and distribution of the QMLE of the factor loadings. The idea of the proof is inspired by that for the fourth identification scheme (i.e., IC4) in Theorem 5.2 of Bai and Li (2012). Note that the asymptotic variance of $\hat{\boldsymbol{\lambda}}_{k,j}$ depends on Γ and $\Sigma_{ee}^{\dagger}$. The matrix $\Sigma_{ee}^{\dagger}$ contains the idiosyncratic variances of the first four assets in each continent. When computing Γ, we often need to invert submatrices of the factor loadings of the first four assets in each continent (see (A.47) for example). Thus, ordering assets, with smaller idiosyncratic variances and less multicollinearity of the factor loadings, as the first four assets in each continent results in a $\hat{\boldsymbol{\lambda}}_{k,j}$ with smaller asymptotic variances.

Theorem 4.4. *Suppose that the assumptions of Proposition 4.1 hold.*

(i)

$$\sqrt{T_f}\,\mathrm{vech}(\hat{M}-M) = -2D_{14}^{+}(I_{14}\otimes M)\Gamma\frac{1}{\sqrt{T_f}}\sum_{t=1}^{T_f}\left(\boldsymbol{e}_t^{\dagger}\otimes\boldsymbol{f}_t\right)+o_p(1),$$

where D_{14}^{+} is defined in Section 1.1.

(ii) As $N, T_f\to\infty$,

$$\sqrt{T_f}\,\mathrm{vech}(\hat{M}-M)\xrightarrow{d}N\left(0,\mathcal{M}\right)$$

where $\mathcal{M}$ is 105×105 and defined as

$$\mathcal{M}:=4D_{14}^{+}(I_{14}\otimes M)\Gamma(\Sigma_{ee}^{\dagger}\otimes M)\Gamma^{\intercal}(I_{14}\otimes M)D_{14}^{+\intercal}.$$

Theorem 4.4 presents the asymptotic representation and distribution of the QMLE of M; it is similar to Theorem 5.3 of Bai and Li (2012).

We could estimate $\boldsymbol{f}_t$ by the generalized least squares (GLS), as Bai and Li (2012) have done in their Theorem 6.1:

$$\hat{\boldsymbol{f}}_t=\left(\hat{\Lambda}^{\intercal}\hat{\Sigma}_{ee}^{-1}\hat{\Lambda}\right)^{-1}\hat{\Lambda}^{\intercal}\hat{\Sigma}_{ee}^{-1}\mathring{\boldsymbol{y}}_t.$$

Theorem 4.5. *Suppose that the assumptions of Proposition 4.1 hold and $\sqrt{N}/T_f\to 0$, $N/T_f\to\Delta\in[0,\infty)$. Then we have*

(i)

$$\sqrt{N}(\hat{\boldsymbol{f}}_t-\boldsymbol{f}_t)=-\sqrt{\Delta}(\boldsymbol{f}_t^{\intercal}\otimes I_{14})K_{14,14}\Gamma\frac{1}{\sqrt{T_f}}\sum_{t=1}^{T_f}\left(\boldsymbol{e}_t^{\dagger}\otimes\boldsymbol{f}_t\right)+Q^{-1}\frac{1}{\sqrt{N}}\Lambda^{\intercal}\Sigma_{ee}^{-1}\boldsymbol{e}_t+o_p(1),$$

where $K_{14,14}$ is the commutation matrix.

(ii)

$$\sqrt{N}(\hat{\boldsymbol{f}}_t-\boldsymbol{f}_t)|\boldsymbol{f}_t\xrightarrow{d}N\left(\boldsymbol{0},\Delta(\boldsymbol{f}_t^{\intercal}\otimes I_{14})K_{14,14}\Gamma(\Sigma_{ee}^{\dagger}\otimes M)\Gamma^{\intercal}K_{14,14}(\boldsymbol{f}_t\otimes I_{14})+Q^{-1}\right),$$

where Q is defined in Assumption 4.1.

Theorem 4.5 gives the asymptotic representation and conditional distribution of the GLS $\hat{\boldsymbol{f}}_t$. Unlike Theorem 6.1 of Bai and Li (2012), since we treat $\boldsymbol{f}_t$ as random, the asymptotic normal distribution in Theorem 4.5(ii) is for $\sqrt{N}(\hat{\boldsymbol{f}}_t-\boldsymbol{f}_t)$ conditioning on $\boldsymbol{f}_t$.

4.2 QMLE-md

We then present the large sample theories of the QMLE-md. Recall that $\hat{\boldsymbol{h}}$ is a vector of finite length of the QMLE estimator $\{\hat{\boldsymbol{\lambda}}_{k,j}, \hat{\sigma}_{k,j}, \hat{M} : k = 1, \ldots, 6, j = 1, \ldots, N\}$. Relying on the asymptotic representations of the QMLE (Theorems 4.2, 4.3, 4.4), one could easily establish the asymptotic distribution of $\hat{\boldsymbol{h}}$, say,

$$\sqrt{T_f}(\hat{\boldsymbol{h}} - h(\boldsymbol{\theta}_m)) \xrightarrow{d} N(\mathbf{0}, \mathcal{H}).$$

Since the choice of $\hat{\boldsymbol{h}}$ varies, we omit the formula for $\mathcal{H}$.

Theorem 4.6. *Suppose that the assumptions of Proposition 4.1 hold. Then we have*

$$\sqrt{T_f}(\check{\boldsymbol{\theta}}_m - \boldsymbol{\theta}_m) \xrightarrow{d} N(\mathbf{0}, \mathcal{O}),$$

where

$$\mathcal{O} := \left[\frac{\partial h(\boldsymbol{\theta}_m)}{\partial \boldsymbol{\theta}_m^{\intercal}} W \frac{\partial h(\boldsymbol{\theta}_m)}{\partial \boldsymbol{\theta}_m}\right]^{-1} \frac{\partial h(\boldsymbol{\theta}_m)}{\partial \boldsymbol{\theta}_m^{\intercal}} W \mathcal{H} W \frac{\partial h(\boldsymbol{\theta}_m)}{\partial \boldsymbol{\theta}_m} \left[\frac{\partial h(\boldsymbol{\theta}_m)}{\partial \boldsymbol{\theta}_m^{\intercal}} W \frac{\partial h(\boldsymbol{\theta}_m)}{\partial \boldsymbol{\theta}_m}\right]^{-1}.$$

In the preceding theorem, choosing $W = \mathcal{H}^{-1}$ gives the most efficient minimum distance estimator. In that case, $\mathcal{O}$ is reduced to $\left[\frac{\partial h(\boldsymbol{\theta}_m)}{\partial \boldsymbol{\theta}_m^{\intercal}} \mathcal{H}^{-1} \frac{\partial h(\boldsymbol{\theta}_m)}{\partial \boldsymbol{\theta}_m}\right]^{-1}$.

5 Inference Procedures for the MLE-one day and the QMLE-res

In classical factor analysis (i.e., fixed N large T_f), the asymptotic variances of the MLE-one day and the QMLE-res defined in Section 3 could be approximated using the numerical Hessian method as researchers have shown that the MLE of a standard factor model is asymptotically normal but has very complicated expressions for the asymptotic covariance matrices (Anderson (2003, p.583)). The large sample theories of the MLE-one day and the QMLE-res in the large N large T_f case remain as a formidable, if not impossible, task to be completed in the future research. In this section, we provide a heuristic procedure to *approximate* the standard errors of the MLE-one day and the QMLE-res in the large N large T_f case.

5.1 MLE-one day

5.1.1 For $\tilde{\phi}$

By relying on the first-order condition of the MLE with respect to ϕ, we propose a parametric bootstrap to approximate the sampling distribution of $\tilde{\phi}$ defined in (3.4). It can be shown that the first-order condition of the MLE with respect to ϕ is

$$\frac{\partial \frac{1}{NT_f} \ell\left(\{\mathring{\boldsymbol{y}}_t\}_{t=1}^{T_f}; \boldsymbol{\theta}\right)}{\partial \phi} = \frac{1}{2N} \operatorname{tr}\left[\Lambda^{\intercal}(\Sigma_{yy}^{-1} S_{yy} \Sigma_{yy}^{-1} - \Sigma_{yy}^{-1})\Lambda K(\phi)\right]$$

where $K(\phi)$ is the 14×14 derivative matrix $\partial M/\partial \phi$. Given the MLE-one day estimator $\{\tilde{Z}^c, \tilde{\Sigma}_c, \tilde{\phi} : c = A, E, U\}$, in each bootstrap replication, simulate $\{y_t\}$ and hence calculate S_{yy}. Then use the numeric means to find $\tilde{\phi}_b$ which solves

$$\frac{1}{2N} \operatorname{tr} \left[\tilde{\Lambda}_b^\intercal (\tilde{\Sigma}_{yy,b}^{-1} S_{yy} \tilde{\Sigma}_{yy,b}^{-1} - \tilde{\Sigma}_{yy,b}^{-1}) \tilde{\Lambda}_b K(\tilde{\phi}_b)\right] = 0.$$

where $\tilde{\Lambda}_b$ and $\tilde{\Sigma}_{yy,b}$ are estimates of Λ and Σ_{yy} computed using $\{\tilde{Z}^c, \tilde{\Sigma}_c : c = A, E, U\}$ and $\tilde{\phi}_b$. If we have B parametric bootstrap replications, then the standard deviation of $\{\tilde{\phi}_b\}_{b=1}^B$ could be used as the approximation for the standard error of $\tilde{\phi}$.

5.1.2 For $\tilde{Z}^c$

We now approximate the standard errors of $\tilde{Z}^c$ for $c = A, E, U$ defined in (3.3). The idea is to use *all* the restrictions implied by Λ defined in (2.5) to reduce the standard error of the QMLE. Since the reduced standard error has incorporated all the information contained in Λ, it should approximate that of $\tilde{Z}^c$. The machinery which we shall employ is Bayes theorem.

In the proof of Theorem 4.3, we show that (see (A.53))

$$\begin{aligned}\sqrt{T_f}(\hat{\boldsymbol{\lambda}}_{k,j} - \boldsymbol{\lambda}_{k,j}) &= (\boldsymbol{\lambda}_{k,j}^\intercal \otimes I_{14})\Gamma \frac{1}{\sqrt{T_f}} \sum_{t=1}^{T_f} \left(\boldsymbol{e}_t^\dagger \otimes \boldsymbol{f}_t\right) + M^{-1} \frac{1}{\sqrt{T_f}} \sum_{t=1}^{T_f} \boldsymbol{f}_t e_{(k-1)N+j,t} + o_p(1) \\ &=: \boldsymbol{a}^* + M^{-1} \frac{1}{\sqrt{T_f}} \sum_{t=1}^{T_f} \boldsymbol{f}_t e_{(k-1)N+j,t} + o_p(1).\end{aligned}$$

Repeating the arguments in (A.54), we have for $k = 1, 2, 3$,

$$\sqrt{T_f} \begin{pmatrix} \hat{\boldsymbol{\lambda}}_{k,j} - \boldsymbol{\lambda}_{k,j} \\ \hat{\boldsymbol{\lambda}}_{k+3,j} - \boldsymbol{\lambda}_{k+3,j} \end{pmatrix} \Bigg| \boldsymbol{a}^* \xrightarrow{d} N \left(\begin{bmatrix} \boldsymbol{a}^* \\ \boldsymbol{a}^* \end{bmatrix}, \begin{bmatrix} \sigma_{k,j}^2 M^{-1} & 0 \\ 0 & \sigma_{k,j}^2 M^{-1} \end{bmatrix} \right) \tag{5.1}$$

$$\boldsymbol{a}^* \xrightarrow{d} N(0, W_a) \tag{5.2}$$

where

$$W_a := (\boldsymbol{\lambda}_{k,j}^\intercal \otimes I_{14})\Gamma(\Sigma_{ee}^\dagger \otimes M)\Gamma^\intercal(\boldsymbol{\lambda}_{k,j} \otimes I_{14}).$$

Define

$$\underbrace{G_k}_{24\times 28} := \begin{pmatrix} S_{k,z} & \mathbf{0} \\ \mathbf{0} & S_{k+3,z} \\ S_{k,nz} & -S_{k+3,nz} \end{pmatrix},$$

where $\{S_{k,nz}\}_{k=1}^6$ are 4×14 with $S_{k,nz}$ being the $7-k, 8-k, 9-k, 15-k$th rows of I_{14} for $k = 1, \ldots, 6$, and $\{S_{k,z}\}_{k=1}^6$ are 10×14 with $S_{k,z}$ being the submatrix of I_{14} after deleting its $7-k, 8-k, 9-k, 15-k$th rows for $k = 1, \ldots, 6$. In other words, $S_{k,z}$ denotes the 10×14 selection matrix which extracts out the 10 zero elements of $\boldsymbol{\lambda}_{k,j}$, while $S_{k,nz}$ denotes the 4×14 selection matrix that extracts out the 4 non-zero elements of $\boldsymbol{\lambda}_{k,j}$. Note that $G_k(\boldsymbol{\lambda}_{k,j}^\intercal, \boldsymbol{\lambda}_{k+3,j}^\intercal)^\intercal = \mathbf{0}$. Let

$$\hat{\boldsymbol{\omega}}_{k,j} := \sqrt{T_f} \begin{pmatrix} \hat{\boldsymbol{\lambda}}_{k,j} \\ \hat{\boldsymbol{\lambda}}_{k+3,j} \end{pmatrix}.$$

We hence have

$$G_k\hat{\boldsymbol{\omega}}_{k,j}\big|\,\boldsymbol{a}^* = G_k\sqrt{T_f}\left(\begin{array}{c}\hat{\boldsymbol{\lambda}}_{k,j}-\boldsymbol{\lambda}_{k,j}\\ \hat{\boldsymbol{\lambda}}_{k+3,j}-\boldsymbol{\lambda}_{k+3,j}\end{array}\right)\Bigg|\,\boldsymbol{a}^* \xrightarrow{d} N\left(G_k\left[\begin{array}{c}\boldsymbol{a}^*\\ \boldsymbol{a}^*\end{array}\right], \sigma_{k,j}^2 G_k(I_2\otimes M^{-1})G_k^{\intercal}\right).$$

Recall that $p(\cdot)$ denotes the asymptotic density function. Then we have

$$\begin{aligned}
&p\left(\boldsymbol{a}^*\right) \propto \exp\left(-\frac{1}{2}\boldsymbol{a}^{*\intercal}W_a^{-1}\boldsymbol{a}^*\right)\\
&p(G_k\hat{\boldsymbol{\omega}}_{k,j}|\boldsymbol{a}^*) = (2\pi)^{-\frac{24}{2}}\left|\sigma_{k,j}^2 G_k(I_2\otimes M^{-1})G_k^{\intercal}\right|^{-1/2}\times\\
&\exp\left\{-\frac{1}{2}\left(G_k\hat{\boldsymbol{\omega}}_{k,j}-G_k\left[\begin{array}{c}\boldsymbol{a}^*\\ \boldsymbol{a}^*\end{array}\right]\right)^{\intercal}\sigma_{k,j}^{-2}\left[G_k(I_2\otimes M^{-1})G_k^{\intercal}\right]^{-1}\left(G_k\hat{\boldsymbol{\omega}}_{k,j}-G_k\left[\begin{array}{c}\boldsymbol{a}^*\\ \boldsymbol{a}^*\end{array}\right]\right)\right\}\\
&\propto\exp\left\{-\frac{1}{2}\left(G_k\hat{\boldsymbol{\omega}}_{k,j}-G_k\left[\begin{array}{c}\boldsymbol{a}^*\\ \boldsymbol{a}^*\end{array}\right]\right)^{\intercal}\sigma_{k,j}^{-2}\left[G_k(I_2\otimes M^{-1})G_k^{\intercal}\right]^{-1}\left(G_k\hat{\boldsymbol{\omega}}_{k,j}-G_k\left[\begin{array}{c}\boldsymbol{a}^*\\ \boldsymbol{a}^*\end{array}\right]\right)\right\}.
\end{aligned}$$

Note that $\{G_k\hat{\boldsymbol{\omega}}_{k,j}\}_{k=1,j=1}^{k=3,j=N}$ are conditionally (given $\boldsymbol{a}^*$) asymptotically independent across k and j because of the asymptotic normality and uncorrelatedness by Assumption 2.1. Then we have

$$\begin{aligned}
&p\left(\{G_k\hat{\boldsymbol{\omega}}_{k,j}\}_{k=1,j=1}^{k=3,j=N}|\boldsymbol{a}^*\right) = \prod_{k=1}^{3}\prod_{j=1}^{N}p\left(G_k\hat{\boldsymbol{\omega}}_{k,j}|\boldsymbol{a}^*\right)\propto\\
&\exp\left\{-\frac{1}{2}\sum_{k,j}\left(G_k\hat{\boldsymbol{\omega}}_{k,j}-G_k\left[\begin{array}{c}I_{14}\\ I_{14}\end{array}\right]\boldsymbol{a}^*\right)^{\intercal}\sigma_{k,j}^{-2}\left[G_k(I_2\otimes M^{-1})G_k^{\intercal}\right]^{-1}\left(G_k\hat{\boldsymbol{\omega}}_{k,j}-G_k\left[\begin{array}{c}I_{14}\\ I_{14}\end{array}\right]\boldsymbol{a}^*\right)\right\}.
\end{aligned}$$

Thus the asymptotic posterior distribution of $\boldsymbol{a}^*$ given $\{G_k\hat{\boldsymbol{\omega}}_{k,j}\}_{k=1,j=1}^{k=3,j=N}$ is

$$\begin{aligned}
&p\left(\boldsymbol{a}^*|\{G_k\hat{\boldsymbol{\omega}}_{k,j}\}_{k=1,j=1}^{k=3,j=N}\right)\propto p\left(\{G_k\hat{\boldsymbol{\omega}}_{k,j}\}_{k=1,j=1}^{k=3,j=N}|\boldsymbol{a}^*\right)p\left(\boldsymbol{a}^*\right)\\
&\propto\exp\left(-\frac{1}{2}\boldsymbol{a}^{*\intercal}W_a^{-1}\boldsymbol{a}^*\right)\times\\
&\exp\left\{-\frac{1}{2}\sum_{k,j}\left(G_k\hat{\boldsymbol{\omega}}_{k,j}-G_k\left[\begin{array}{c}I_{14}\\ I_{14}\end{array}\right]\boldsymbol{a}^*\right)^{\intercal}\sigma_{k,j}^{-2}\left[G_k(I_2\otimes M^{-1})G_k^{\intercal}\right]^{-1}\left(G_k\hat{\boldsymbol{\omega}}_{k,j}-G_k\left[\begin{array}{c}I_{14}\\ I_{14}\end{array}\right]\boldsymbol{a}^*\right)\right\}\\
&\propto\exp\left(\Psi_1\boldsymbol{a}^*-\frac{1}{2}\boldsymbol{a}^{*\intercal}\Psi_2\boldsymbol{a}^*\right) \qquad (5.3)
\end{aligned}$$

where

$$\Psi_1 := \sum_{k=1}^{3}\sum_{j=1}^{N}(G_k\hat{\boldsymbol{\omega}}_{k,j})^{\intercal}\sigma_{k,j}^{-2}\left[G_k(I_2\otimes M^{-1})G_k^{\intercal}\right]^{-1}G_k\left[\begin{array}{c}I_{14}\\ I_{14}\end{array}\right]$$

$$\Psi_2 := \left\{W_a^{-1}+\sum_{k=1}^{3}\sum_{j=1}^{N}\left[\begin{array}{c}I_{14}\\ I_{14}\end{array}\right]^{\intercal}G_k^{\intercal}\sigma_{k,j}^{-2}\left[G_k(I_2\otimes M^{-1})G_k^{\intercal}\right]^{-1}G_k\left[\begin{array}{c}I_{14}\\ I_{14}\end{array}\right]\right\}.$$

Thus (5.3) implies that $\boldsymbol{a}^*|\{G_k\hat{\boldsymbol{\omega}}_{k,j}\}_{k=1,j=1}^{k=3,j=N}$ is asymptotically distributed as $N\left(\Psi_1\Psi_2^{-1},\Psi_2^{-1}\right)$. In particular, $\boldsymbol{a}^*|\{G_k\hat{\boldsymbol{\omega}}_{k,j}=\mathbf{0}\}_{k=1,j=1}^{k=3,j=N}$ is asymptotically distributed as $N\left(\mathbf{0},\Psi_2^{-1}\right)$. We know that for $\ell > 4$, say, the asymptotic variance of $T_f^{1/2}(\hat{\boldsymbol{\lambda}}_{k,\ell} - \boldsymbol{\lambda}_{k,\ell})$ is the asymptotic variance of $\boldsymbol{a}^*$ plus $\sigma_{k,\ell}^2 M^{-1}$. Conditioning on $\{G_k \hat{\boldsymbol{\omega}}_{k,j} = \mathbf{0}\}_{k=1,j=1}^{k=3,j=N}$, the asymptotic variance of $\boldsymbol{a}^*$ is reduced to Ψ_2^{-1}. Thus the square roots of the diagonal elements of $\Psi_2^{-1} + \sigma_{k,\ell}^2 M^{-1}$, whose positions correspond to the non-zero elements of $\boldsymbol{\lambda}_{k,\ell}$, could be used as approximations of the standard errors of the MLE-one day estimators of those loadings.

5.1.3 For $\tilde{\Sigma}_c$

We shall approximate the standard error of $\tilde{\sigma}_{c,j}^2$ by $\tilde{\sigma}_{c,j}^2 \sqrt{2/T_f}$ (see Theorem 4.2). This is because the QMLE of Σ_c has already incorporated diagonality of Σ_c for $c = A, E, U$, so the standard error of the QMLE of $\sigma_{c,j}^2$ should approximate that of the corresponding MLE-one day estimator reasonably well.

5.2 QMLE-res

The standard errors of the QMLE-res are calculated in the same way as those of the MLE-one day. The idea is that since the QMLE-res has incorporated all the restrictions implied by Λ and M, its standard errors should be close to those of the MLE-one day.

6 Monte Carlo Simulations

In this section, we shall conduct Monte Carlo simulations to evaluate the performances of our proposed estimators. We specify the following values for the parameters: $N = 50, 200$; $T = 750$ (around one year's trading data), 1500, 2250; $\phi = 0.3$. For $c = A, E, U$, $\Sigma_{c,ii}$ are drawn from uniform$[0.2, 2]$ for $i = 1, \ldots, N$, and

$$z_{i,j}^c = 0.6a_{c,i,j} + 0.6d_{c,j} - 0.2$$

where $\{a_{c,i,j}\}_{i=1,j=0}^{i=N,j=3}$ and $\{d_{c,j}\}_{j=0}^{3}$ are all drawn from uniform$[0, 1]$. For each continent, we put the assets with the smallest four idiosyncratic variances as the first four assets. After the logarithmic 24-hr returns are generated, the econometrician only observes those logarithmic 24-hr returns whose ts correspond to the closing times of their belonging continents. The econometrician is aware of the structure of the true model (2.3), but does not know the values of those parameters. In particular, he is aware of diagonality of Σ_c.

We estimate the model using the MLE-one day, the QMLE-res and the QMLE-md. To initialize the EM algorithm for the MLE-one day and the QMLE-res, the starting values of the parameters are estimated according to the procedure mentioned in Section 3.1. We now briefly explain how to select $\hat{\boldsymbol{h}}$ for the QMLE-md. Take the factor loading of Asia's jth asset (i.e., $z_{j,0}^A, z_{j,1}^A, z_{j,2}^A, z_{j,3}^A$) as an example: Select $\hat{\boldsymbol{h}} = (\hat{\boldsymbol{\lambda}}_{1,j}^\intercal, \hat{\boldsymbol{\lambda}}_{4,j}^\intercal, \hat{\boldsymbol{\lambda}}_{2,1}^\intercal, \hat{\boldsymbol{\lambda}}_{5,1}^\intercal, \hat{\boldsymbol{\lambda}}_{2,5}^\intercal, \hat{\boldsymbol{\lambda}}_{5,5}^\intercal, \hat{\boldsymbol{\lambda}}_{3,1}^\intercal, \hat{\boldsymbol{\lambda}}_{6,1}^\intercal, \hat{\boldsymbol{\lambda}}_{3,5}^\intercal, \hat{\boldsymbol{\lambda}}_{6,5}^\intercal)^\intercal$. For ϕ: Select

$$\hat{\boldsymbol{h}} = (\hat{\boldsymbol{\lambda}}_{1,1}^\intercal, \hat{\boldsymbol{\lambda}}_{4,1}^\intercal, \hat{\boldsymbol{\lambda}}_{2,1}^\intercal, \hat{\boldsymbol{\lambda}}_{5,1}^\intercal, \hat{\boldsymbol{\lambda}}_{3,1}^\intercal, \hat{\boldsymbol{\lambda}}_{6,1}^\intercal, (\operatorname{vech} M)^\intercal)^\intercal.$$

For $\sigma_{k,j}^2$, set its QMLE-md to the QMLE.

The number of the Monte Carlo samples is chosen to be 200. From these 200 Monte Carlo samples, we calculate the following three quantities for evaluation:

(i) the root mean square errors (RMSE),

(ii) the average of the standard errors (Ave.se) across the Monte Carlo samples.

(iii) the coverage probability (Cove) of the confidence interval formed by the point estimate $\pm$ 1.96$\times$the standard error. The standard errors differ across the Monte Carlo samples.

For a particular j and c, it is impossible to present a evaluation criterion of $z^c_{i,j}$ for all i, so we only report the average value for the vector $\boldsymbol{z}^c_j$. Likewise, we report the average value for the three diagonals of $\{\Sigma_c : c = A, E, U\}$. Tables 1 and 2 report these results. To save space, we only present the results for $T = 750, 2250$ (the results for $T = 1500$ are available upon request). We see that the MLE-one day and QMLE-res estimators are very similar in terms of the three evaluation criteria. In terms of RMSE, the MLE-one day and the QMLE-res are better than the QMLE-md, but the gap quickly narrows when N or T increases. In terms of Ave.se, the QMLE-md has slightly larger standard errors than those of the MLE-one day or the QMLE-res. This is probably because the QMLE, the first-step estimator in the QMLE-md, has large standard errors. When N and T increase, we obtain smaller RMSE, smaller Ave.se, and better coverage in general for all estimators.

7 Empirical Work

In this section, we present two empirical applications of our model. Section 7.1 is about modelling equity portfolio returns from Japan, Europe and the US. That is, one market per continent. Section 7.2 studies MSCI equity indices of the developed and emerging markets (41 markets across three continents). We first list the trading hours of the world's top ten stock exchanges in terms of market capitalisation in Table 3. This shows the overlaps and lack of overlaps.

7.1 An Empirical Study of Three Markets

We now apply our model to equity portfolios of three continents/markets: Japan, Europe and the US. Take Japan as an example. First, we consider six equity portfolios constructed by intersections of 2 size groups (small (S) and big (B)) and 3 book-to-market equity ratio (B/M) groups (growth (G), neutral (N) and value (V)), in the spirit of Fama and French (1993); we denote the six portfolios SG, SN, SV, BG, BN and BV. Second, in a similar manner we consider six equity portfolios constructed by intersections of 2 size groups (small (S) and big (B)) and 3 momentum groups (loser (L), neutral (N) and winner (W)); we denote these six portfolios SL, SN, SW, BL, BN and BW. We downloaded the daily value-weighted portfolio returns (in percentage points) from Kenneth R. French's website.[4] Note that these returns are not logarithmic returns, so strictly speaking our model does not apply. Moreover we demeaned and standardised the daily value-weighted portfolio returns so that the returns have sample variances of one. In SM B.4, we show that our model could still be applied by making some innocuous approximations. Since we do not have so many assets in this application, we shall use the

[4] https://mba.tuck.dartmouth.edu/pages/faculty/ken.french/

	N=50, T=750								
	MLE-one day			QMLE-res			QMLE-md		
	RMSE	Ave.se	Cove	RMSE	Ave.se	Cove	RMSE	Ave.se	Cove
$\boldsymbol{z}_0^A$	0.0811	0.1072	0.9868	0.0864	0.1062	0.9802	0.2665	0.1532	0.7663
$\boldsymbol{z}_1^A$	0.0869	0.1078	0.9845	0.0907	0.1073	0.9799	0.3885	0.1641	0.7805
$\boldsymbol{z}_2^A$	0.0789	0.1016	0.9869	0.0871	0.1020	0.9757	0.3600	0.1396	0.7708
$\boldsymbol{z}_3^A$	0.0985	0.1010	0.9562	0.1038	0.1011	0.9466	0.2775	0.1204	0.6615
$\boldsymbol{z}_0^E$	0.0909	0.1052	0.9746	0.0938	0.1052	0.9698	0.2889	0.1440	0.8072
$\boldsymbol{z}_1^E$	0.0946	0.1076	0.9704	0.0998	0.1063	0.9596	0.2999	0.1640	0.7806
$\boldsymbol{z}_2^E$	0.0855	0.1049	0.9806	0.0879	0.1045	0.9768	0.4058	0.1473	0.7735
$\boldsymbol{z}_3^E$	0.0833	0.0975	0.9751	0.0848	0.0976	0.9710	0.1820	0.0942	0.7001
$\boldsymbol{z}_0^U$	0.0881	0.1044	0.9761	0.0985	0.1044	0.9545	0.3451	0.1854	0.7926
$\boldsymbol{z}_1^U$	0.0981	0.1067	0.9643	0.1040	0.1069	0.9508	0.5039	0.2274	0.8087
$\boldsymbol{z}_2^U$	0.0937	0.1067	0.9682	0.1007	0.1049	0.9549	0.3402	0.1980	0.7924
$\boldsymbol{z}_3^U$	0.0812	0.0960	0.9759	0.0831	0.0959	0.9727	0.2253	0.1368	0.7623
Σ_c	0.1143	0.1062	0.9114	0.1149	0.1060	0.9070	0.2417	0.1294	0.6342
ϕ	0.0447	0.0386	0.9400	0.0542	0.0387	0.8750	0.0491	0.0474	0.9500
	N=50, T=2250								
	MLE-one day			QMLE-res			QMLE-md		
	RMSE	Ave.se	Cove	RMSE	Ave.se	Cove	RMSE	Ave.se	Cove
$\boldsymbol{z}_0^A$	0.0465	0.0619	0.9894	0.0484	0.0614	0.9845	0.1244	0.0869	0.8519
$\boldsymbol{z}_1^A$	0.0495	0.0626	0.9835	0.0514	0.0622	0.9806	0.2010	0.0984	0.8516
$\boldsymbol{z}_2^A$	0.0446	0.0595	0.9908	0.0489	0.0595	0.9818	0.1510	0.0852	0.8396
$\boldsymbol{z}_3^A$	0.0531	0.0589	0.9682	0.0542	0.0589	0.9636	0.1273	0.0767	0.7390
$\boldsymbol{z}_0^E$	0.0513	0.0623	0.9783	0.0527	0.0619	0.9739	0.1459	0.0916	0.8564
$\boldsymbol{z}_1^E$	0.0544	0.0633	0.9723	0.0576	0.0627	0.9618	0.1462	0.0952	0.8580
$\boldsymbol{z}_2^E$	0.0499	0.0614	0.9811	0.0511	0.0609	0.9788	0.2176	0.0834	0.8533
$\boldsymbol{z}_3^E$	0.0467	0.0571	0.9814	0.0473	0.0571	0.9802	0.0850	0.0558	0.8043
$\boldsymbol{z}_0^U$	0.0513	0.0610	0.9765	0.0568	0.0605	0.9578	0.1707	0.1215	0.8618
$\boldsymbol{z}_1^U$	0.0572	0.0636	0.9697	0.0596	0.0630	0.9614	0.2290	0.1386	0.8660
$\boldsymbol{z}_2^U$	0.0540	0.0625	0.9742	0.0571	0.0617	0.9602	0.1600	0.1069	0.8592
$\boldsymbol{z}_3^U$	0.0462	0.0560	0.9816	0.0472	0.0559	0.9771	0.1048	0.0804	0.8483
Σ_c	0.0649	0.0624	0.9215	0.0651	0.0624	0.9189	0.1115	0.0842	0.8253
ϕ	0.0238	0.0216	0.9100	0.0277	0.0218	0.8450	0.0283	0.0223	0.8900

Table 1: RMSE, Ave.se and Cove stand for the root mean square errors, the average of the standard errors across the Monte Carlo samples, and the coverage probability of the confidence interval formed by the point estimate $\pm$ 1.96$\times$the standard error, respectively.

	N=200, T=750								
	MLE-one day			QMLE-res			QMLE-md		
	RMSE	Ave.se	Cove	RMSE	Ave.se	Cove	RMSE	Ave.se	Cove
$\boldsymbol{z}_0^A$	0.0823	0.0924	0.9635	0.0860	0.0923	0.9538	0.1638	0.1435	0.9278
$\boldsymbol{z}_1^A$	0.0835	0.0933	0.9660	0.0859	0.0932	0.9592	0.2047	0.2034	0.9291
$\boldsymbol{z}_2^A$	0.0796	0.0897	0.9672	0.0837	0.0896	0.9561	0.1680	0.1586	0.9325
$\boldsymbol{z}_3^A$	0.0715	0.0884	0.9807	0.0724	0.0884	0.9791	0.1244	0.1080	0.9120
$\boldsymbol{z}_0^E$	0.0817	0.0916	0.9637	0.0829	0.0915	0.9604	0.1714	0.1570	0.9322
$\boldsymbol{z}_1^E$	0.0843	0.0925	0.9588	0.0888	0.0924	0.9484	0.1737	0.1749	0.9497
$\boldsymbol{z}_2^E$	0.0816	0.0916	0.9630	0.0829	0.0916	0.9592	0.1435	0.1322	0.9345
$\boldsymbol{z}_3^E$	0.0694	0.0872	0.9823	0.0696	0.0872	0.9819	0.1169	0.1043	0.9147
$\boldsymbol{z}_0^U$	0.0806	0.0957	0.9745	0.0839	0.0955	0.9678	0.1733	0.1646	0.9455
$\boldsymbol{z}_1^U$	0.0834	0.0964	0.9714	0.0866	0.0964	0.9644	0.2480	0.2438	0.9392
$\boldsymbol{z}_2^U$	0.0785	0.0969	0.9812	0.0825	0.0968	0.9734	0.1809	0.2050	0.9569
$\boldsymbol{z}_3^U$	0.0717	0.0916	0.9852	0.0747	0.0916	0.9798	0.1570	0.1432	0.9232
Σ_c	0.0953	0.0924	0.9444	0.0954	0.0924	0.9437	0.1777	0.1173	0.7849
ϕ	0.0392	0.0322	0.8800	0.0413	0.0321	0.8850	0.0411	0.0419	0.9500
	N=200, T=2250								
	MLE-one day			QMLE-res			QMLE-md		
	RMSE	Ave.se	Cove	RMSE	Ave.se	Cove	RMSE	Ave.se	Cove
$\boldsymbol{z}_0^A$	0.0465	0.0539	0.9703	0.0502	0.0538	0.9550	0.0874	0.0842	0.9481
$\boldsymbol{z}_1^A$	0.0477	0.0545	0.9691	0.0493	0.0543	0.9624	0.1119	0.1177	0.9471
$\boldsymbol{z}_2^A$	0.0447	0.0523	0.9723	0.0469	0.0523	0.9656	0.0911	0.0963	0.9508
$\boldsymbol{z}_3^A$	0.0408	0.0516	0.9836	0.0410	0.0516	0.9831	0.0659	0.0651	0.9377
$\boldsymbol{z}_0^E$	0.0473	0.0532	0.9660	0.0480	0.0531	0.9613	0.0932	0.0966	0.9517
$\boldsymbol{z}_1^E$	0.0495	0.0539	0.9568	0.0546	0.0538	0.9329	0.0907	0.1026	0.9677
$\boldsymbol{z}_2^E$	0.0448	0.0533	0.9749	0.0460	0.0531	0.9706	0.0744	0.0806	0.9600
$\boldsymbol{z}_3^E$	0.0405	0.0505	0.9823	0.0406	0.0505	0.9819	0.0620	0.0630	0.9419
$\boldsymbol{z}_0^U$	0.0451	0.0557	0.9810	0.0471	0.0555	0.9750	0.0914	0.0999	0.9662
$\boldsymbol{z}_1^U$	0.0482	0.0563	0.9723	0.0502	0.0561	0.9648	0.1352	0.1531	0.9576
$\boldsymbol{z}_2^U$	0.0457	0.0564	0.9800	0.0488	0.0562	0.9694	0.0943	0.1131	0.9726
$\boldsymbol{z}_3^U$	0.0410	0.0532	0.9874	0.0423	0.0532	0.9845	0.0810	0.0845	0.9515
Σ_c	0.0547	0.0540	0.9523	0.0547	0.0540	0.9519	0.0860	0.0738	0.9088
ϕ	0.0216	0.0225	0.9650	0.0223	0.0225	0.9450	0.0232	0.0230	0.9550

Table 2: RMSE, Ave.se and Cove stand for the root mean square errors, the average of the standard errors across the Monte Carlo samples, and the coverage probability of the confidence interval formed by the point estimate ± 1.96×the standard error, respectively.

	Market	Market Cap (US trillions)	Trading Hours (Beijing Time)	Country
1	New York S.E.	\$26.91	21:30-04:00 (+1)†	US
2	NASDAQ S.E.	\$23.46	21:30-04:00 (+1)†	US
3	Shanghai S.E.	\$7.69	09:30-11:30, 13:00-15:00	China
4	Tokyo S.E.	\$6.79	08:00-10:30, 11:30-14:00	Japan
5	Hong Kong S.E.	\$6.02	10:00-12:30, 14:30-16:00	China
6	Shenzhen S.E.	\$5.74	09:30-11:30, 13:00-15:00	China
7	London S.E.	\$3.83	15:00-19:00, 19:02-23:30†	UK
8	India National S.E.	\$3.4	11:45-18:00	India
9	Toronto S.E.	\$3.18	21:30-04:00 (+1)†	Canada
10	Frankfurt S.E.	\$2.63	15:00-23:30†	Germany

Table 3: S.E. stands for Stock Exchange. Market Capitalisations were measured on October 29th 2021. (+1) indicates +one day. † means summer time; its corresponding winter trading hours are summer trading hours plus one hour. Data source: `https://www.tradinghours.com/markets`.

MLE-one day estimator with its standard errors approximated by the numerical Hessian method.

We next discuss how to interpret the factor loadings of our model. Let $\dot{y}^c_{i,t}$ denote the standardised return of portfolio i of market c on period t. Recall that

$$\dot{y}^c_{i,t} = \sum_{j=0}^{2} z^c_{i,j} f_{g,t-j} + z^c_{i,3} f_{C,t} + e^c_{i,t},$$

where $\mathrm{var}(f_{g,t}) = (1-\phi^2)^{-1}$ and $\mathrm{var}\, f_{C,t} = 1$. Thus an additional standard-deviation increase in $f_{g,t-1}$ predicts $z^c_{i,1}/\sqrt{1-\phi^2}$ standard-deviation increase in the standardised return $\dot{y}^c_{i,t}$, while an additional standard-deviation increase in $f_{C,t}$ predicts $z^c_{i,3}$ standard-deviation increase in the standardised return $\dot{y}^c_{i,t}$.

We then give a formula for variance decomposition. Recall (2.3): $\boldsymbol{\alpha}_{t+1} = \mathcal{T}\boldsymbol{\alpha}_t + R\boldsymbol{\eta}_t$, $\boldsymbol{\eta}_t \sim N(0, I_2)$. We first calculate the unconditional variance of $\boldsymbol{\alpha}_t$; it can be shown that

$$\mathrm{var}(\boldsymbol{\alpha}_t) = \mathrm{unvec}\left\{[I_{16} - \mathcal{T}\otimes\mathcal{T}]^{-1}(R\otimes R)\,\mathrm{vec}(I_2)\right\} = \begin{bmatrix} \frac{1}{1-\phi^2} & \frac{\phi}{1-\phi^2} & \frac{\phi^2}{1-\phi^2} & 0 \\ \frac{\phi}{1-\phi^2} & \frac{1}{1-\phi^2} & \frac{\phi}{1-\phi^2} & 0 \\ \frac{\phi^2}{1-\phi^2} & \frac{\phi}{1-\phi^2} & \frac{1}{1-\phi^2} & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix}.$$

Since $\dot{y}^c_{i,t} = \begin{bmatrix} z^c_{i,0} & z^c_{i,1} & z^c_{i,2} & z^c_{i,3} \end{bmatrix} \boldsymbol{\alpha}_t + e^c_{i,t}$, we have

$$\begin{aligned}
\operatorname{var}(\dot{y}^c_{i,t}) &= \begin{bmatrix} z^c_{i,0} & z^c_{i,1} & z^c_{i,2} & z^c_{i,3} \end{bmatrix} \begin{bmatrix} \frac{1}{1-\phi^2} & \frac{\phi}{1-\phi^2} & \frac{\phi^2}{1-\phi^2} & 0 \\ \frac{\phi}{1-\phi^2} & \frac{1}{1-\phi^2} & \frac{\phi}{1-\phi^2} & 0 \\ \frac{\phi^2}{1-\phi^2} & \frac{\phi}{1-\phi^2} & \frac{1}{1-\phi^2} & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} z^c_{i,0} \\ z^c_{i,1} \\ z^c_{i,2} \\ z^c_{i,3} \end{bmatrix} + \operatorname{var}(e^c_{i,t}) \\
&= \underbrace{\frac{1}{1-\phi^2}\left[z^{c,2}_{i,0} + z^{c,2}_{i,1} + z^{c,2}_{i,2} + 2\phi z^c_{i,1} z^c_{i,0} + 2\phi z^c_{i,1} z^c_{i,2} + 2\phi^2 z^c_{i,2} z^c_{i,0}\right]}_{\text{variance due to the global factor}} \\
&\quad + \underbrace{z^{c,2}_{i,3}}_{\text{variance due to the continental factor}} + \underbrace{\sigma^2_{c,i}}_{\text{variance due to the idiosyncratic error}} .
\end{aligned} \tag{7.1}$$

For fixed $\operatorname{var}(\dot{y}^c_{i,t})$ and $\sigma^2_{c,i}$, a small $z^{c,2}_{i,3}$ means that the variance of the return is largely explained by the global factor.

7.1.1 Two Periods of Five-Year Data

We estimate our model twice using two periods of data (20110103-20151231; 20160104-20201231). We take care of the missing returns due to the continent-specific reasons using the technique outlined in Section B.1. The starting values of the parameters for the EM algorithm are estimated according to Section 3.

The MLE-one day estimates of the factor loading matrices are reported in Table 4. We first examine the six portfolios constructed by intersections of size and book-to-market equity ratio (B/M) groups. In 2016-2020 the Japanese standardised returns were more likely to be affected by the global factor during the US trading time ($\boldsymbol{z}^A_1$) and less likely to be affected by the global factor during the European trading time ($\boldsymbol{z}^A_2$) than they were in 2011-2015. Take the Japanese SG portfolio as an example. In 2011-2015, an additional standard-deviation increase in $f_{g,t-1}$ predicts $0.24/\sqrt{1-0.1654^2} = 0.2424$ standard-deviation increase in the standardised return $\dot{y}^A_{SG,t}$, while an additional standard-deviation increase in $f_{A,t-2}$ predicts $0.51/\sqrt{1-0.1654^2} = 0.5171$ standard-deviation increase in the standardised return $\dot{y}^A_{SG,t}$. In 2016-2020, an additional standard-deviation increase in $f_{g,t-1}$ predicts $0.71/\sqrt{1-0.2323^2} = 0.73$ standard-deviation increase in the standardised return $\dot{y}^A_{SG,t}$, while an additional standard-deviation increase in $f_{A,t-2}$ predicts $0.26/\sqrt{1-0.2323^2} = 0.2673$ standard-deviation increase in the standardised return $\dot{y}^A_{SG,t}$. In 2011-2015, an additional standard-deviation increase in $f_{C,t}$ predicts 0.75 standard-deviation increase in the standardised return, while in 2016-2020, an additional standard-deviation increase in $f_{C,t}$ only predicts 0.63 standard-deviation increase in the standardised return.

For the European portfolios, the standardised returns were less likely to be affected by the global factor during the European and US trading times ($\boldsymbol{z}^E_0, \boldsymbol{z}^E_2$), and more likely to be affected by the continental factor ($\boldsymbol{z}^E_3$) than they were in 2011-2015. Because of the much larger continental loadings in 2016-2020, one could argue that the European portfolios became less integrated into the global market than they were in 2011-2015. Take the European BG portfolio as an example. In 2011-2015, an additional standard-deviation increase in $f_{g,t}$ predicts $0.47/\sqrt{1-0.1654^2} = 0.4766$ standard-deviation increase in the standardised return. In 2016-2020, an additional standard-deviation increase in $f_{g,t}$ predicts $0.18/\sqrt{1-0.2323^2} = 0.1851$ standard-deviation increase in the standardised return.

	Size and B/M portfolios in 2011-2015											
	z_0^A	z_1^A	z_2^A	z_3^A	z_0^E	z_1^E	z_2^E	z_3^E	z_0^U	z_1^U	z_2^U	z_3^U
SG	0.16	0.24	0.51	0.75	0.43	0.54	0.57	0.48	0.45	0.46	0.49	0.48
SN	0.18	0.30	0.63	0.73	0.44	0.54	0.57	0.52	0.47	0.48	0.52	0.48
SV	0.17	0.31	0.66	0.68	0.43	0.53	0.57	0.48	0.47	0.48	0.52	0.46
BG	0.20	0.39	0.85	0.34	0.47	0.58	0.70	0.16	0.41	0.56	0.66	0.15
BN	0.18	0.38	0.89	0.31	0.50	0.60	0.71	0.17	0.41	0.58	0.72	0.09
BV	0.18	0.37	0.85	0.34	0.51	0.56	0.65	0.20	0.42	0.55	0.63	0.18
	Size and B/M portfolios in 2016-2020											
	z_0^A	z_1^A	z_2^A	z_3^A	z_0^E	z_1^E	z_2^E	z_3^E	z_0^U	z_1^U	z_2^U	z_3^U
SG	0.30	0.71	0.26	0.63	0.23	0.53	0.26	0.77	0.50	0.29	0.60	0.42
SN	0.31	0.82	0.47	0.55	0.34	0.54	0.24	0.78	0.40	0.50	0.64	0.47
SV	0.30	0.83	0.56	0.46	0.42	0.55	0.23	0.73	0.30	0.62	0.64	0.42
BG	0.33	0.88	0.36	0.34	0.18	0.54	0.21	0.68	0.55	0.18	0.72	0.01†
BN	0.35	0.92	0.57	0.20	0.36	0.58	0.18	0.68	0.41	0.54	0.74	0.12
BV	0.33	0.85	0.70	0.08	0.52	0.57	0.16	0.60	0.30	0.71	0.72	0.12
	Size and momentum portfolios in 2011-2015											
	z_0^A	z_1^A	z_2^A	z_3^A	z_0^E	z_1^E	z_2^E	z_3^E	z_0^U	z_1^U	z_2^U	z_3^U
SL	0.08	0.27	0.27	0.89	0.37	0.53	0.72	0.26	0.37	0.41	0.69	0.42
SN	0.06	0.25	0.27	0.91	0.42	0.44	0.79	0.16	0.39	0.53	0.53	0.46
SW	0.05♣	0.22	0.26	0.87	0.47	0.34	0.80	0.06♣	0.43	0.59	0.35	0.55
BL	0.12	0.28	0.33	0.79	0.39	0.57	0.64	0.14	0.30	0.48	0.74	0.15
BN	0.08	0.27	0.36	0.81	0.49	0.47	0.71	-0.05†	0.34	0.74	0.49	0.12
BW	0.05	0.26	0.36	0.78	0.55	0.31	0.75	-0.24	0.38	0.77	0.25	0.28
	Size and momentum portfolios in 2016-2020											
	z_0^A	z_1^A	z_2^A	z_3^A	z_0^E	z_1^E	z_2^E	z_3^E	z_0^U	z_1^U	z_2^U	z_3^U
SL	0.23	0.47	0.34	0.84	0.63	0.85	0.35	-0.19	0.52	0.75	0.37	0.43
SN	0.21	0.47	0.24	0.88	0.41	0.95	0.44	-0.25	0.59	0.60	0.42	0.47
SW	0.23	0.45	0.07	0.84	0.24	0.94	0.47	-0.20	0.65	0.34	0.43	0.44
BL	0.19	0.46	0.44	0.73	0.68	0.84	0.28	0.05	0.56	0.83	0.35	0.22
BN	0.22	0.50	0.29	0.80	0.44	0.97	0.41	0.17	0.71	0.66	0.40	0.16
BW	0.21	0.49	0.04†	0.77	0.14	0.94	0.45	0.11	0.87	0.25	0.34	0.01†

Table 4: The MLE-one day of the factor loadings. To save space, we do not report the standard errors, which are around 0.02, with a minimum 0.0084 and maximum 0.0404. Almost all the estimates are significant at 1% significance level. Entries with ♣ are significant only at 5% significance level; entries with † are insignificant at 10%.

In 2011-2015, an additional standard-deviation increase in $f_{C,t}$ predicts 0.16 standard-deviation increase in the standardised return, while in 2016-2020 an additional standard-deviation increase in $f_{C,t}$ predicts 0.68 standard-deviation increase in the standardised return.

For the US portfolios, the standardised returns in 2015-2020 to a large extent became slightly less affected by the global factor during the US trading time ($\boldsymbol{z}_0^U$), and became slightly more affected by the global factor during the Asian trading time ($\boldsymbol{z}_1^U$) than they were in 2011-2015. Take the US SN portfolio as an example. In 2011-2015, an additional standard-deviation increase in $f_{g,t}$ predicts $0.47/\sqrt{1-0.1654^2} = 0.4766$ standard-deviation increase in the standardised return, while an additional standard-deviation increase in $f_{g,t-2}$ predicts $0.52/\sqrt{1-0.1654^2} = 0.5273$ standard-deviation increase in the standardised return. In 2016-2020, an additional standard-deviation increase in $f_{g,t}$ predicts $0.40/\sqrt{1-0.2323^2} = 0.4113$ standard-deviation increase in the standardised return, while an additional standard-deviation increase in $f_{g,t-2}$ predicts $0.64/\sqrt{1-0.2323^2} = 0.6580$ standard-deviation increase in the standardised return. For the US portfolios, the loadings for the continental factor have decreased slightly; in particular the continental loading of the BG portfolio has decreased from 0.15 to something statistically insignificant.

Over the two periods of five years, a few general patterns emerge. First, within the same B/M ratio category, the big portfolio has much smaller loadings for the continental factor but larger loadings for the global factor than the small portfolio. In particular, the variances of the standardised returns of the US big portfolios could largely be explained by the global factor in light of (7.1). Second, within the same size category, the value portfolio is more affected by the global factor during the European trading time than the growth portfolio across the three continents.

We next examine the six portfolios constructed by intersections of size and momentum groups. The Japanese portfolios in general were more affected by the global factor during the Asian trading time ($\boldsymbol{z}_0^A$) than they were in 2011-2015. The effect of the global factor during the US trading time on the Japanese portfolios ($\boldsymbol{z}_1^A$) has almost doubled in 2016-2020. For example, in 2011-2015 an additional standard-deviation increase in $f_{g,t-1}$ predicts 0.27 standard-deviation increase in the standardised return of the Japanese BN portfolio, while in 2016-2020, the same increase predicts $0.5/\sqrt{1-0.3079^2} = 0.5255$ standard-deviation increase in the standardised return.

For the European portfolios, the standardised returns in 2015-2020 became more affected by the global factor during the Asian trading time ($\boldsymbol{z}_1^E$), and became less affected by the global factor during the US trading time ($\boldsymbol{z}_2^E$) than they were in 2011-2015. Take the European BW portfolio as an example. In 2011-2015, an additional standard-deviation increase in $f_{g,t-1}$ predicts 0.31 standard-deviation increase in the standardised return, while an additional standard-deviation increase in $f_{g,t-2}$ predicts 0.75 standard-deviation increase in the standardised return. In 2016-2020, an additional standard-deviation increase in $f_{g,t-1}$ predicts $0.94/\sqrt{1-0.3079^2} = 0.9880$ standard-deviation increase in the standardised return, while an additional standard-deviation increase in $f_{g,t-2}$ predicts $0.45/\sqrt{1-0.3079^2} = 0.4730$ standard-deviation increase in the standardised return.

For the US portfolios, the standardised returns in 2015-2020 became more affected by the global factor during the US trading time ($\boldsymbol{z}_0^U$). Take the US BL portfolio as an example. In 2011-2015, an additional standard-deviation increase in $f_{g,t}$ predicts 0.30 standard-deviation increase in the standardised return, while in 2016-2020 the same increase predicts $0.56/\sqrt{1-0.3079^2} = 0.5886$ standard-deviation increase in the stan-

	Size and B/M		Size and momentum	
	2011-2015	2016-2020	2011-2015	2016-2020
$\tilde{\phi}$	-0.1654	-0.2323	0.0051	-0.3079
	(0.0266)	(0.0219)	(0.0273)	(0.0229)

Table 5: The MLE-one day. The standard errors are in parentheses.

dardised return.

For the size-momentum portfolios, one consistent pattern across the three continents is that in 2011-2015 within the same size category, the winner (W) portfolio was less affected by the global factor during the Asian trading time than the loser (L) portfolio. In 2016-2020, again within the same size category, the winner portfolio was less affected by the global factor during the European trading time than the loser portfolio.

The $\tilde{\phi}$ in the application of size-B/M portfolios is significantly negative in both five-year periods. The value is -0.1654 with a standard error of 0.0266 in 2011-2015, and -0.2323 with a standard error of 0.0219 in 2016-2020. The $\tilde{\phi}$ in the application of size-momentum portfolios is statistically insignificant with a point estimate of 0.0051 in 2011-2015, and significantly negative in 2016-2020, with a value of -0.3079 and a standard error of 0.0229.

7.1.2 Time Series Patterns

In this subsection, we estimate the model using twenty periods of one-year data (1991-2020). For simplicity, we only consider the size-B/M portfolios. The detailed point estimates and their standard errors are available upon request; here we only discuss the main findings.

First, across the three continents, the big portfolios tended to have smaller loadings for the continental factor but larger loadings for the global factor than the small portfolios. The Japanese idiosyncratic variances are in general larger than those of the US and Europe. This is especially so in 1999-2001 and 2019-2020. These observations are consistent with the observations based on five-year data reported in the previous subsection.

Second, we discuss some year-specific patterns:

(i) In 1998, the Japanese portfolios have particularly large loadings for the global factor during the Asian trading time, but small loadings for the global factor during the US trading time. This could be interpreted as the effect of the Asian financial crisis.

(ii) During the 2007-2008 financial crisis, the Japanese portfolios have large loadings for the global factors during the European trading time. The European portfolios have small loadings for the continental factor but large loadings for the global factor during the US trading time. This could be interpreted as the spread of the US subprime mortgage crisis.

(iii) In 2017-2018, the Japanese portfolios have large loadings for the global factor during the US trading time but small loadings for the continental factor. The US portfolios have large loadings for the global factor during the Asian trading time. The European portfolios have large loadings for the continental factor but small loadings for

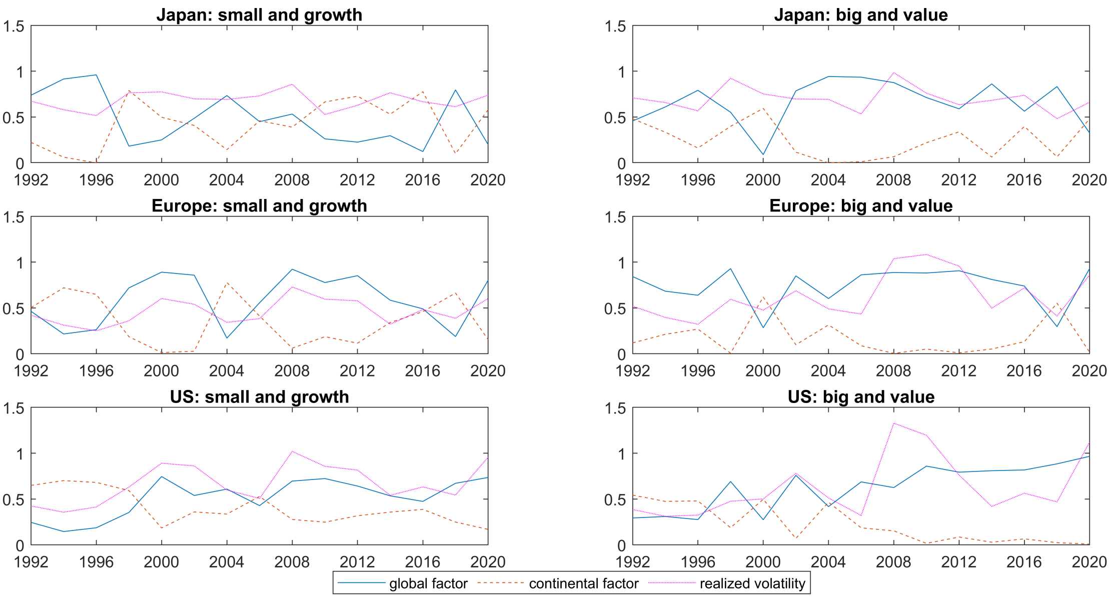

Figure 1: Variance decomposition.

the global factor during the Asian and US trading times. This could be interpreted as Japan and US markets being more integrated during this period but not so for the Europe. This could be due to the Sino-US trade war.

(iv) In 2020, the European portfolios have quite small loadings for the continental factor. Compared with the Japanese and European portfolios, the US portfolios have relatively constant loadings for the global factor during the three trading periods of a day.

Last, we re-estimate the model using fifteen periods of two-year data (1991-2020) and compute the variance decomposition using (7.1). The decompositions are plotted in Figure 1. The blue solid and red dashed lines depict the variance proportions of the global and continental factors, respectively. The magenta dotted lines represent the realized volatilities computed using the standardised portfolio returns (divided by two for a better layout). We find that the continental factor accounts for a decreasing share of variance of the US BV standardised portfolio returns in the past 30 years. In the 1990s, the global factor only accounted for small shares of variances of the US standardized portfolio returns. During the turbulent years such as the 2008 financial crisis, the global factor tended to account for larger shares of variances of the European and US standardized portfolio returns.

7.2 An Empirical Study of Many Markets

We now apply our model to MSCI equity indices of the developed and emerging markets (41 markets in total). The daily indices are obtained from https://www.msci.com/end-of-day-data-search. There are 6 indices for each market: Large-Growth, Mid-Growth, Small-Growth, Large-Value, Mid-Value, and Small-Value, all in USD currency. According to the closing time of each market, we categorize these markets into 3 continents: Asia-Pacific, Europe and America. Since the closing time of the Israeli market is both far away from Asia-Pacific and Europe, we exclude it from our sample. We use the data of the period from January 1st 2018 to February 21st 2022. Indices starting after January 1st 2018 are excluded. We estimate the model using the QMLE-md, with choices of $\hat{\boldsymbol{h}}$ similar to those mentioned in Section 6.[5] The estimated ϕ is 0.338 with a standard error 0.0262.

Table 6 reports the estimates of the factor loadings and idiosyncratic variances for the Asian-Pacific continent. We present all the indices for Mainland China, Hong Kong and Japan, but only Middle-Value and Middle-Growth indices for other Asian-Pacific markets in the interest of space. There are several findings. First, Mainland China and Hong Kong have particularly high loadings on the global factors during the US trading time (i.e, $\boldsymbol{z}_1^A$). Second, Japan has high loadings on the continental factor (i.e, $\boldsymbol{z}_3^A$) but small idiosyncratic variances. Third, the growth indices in general have larger idiosyncratic variances than the value indices. Fourth, most other Asian-Pacific markets have large loadings on the global factor during the US trading time (i.e, $\boldsymbol{z}_1^A$), but small and insignificant loadings on the continental factor (i.e, $\boldsymbol{z}_3^A$).

Table 7 reports the estimates of the factor loadings and idiosyncratic variances for the European continent. We present all the indices for the UK, but only Mid-Value and Mid-Growth indices for other European markets in the interest of space. Most European markets have the largest loadings on the global factor during the Asian trading time (i.e., $\boldsymbol{z}_1^E$). The developed European markets have large and positive loadings on the continental factor (i.e., $\boldsymbol{z}_3^E$), but the emerging European markets have small or negative loadings on the continental factor (i.e., $\boldsymbol{z}_3^E$).

Table 8 reports the estimates of the factor loadings and idiosyncratic variance for the American continent. The US and Canada have statistically insignificant factor loadings on the continental factor (i.e., $\boldsymbol{z}_3^U$), while Brazil has large factor loadings on the continental factor (i.e., $\boldsymbol{z}_3^U$ with point estimates greater than 1). Moreover, some emerging American markets (e.g., Mexico) have higher loadings on the global factor during the Asian trading time (i.e., $\boldsymbol{z}_2^U$) but small (and possibly insignificant) loadings on the global factor during the American and European trading times (i.e., $\boldsymbol{z}_0^U, \boldsymbol{z}_1^U$), while this pattern does not hold for the US market.

8 Conclusion

In this article we propose a new framework of using a statistical dynamic factor model to model a large number of daily stock returns across different time zones. The presence of global and continental factors describes a situation in which all the new information represented by the global and continental factors accumulated since the last closure of a continent will have an impact on the upcoming observed logarithmic 24-hr returns of that continent. Our model is identified under a mild fixed-signs assumption and hence has a structural interpretation. Several estimator are outlined: the MLE-one day, the QMLE-res, the QMLE, the QMLE-md and the Bayesian. The asymptotic theories of the QMLE and the QMLE-md are carefully derived. In addition, we propose a way to approximate the standard errors of the MLE-one day and the QMLE-res. Monte Carlo simulations show good performance of the MLE-one day, the QMLE-res and the QMLE-md. Last,

[5] We also estimate the model using the MLE-one day (results available upon request). Its point estimates are close to those of the QMLE-md; its standard errors are slightly smaller than those of the QMLE-md.

	z_0^A	z_1^A	z_2^A	z_3^A	$\sigma_{A,i}^2$
JAPAN LG	0.167 (0.118)	0.658 (0.216)	0.179 (0.125)	0.620 (0.054)	0.119 (0.010)
JAPAN LV	0.037 (0.075)	0.520 (0.198)	0.107 (0.103)	0.724 (0.032)	0.086 (0.008)
JAPAN MG	0.133 (0.121)	0.689 (0.232)	0.204 (0.130)	0.725 (0.051)	0.089 (0.008)
JAPAN MV	0.070 (0.084)	0.495 (0.206)	0.125 (0.109)	0.740 (0.034)	0.042 (0.004)
JAPAN SG	0.099 (0.118)	0.699 (0.239)	0.248 (0.132)	0.776 (0.050)	0.121 (0.011)
JAPAN SV	0.039 (0.077)	0.502 (0.209)	0.154 (0.109)	0.789 (0.031)	0.048 (0.004)
CHINA (Mainland) LG	0.486 (0.612)	1.014 (0.229)	-0.362 (0.234)	-0.406 (0.277)	0.592 (0.052)
CHINA (Mainland) LV	0.300 (0.595)	0.984 (0.222)	-0.586 (0.218)	-0.249 (0.252)	0.188 (0.016)
CHINA (Mainland) MG	0.294 (0.676)	1.229 (0.254)	-0.452 (0.254)	-0.305 (0.304)	1.060 (0.093)
CHINA (Mainland) MV	0.664 (0.688)	0.993 (0.253)	-0.269 (0.273)	0.003 (0.317)	0.343 (0.030)
CHINA (Mainland) SG	0.426 (0.578)	0.943 (0.212)	-0.359 (0.215)	-0.203 (0.249)	0.280 (0.025)
CHINA (Mainland) SV	0.491 (0.539)	0.816 (0.198)	-0.231 (0.213)	-0.007 (0.245)	0.156 (0.014)
HONG KONG LG	0.435 (0.706)	1.049 (0.258)	-0.282 (0.267)	0.008 (0.325)	0.423 (0.037)
HONG KONG LV	0.336 (0.444)	0.730 (0.167)	-0.420 (0.163)	-0.120 (0.192)	0.283 (0.025)
HONG KONG MG	0.408 (0.551)	0.872 (0.206)	-0.453 (0.220)	-0.008 (0.263)	0.508 (0.044)
HONG KONG MV	0.214 (0.422)	0.695 (0.159)	-0.345 (0.158)	0.006 (0.183)	0.274 (0.024)
HONG KONG SG	0.518 (0.582)	0.836 (0.216)	-0.196 (0.224)	0.052 (0.259)	0.256 (0.022)
HONG KONG SV	0.359 (0.441)	0.640 (0.162)	-0.101 (0.173)	0.038 (0.201)	0.163 (0.014)
AUSTRALIA MG	0.501 (0.102)	0.174 (0.105)	0.086 (0.102)	0.080 (0.072)	0.450 (0.039)
NEW ZEALAND MG	0.434 (0.185)	0.314 (0.120)	0.152 (0.126)	-0.099 (0.106)	1.101 (0.096)
SINGAPORE MG	0.450 (0.299)	0.474 (0.119)	-0.221 (0.143)	-0.056 (0.155)	0.484 (0.042)
INDIA MG	-0.055 (0.107)	0.168 (0.082)	0.035 (0.089)	0.164 (0.092)	2.322 (0.203)
INDONESIA MG	0.217 (0.137)	0.057 (0.128)	0.140 (0.131)	-0.381 (0.130)	6.538 (0.572)
KOREA MG	0.181 (0.156)	0.081 (0.102)	-0.010 (0.112)	0.109 (0.120)	3.503 (0.307)
MALAYSIA MG	0.023 (0.126)	0.200 (0.083)	0.078 (0.087)	0.066 (0.085)	1.460 (0.128)
PHILIPPINES MG	0.245 (0.280)	0.437 (0.126)	-0.120 (0.131)	-0.131 (0.137)	2.259 (0.198)
TAIWAN MG	0.500 (0.457)	0.840 (0.190)	0.046 (0.213)	0.004 (0.220)	0.737 (0.065)
THAILAND MG	0.323 (0.163)	0.213 (0.078)	-0.169 (0.093)	0.033 (0.103)	0.855 (0.075)
AUSTRALIA MV	0.411 (0.107)	0.226 (0.090)	-0.025 (0.095)	0.059 (0.076)	0.550 (0.048)
NEW ZEALAND MV	0.296 (0.056)	0.065 (0.060)	0.102 (0.060)	-0.026 (0.052)	0.886 (0.078)
SINGAPORE MV	0.413 (0.210)	0.403 (0.091)	-0.217 (0.107)	-0.124 (0.110)	0.307 (0.027)
INDIA MV	0.456 (0.221)	0.309 (0.114)	-0.199 (0.126)	0.103 (0.141)	1.651 (0.145)
INDONESIA MV	0.447 (0.368)	0.718 (0.164)	-0.537 (0.185)	-0.110 (0.195)	1.854 (0.162)
KOREA MV	0.398 (0.405)	0.810 (0.181)	-0.433 (0.178)	0.099 (0.189)	0.817 (0.072)
MALAYSIA MV	0.296 (0.241)	0.373 (0.111)	-0.067 (0.125)	0.163 (0.125)	0.729 (0.064)
PHILIPPINES MV	0.224 (0.237)	0.388 (0.104)	0.018 (0.105)	-0.168 (0.106)	1.338 (0.117)
TAIWAN MV	0.366 (0.346)	0.651 (0.141)	0.019 (0.154)	-0.050 (0.159)	0.254 (0.022)
THAILAND MV	0.325 (0.163)	0.270 (0.076)	-0.231 (0.093)	0.020 (0.099)	0.620 (0.054)

Table 6: Selected QMLE-md estimates of the factor loadings and idiosyncratic variances. LV, LG, MV, MG, SV, SG stand for Large-Value, Large-Growth, Middle-Value, Middle-Growth, Small-Value and Small-Growth, respectively. The standard errors are in parentheses. The z_0^A, z_1^A and z_2^A stand for the factor loadings on the global factor during the Asian-Pacific, American, and European trading times, respectively. The z_3^A stands for the factor loadings on the Asian-Pacific continental factor.

	$\boldsymbol{z}_0^E$	$\boldsymbol{z}_1^E$	$\boldsymbol{z}_2^E$	$\boldsymbol{z}_3^E$	$\sigma_{E,i}^2$
U.K. LG	0.100 (0.039)	0.264 (0.091)	0.218 (0.055)	0.308 (0.031)	0.277 (0.024)
U.K. LV	0.101 (0.046)	0.251 (0.128)	0.101 (0.080)	0.357 (0.045)	0.222 (0.019)
U.K. MG	0.218 (0.055)	0.486 (0.158)	0.269 (0.085)	0.508 (0.034)	0.173 (0.015)
U.K. MV	0.040 (0.057)	0.477 (0.131)	0.176 (0.079)	0.552 (0.042)	0.441 (0.039)
U.K. SG	0.171 (0.054)	0.479 (0.161)	0.255 (0.086)	0.472 (0.035)	0.310 (0.027)
U.K. SV	0.055 (0.054)	0.519 (0.124)	0.138 (0.074)	0.494 (0.039)	0.357 (0.031)
AUSTRIA MG	0.142 (0.074)	0.661 (0.103)	0.083 (0.084)	0.373 (0.065)	1.512 (0.132)
BELGIUM MG	0.222 (0.079)	0.905 (0.105)	-0.138 (0.089)	0.395 (0.067)	0.890 (0.078)
DENMARK MG	0.238 (0.056)	0.452 (0.109)	0.140 (0.069)	0.369 (0.041)	0.487 (0.043)
FINLAND MG	0.228 (0.072)	0.699 (0.130)	0.048 (0.088)	0.417 (0.055)	0.668 (0.058)
FRANCE MG	0.293 (0.057)	0.579 (0.143)	0.162 (0.079)	0.499 (0.033)	0.113 (0.010)
GERMANY MG	0.280 (0.061)	0.640 (0.125)	0.082 (0.075)	0.501 (0.039)	0.173 (0.015)
IRELAND MG	0.158 (0.059)	0.521 (0.111)	0.099 (0.074)	0.223 (0.050)	0.753 (0.066)
ITALY MG	0.228 (0.077)	0.782 (0.154)	0.088 (0.092)	0.548 (0.049)	0.461 (0.040)
NETHERLANDS MG	0.059 (0.047)	0.460 (0.071)	0.120 (0.054)	0.323 (0.038)	0.360 (0.032)
NORWAY MG	0.142 (0.060)	0.791 (0.097)	0.037 (0.078)	0.236 (0.056)	0.605 (0.053)
PORTUGAL MG	0.043 (0.060)	0.823 (0.082)	-0.120 (0.075)	0.250 (0.059)	0.745 (0.065)
SPAIN MG	0.227 (0.066)	0.471 (0.121)	0.173 (0.078)	0.495 (0.050)	0.893 (0.078)
SWEDEN MG	0.195 (0.062)	0.736 (0.167)	0.234 (0.096)	0.471 (0.045)	0.337 (0.030)
SWITZERLAND MG	0.216 (0.048)	0.454 (0.128)	0.239 (0.072)	0.378 (0.032)	0.157 (0.014)
GREECE MG	-0.022 (0.081)	0.894 (0.139)	-0.061 (0.101)	0.086 (0.083)	1.971 (0.173)
POLAND MG	0.060 (0.066)	1.116 (0.162)	0.151 (0.097)	-0.082 (0.073)	1.132 (0.099)
RUSSIA MG	-0.128 (0.063)	0.860 (0.071)	-0.091 (0.073)	-0.264 (0.070)	1.096 (0.096)
SOUTH AFRICA MG	-0.192 (0.095)	1.557 (0.117)	0.128 (0.114)	-0.367 (0.106)	1.140 (0.100)
TURKEY MG	-0.229 (0.107)	1.465 (0.157)	-0.510 (0.123)	-0.635 (0.116)	2.982 (0.261)
AUSTRIA MV	0.058 (0.073)	1.002 (0.114)	-0.114 (0.093)	0.362 (0.068)	0.975 (0.085)
BELGIUM MV	-0.030 (0.047)	0.519 (0.091)	0.044 (0.062)	0.391 (0.039)	0.284 (0.025)
DENMARK MV	-0.037 (0.094)	0.588 (0.109)	-0.084 (0.099)	0.280 (0.090)	3.513 (0.308)
FINLAND MV	0.161 (0.067)	0.700 (0.143)	0.101 (0.091)	0.399 (0.055)	1.007 (0.088)
FRANCE MV	0.049 (0.049)	0.587 (0.092)	-0.022 (0.062)	0.397 (0.037)	0.234 (0.020)
GERMANY MV	0.110 (0.051)	0.520 (0.086)	0.034 (0.060)	0.438 (0.037)	0.204 (0.018)
IRELAND MV	0.003 (0.087)	0.619 (0.112)	-0.036 (0.089)	0.746 (0.068)	1.823 (0.160)
ITALY MV	-0.021 (0.067)	0.685 (0.081)	-0.056 (0.069)	0.582 (0.051)	0.489 (0.043)
NETHERLANDS MV	0.157 (0.061)	0.593 (0.117)	0.072 (0.075)	0.454 (0.043)	0.256 (0.022)
NORWAY MV	0.103 (0.062)	0.800 (0.085)	-0.063 (0.075)	0.256 (0.058)	0.807 (0.071)
SPAIN MV	-0.057 (0.058)	0.686 (0.075)	-0.097 (0.062)	0.534 (0.043)	0.272 (0.024)
SWEDEN MV	0.132 (0.059)	0.573 (0.122)	0.126 (0.077)	0.467 (0.045)	0.457 (0.040)
SWITZERLAND MV	0.024 (0.042)	0.387 (0.099)	0.191 (0.060)	0.442 (0.031)	0.212 (0.019)
CZECH MV	-0.093 (0.048)	0.456 (0.069)	0.080 (0.057)	0.059 (0.050)	0.796 (0.070)
EGYPT MV	0.176 (0.083)	-0.244 (0.168)	0.380 (0.097)	0.121 (0.083)	2.634 (0.231)
GREECE MV	-0.091 (0.085)	0.867 (0.158)	-0.036 (0.110)	0.329 (0.083)	2.448 (0.214)
POLAND MV	-0.077 (0.070)	0.907 (0.094)	-0.002 (0.083)	-0.095 (0.076)	1.366 (0.120)
RUSSIA MV	-0.199 (0.067)	0.862 (0.074)	-0.111 (0.075)	-0.227 (0.072)	1.329 (0.116)
SOUTH AFRICA MV	-0.130 (0.076)	1.430 (0.105)	0.234 (0.098)	-0.293 (0.090)	0.709 (0.062)
TURKEY MV	-0.474 (0.148)	1.894 (0.243)	-0.863 (0.169)	-0.835 (0.156)	5.356 (0.469)

Table 7: Selected QMLE-md estimates of the factor loadings and idiosyncratic variances. LV, LG, MV, MG, SV, SG stand for Large-Value, Large-Growth, Middle-Value, Middle-Growth, Small-Value and Small-Growth, respectively. The standard errors are in parentheses. The $\boldsymbol{z}_0^E$, $\boldsymbol{z}_1^E$ and $\boldsymbol{z}_2^E$ stand for the factor loadings on the global factor during the European, Asian-Pacific and American trading times, respectively. The $\boldsymbol{z}_3^E$ stands for the factor loadings on the European continental factor.

	z_0^U	z_1^U	z_2^U	z_3^U	$\sigma^2_{U,i}$
USA LG	0.117 (0.231)	0.809 (0.131)	0.174 (0.097)	0.051 (0.136)	0.135 (0.012)
USA LV	-0.232 (0.165)	0.380 (0.093)	0.348 (0.085)	-0.186 (0.116)	0.079 (0.007)
USA MG	0.238 (0.182)	0.780 (0.121)	0.141 (0.090)	0.082 (0.131)	0.060 (0.005)
USA MV	-0.083 (0.166)	0.496 (0.099)	0.297 (0.076)	-0.125 (0.106)	0.066 (0.006)
USA SG	-0.154 (0.225)	0.673 (0.148)	0.550 (0.189)	-0.271 (0.217)	0.080 (0.007)
USA SV	0.084 (0.150)	0.374 (0.114)	0.367 (0.146)	-0.177 (0.174)	0.101 (0.009)
CANADA LG	-0.133 (0.104)	0.312 (0.060)	0.474 (0.074)	0.001 (0.116)	0.192 (0.017)
CANADA LV	-0.122 (0.094)	0.208 (0.062)	0.517 (0.085)	-0.032 (0.087)	0.175 (0.015)
CANADA MG	0.002 (0.121)	0.369 (0.075)	0.370 (0.075)	0.193 (0.107)	0.271 (0.024)
CANADA MV	-0.081 (0.128)	0.333 (0.080)	0.602 (0.095)	0.148 (0.140)	0.364 (0.032)
CANADA SG	-0.051 (0.097)	0.248 (0.064)	0.406 (0.075)	0.102 (0.105)	0.334 (0.029)
CANADA SV	-0.162 (0.078)	0.186 (0.051)	0.483 (0.066)	0.177 (0.093)	0.264 (0.023)
BRAZIL SG	-0.138 (0.130)	0.502 (0.269)	0.364 (0.180)	1.480 (0.108)	0.333 (0.029)
BRAZIL SV	-0.161 (0.141)	0.542 (0.273)	0.421 (0.175)	1.374 (0.126)	0.131 (0.011)
BRAZIL MV	-0.072 (0.124)	0.241 (0.224)	0.311 (0.166)	1.512 (0.105)	0.413 (0.036)
BRAZIL MG	-0.011 (0.110)	0.227 (0.202)	0.369 (0.152)	1.327 (0.094)	0.256 (0.022)
BRAZIL LG	-0.000 (0.121)	0.333 (0.215)	0.451 (0.159)	1.348 (0.107)	0.354 (0.031)
BRAZIL LV	-0.050 (0.151)	0.335 (0.262)	0.521 (0.182)	1.538 (0.135)	0.731 (0.064)
MEXICO SG	0.039 (0.105)	0.211 (0.096)	0.571 (0.082)	0.426 (0.110)	0.696 (0.061)
MEXICO SV	0.083 (0.116)	0.317 (0.106)	0.564 (0.087)	0.355 (0.112)	0.726 (0.064)
MEXICO MV	0.038 (0.106)	0.223 (0.101)	0.664 (0.086)	0.421 (0.115)	0.726 (0.064)
MEXICO MG	0.047 (0.097)	-0.151 (0.096)	0.169 (0.092)	0.213 (0.105)	3.153 (0.276)
MEXICO LG	-0.013 (0.118)	0.248 (0.108)	0.678 (0.091)	0.414 (0.120)	0.815 (0.071)
MEXICO LV	-0.021 (0.125)	0.268 (0.115)	0.781 (0.099)	0.349 (0.133)	0.866 (0.076)
CHILE SG	-0.125 (0.097)	-0.031 (0.087)	0.875 (0.099)	0.319 (0.151)	0.864 (0.076)
CHILE SV	-0.122 (0.116)	0.065 (0.097)	0.918 (0.102)	0.300 (0.157)	1.058 (0.093)
CHILE MV	-0.051 (0.101)	0.043 (0.092)	0.842 (0.089)	0.233 (0.137)	0.799 (0.070)
CHILE MG	-0.010 (0.110)	0.188 (0.104)	0.759 (0.093)	0.112 (0.138)	1.478 (0.129)
CHILE LG	0.072 (0.122)	0.319 (0.111)	0.767 (0.097)	0.301 (0.136)	0.984 (0.086)
CHILE LV	0.042 (0.114)	0.249 (0.102)	0.697 (0.091)	0.204 (0.129)	0.822 (0.072)
COLOMBIA SG	-0.027 (0.119)	0.416 (0.110)	0.418 (0.090)	0.179 (0.115)	1.551 (0.136)
COLOMBIA SV	-0.065 (0.126)	0.200 (0.106)	0.480 (0.089)	0.138 (0.124)	1.653 (0.145)
COLOMBIA LG	-0.482 (0.158)	0.267 (0.104)	0.691 (0.129)	0.617 (0.175)	1.800 (0.158)
COLOMBIA LV	-0.095 (0.134)	0.359 (0.104)	0.716 (0.093)	0.257 (0.136)	0.873 (0.076)

Table 8: Selected QMLE-md estimates of the factor loadings and idiosyncratic variances. LV, LG, MV, MG, SV, SG stand for Large-Value, Large-Growth, Middle-Value, Middle-Growth, Small-Value and Small-Growth, respectively. The standard errors are in parentheses. The z_0^U, z_1^U and z_2^U stand for the factor loadings on the global factor during the American, European and Asian-Pacific trading times, respectively. The z_3^U stands for the factor loadings on the American continental factor.

we present two empirical applications of our model. One future research direction is to work out the asymptotic theories of the MLE-one day and the QMLE-res in the large N large T_f case. Perhaps one could try to adapt the results of Barigozzi and Luciani (2022).

A Appendix

A.1 Proof of Lemma 2.1

Proof of Lemma 2.1. This proof is inspired by that of Bai and Wang (2015). Suppose that Assumption 2.2 hold. *Fix* a particular t. Recall (2.3):

$$\boldsymbol{y}_t = Z_t \begin{pmatrix} f_{g,t} \\ f_{g,t-1} \\ f_{g,t-2} \\ f_{C,t} \end{pmatrix} + \boldsymbol{\varepsilon}_t \quad f_{g,t+1} = \phi f_{g,t} + \eta_{g,t} \quad f_{C,t+1} = \eta_{C,t}.$$

Note that

$$\begin{aligned} f_{g,t} &= \phi^3 f_{g,t-3} + \phi^2 \eta_{g,t-3} + \phi \eta_{g,t-2} + \eta_{g,t-1} \\ f_{g,t-1} &= \phi^2 f_{g,t-3} + \phi \eta_{g,t-3} + \eta_{g,t-2} \\ f_{g,t-2} &= \phi f_{g,t-3} + \eta_{g,t-3}. \end{aligned}$$

Since Z_t assumes one of $\{Z^A, Z^E, Z^U\}$, we need to consider three 4×4 rotation matrices represented by:

$$\Delta_1 := \begin{bmatrix} A_1 & B_1 & C_1 & O_1 \\ D_1 & E_1 & F_1 & P_1 \\ G_1 & H_1 & I_1 & Q_1 \\ R_1 & S_1 & T_1 & W_1 \end{bmatrix}, \Delta_2 := \begin{bmatrix} A_2 & B_2 & C_2 & O_2 \\ D_2 & E_2 & F_2 & P_2 \\ G_2 & H_2 & I_2 & Q_2 \\ R_2 & S_2 & T_2 & W_2 \end{bmatrix}, \Delta_3 := \begin{bmatrix} A_3 & B_3 & C_3 & O_3 \\ D_3 & E_3 & F_3 & P_3 \\ G_3 & H_3 & I_3 & Q_3 \\ R_3 & S_3 & T_3 & W_3 \end{bmatrix}.$$

Consider

$$\begin{bmatrix} A_1 & B_1 & C_1 & O_1 \\ D_1 & E_1 & F_1 & P_1 \\ G_1 & H_1 & I_1 & Q_1 \\ R_1 & S_1 & T_1 & W_1 \end{bmatrix} \begin{bmatrix} f_{g,t} \\ f_{g,t-1} \\ f_{g,t-2} \\ f_{C,t} \end{bmatrix} = \begin{bmatrix} \tilde{f}_{g,t} \\ \tilde{f}_{g,t-1} \\ \tilde{f}_{g,t-2} \\ \tilde{f}_{C,t} \end{bmatrix} \tag{A.1}$$

$$\begin{bmatrix} A_3 & B_3 & C_3 & O_3 \\ D_3 & E_3 & F_3 & P_3 \\ G_3 & H_3 & I_3 & Q_3 \\ R_3 & S_3 & T_3 & W_3 \end{bmatrix} \begin{bmatrix} f_{g,t-1} \\ f_{g,t-2} \\ f_{g,t-3} \\ f_{C,t-1} \end{bmatrix} = \begin{bmatrix} \tilde{f}_{g,t-1} \\ \tilde{f}_{g,t-2} \\ \tilde{f}_{g,t-3} \\ \tilde{f}_{C,t-1} \end{bmatrix} \tag{A.2}$$

$$\begin{bmatrix} A_2 & B_2 & C_2 & O_2 \\ D_2 & E_2 & F_2 & P_2 \\ G_2 & H_2 & I_2 & Q_2 \\ R_2 & S_2 & T_2 & W_2 \end{bmatrix} \begin{bmatrix} f_{g,t-2} \\ f_{g,t-3} \\ f_{g,t-4} \\ f_{C,t-2} \end{bmatrix} = \begin{bmatrix} \tilde{f}_{g,t-2} \\ \tilde{f}_{g,t-3} \\ \tilde{f}_{g,t-4} \\ \tilde{f}_{C,t-2} \end{bmatrix}. \tag{A.3}$$

Considering (A.1) and (A.2), we have

$$\tilde{f}_{g,t-1} = D_1 f_{g,t} + E_1 f_{g,t-1} + F_1 f_{g,t-2} + P_1 f_{C,t} = A_3 f_{g,t-1} + B_3 f_{g,t-2} + C_3 f_{g,t-3} + O_3 f_{C,t-1}$$

whence we have

$$0=\left[D_1\phi^3+(E_1-A_3)\phi^2+(F_1-B_3)\phi-C_3\right]f_{g,t-3}+\left[D_1\phi^2+(E_1-A_3)\phi+(F_1-B_3)\right]\eta_{g,t-3}\\ +\left[D_1\phi+(E_1-A_3)\right]\eta_{g,t-2}+D_1\eta_{g,t-1}+P_1\eta_{C,t-1}-O_3\eta_{C,t-2}.$$

Note that each of $\eta_{g,t-1},\eta_{C,t-1},\eta_{C,t-2}$ is uncorrelated with any other term on the right hand side of the preceding display. We necessarily have $D_1\eta_{g,t-1}=0$, $P_1\eta_{C,t-1}=0$ and $O_3\eta_{C,t-2}=0$ because of the non-zero variance. Equivalently, we have $D_1=P_1=O_3=0$. Likewise, we deduce that $E_1=A_3$, $F_1=B_3$ and $C_3=0$. Next, note that

$$\tilde{f}_{g,t-2}=G_1f_{g,t}+H_1f_{g,t-1}+I_1f_{g,t-2}+Q_1f_{C,t}=D_3f_{g,t-1}+E_3f_{g,t-2}+F_3f_{g,t-3}+P_3f_{C,t-1}$$

whence we have

$$0=\left[G_1\phi^3+(H_1-D_3)\phi^2+(I_1-E_3)\phi-F_3\right]f_{g,t-3}+\left[G_1\phi^2+(H_1-D_3)\phi+(I_1-E_3)\right]\eta_{g,t-3}\\ \left[G_1\phi+(H_1-D_3)\right]\eta_{g,t-2}+G_1\eta_{g,t-1}+Q_1\eta_{C,t-1}-P_3\eta_{C,t-2}.$$

Using a similar trick, we deduce that

$$G_1=Q_1=P_3=0,\qquad H_1=D_3=0,\qquad I_1=E_3,\qquad F_3=0.$$

The rotation matrices Δ_1,Δ_3 are deduced to

$$\Delta_1:=\begin{bmatrix}A_1&B_1&C_1&O_1\\0&E_1&F_1&0\\0&H_1&I_1&0\\R_1&S_1&T_1&W_1\end{bmatrix}\tag{A.4}$$

$$\Delta_3:=\begin{bmatrix}E_1&F_1&0&0\\H_1&I_1&0&0\\G_3&H_3&I_3&Q_3\\R_3&S_3&T_3&W_3\end{bmatrix}.\tag{A.5}$$

Applying the trick to (A.2) and (A.3), we deduce

$$\Delta_3:=\begin{bmatrix}A_3&B_3&C_3&O_3\\0&E_3&F_3&0\\0&H_3&I_3&0\\R_3&S_3&T_3&W_3\end{bmatrix}\tag{A.6}$$

$$\Delta_2:=\begin{bmatrix}E_3&F_3&0&0\\H_3&I_3&0&0\\G_2&H_2&I_2&Q_2\\R_2&S_2&T_2&W_2\end{bmatrix}.\tag{A.7}$$

Applying the trick to (A.3) and (A.1), we deduce

$$\Delta_2 := \begin{bmatrix} A_2 & B_2 & C_2 & O_2 \\ 0 & E_2 & F_2 & 0 \\ 0 & H_2 & I_2 & 0 \\ R_2 & S_2 & T_2 & W_2 \end{bmatrix} \tag{A.8}$$

$$\Delta_1 := \begin{bmatrix} E_2 & F_2 & 0 & 0 \\ H_2 & I_2 & 0 & 0 \\ G_1 & H_1 & I_1 & Q_1 \\ R_1 & S_1 & T_1 & W_1 \end{bmatrix}. \tag{A.9}$$

Comparing (A.4) and (A.9), we have

$$A_1 = E_2, \quad B_1 = F_2, \quad H_2 = 0, \quad I_2 = E_1, \quad G_1 = 0, \quad F_1 = 0.$$

Comparing (A.5) and (A.6), we have

$$H_1 = 0, \quad F_3 = 0, \quad I_3 = E_2 = A_1.$$

Comparing (A.7) and (A.8), we have

$$A_2 = E_3 = I_1, \quad B_2 = F_3 = 0, \quad E_2 = A_1, \quad I_2 = E_1, \quad H_3 = 0, \quad F_2 = 0.$$

Thus the rotation matrices $\Delta_1, \Delta_2, \Delta_3$ are reduced to

$$\Delta_1 := \begin{bmatrix} A_1 & 0 & 0 & 0 \\ 0 & E_1 & 0 & 0 \\ 0 & 0 & I_1 & 0 \\ R_1 & S_1 & T_1 & W_1 \end{bmatrix}, \Delta_2 := \begin{bmatrix} I_1 & 0 & 0 & 0 \\ 0 & A_1 & 0 & 0 \\ 0 & 0 & E_1 & 0 \\ R_2 & S_2 & T_2 & W_2 \end{bmatrix}, \Delta_3 := \begin{bmatrix} E_1 & 0 & 0 & 0 \\ 0 & I_1 & 0 & 0 \\ 0 & 0 & A_1 & 0 \\ R_3 & S_3 & T_3 & W_3 \end{bmatrix}.$$

Note that

$$R_1 f_{g,t} + S_1 f_{g,t-1} + T_1 f_{g,t-2} + W_1 f_{C,t} = \tilde{f}_{c,t},$$

whence we have

$$\left(R_1\phi^3 + S_1\phi^2 + T_1\phi\right) f_{g,t-3} + \left(R_1\phi^2 + S_1\phi + T_1\right)\eta_{g,t-3} + (R_1\phi + S_1)\eta_{g,t-2} + R_1\eta_{g,t-1} + W_1\eta_{C,t-1} \\ = \tilde{\eta}_{c,t-1}.$$

Since $\eta_{g,t}$ and $\eta_{C,t}$ are uncorrelated, we have $R_1 = S_1 = T_1 = 0$. Then Δ_1 is reduced to

$$\Delta_1 = \begin{bmatrix} A_1 & 0 & 0 & 0 \\ 0 & E_1 & 0 & 0 \\ 0 & 0 & I_1 & 0 \\ 0 & 0 & 0 & W_1 \end{bmatrix}.$$

Applying the similar trick, we have

$$\Delta_2 := \begin{bmatrix} I_1 & 0 & 0 & 0 \\ 0 & A_1 & 0 & 0 \\ 0 & 0 & E_1 & 0 \\ 0 & 0 & 0 & W_2 \end{bmatrix}, \qquad \Delta_3 := \begin{bmatrix} E_1 & 0 & 0 & 0 \\ 0 & I_1 & 0 & 0 \\ 0 & 0 & A_1 & 0 \\ 0 & 0 & 0 & W_3 \end{bmatrix}.$$

We have

$$
\begin{aligned}
\tilde{f}_{g,t} &= A_1 f_{g,t} = A_1(\phi f_{g,t-1} + \eta_{g,t-1}) = A_1\phi f_{g,t-1} + A_1\eta_{g,t-1} \qquad \operatorname{var}(A_1\eta_{g,t-1}) = 1\\
\tilde{f}_{g,t-1} &= E_1 f_{g,t-1} = E_1(\phi f_{g,t-2} + \eta_{g,t-2}) = E_1\phi f_{g,t-2} + E_1\eta_{g,t-2} \qquad \operatorname{var}(E_1\eta_{g,t-2}) = 1\\
\tilde{f}_{g,t-2} &= I_1 f_{g,t-2} = I_1(\phi f_{g,t-3} + \eta_{g,t-3}) = I_1\phi f_{g,t-3} + I_1\eta_{g,t-3} \qquad \operatorname{var}(I_1\eta_{g,t-3}) = 1\\
\tilde{f}_{c,t} &= W_1 f_{C,t} = W_1\eta_{C,t-1} \qquad \operatorname{var}(W_1\eta_{C,t-1}) = 1\\
\tilde{f}_{c,t-1} &= W_3 f_{C,t-1} = W_3\eta_{C,t-2} \qquad \operatorname{var}(W_3\eta_{C,t-2}) = 1\\
\tilde{f}_{c,t-2} &= W_2 f_{C,t-2} = W_2\eta_{C,t-3} \qquad \operatorname{var}(W_2\eta_{C,t-3}) = 1
\end{aligned}
$$

We hence deduce that $A_1 = \pm 1$, $E_1 = \pm 1$, $I_1 = \pm 1$ and $W_i = \pm 1$ for $i = 1, 2, 3$. Requiring that estimators of $\boldsymbol{z}_0^A, \boldsymbol{z}_1^A, \boldsymbol{z}_2^A, \boldsymbol{z}_3^A, \boldsymbol{z}_3^E, \boldsymbol{z}_3^U$ have the same column signs as those of $\boldsymbol{z}_0^A, \boldsymbol{z}_1^A, \boldsymbol{z}_2^A, \boldsymbol{z}_3^A, \boldsymbol{z}_3^E, \boldsymbol{z}_3^U$ ensures that $A_1 = 1$, $E_1 = 1$, $I_1 = 1$ and $W_i = 1$ for $i = 1, 2, 3$. Thus $\Delta_1, \Delta_2, \Delta_3$ are reduced to identity matrices. Note that the proof works for both $\phi = 0$ and $\phi \neq 0$. $\square$

A.2 Bayesian Estimation Using the Gibbs Sampling

In this subsection, we outline a Bayesian procedure (i.e., the Gibbs sampling) to estimate the factor model (2.3). The Gibbs sampling consists of the following steps:

1. Get the starting values of the model parameters for the Gibbs sampling.
2. Conditional on the model parameters and the observed data, draw the factors.
3. Conditional on the factors and the observed data, draw the model parameters.
4. Return to step 2.

After a large number of steps, the collection of drawn factors will give the posterior distribution of the factors given the data, while the collection of drawn parameters will give the posterior distribution of the parameters given the data. We shall now explain steps 2-3 in detail.

A.2.1 Step 2

Let $Y_{1:t} := \{\boldsymbol{y}_t^\intercal, \dots, \boldsymbol{y}_1^\intercal\}^\intercal$ and $Y_0 = \emptyset$. For

$$
\boldsymbol{\alpha}_t = \begin{bmatrix} f_{g,t} \\ f_{g,t-1} \\ f_{g,t-2} \\ f_{C,t} \end{bmatrix}, \qquad \boldsymbol{\alpha}_{t+1} := \begin{bmatrix} f_{g,t+1} \\ f_{g,t} \\ f_{g,t-1} \\ f_{C,t+1} \end{bmatrix},
$$

only the last 2 elements of $\boldsymbol{\alpha}_t$ are unknown, i.e., $f_{g,t-2}, f_{C,t}$, once $\boldsymbol{\alpha}_{t+1}$ is given. We want to sample $\boldsymbol{\alpha}_t$ from the joint distribution

$$
\begin{aligned}
&p(\boldsymbol{\alpha}_T, \boldsymbol{\alpha}_{T-1}, \dots, \boldsymbol{\alpha}_1 | Y_{1:T})\\
&= p(\boldsymbol{\alpha}_T | Y_{1:T}) p(\boldsymbol{\alpha}_{T-1} | \boldsymbol{\alpha}_T, Y_{1:T}) p(\boldsymbol{\alpha}_{T-2} | \boldsymbol{\alpha}_{T-1}, \boldsymbol{\alpha}_T, Y_{1:T}) \cdots p(\boldsymbol{\alpha}_1 | \boldsymbol{\alpha}_2, \dots, \boldsymbol{\alpha}_T, Y_{1:T})\\
&= p(\boldsymbol{\alpha}_T | Y_{1:T}) \prod_{t=1}^{T-1} p(\boldsymbol{\alpha}_t | \boldsymbol{\alpha}_{t+1}, \dots, \boldsymbol{\alpha}_T, Y_{1:T}) = p(\boldsymbol{\alpha}_T | Y_{1:T}) \prod_{t=1}^{T-1} p(\boldsymbol{\alpha}_t | \boldsymbol{\alpha}_{t+1}, Y_{1:t})
\end{aligned}
$$

where the last equality is due to the Markov structure of the state space system (Carter and Kohn (1994) Lemma 2.1).

We first sample

$$\boldsymbol{\alpha}_T|Y_{1:T} \sim N(a_{T|T}, P_{T|T})$$

where $a_{T|T} := \mathbb{E}[\boldsymbol{\alpha}_T|Y_{1:T}]$ and $P_{T|T} := \mathrm{var}(\boldsymbol{\alpha}_T|Y_{1:T})$. These are obtained from Kalman filter (please refer to Section A.3 for details). Then we may sample $\boldsymbol{\alpha}_t$ from $p(\boldsymbol{\alpha}_t|\boldsymbol{\alpha}_{t+1}, Y_{1:t})$, $t = T-1, \ldots, 3$. Given $\boldsymbol{\alpha}_{t+1}$, only the last 2 elements of $\boldsymbol{\alpha}_t$, i.e., $f_{g,t-2}, f_{C,t}$, are random, which can be drawn from $p(f_{g,t-2}, f_{C,t}|\boldsymbol{\alpha}_{t+1}, Y_{1:t})$ for $t = T-1, \ldots, 3$. The conditional density $p(f_{g,t-2}, f_{C,t}|\boldsymbol{\alpha}_{t+1}, Y_{1:t})$ can be written as

$$\begin{aligned} &p(f_{g,t-2}, f_{C,t}|\boldsymbol{\alpha}_{t+1}, Y_{1:t}) = p(f_{g,t-2}, f_{C,t}|f_{g,t+1}, f_{g,t}, f_{g,t-1}, f_{C,t+1}, Y_{1:t}) \\ &= p(f_{g,t-2}|f_{C,t}, f_{g,t+1}, f_{g,t}, f_{g,t-1}, f_{C,t+1}, Y_{1:t})p(f_{C,t}|f_{g,t+1}, f_{g,t}, f_{g,t-1}, f_{C,t+1}, Y_{1:t}). \end{aligned} \tag{A.10}$$

We consider the first term of (A.10).

$$\begin{aligned} &p(f_{g,t-2}|f_{C,t}, f_{g,t+1}, f_{g,t}, f_{g,t-1}, f_{C,t+1}, Y_{1:t}) = \frac{p(f_{g,t-2}, f_{g,t+1}|f_{C,t}, f_{g,t}, f_{g,t-1}, f_{C,t+1}, Y_{1:t})}{p(f_{g,t+1}|f_{C,t}, f_{g,t}, f_{g,t-1}, f_{C,t+1}, Y_{1:t})} \\ &\propto p(f_{g,t-2}, f_{g,t+1}|f_{C,t}, f_{g,t}, f_{g,t-1}, f_{C,t+1}, Y_{1:t}) \\ &= p(f_{g,t+1}|f_{g,t-2}, f_{C,t}, f_{g,t}, f_{g,t-1}, f_{C,t+1}, Y_{1:t})p(f_{g,t-2}|f_{C,t}, f_{g,t}, f_{g,t-1}, f_{C,t+1}, Y_{1:t}) \end{aligned} \tag{A.11}$$

Consider the second term of (A.11) first.

$$p(f_{g,t-2}|f_{C,t}, f_{g,t}, f_{g,t-1}, f_{C,t+1}, Y_{1:t}) = p(f_{g,t-2}|f_{C,t}, f_{g,t}, f_{g,t-1}, Y_{1:t}).$$

Note that we can obtain the distribution $\boldsymbol{\alpha}_t|Y_{1:t}$ via Kalman filter:

$$\left.\begin{bmatrix} f_{g,t} \\ f_{g,t-1} \\ f_{g,t-2} \\ f_{C,t} \end{bmatrix}\right|_{Y_{1:t}} = \boldsymbol{\alpha}_t|Y_{1:t} \sim N(\boldsymbol{a}_{t|t}, P_{t|t}) =: N\left(\begin{bmatrix} a_{t|t,1} \\ a_{t|t,2} \\ a_{t|t,3} \\ a_{t|t,4} \end{bmatrix} \begin{bmatrix} P_{t|t,11} & P_{t|t,12} & P_{t|t,13} & P_{t|t,14} \\ P_{t|t,21} & P_{t|t,22} & P_{t|t,23} & P_{t|t,24} \\ P_{t|t,31} & P_{t|t,32} & P_{t|t,33} & P_{t|t,34} \\ P_{t|t,41} & P_{t|t,42} & P_{t|t,43} & P_{t|t,44} \end{bmatrix}\right),$$

for $t = T-1, \ldots, 3$. Reshuffle the elements of $\boldsymbol{\alpha}_t$ a little bit:

$$\left.\begin{bmatrix} f_{C,t} \\ f_{g,t} \\ f_{g,t-1} \\ f_{g,t-2} \end{bmatrix}\right|_{Y_{1:t}} \sim N(P_e\boldsymbol{a}_{t|t}, P_eP_{t|t}P_e^{\intercal}) = N\left(\begin{bmatrix} a_{t|t,4} \\ a_{t|t,1} \\ a_{t|t,2} \\ a_{t|t,3} \end{bmatrix} \begin{bmatrix} P_{t|t,44} & P_{t|t,41} & P_{t|t,42} & P_{t|t,43} \\ P_{t|t,14} & P_{t|t,11} & P_{t|t,12} & P_{t|t,13} \\ P_{t|t,24} & P_{t|t,21} & P_{t|t,22} & P_{t|t,23} \\ P_{t|t,34} & P_{t|t,31} & P_{t|t,32} & P_{t|t,33} \end{bmatrix}\right), \tag{A.12}$$

where P_e is a permutation matrix

$$P_e = \begin{bmatrix} 0 & 0 & 0 & 1 \\ 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \end{bmatrix}.$$

Partition (A.12) accordingly:

$$\left[\begin{array}{c} f_{C,t} \\ f_{g,t} \\ f_{g,t-1} \\ f_{g,t-2} \end{array}\right]\Bigg|_{Y_{1:t}} \sim N\left(\left(\begin{array}{c} \boldsymbol{a}_{t|t,a} \\ a_{t|t,b} \end{array}\right)\left[\begin{array}{cc} P_{t|t,aa} & P_{t|t,ab} \\ P_{t|t,ba} & P_{t|t,bb} \end{array}\right]\right)$$

where $\boldsymbol{a}_{t|t,a} := (a_{t|t,4}, a_{t|t,1}, a_{t|t,2})^\intercal$, $a_{t|t,b} := a_{t|t,3}$, $P_{t|t,bb} := P_{t|t,33}$, $P_{t|t,ab} := (P_{t|t,43}, P_{t|t,13}, P_{t|t,23})^\intercal$, $P_{t|t,ba} = P_{t|t,ab}^\intercal$ and

$$P_{t|t,aa} := \left(\begin{array}{ccc} P_{t|t,44} & P_{t|t,41} & P_{t|t,42} \\ P_{t|t,14} & P_{t|t,11} & P_{t|t,12} \\ P_{t|t,24} & P_{t|t,21} & P_{t|t,22} \end{array}\right).$$

Then

$$f_{g,t-2}|f_{C,t}, f_{g,t}, f_{g,t-1}, Y_{1:t} \sim N(c_t, C_t) \qquad t = T-1, \dots, 3$$

where

$$\begin{aligned} c_t &= a_{t|t,b} + P_{t|t,ba} P_{t|t,aa}^{-1} \left[\left(\begin{array}{c} f_{C,t} \\ f_{g,t} \\ f_{g,t-1} \end{array}\right) - \boldsymbol{a}_{t|t,a}\right] \\ C_t &= P_{t|t,bb} - P_{t|t,ba} P_{t|t,aa}^{-1} P_{t|t,ab}. \end{aligned}$$

We now consider the first term of (A.11). Note that $f_{g,t+1} = \phi f_{g,t} + \eta_{g,t}$, $\eta_{g,t} \sim N(0,1)$. We hence have $f_{g,t+1}|f_{g,t-2}, f_{C,t}, f_{g,t}, f_{g,t-1}, f_{C,t+1}, Y_{1:t} \sim N(\phi f_{g,t}, 1)$, which is independent of $f_{g,t-2}$. We hence have

$$p(f_{g,t-2}|f_{C,t}, f_{g,t+1}, f_{g,t}, f_{g,t-1}, f_{C,t+1}, Y_{1:t}) \propto N(c_t, C_t). \tag{A.13}$$

Consider the second term of (A.10).

$$p(f_{C,t}|f_{g,t+1}, f_{g,t}, f_{g,t-1}, f_{C,t+1}, Y_{1:t}) = p(f_{C,t}|f_{g,t+1}, f_{g,t}, f_{g,t-1}, Y_{1:t}) = p(f_{C,t}|f_{g,t}, f_{g,t-1}, Y_{1:t}).$$

Using (A.12), we have

$$\left[\begin{array}{c} f_{C,t} \\ f_{g,t} \\ f_{g,t-1} \end{array}\right]\Bigg|_{Y_{1:t}} \sim N\left(\left[\begin{array}{c} a_{t|t,4} \\ a_{t|t,1} \\ a_{t|t,2} \end{array}\right]\left[\begin{array}{ccc} P_{t|t,44} & P_{t|t,41} & P_{t|t,42} \\ P_{t|t,14} & P_{t|t,11} & P_{t|t,12} \\ P_{t|t,24} & P_{t|t,21} & P_{t|t,22} \end{array}\right]\right) =: N\left(\left(\begin{array}{c} a_{t|t,c} \\ \boldsymbol{a}_{t|t,d} \end{array}\right)\left[\begin{array}{cc} P_{t|t,cc} & P_{t|t,cd} \\ P_{t|t,dc} & P_{t|t,dd} \end{array}\right]\right),$$

where $a_{t|t,c} = a_{t|t,4}$, $\boldsymbol{a}_{t|t,d} := (a_{t|t,1}, a_{t|t,2})^\intercal$, $P_{t|t,cc} := P_{t|t,44}$, $P_{t|t,cd} := (P_{t|t,41}, P_{t|t,42})$, $P_{t|t,dc} = P_{t|t,cd}^\intercal$, and

$$P_{t|t,dd} := \left(\begin{array}{cc} P_{t|t,11} & P_{t|t,12} \\ P_{t|t,21} & P_{t|t,22} \end{array}\right).$$

Then

$$f_{C,t}|f_{g,t}, f_{g,t-1}, Y_{1:t} \sim N(d_t, D_t) \qquad t = T-1, \dots, 3 \tag{A.14}$$

where

$$\begin{aligned} d_t &= a_{t|t,c} + P_{t|t,cd} P_{t|t,dd}^{-1} \left[\begin{pmatrix} f_{g,t} \\ f_{g,t-1} \end{pmatrix} - \boldsymbol{a}_{t|t,d} \right] \\ D_t &= P_{t|t,cc} - P_{t|t,cd} P_{t|t,dd}^{-1} P_{t|t,dc}. \end{aligned}$$

In sum, conditional on the model's parameters, the following recursion describes how to draw from $p(\boldsymbol{\alpha}_T, \boldsymbol{\alpha}_{T-1}, \dots, \boldsymbol{\alpha}_1 | Y_{1:T})$:

(a) We first sample

$$\boldsymbol{\alpha}_T | Y_{1:T} \sim N(\boldsymbol{a}_{T|T}, P_{T|T})$$

where $\boldsymbol{a}_{T|T} := \mathbb{E}[\boldsymbol{\alpha}_T | Y_{1:T}]$ and $P_{T|T} := \text{var}(\boldsymbol{\alpha}_T | Y_{1:T})$.

(b) Consider $t = T-1, \dots, 3$. For each fixed t, first draw $f_{C,t}$ from $N(d_t, D_t)$ as in (A.14), and then draw $f_{g,t-2}$ from $N(c_t, C_t)$ as in (A.13).

(c) For $t = 2$, draw $f_{C,2}$ from $N(d_2, D_2)$ as in (A.14).

(d) For $t = 1$, note that

$$\left. \begin{bmatrix} f_{C,1} \\ f_{g,1} \end{bmatrix} \right|_{Y_1} \sim N \left(\begin{pmatrix} a_{1|1,4} \\ a_{1|1,1} \end{pmatrix} \begin{bmatrix} P_{1|1,44} & P_{1|1,41} \\ P_{1|1,14} & P_{1|1,11} \end{bmatrix} \right),$$

whence we have

$$f_{C,1}|f_{g,1}, f_{g,0}, Y_1 = f_{C,1}|f_{g,1}, Y_1 \sim N(d_1, D_1),$$

where

$$\begin{aligned} d_1 &= a_{1|1,4} + P_{1|1,41} P_{1|1,11}^{-1} (f_{g,1} - a_{1|1,1}) \\ D_1 &= P_{1|1,44} - P_{1|1,41} P_{1|1,11}^{-1} P_{1|1,14}. \end{aligned}$$

Thus draw $f_{C,1}$ from $N(d_1, D_1)$.

Steps (a)-(d) finish one round of sampling from $p(\boldsymbol{\alpha}_T, \boldsymbol{\alpha}_{T-1}, \dots, \boldsymbol{\alpha}_1 | Y_{1:T})$.

A.2.2 Step 3

In this subsubsection, we consider how to draw model parameters conditional on the factors and observed data. We first provide two auxiliary results.

Lemma A.1.

$$tr(Z^\intercal B Z C) = (vec Z)^\intercal (C^\intercal \otimes B) vec Z.$$

Proposition A.1. *If a positive random variable X has a density that is proportional to*

$$x^{-a-1}e^{-\frac{1}{bx}},$$

then $X \sim$ inverse-gamma(a,b) and $Y := 1/X \sim$ gamma(a,b). The matlab syntax is

y = random('Gamma',a,b)

x = 1/y

Draw $Z^A, Z^E, Z^U, \Sigma_A, \Sigma_E, \Sigma_U$ We first consider how to draw $Z^A, Z^E, Z^U, \Sigma_A, \Sigma_E, \Sigma_U$. Consider

$$\begin{aligned} \underbrace{\boldsymbol{y}_t^A}_{N_A\times 1} &= Z_t\boldsymbol{\alpha}_t + \boldsymbol{e}_t^A = Z^A\boldsymbol{\alpha}_t + \boldsymbol{e}_t^A \qquad t \in T_A \\ \boldsymbol{y}_t^E &= Z_t\boldsymbol{\alpha}_t + \boldsymbol{e}_t^E = Z^E\boldsymbol{\alpha}_t + \boldsymbol{e}_t^E \qquad t \in T_E \\ \boldsymbol{y}_t^U &= Z_t\boldsymbol{\alpha}_t + \boldsymbol{e}_t^U = Z^U\boldsymbol{\alpha}_t + \boldsymbol{e}_t^U \qquad t \in T_U. \end{aligned} \tag{A.15}$$

We shall use (A.15) to illustrate the procedure. We assume that little is known a priori about Z^A and Σ_A. We hence use Jeffrey's priors:

$$p(Z^A) = \text{constant} \qquad\qquad p(\Sigma_A) \propto |\Sigma_A|^{-1}.$$

Moreover, we assume that Z^A and Σ_A are independent, so

$$\begin{aligned} &p(Z^A, \Sigma_A) = p(Z^A)p(\Sigma_A) \propto |\Sigma_A|^{-1} \\ &p(Z^A|\Sigma_A) \propto 1. \end{aligned}$$

Stacking (A.15), we have

$$\underbrace{Y^A}_{\frac{T}{3}\times N_A} = \begin{bmatrix} \boldsymbol{y}_1^{A\intercal} \\ \boldsymbol{y}_4^{A\intercal} \\ \boldsymbol{y}_7^{A\intercal} \\ \vdots \\ \boldsymbol{y}_{T-2}^{A\intercal} \end{bmatrix} = \begin{bmatrix} \boldsymbol{\alpha}_1^\intercal \\ \boldsymbol{\alpha}_4^\intercal \\ \boldsymbol{\alpha}_7^\intercal \\ \vdots \\ \boldsymbol{\alpha}_{T-2}^\intercal \end{bmatrix} Z^{A\intercal} + \begin{bmatrix} \boldsymbol{e}_1^{A\intercal} \\ \boldsymbol{e}_4^{A\intercal} \\ \boldsymbol{e}_7^{A\intercal} \\ \vdots \\ \boldsymbol{e}_{T-2}^{A\intercal} \end{bmatrix} =: \underbrace{\Xi_A}_{\frac{T}{3}\times 4} \underbrace{Z^{A\intercal}}_{4\times N_A} + \underbrace{E_A}_{\frac{T}{3}\times N_A} \tag{A.16}$$

Suppose that we observe $\{\boldsymbol{\alpha}_t\}_{t=1}^T$. How do we estimate $Z^{A\intercal}$? The OLS estimate would be

$$\ddot{Z}^{A\intercal} = \left(\sum_{t\in T_A} \boldsymbol{\alpha}_t\boldsymbol{\alpha}_t^\intercal\right)^{-1} \sum_{t\in T_A} \boldsymbol{\alpha}_t \boldsymbol{y}_t^{A\intercal} = (\Xi_A^\intercal\Xi_A)^{-1}\Xi_A^\intercal Y^A.$$

We can calculate

$$\begin{aligned} p(Y^A|\Xi_A, Z^A, \Sigma_A) &= \prod_{t\in T_A} \frac{1}{(2\pi)^{\frac{N_A}{2}}|\Sigma_A|^{1/2}} \exp\left[-\frac{1}{2}(\boldsymbol{y}_t^A - Z^A\boldsymbol{\alpha}_t)^\intercal\Sigma_A^{-1}(\boldsymbol{y}_t^A - Z^A\boldsymbol{\alpha}_t)\right] \\ &= (2\pi)^{-\frac{N_AT}{6}}|\Sigma_A|^{-\frac{T}{6}} \exp\left[-\frac{1}{2}\operatorname{tr}\left(\sum_{t\in T_A}(\boldsymbol{y}_t^A - Z^A\boldsymbol{\alpha}_t)(\boldsymbol{y}_t^A - Z^A\boldsymbol{\alpha}_t)^\intercal\Sigma_A^{-1}\right)\right] \\ &= (2\pi)^{-\frac{N_AT}{6}}|\Sigma_A|^{-\frac{T}{6}} \exp\left[-\frac{1}{2}\operatorname{tr}\left(\sum_{t\in T_A}(\boldsymbol{y}_t^A - \ddot{Z}^A\boldsymbol{\alpha}_t)(\boldsymbol{y}_t^A - \ddot{Z}^A\boldsymbol{\alpha}_t)^\intercal\Sigma_A^{-1}\right)\right] \\ &\quad \cdot \exp\left[-\frac{1}{2}\operatorname{tr}\left((\ddot{Z}^A - Z^A)\Xi_A^\intercal\Xi_A(\ddot{Z}^A - Z^A)^\intercal\Sigma_A^{-1}\right)\right] \\ &= (2\pi)^{-\frac{N_AT}{6}}|\Sigma_A|^{-\frac{T}{6}} \exp\left[-\frac{1}{2}\operatorname{tr}\left(S_A\Sigma_A^{-1}\right)\right] \cdot \exp\left[-\frac{1}{2}\operatorname{tr}\left((\ddot{Z}^A - Z^A)\Xi_A^\intercal\Xi_A(\ddot{Z}^A - Z^A)^\intercal\Sigma_A^{-1}\right)\right], \end{aligned}$$

where

$$S_A := \sum_{t\in T_A}(\boldsymbol{y}_t^A - \ddot{Z}^A\boldsymbol{\alpha}_t)(\boldsymbol{y}_t^A - \ddot{Z}^A\boldsymbol{\alpha}_t)^\intercal = \ddot{E}_A^\intercal \ddot{E}_A = (Y^A - \Xi_A\ddot{Z}^{A\intercal})^\intercal(Y^A - \Xi_A\ddot{Z}^{A\intercal}).$$

We can hence calculate the posterior

$$\begin{aligned}
&p(Z^A,\Sigma_A|Y^A,\Xi_A) \propto p(Z^A,\Sigma_A,Y^A|\Xi_A) = p(Z^A,\Sigma_A|\Xi_A)p(Y^A|Z^A,\Sigma_A,\Xi_A)\\
&\propto p(Z^A,\Sigma_A)p(Y^A|Z^A,\Sigma_A,\Xi_A)\\
&\propto |\Sigma_A|^{-\frac{T/3+2}{2}}\exp\left[-\frac{1}{2}\operatorname{tr}\left(S_A\Sigma_A^{-1}\right)\right]\cdot\exp\left[-\frac{1}{2}\operatorname{tr}\left((\ddot{Z}^A - Z^A)\Xi_A^\intercal\Xi_A(\ddot{Z}^A - Z^A)^\intercal\Sigma_A^{-1}\right)\right],
\end{aligned}$$

where the second $\propto$ is due to that knowing Ξ_A does not shed any information on Z^A,Σ_A.

Viewing the preceding display as a prior probability density function (pdf), we see that it factors into a normal part for Z^A given Σ_A and a marginal pdf for Σ_A:[6]

$$\begin{aligned}
p(Z^A|\Sigma_A,Y^A,\Xi_A) &\propto |\Sigma_A|^{-\frac{4}{2}}\exp\left[-\frac{1}{2}\operatorname{tr}\left((\ddot{Z}^A - Z^A)\Xi_A^\intercal\Xi_A(\ddot{Z}^A - Z^A)^\intercal\Sigma_A^{-1}\right)\right]\\
&= |\Sigma_A|^{-\frac{4}{2}}\exp\left[-\frac{1}{2}\left[\operatorname{vec}(\ddot{Z}^{A\intercal} - Z^{A\intercal})\right]^\intercal\left(\Sigma_A^{-1}\otimes\Xi_A^\intercal\Xi_A\right)\left[\operatorname{vec}(\ddot{Z}^{A\intercal} - Z^{A\intercal})\right]\right]
\end{aligned}$$

via Lemma A.1 and

$$\begin{aligned}
p(\Sigma_A|Y^A,\Xi_A) &\propto |\Sigma_A|^{-\frac{T/3+2-4}{2}}\exp\left[-\frac{1}{2}\operatorname{tr}\left(S_A\Sigma_A^{-1}\right)\right] = \prod_{i=1}^{N_A}(\sigma_{A,i}^2)^{-\frac{T/3+2-4}{2}}\exp\left[-\frac{1}{2}\frac{S_{A,ii}}{\sigma_{A,i}^2}\right]\\
&= \prod_{i=1}^{N_A}(\sigma_{A,i}^2)^{-\frac{T/3-4}{2}-1}\exp\left[-\frac{1}{\sigma_{A,i}^2\cdot(2/S_{A,ii})}\right]
\end{aligned}$$

whence we have

$$\begin{aligned}
\operatorname{vec}(Z^{A\intercal})|\Sigma_A,Y^A,\Xi_A &\sim N\left(\operatorname{vec}(\ddot{Z}^{A\intercal}),\Sigma_A\otimes(\Xi_A^\intercal\Xi_A)^{-1}\right)\\
p(\sigma_{A,i}^2|Y^A,\Xi_A) &\propto (\sigma_{A,i}^2)^{-\frac{T/3-4}{2}-1}\exp\left[-\frac{1}{\sigma_{A,i}^2(2/S_{A,ii})}\right]
\end{aligned}$$

for $i = 1,\ldots,N_A$. Thus,

$$\begin{aligned}
\sigma_{A,i}^2|Y^A,\Xi_A &\sim \text{inverse-gamma}\left(\frac{T/3-4}{2},\frac{2}{S_{A,ii}}\right)\\
\sigma_{A,i}^{-2}|Y^A,\Xi_A &\sim \text{gamma}\left(\frac{T/3-4}{2},\frac{2}{S_{A,ii}}\right).
\end{aligned}$$

via Proposition A.1.

[6]To see why the exponent of $|\Sigma_A|$ in $p(Z^A|\Sigma_A,Y^A,\Xi_A)$ is $-4/2$: The covariance matrix of $\operatorname{vec}(Z^{A\intercal})$ is $\Sigma_A\otimes(\Xi_A^\intercal\Xi_A)^{-1}$. When calculating $p(Z^A|\Sigma_A,Y^A,\Xi_A)$, we have a term:

$$\frac{1}{\left|\Sigma_A\otimes(\Xi_A^\intercal\Xi_A)^{-1}\right|^{\frac{1}{2}}} = \left|\Sigma_A\otimes(\Xi_A^\intercal\Xi_A)^{-1}\right|^{-\frac{1}{2}} \propto |\Sigma_A|^{-\frac{4}{2}}.$$

Draw ϕ We now consider how to draw ϕ conditional on the factors and observed data. Recall (2.2): $f_{g,t+1} = \phi f_{g,t} + \eta_{g,t}$. Define

$$\ddot{\phi} := \left(\sum_{t=1}^{T-1} f_{g,t}^2 \right)^{-1} \sum_{t=1}^{T-1} f_{g,t} f_{g,t+1}.$$

Recalling (3.1), we have

$$p(\Xi|\phi) \propto \prod_{t=0}^{T-1} \frac{1}{\sqrt{2\pi}} \exp\left[-\frac{(f_{g,t+1} - \phi f_{g,t})^2}{2} \right] = (2\pi)^{-\frac{T}{2}} \exp\left[-\frac{1}{2} S_{f,g} \right] \exp\left[-\frac{1}{2} (\ddot{\phi} - \phi)^2 \sum_{t=0}^{T-1} f_{g,t}^2 \right],$$

where $S_{f,g} := \sum_{t=0}^{T-1} (f_{g,t+1} - \ddot{\phi} f_{g,t})^2$, and the last equality is due to the identity:

$$\sum_{t=0}^{T-1} (f_{g,t+1} - \phi f_{g,t})^2 = \sum_{t=0}^{T-1} (f_{g,t+1} - \ddot{\phi} f_{g,t})^2 + \sum_{t=0}^{T-1} (\ddot{\phi} - \phi)^2 f_{g,t}^2.$$

Assume that $p(\phi) =$ constant. We can calculate

$$\begin{aligned} p(\phi, \Xi) \propto p(\Xi|\phi) &\propto (2\pi)^{-\frac{T}{2}} \exp\left[-\frac{1}{2} S_{f,g} \right] \exp\left[-\frac{1}{2} (\ddot{\phi} - \phi)^2 \sum_{t=0}^{T-1} f_{g,t}^2 \right] \\ &\propto \exp\left[-\frac{1}{2} \cdot \frac{(\ddot{\phi} - \phi)^2}{1/\sum_{t=0}^{T-1} f_{g,t}^2} \right], \end{aligned}$$

whence we have

$$\phi|\Xi \sim N\left(\ddot{\phi}, \frac{1}{\sum_{t=0}^{T-1} f_{g,t}^2} \right).$$

We will discard the draw if the stationarity condition $|\phi| < 1$ is not satisfied and re-draw from the preceding display until we obtain a draw satisfying the stationarity condition.

A.2.3 An Application

The Bayesian approach is computationally intensive and feasible only for small N_c. In some unreported Monte Carlo simulations, we found that the proposed Bayesian estimator works well for $N_c = 20$ but not so well for $N_c = 200$ for $c = A, E, U$.

Here we use the proposed Bayesian estimator to re-estimate the model for the empirical study in Section 7.1. The starting values of the model parameters for the Gibbs sampling are set to those of the EM algorithm. The length of the Markov chain is chosen to be 50,000 with a burn-in period of 40,000. After the burn-in period, we store the every 10th draw; the posterior distributions are formed based on these stored 1000 draws. The results are reported in Tables 9 and 10. One can see that these results are similar to those obtained by the MLE-one day in Section 7.1. The only major difference is that the Bayesian estimator gives a significant -0.1495 for ϕ for size-momentum portfolios in 2011-2015 (see Table 10). However, it is important to stress that two valid estimation procedures cannot produce identical results in the finite samples.

Size and B/M portfolios in 2011-2015

	$\boldsymbol{z}_0^A$	$\boldsymbol{z}_1^A$	$\boldsymbol{z}_2^A$	$\boldsymbol{z}_3^A$	$\boldsymbol{z}_0^E$	$\boldsymbol{z}_1^E$	$\boldsymbol{z}_2^E$	$\boldsymbol{z}_3^E$	$\boldsymbol{z}_0^U$	$\boldsymbol{z}_1^U$	$\boldsymbol{z}_2^U$	$\boldsymbol{z}_3^U$
SG	0.26	0.23	0.58	0.73	0.53	0.73	0.25	0.45	0.56	0.45	0.59	0.50
SN	0.29	0.27	0.69	0.71	0.54	0.74	0.24	0.49	0.57	0.48	0.63	0.49
SV	0.28	0.28	0.73	0.65	0.53	0.73	0.23	0.45	0.57	0.48	0.62	0.48
BG	0.31	0.35	0.89	0.31	0.60	0.83	0.30	0.11	0.65	0.53	0.64	0.16
BN	0.28	0.34	0.94	0.27	0.63	0.85	0.28	0.12	0.69	0.53	0.68	0.10
BV	0.28	0.33	0.90	0.30	0.63	0.79	0.24	0.16	0.62	0.52	0.64	0.19

Size and B/M portfolios in 2016-2020

	$\boldsymbol{z}_0^A$	$\boldsymbol{z}_1^A$	$\boldsymbol{z}_2^A$	$\boldsymbol{z}_3^A$	$\boldsymbol{z}_0^E$	$\boldsymbol{z}_1^E$	$\boldsymbol{z}_2^E$	$\boldsymbol{z}_3^E$	$\boldsymbol{z}_0^U$	$\boldsymbol{z}_1^U$	$\boldsymbol{z}_2^U$	$\boldsymbol{z}_3^U$
SG	0.22	0.47	0.12	0.76	0.22	0.62	0.20	0.74	0.16	0.26	0.84	0.33
SN	0.24	0.56	0.33	0.75	0.34	0.61	0.20	0.75	0.08	0.49	0.85	0.35
SV	0.23	0.56	0.42	0.69	0.43	0.59	0.20	0.71	0.02$^{\clubsuit}$	0.62	0.80	0.31
BG	0.26	0.74	0.22	0.49	0.17	0.65	0.18	0.63	0.21	0.14	0.89	-0.06
BN	0.29	0.80	0.46	0.39	0.36	0.68	0.16	0.63	0.11	0.49	0.89	-0.01$^{\clubsuit}$
BV	0.27	0.67	0.60	0.36	0.53	0.62	0.15	0.56	0.05	0.68	0.81	0.01$^{\clubsuit}$

Size and momentum portfolios in 2011-2015

	$\boldsymbol{z}_0^A$	$\boldsymbol{z}_1^A$	$\boldsymbol{z}_2^A$	$\boldsymbol{z}_3^A$	$\boldsymbol{z}_0^E$	$\boldsymbol{z}_1^E$	$\boldsymbol{z}_2^E$	$\boldsymbol{z}_3^E$	$\boldsymbol{z}_0^U$	$\boldsymbol{z}_1^U$	$\boldsymbol{z}_2^U$	$\boldsymbol{z}_3^U$
SL	0.09	0.16	0.35	0.89	0.54	0.53	0.48	0.45	0.26	0.78	0.48	0.40
SN	0.07	0.16	0.33	0.92	0.54	0.42	0.52	0.53	0.28	0.83	0.32	0.43
SW	0.04$^{\clubsuit}$	0.13	0.30	0.89	0.57	0.29	0.51	0.49	0.29	0.84	0.13	0.47
BL	0.12	0.20	0.42	0.78	0.57	0.58	0.56	0.20	0.24	0.81	0.57	0.11
BN	0.08	0.20	0.42	0.81	0.62	0.45	0.67	0.17	0.29	0.94	0.29	0.08
BW	0.03$^{\clubsuit}$	0.18	0.40	0.79	0.63	0.24	0.62	0.23	0.29	0.92	0.01$^{\clubsuit}$	0.17

Size and momentum portfolios in 2016-2020

	$\boldsymbol{z}_0^A$	$\boldsymbol{z}_1^A$	$\boldsymbol{z}_2^A$	$\boldsymbol{z}_3^A$	$\boldsymbol{z}_0^E$	$\boldsymbol{z}_1^E$	$\boldsymbol{z}_2^E$	$\boldsymbol{z}_3^E$	$\boldsymbol{z}_0^U$	$\boldsymbol{z}_1^U$	$\boldsymbol{z}_2^U$	$\boldsymbol{z}_3^U$
SL	0.23	0.39	0.29	0.86	0.65	0.83	0.32	-0.20	0.58	0.74	0.40	0.27
SN	0.22	0.39	0.18	0.90	0.43	0.94	0.39	-0.26	0.65	0.59	0.48	0.25
SW	0.24	0.37	0.02$^{\clubsuit}$	0.86	0.24	0.95	0.41	-0.23	0.64	0.31	0.53	0.31
BL	0.19	0.37	0.39	0.76	0.70	0.82	0.27	0.04	0.56	0.79	0.41	0.02$^{\clubsuit}$
BN	0.23	0.41	0.23	0.82	0.44	0.97	0.37	0.16	0.68	0.59	0.50	-0.12
BW	0.22	0.38	-0.04$^{\clubsuit}$	0.80	0.13	0.96	0.39	0.08	0.65	0.15	0.52	-0.07

Table 9: The factor loadings estimated by the Bayesian posterior means. To save space, we do not report the standard deviations of the posterior distributions, with a mean 0.0151, a minimum 2.0699×10^{-6} and a maximum 0.0359. An entry with superscript $\clubsuit$ means that zero is within the interval of this entry (i.e., the posterior mean) $\pm 1.96\times$the standard deviation of the posterior distribution.

	Size and B/M		Size and momentum	
	2011-2015	2016-2020	2011-2015	2016-2020
ϕ	-0.1599	-0.1492	-0.1495	-0.2399
	(0.0178)	(0.0238)	(0.0367)	(0.0209)

Table 10: Parameter ϕ estimated by the Bayesian posterior means. The standard deviations of the posterior distributions are in parentheses.

A.3 Formulas for the Kalman Filter and Smoother

In this subsection, we shall give the recursive formulas for the Kalman filter and smoother, which will be used in the Bayesian estimation and EM algorithm, respectively.

A.3.1 The Kalman Filter

Recall the model (2.3). Let $Y_{1:t} := \{\boldsymbol{y}_t^\intercal, \ldots, \boldsymbol{y}_1^\intercal\}^\intercal$ and $Y_0 = \emptyset$. Define

$$\begin{aligned}\boldsymbol{\alpha}_{t|t-1} &:= \mathbb{E}[\boldsymbol{\alpha}_t | Y_{1:t-1}] \qquad P_{t|t-1} := \mathrm{var}(\boldsymbol{\alpha}_t | Y_{1:t-1}) \\ \boldsymbol{\alpha}_{t|t} &:= \mathbb{E}[\boldsymbol{\alpha}_t | Y_{1:t}] \qquad P_{t|t} := \mathrm{var}(\boldsymbol{\alpha}_t | Y_{1:t}).\end{aligned}$$

Given that $\boldsymbol{\alpha}_0 = 0$, it can be calculated that

$$\boldsymbol{\alpha}_{1|0} = 0 \qquad P_{1|0} = RR^\intercal.$$

Define

$$\boldsymbol{v}_t := \boldsymbol{y}_t - \mathbb{E}[\boldsymbol{y}_t | Y_{1:t-1}] = \boldsymbol{y}_t - Z_t \boldsymbol{\alpha}_{t|t-1} = Z_t(\boldsymbol{\alpha}_t - \boldsymbol{\alpha}_{t|t-1}) + \boldsymbol{\varepsilon}_t.$$

That is, $\boldsymbol{v}_t$ is the one-step ahead forecast error of $\boldsymbol{y}_t$ given $Y_{1:t-1}$. When $Y_{1:t-1}$ and $\boldsymbol{v}_t$ are fixed, then $Y_{1:t}$ is fixed. When $Y_{1:t}$ is fixed, then $Y_{1:t-1}$ and $\boldsymbol{v}_t$ are fixed. Thus

$$\begin{aligned}\boldsymbol{\alpha}_{t|t} &= \mathbb{E}[\boldsymbol{\alpha}_t | Y_{1:t}] = \mathbb{E}[\boldsymbol{\alpha}_t | Y_{1:t-1}, \boldsymbol{v}_t] \\ \boldsymbol{\alpha}_{t+1|t} &= \mathbb{E}[\boldsymbol{\alpha}_{t+1} | Y_{1:t-1}, \boldsymbol{v}_t] \\ \mathbb{E}[\boldsymbol{v}_t | Y_{1:t-1}] &= \mathbb{E}\left[Z_t(\boldsymbol{\alpha}_t - \boldsymbol{\alpha}_{t|t-1}) + \boldsymbol{\varepsilon}_t | Y_{1:t-1}\right] = 0.\end{aligned}$$

We have

$$\left.\begin{matrix}\boldsymbol{\alpha}_t \\ \boldsymbol{v}_t\end{matrix}\right|_{Y_{1:t-1}} \sim N\left(\begin{pmatrix}\boldsymbol{\alpha}_{t|t-1} \\ 0\end{pmatrix}, \begin{pmatrix}P_t & \mathrm{cov}(\boldsymbol{\alpha}_t, \boldsymbol{v}_t | Y_{1:t-1}) \\ \mathrm{cov}(\boldsymbol{\alpha}_t, \boldsymbol{v}_t | Y_{1:t-1})^\intercal & F_t\end{pmatrix}\right)$$

where

$$\begin{aligned}F_t &:= \mathrm{var}(\boldsymbol{v}_t | Y_{1:t-1}) = \mathrm{var}\left(Z_t(\boldsymbol{\alpha}_t - \boldsymbol{\alpha}_{t|t-1}) + \boldsymbol{\varepsilon}_t | Y_{1:t-1}\right) = Z_t P_t Z_t^\intercal + \Sigma_t \\ \mathrm{cov}(\boldsymbol{\alpha}_t, \boldsymbol{v}_t | Y_{1:t-1}) &= \mathbb{E}\left[\boldsymbol{\alpha}_t v_t^\intercal | Y_{1:t-1}\right] = \mathbb{E}\left[\boldsymbol{\alpha}_t \left(Z_t(\boldsymbol{\alpha}_t - \boldsymbol{\alpha}_{t|t-1}) + \boldsymbol{\varepsilon}_t\right)^\intercal | Y_{1:t-1}\right] = P_t Z_t^\intercal.\end{aligned}$$

Thus invoking a lemma on multivariate normal, we have

$$\begin{aligned}\boldsymbol{\alpha}_{t|t} &= \boldsymbol{\alpha}_{t|t-1} + P_t Z_t^\intercal F_t^{-1} \boldsymbol{v}_t \\ P_{t|t} &= P_t - P_t Z_t^\intercal F_t^{-1} Z_t P_t.\end{aligned} \tag{A.17}$$

We now develop the recursions for $\boldsymbol{\alpha}_{t+1|t}$ and P_{t+1}.

$$\begin{aligned}\boldsymbol{\alpha}_{t+1|t} &= \mathbb{E}[\boldsymbol{\alpha}_{t+1}|Y_{1:t}] = \mathbb{E}[\mathcal{T}\boldsymbol{\alpha}_t + R\boldsymbol{\eta}_t|Y_{1:t}] = \mathcal{T}\boldsymbol{\alpha}_{t|t} \\ P_{t+1} &= \operatorname{var}(\boldsymbol{\alpha}_{t+1}|Y_{1:t}) = \operatorname{var}(\mathcal{T}\boldsymbol{\alpha}_t + R\boldsymbol{\eta}_t|Y_{1:t}) = \mathcal{T}P_{t|t}\mathcal{T}^\intercal + RR^\intercal\end{aligned}$$

for $t = 1, \dots, T-1$.

A.3.2 The Kalman Smoother

We present the formulas for the Kalman smoother here. For the derivations, see Durbin and Koopman (2012).

$$\begin{aligned}L_t &:= \mathcal{T} - (\mathcal{T}P_t Z_t^\intercal F_t^{-1})Z_t \\ \varsigma_{t-1} &:= Z_t^\intercal F_t^{-1}\boldsymbol{v}_t + L_t^\intercal \varsigma_t \\ N_{t-1} &:= Z_t^\intercal F_t^{-1} Z_t + L_t^\intercal N_t L_t \\ \boldsymbol{\alpha}_{t|T} &:= \mathbb{E}[\boldsymbol{\alpha}_t|Y_{1:T}] = \boldsymbol{\alpha}_{t|t-1} + P_t\varsigma_{t-1} \\ P_{t|T} &:= \operatorname{var}(\boldsymbol{\alpha}_t|Y_{1:T}) = P_t - P_t N_{t-1} P_t \\ \operatorname{cov}(\boldsymbol{\alpha}_t, \boldsymbol{\alpha}_{t+1}|Y_{1:T}) &= P_t L_t^\intercal (I_4 - N_t P_{t+1});\end{aligned}$$

for $t = T, \dots, 1$, initialised with $\varsigma_T = 0$ and $N_T = 0$.

A.4 First-Order Conditions of (3.6)

In this subsection, we derive (3.7). Note that we only utilise information that M is symmetric, positive definite and that Σ_{ee} is diagonal to derive the first-order conditions; no specific knowledge of Λ or M is utilised to derive the first-order conditions. Cholesky decompose M: $M = LL^\intercal$, where L is the unique lower triangular matrix with positive diagonal entries. Thus

$$\Sigma_{yy} = \Lambda M \Lambda^\intercal + \Sigma_{ee} = \Lambda L L^\intercal \Lambda^\intercal + \Sigma_{ee} = BB^\intercal + \Sigma_{ee}, \tag{A.18}$$

where $B := \Lambda L$. Recall the log-likelihood function (3.6) omitting the constant:

$$-\frac{1}{2N}\log|\Sigma_{yy}| - \frac{1}{2N}\operatorname{tr}(S_{yy}\Sigma_{yy}^{-1}) = -\frac{1}{2N}\log|BB^\intercal + \Sigma_{ee}| - \frac{1}{2N}\operatorname{tr}\left(S_{yy}\left[BB^\intercal + \Sigma_{ee}\right]^{-1}\right).$$

Take the derivatives of the preceding display with respect to B and Σ_{ee}. The FOC of Σ_{ee} is:

$$\operatorname{diag}(\hat{\Sigma}_{yy}^{-1}) = \operatorname{diag}(\hat{\Sigma}_{yy}^{-1} S_{yy} \hat{\Sigma}_{yy}^{-1}), \tag{A.19}$$

where $\hat{\Sigma}_{yy} := \hat{B}\hat{B}^\intercal + \hat{\Sigma}_{ee}$. The FOC of B is

$$\hat{B}^\intercal \hat{\Sigma}_{yy}^{-1}(S_{yy} - \hat{\Sigma}_{yy}) = 0. \tag{A.20}$$

Note that (A.19, A.20) has $6N(14+1)$ equations, while B, Σ_{ee} has $6N(14+1)$ parameters. Thus $\hat{B}, \hat{\Sigma}_{ee}$ can be uniquely solved. Then we need identification conditions to kick in. Even though we could uniquely determine $\hat{B}$, we cannot uniquely determine $\hat{\Lambda}, \hat{M}$. This is because

$$\hat{B}\hat{B}^\intercal = \tilde{\Lambda}\tilde{M}\tilde{\Lambda}^\intercal = \mathring{\Lambda}\mathring{M}\mathring{\Lambda}^\intercal = \tilde{\Lambda}CC^{-1}\tilde{M}(C^{-1})^\intercal C^\intercal \tilde{\Lambda}^\intercal$$

where $\mathring{\Lambda} := \tilde{\Lambda}C$ and $\mathring{M} := C^{-1}\tilde{M}(C^{-1})^{\intercal}$ for any 14×14 invertible C.[7] We hence need to impose 14^2 identification restrictions on the estimates of Λ and M to rule out the rotational indeterminacy. After imposing the 14^2 restrictions, we obtain the unique estimates, say, $\hat{\Lambda}, \hat{M}$ (and hence $\hat{L}$). Substituting $\hat{B} = \hat{\Lambda}\hat{L}$ into (A.20), we have

$$\hat{\Lambda}^{\intercal}\hat{\Sigma}_{yy}^{-1}(S_{yy} - \hat{\Sigma}_{yy}) = 0.$$

A.5 Proof of Proposition 4.1

As Bai and Li (2012) did, we use a superscript "*" to denote the true parameters, $\Lambda^*, \Sigma^*_{ee}, M^*$ etc. The parameters without the superscript "*" denote the generic parameters in the likelihood function. Note that the proof of (4.2) is exactly the same as that of Bai and Li (2012), so we omit the details here.

Define

$$\begin{aligned} \hat{H} &:= (\hat{\Lambda}^{\intercal}\hat{\Sigma}_{ee}^{-1}\hat{\Lambda})^{-1} \\ A &:= (\hat{\Lambda} - \Lambda^*)^{\intercal}\hat{\Sigma}_{ee}^{-1}\hat{\Lambda}\hat{H} \\ K &:= \hat{M}^{-1}(MA - A^{\intercal}MA). \end{aligned} \tag{A.21}$$

Our assumptions satisfy those of Bai and Li (2012), so (A17) of Bai and Li (2012) still holds (in our notation):

$$\hat{\boldsymbol{\lambda}}_{k,j} - \boldsymbol{\lambda}^*_{k,j} = K\boldsymbol{\lambda}^*_{k,j} + o_p(1), \tag{A.22}$$

for $k = 1, \dots, 6$ and $j = 1, \dots, N$. As mentioned before, Λ^* (i.e., $\{\boldsymbol{\lambda}^*_{k,j} : k = 1, \dots, 6, j = 1, \dots, N\}$) defined (2.5) gives more than 14^2 restrictions, but in order to utilise the theories of Bai and Li (2012) we shall only impose 14^2 restrictions on $\{\hat{\boldsymbol{\lambda}}_{k,j} : k = 1, \dots, 6, j = 1, \dots, N\}$. How to select these 14^2 restrictions from those implied by $\{\boldsymbol{\lambda}^*_{k,j}\}$ are crucial because we cannot afford imposing a restriction which is not instrumental for the proofs later. The idea is that one restriction should pin down one free parameter in K. We shall now explain our procedure. Write (A.22) in matrix form:

$$\hat{\Lambda}_k - \Lambda^*_k = K\Lambda^*_k + o_p(1), \tag{A.23}$$

where $\hat{\Lambda}_k := (\hat{\boldsymbol{\lambda}}_{k,1}, \dots, \hat{\boldsymbol{\lambda}}_{k,N})$ and $\Lambda^*_k := (\boldsymbol{\lambda}^*_{k,1}, \dots, \boldsymbol{\lambda}^*_{k,N})$ are $14 \times N$ matrices. For a generic matrix C, let $C_{x,y}$ denote the matrix obtained by intersecting the rows and columns whose indices are in x and y, respectively; let $C_{x,\bullet}$ denote the matrix obtained by extracting the rows whose indices are in x while $C_{\bullet,y}$ denote the matrix obtained by extracting the columns whose indices are in y.

A.5.1 Step I Impose Some Zero Restrictions in $\{\Lambda^*_k\}_{k=1}^6$

Let $a \subset \{1, 2, \dots, 14\}$ and $c \subset \{1, \dots, N\}$ be two vectors of indices, whose identities vary from place to place. From (A.23), we have

$$\begin{aligned} \left(\hat{\Lambda}_k - \Lambda^*_k\right)_{a,c} &= K_{a,\bullet}\Lambda^*_{k,\bullet,c} + o_p(1) = K_{a,b}\Lambda^*_{k,b,c} + K_{a,-b}\Lambda^*_{k,-b,c} + o_p(1) = K_{a,b}\Lambda^*_{k,b,c} + o_p(1) \\ &= K_{a,b_1}\Lambda^*_{k,b_1,c} + K_{a,b_2}\Lambda^*_{k,b_2,c} + o_p(1) \end{aligned} \tag{A.24}$$

[7] Note that we could find at least one pair $(\tilde{\Lambda}, \tilde{M})$ satisfying $\hat{B}\hat{B}^{\intercal} = \tilde{\Lambda}\tilde{M}\tilde{\Lambda}^{\intercal}$: $\tilde{\Lambda} = \hat{B}$ and $\tilde{M} = I_{14}$.

Step	k	c	a	b_1	b_2
I.1	1	{1,2,3,4}	{1,2,3,4,5,9,10,11,12,13}	$\emptyset$	$\{6,7,8,14\}$
I.2	6	{1,2,3,4}	{4,5,6,7,8,10,11,12,13,14}	$\emptyset$	{1,2,3,9}
I.3	4	{1,2,3}	{6,7,8,10,12,13,14}	3	{4,5,11}
I.4	4	{1,2,3,4}	{1,2,9}	$\emptyset$	{3,4,5,11}
I.5	3	{1,2,3}	3	6	{4,5,12}
I.6	3	{1,2}	{7,8,14}	{4,5}	{6,12}
I.7	2	{1,2}	{4,11}	{6,7}	{5,13}
I.8	2	{1,2}	{8,14}	{5,6}	{7,13}
I.9	5	{1,2}	{1,9}	{3,4}	{2,10}
I.10	5	{1,2}	{5}	{2,3}	{4,10}
I.11	5	{1,2}	{11}	{2,3}	{4,10}
I.12	5	1	{6,7,8,12,13,14}	{2,3,4}	10
I.13	3	1	{1,2,9,10,11,13}	{4,5,6}	12
I.14	2	1	{1,2,3,9,10,12}	{5,6,7}	13

Table 11: Step I.

where $b \subset \{1, 2, \ldots, 14\}$ is chosen in such a way such that $\Lambda^*_{k,-b,c} = 0$ for each of the steps below, $-b$ denotes the complement of b, and $b_1 \cup b_2 = b$ with the cardinality of b_2 equal to the cardinality of c.

In each of the sub-step of step I, we shall impose $\hat{\Lambda}_{k,a,c} = 0$. Step I is detailed in Table 11, and we shall use step I.1 to illustrate. For step I.1, $\hat{\Lambda}_{k,a,c} = 0$ means

$$\begin{bmatrix} \hat{\boldsymbol{\lambda}}_{1,1}^\intercal \\ \hat{\boldsymbol{\lambda}}_{1,2}^\intercal \\ \hat{\boldsymbol{\lambda}}_{1,3}^\intercal \\ \hat{\boldsymbol{\lambda}}_{1,4}^\intercal \end{bmatrix} = \begin{bmatrix} 0 & 0 & 0 & 0 & 0 & - & - & - & 0 & 0 & 0 & 0 & 0 & - \\ 0 & 0 & 0 & 0 & 0 & - & - & - & 0 & 0 & 0 & 0 & 0 & - \\ 0 & 0 & 0 & 0 & 0 & - & - & - & 0 & 0 & 0 & 0 & 0 & - \\ 0 & 0 & 0 & 0 & 0 & - & - & - & 0 & 0 & 0 & 0 & 0 & - \end{bmatrix}.$$

This means (A.24) holds with LHS being $\left(\hat{\Lambda}_k - \Lambda^*_k\right)_{a,c} = 0$, where

$$k = 1, \quad c = \{1,2,3,4\}, \quad a = \{1,2,3,4,5,9,10,11,12,13\}, \quad b_1 = \emptyset, \quad b_2 = \{6,7,8,14\}.$$

Note that $c = \{1, 2, 3, 4\}$ is arbitrary and could be replaced with any other $c \subset \{1, \ldots, N\}$ with cardinality being 4. The crucial point is that c needs to be chosen such that $\Lambda^*_{k,b_2,c}$ is invertible. This is an innocuous requirement given large N, so we shall make this assumption implicitly for the rest of the article. Solving (A.24) gives $K_{a,b_2} = o_p(1)$.

A.5.2 Step II Impose Some Equality Restrictions in $\{\Lambda^*_k\}_{k=1}^6$

(II.1) Note that (A.23) implies

$$\hat{\Lambda}_{6,9,c} - \Lambda^*_{6,9,c} = K_{9,x}\Lambda^*_{6,x,c} + o_p(1) = K_{9,1}\Lambda^*_{6,1,c} + K_{9,9}\Lambda^*_{6,9,c} + o_p(1) \qquad x = \{1,2,3,9\}$$
$$\hat{\Lambda}_{3,12,c} - \Lambda^*_{3,12,c} = K_{12,y}\Lambda^*_{3,y,c} + o_p(1) = K_{12,12}\Lambda^*_{3,12,c} + o_p(1) \qquad y = \{4,5,6,12\}.$$

We then impose $\hat{\Lambda}_{6,9,c} = \hat{\Lambda}_{3,12,c}$ for $c = \{1, 2\}$. The preceding display implies

$$K_{9,1}\Lambda^*_{6,1,c} + \left(K_{9,9} - K_{12,12}\right)\Lambda^*_{6,9,c} = o_p(1)$$

whence we have $K_{9,1} = o_p(1)$ and $K_{9,9} - K_{12,12} = o_p(1)$.

(II.2) We impose $\hat{\Lambda}_{4,11,c} = \hat{\Lambda}_{1,14,c}$ for $c = \{1,2\}$. Repeating the procedure in step II.1, we have $K_{14,8} = o_p(1)$ and $K_{11,11} - K_{14,14} = o_p(1)$ in the same way.

(II.3) We impose $\hat{\Lambda}_{6,1,c} = \hat{\Lambda}_{3,4,c}$ for $c = \{1,2\}$, and have $K_{1,1} - K_{4,4} = o_p(1), K_{1,9} - K_{4,12} = o_p(1)$.

(II.4) We impose $\hat{\Lambda}_{6,2,c} = \hat{\Lambda}_{3,5,c}$ for $c = \{1,2,3\}$, and have $K_{2,1} = o_p(1), K_{2,2} - K_{5,5} = o_p(1), K_{2,9} - K_{5,12} = o_p(1)$.

(II.5) We impose $\hat{\Lambda}_{6,3,c} = \hat{\Lambda}_{3,6,c}$ for $c = \{1,2,3,4\}$, and have $K_{3,1} = o_p(1), K_{3,2} = o_p(1), K_{3,3} - K_{6,6} = o_p(1), K_{3,9} - K_{6,12} = o_p(1)$.

(II.6) We impose $\hat{\Lambda}_{1,8,c} = \hat{\Lambda}_{4,5,c}$ for $c = \{1,2\}$, and have $K_{8,8} - K_{5,5} = o_p(1), K_{8,14} - K_{5,11} = o_p(1)$.

(II.7) We impose $\hat{\Lambda}_{1,7,c} = \hat{\Lambda}_{4,4,c}$ for $c = \{1,2,3\}$, and have $K_{7,8} = o_p(1), K_{7,7} - K_{4,4} = o_p(1), K_{7,14} - K_{4,11} = o_p(1)$.

(II.8) We impose $\hat{\Lambda}_{1,6,c} = \hat{\Lambda}_{4,3,c}$ for $c = \{1,2,3\}$, and have $K_{6,7} = o_p(1), K_{6,8} = o_p(1), K_{6,14} - K_{3,11} = o_p(1)$.

A.5.3 Step III Impose Some Restrictions in M^*

After steps I and II, K is reduced to

$$K = \begin{bmatrix} \overline{K}_{11} & \overline{K}_{12} \\ 0 & \overline{K}_{22} \end{bmatrix} + o_p(1),$$

where

$$\overline{K}_{11} = \begin{bmatrix} K_{1,1} & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & K_{2,2} & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & K_{3,3} & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & K_{1,1} & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & K_{2,2} & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & K_{3,3} & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & K_{1,1} & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & K_{2,2} \end{bmatrix} \tag{A.25}$$

$$\overline{K}_{12} = \begin{bmatrix} K_{4,12} & 0 & 0 & 0 & 0 & 0 \\ K_{5,12} & K_{2,10} & 0 & 0 & 0 & 0 \\ K_{6,12} & K_{3,10} & K_{3,11} & 0 & 0 & 0 \\ 0 & K_{4,10} & K_{4,11} & K_{4,12} & 0 & 0 \\ 0 & 0 & K_{5,11} & K_{5,12} & K_{5,13} & 0 \\ 0 & 0 & 0 & K_{6,12} & K_{6,13} & K_{3,11} \\ 0 & 0 & 0 & 0 & K_{7,13} & K_{4,11} \\ 0 & 0 & 0 & 0 & 0 & K_{5,11} \end{bmatrix}, \overline{K}_{22} = \begin{bmatrix} K_{12,12} & 0 & 0 & 0 & 0 & 0 \\ 0 & K_{10,10} & 0 & 0 & 0 & 0 \\ 0 & 0 & K_{11,11} & 0 & 0 & 0 \\ 0 & 0 & 0 & K_{12,12} & 0 & 0 \\ 0 & 0 & 0 & 0 & K_{13,13} & 0 \\ 0 & 0 & 0 & 0 & 0 & K_{11,11} \end{bmatrix}.$$

In the paragraph above (A16) of Bai and Li (2012), they showed $A = O_p(1)$. Given Assumption 4.1(iii), we have $K = O_p(1)$. Next, (A16) of Bai and Li (2012) still holds and could be written as

$$\hat{M} - (I_{14} - A^\intercal)M^*(I_{14} - A) = o_p(1). \tag{A.26}$$

Since M^* and $\hat{M}^*$ are of full rank (Assumption 4.1(iii)), (A.26) implies that $I_{14} - A$ is of full rank. Write (A.26) as

$$\hat{M}(K + I_{14}) - (I_{14} - A^\intercal)M^* = o_p(1).$$

As Bai and Li (2012) did in their (A20), we could premultiply the preceding display by $\left[(I_{14} - A^\intercal)M^*\right]^{-1}$ to arrive at (after some algebra and relying on (A.26)):

$$(I_{14} - A)(K + I_{14}) - I_{14} = o_p(1). \tag{A.27}$$

Likewise, partition A into 8×8, 8×6, 6×8 and 6×6 submatrices:

$$A = \begin{bmatrix} A_{11} & A_{12} \\ A_{21} & A_{22} \end{bmatrix}.$$

Then (A.27) could be written into

$$(I_8 - A_{11})(\overline{K}_{11} + I_8) - I_8 = o_p(1) \tag{A.28}$$
$$(I_8 - A_{11})\overline{K}_{12} - A_{12}(\overline{K}_{22} + I_6) = o_p(1) \tag{A.29}$$
$$-A_{21}(\overline{K}_{11} + I_8) = o_p(1) \tag{A.30}$$
$$-A_{21}\overline{K}_{12} + (I_6 - A_{22})(\overline{K}_{22} + I_6) - I_6 = o_p(1). \tag{A.31}$$

Consider (A.28) first. Since $I_8 + \overline{K}_{11}$ is diagonal, we deduce that the diagonal elements of $I_8 + \overline{K}_{11}$ could not converge to 0, and A_{11} converges to a diagonal matrix. Using the fact that the diagonal elements of $I_8 + \overline{K}_{11}$ could not converge to 0, (A.30) implies

$$A_{21} = o_p(1),$$

and $A_{21}\overline{K}_{12} = o_p(1)O_p(1) = o_p(1)$. Then (A.31) is reduced to

$$(I_6 - A_{22})(\overline{K}_{22} + I_6) - I_6 = o_p(1).$$

Since $\overline{K}_{22} + I_6$ is diagonal, we deduce that the diagonal elements of $\overline{K}_{22} + I_6$ could not converge to 0, and A_{22} should converge to a diagonal matrix as well. To sum up

$$I_8 + \overline{K}_{11} \qquad I_8 - A_{11} \qquad I_6 + \overline{K}_{22} \qquad I_6 - A_{22}$$

are diagonal or diagonal in the limit, and invertible in the limit (i.e., none of the diagonal elements is zero in the limit). Moreover, (A.28) implies

$$(I_8 + \overline{K}_{11})^{-1} = (I_8 - A_{11}) + o_p(1). \tag{A.32}$$

Via (A.27), (A.26) implies

$$(I_{14} + K^\intercal)\hat{M}(I_{14} + K) - M^* = o_p(1). \tag{A.33}$$

Partition $\hat{M}$ into 8×8, 8×6, 6×8 and 6×6 submatrices:

$$\hat{M} = \begin{bmatrix} \hat{\overline{M}}_{11} & \hat{\overline{M}}_{12} \\ \hat{\overline{M}}_{21} & \hat{\overline{M}}_{22} \end{bmatrix}.$$

Then (A.33) could be written as

$$(I_8 + \overline{K}_{11})\hat{\overline{M}}_{11}(I_8 + \overline{K}_{11}) - \Phi^* = o_p(1) \tag{A.34}$$

$$(I_8 + \overline{K}_{11})\hat{\overline{M}}_{11}\overline{K}_{12} + (I_8 + \overline{K}_{11})\hat{\overline{M}}_{12}(I_6 + \overline{K}_{22}) = o_p(1) \tag{A.35}$$

$$\left[\overline{K}_{12}^{\intercal}\hat{\overline{M}}_{11} + (I_6 + \overline{K}_{22})\hat{\overline{M}}_{21}\right](\overline{K}_{11} + I_8) = o_p(1) \tag{A.36}$$

$$\overline{K}_{12}^{\intercal}\left(\hat{\overline{M}}_{12}(I_6 + \overline{K}_{22}) + \hat{\overline{M}}_{11}\overline{K}_{12}\right) + (I_6 + \overline{K}_{22})\left(\hat{\overline{M}}_{22}(I_6 + \overline{K}_{22}) + \hat{\overline{M}}_{21}\overline{K}_{12}\right) - I_6 = o_p(1). \tag{A.37}$$

Step III.1 Considering (A.34), we have

$$(1 + K_{3,3})^2\hat{M}_{6,6} = \frac{1}{1 - \phi^{*,2}} + o_p(1)$$
$$(1 + K_{1,1})^2\hat{M}_{4,4} = \frac{1}{1 - \phi^{*,2}} + o_p(1)$$
$$(1 + K_{2,2})^2\hat{M}_{5,5} = \frac{1}{1 - \phi^{*,2}} + o_p(1).$$

Imposing $\hat{M}_{4,4} = \hat{M}_{6,6}$, we have

$$\hat{M}_{4,4}\left[(1 + K_{1,1})^2 - (1 + K_{3,3})^2\right] = o_p(1).$$

Since (A.34) implies that $\hat{M}_{4,4} \neq o_p(1)$. The preceding display implies

$$K_{3,3} = K_{1,1} + o_p(1), \qquad \text{or} \qquad 1 + K_{3,3} = -(1 + K_{1,1}) + o_p(1).$$

Likewise, imposing $\hat{M}_{4,4} = \hat{M}_{5,5}$, we have

$$K_{2,2} = K_{1,1} + o_p(1), \qquad \text{or} \qquad 1 + K_{2,2} = -(1 + K_{1,1}) + o_p(1).$$

Thus, there are four cases:

(a) $K_{2,2} = K_{1,1} + o_p(1)$ and $K_{3,3} = K_{1,1} + o_p(1)$

(b) $K_{2,2} = K_{1,1} + o_p(1)$ and $1 + K_{3,3} = -(1 + K_{1,1}) + o_p(1)$

(c) $1 + K_{2,2} = -(1 + K_{1,1}) + o_p(1)$ and $K_{3,3} = K_{1,1} + o_p(1)$

(d) $1 + K_{2,2} = -(1 + K_{1,1}) + o_p(1)$ and $1 + K_{3,3} = -(1 + K_{1,1}) + o_p(1)$.

Irrespective of case, (A.34) is reduced to $(1+K_{1,1})^2\hat{\overline{M}}_{11} = \Phi^* + o_p(1)$, whence we have

$$(1 + K_{1,1})^2\hat{M}_{4,4} = \frac{1}{1 - \phi^{*,2}} + o_p(1)$$
$$(1 + K_{1,1})^2\hat{M}_{6,4} = \frac{\phi^{*,2}}{1 - \phi^{*,2}} + o_p(1).$$

Imposing $\hat{M}_{4,4} - \hat{M}_{6,4} = 1$, we have

$$(1 + K_{1,1})^2 = 1 + o_p(1)$$

whence we have $K_{1,1} = o_p(1)$ or $1+K_{1,1} = -1+o_p(1)$. Suppose that $1+K_{1,1} = -1+o_p(1)$. Then (A.32) implies $(A_{11})_{1,1} = 2 + o_p(1)$. Note that the identification scheme which we employ in Proposition 4.1 only identifies Λ^* up to a column sign change. Thus by assuming that $\hat{\Lambda}$ and Λ^* have the same column signs, we can rule out the case $(A_{11})_{1,1} = 2 + o_p(1)$ (Bai and Li (2012, p.445, p.463)). Thus we have $K_{1,1} = o_p(1)$ and hence rule out cases (b)-(d). To sum up, we have

$$K_{1,1} = o_p(1), \quad \overline{K}_{11} = o_p(1), \quad \hat{\overline{M}}_{11} = \Phi^* + o_p(1), \quad A_{11} = o_p(1).$$

Step III.2 Now (A.35) is reduced to

$$\hat{\overline{M}}_{12}(I_6 + \overline{K}_{22}) = -\Phi^* \overline{K}_{12} + o_p(1)$$

Impose three more restrictions: Assume the 4th-6th elements of the fourth column of $\hat{\overline{M}}_{12}$ are zero; that is $\hat{M}_{4,12} = \hat{M}_{5,12} = \hat{M}_{6,12} = 0$. This implies that the corresponding three elements of $\Phi^*_{12}\overline{K}$ are $o_p(1)$:

$$\frac{1}{1-\phi^{*2}} \begin{bmatrix} K_{4,12} + \phi^* K_{5,12} + \phi^{*2} K_{6,12} \\ \phi^* K_{4,12} + K_{5,12} + \phi^* K_{6,12} \\ \phi^{*2} K_{4,12} + \phi^* K_{5,12} + K_{6,12} \end{bmatrix} = o_p(1)$$

whence we have $K_{4,12} = K_{5,12} = K_{6,12} = o_p(1)$. Similarly, assuming the 2nd-4th elements of the second column of $\hat{\overline{M}}_{12}$ are zero, we could deduce that $K_{2,10} = K_{3,10} = K_{4,10} = o_p(1)$; assuming the 3rd-5th elements of the third column of $\hat{\overline{M}}_{12}$ are zero, we could deduce that $K_{3,11} = K_{4,11} = K_{5,11} = o_p(1)$; assuming $\hat{M}_{5,13} = \hat{M}_{6,13} = \hat{M}_{7,13} = 0$,we could deduce that $K_{5,13} = K_{6,13} = K_{7,13} = o_p(1)$. We hence obtain

$$\overline{K}_{12} = o_p(1).$$

Step III.3 With $\overline{K}_{12} = o_p(1)$, (A.37) is reduced to

$$(I_6 + \overline{K}_{22})\hat{\overline{M}}_{22}(I_6 + \overline{K}_{22}) - I_6 = o_p(1).$$

Since $I_6 + \overline{K}_{22}$ is diagonal, $\hat{\overline{M}}_{22}$ is asymptotically diagonal. Imposing that the 2nd-5th diagonal elements of $\hat{\overline{M}}_{22}$ are 1 (i.e., $\hat{M}_{j,j} = 1$ for $j = 10, 11, 12, 13$), we have

$$\begin{aligned} (1 + K_{10,10})^2 - 1 &= o_p(1) \\ (1 + K_{11,11})^2 - 1 &= o_p(1) \\ (1 + K_{12,12})^2 - 1 &= o_p(1) \\ (1 + K_{13,13})^2 - 1 &= o_p(1) \end{aligned}$$

whence we have $K_{j,j} = o_p(1)$ or $1 + K_{j,j} = -1 + o_p(1)$ for $j = 10, 11, 12, 13$. Likewise, the case $1 + K_{j,j} = -1 + o_p(1)$ is ruled out. Thus

$$\overline{K}_{22} = o_p(1), \quad A_{22} = o_p(1), \quad \hat{\overline{M}}_{22} = I_6 + o_p(1).$$

Then (A.35) and (A.36) imply $\hat{\overline{M}}_{12} = o_p(1)$ and $\hat{M}_{21} = o_p(1)$, respectively. Also (A.29) implies $A_{12} = o_p(1)$. To sum up, we have

$$K = o_p(1), \qquad A = o_p(1), \qquad \hat{M} = M^* + o_p(1).$$

Substituting $K = o_p(1)$ into (A.22), we have

$$\hat{\boldsymbol{\lambda}}_{k,j} - \boldsymbol{\lambda}^*_{k,j} = o_p(1),$$

for $k = 1, \ldots, 6$ and $j = 1, \ldots, N$.

A.6 Proof of Theorem 4.1

Given consistency, we can drop the superscript from the true parameters for simplicity. The proof of Theorem 4.1 resembles that of Theorem 5.1 of Bai and Li (2012). Most of the proof of Theorem 5.1 of Bai and Li (2012) is insensitive to the identification condition; the only exception is their Lemma B5. Thus we only need to prove the result of their Lemma B5 under our identification condition. That is, we want to prove

$$MA = O_p(T_f^{-1/2}) + O_p\left(\left[\frac{1}{6N}\sum_{k=1}^{6}\sum_{j=1}^{N}(\hat{\sigma}^2_{k,j} - \sigma^2_{k,j})^2\right]^{1/2}\right) =: O_p(\Diamond). \tag{A.38}$$

Following the approach of Bai and Li (2012) and using their Lemmas B1, B2, B3, one could show that

$$\hat{\boldsymbol{\lambda}}_{k,j} - \boldsymbol{\lambda}_{k,j} = K\boldsymbol{\lambda}_{k,j} + O_p(\Diamond).$$

Using the same approach we adopted in the proof of Proposition 4.1, we arrive at

$$K = \begin{bmatrix} \overline{K}_{11} & \overline{K}_{12} \\ 0 & \overline{K}_{22} \end{bmatrix} + O_p(\Diamond),$$

where $\overline{K}_{11}$, $\overline{K}_{12}$, $\overline{K}_{22}$ are defined in (A.25), and

$$\hat{M} - (I_{14} - A^\intercal)M(I_{14} - A) = O_p(\Diamond) \tag{A.39}$$

$$(I_{14} - A)(K + I_{14}) - I_{14} = O_p(\Diamond) \tag{A.40}$$

Note that $\sqrt{1 + O_p(\Diamond)} = 1 + O_p(\Diamond)$ and $(1 + O_p(\Diamond))^{-1} = 1 + O_p(\Diamond)$ because of the generalised Binomial theorem and that $O_p(\Diamond) = o_p(1)$. Then one could repeat the argument in the proof of Proposition 4.1 to arrive at

$$K = O_p(\Diamond), \qquad A = O_p(\Diamond), \qquad \hat{M} = M + O_p(\Diamond) \tag{A.41}$$

$$\hat{\boldsymbol{\lambda}}_{k,j} - \boldsymbol{\lambda}_{k,j} = O_p(\Diamond), \qquad MA = O_p(\Diamond) \tag{A.42}$$

for $k = 1, \ldots, 6$ and $j = 1, \ldots, N$.

A.7 Proof of Theorem 4.2

Equation (C4) of the supplement of Bai and Li (2012) still holds in our case since its derivation does not involve identification conditions (page 17 of the supplement of Bai and Li (2012)); in our notation it is

$$\hat{\sigma}_m^2 - \sigma_m^2 = \frac{1}{T_f}\sum_{t=1}^{T_f}(e_{m,t}^2 - \sigma_m^2) + o_p(T_f^{-1/2})$$

for $m = 1, \ldots, 6N$, where the single-index σ_m^2 is defined as $\sigma_m^2 := \sigma^2_{\lceil \frac{m}{N}\rceil, m-\lfloor \frac{m}{N}\rfloor N}$; interpret $\hat{\sigma}_m^2, e_{m,t}$ similarly. Thus theorem 4.2 follows.

A.8 Proof of Theorem 4.3

Pre-multiply $\hat{M}$ to (A14) of Bai and Li (2012) and write in our notation:

$$\begin{aligned}
&\hat{M}\left(\hat{\boldsymbol{\lambda}}_{k,j} - \boldsymbol{\lambda}_{k,j}\right) = M(\hat{\Lambda} - \Lambda)^\intercal \hat{\Sigma}_{ee}^{-1}\hat{\Lambda}\hat{H}\boldsymbol{\lambda}_{k,j} \\
&- \hat{H}\hat{\Lambda}^\intercal\hat{\Sigma}_{ee}^{-1}(\hat{\Lambda} - \Lambda)M(\hat{\Lambda} - \Lambda)^\intercal\hat{\Sigma}_{ee}^{-1}\hat{\Lambda}\hat{H}\boldsymbol{\lambda}_{k,j} \\
&- \hat{H}\hat{\Lambda}^\intercal\hat{\Sigma}_{ee}^{-1}\Lambda\left(\frac{1}{T_f}\sum_{t=1}^{T_f}\boldsymbol{f}_t\boldsymbol{e}_t^\intercal\right)\hat{\Sigma}_{ee}^{-1}\hat{\Lambda}\hat{H}\boldsymbol{\lambda}_{k,j} \\
&- \hat{H}\hat{\Lambda}^\intercal\hat{\Sigma}_{ee}^{-1}\left(\frac{1}{T_f}\sum_{t=1}^{T_f}\boldsymbol{e}_t\boldsymbol{f}_t^\intercal\right)\Lambda^\intercal\hat{\Sigma}_{ee}^{-1}\hat{\Lambda}\hat{H}\boldsymbol{\lambda}_{k,j} \\
&- \hat{H}\left(\sum_{m=1}^{6N}\sum_{\ell=1}^{6N}\frac{1}{\hat{\sigma}_m^2\hat{\sigma}_\ell^2}\hat{\lambda}_m\hat{\lambda}_\ell^\intercal\frac{1}{T_f}\sum_{t=1}^{T_f}\left[e_{m,t}e_{\ell,t} - \mathbb{E}(e_{m,t}e_{\ell,t})\right]\right)\hat{H}\boldsymbol{\lambda}_{k,j} \\
&+ \hat{H}\left(\sum_{m=1}^{6N}\frac{1}{\hat{\sigma}_m^4}\hat{\lambda}_m\hat{\lambda}_m^\intercal(\hat{\sigma}_m^2 - \sigma_m^2)\right)\hat{H}\boldsymbol{\lambda}_{k,j} \\
&+ \hat{H}\hat{\Lambda}^\intercal\hat{\Sigma}_{ee}^{-1}\left(\frac{1}{T_f}\sum_{t=1}^{T_f}\boldsymbol{e}_t\boldsymbol{f}_t^\intercal\right)\boldsymbol{\lambda}_{k,j} + \hat{H}\hat{\Lambda}^\intercal\hat{\Sigma}_{ee}^{-1}\Lambda\left(\frac{1}{T_f}\sum_{t=1}^{T_f}\boldsymbol{f}_t e_{(k-1)N+j,t}\right) \\
&+ \hat{H}\left(\sum_{i=1}^{6N}\frac{1}{\hat{\sigma}_m^2}\hat{\lambda}_m\frac{1}{T_f}\sum_{t=1}^{T_f}\left[e_{m,t}e_{(k-1)N+j,t} - \mathbb{E}(e_{m,t}e_{(k-1)N+j,t})\right]\right) - \hat{H}\hat{\boldsymbol{\lambda}}_{k,j}\frac{1}{\hat{\sigma}_{k,j}^2}(\hat{\sigma}_{k,j}^2 - \sigma_{k,j}^2),
\end{aligned} \tag{A.43}$$

where $e_{i,t}$ denotes the ith element of $\boldsymbol{e}_t$, the single-index $\boldsymbol{\lambda}_m$ is defined as $\boldsymbol{\lambda}_m := \boldsymbol{\lambda}_{\lceil \frac{m}{N}\rceil, m-\lfloor \frac{m}{N}\rfloor N}$; interpret $\hat{\boldsymbol{\lambda}}_m, \sigma_m^2, \hat{\sigma}_m^2$ similarly.

Consider the right hand side of (A.43). The third and fourth terms are $o_p(T_f^{-1/2})$ by Lemma C1(e) of Bai and Li (2012). The fifth term is $o_p(T_f^{-1/2})$ by Lemma C1(d) of Bai and Li (2012). The six term is $o_p(T_f^{-1/2})$ by Lemma C1(f) of Bai and Li (2012). The seventh term is $o_p(T_f^{-1/2})$ by Lemma C1(e) of Bai and Li (2012). The ninth term is $o_p(T_f^{-1/2})$ by Lemma C1(c) of Bai and Li (2012). The tenth term is $o_p(T_f^{-1/2})$ by Theorem 4.2. Thus (A.43) becomes

$$\hat{M}\left(\hat{\boldsymbol{\lambda}}_{k,j} - \boldsymbol{\lambda}_{k,j}\right) = MA\boldsymbol{\lambda}_{k,j} - A^\intercal MA\boldsymbol{\lambda}_{k,j} + \hat{H}\hat{\Lambda}^\intercal\hat{\Sigma}_{ee}^{-1}\Lambda\left(\frac{1}{T_f}\sum_{t=1}^{T_f}\boldsymbol{f}_t e_{(k-1)N+j,t}\right) + o_p(T_f^{-1/2}).$$

Substituting (4.5) into (A.38), we have $O_p(\Diamond) = O_p(T_f^{-1/2})$. Thus $A = O_p(T_f^{-1/2})$ via (A.42). The preceding display hence becomes

$$\hat{\boldsymbol{\lambda}}_{k,j} - \boldsymbol{\lambda}_{k,j} = \hat{M}^{-1} M A \boldsymbol{\lambda}_{k,j} + \hat{M}^{-1} \hat{H} \hat{\Lambda}^\intercal \hat{\Sigma}_{ee}^{-1} \Lambda \left(\frac{1}{T_f} \sum_{t=1}^{T_f} \boldsymbol{f}_t e_{(k-1)N+j,t} \right) + o_p(T_f^{-1/2}). \tag{A.44}$$

Note that

$$\begin{aligned} &\hat{H}\hat{\Lambda}^\intercal \hat{\Sigma}_{ee}^{-1} \Lambda = (\hat{\Lambda}^\intercal \hat{\Sigma}_{ee}^{-1} \hat{\Lambda})^{-1} \hat{\Lambda}^\intercal \hat{\Sigma}_{ee}^{-1} \Lambda = (\hat{\Lambda}^\intercal \hat{\Sigma}_{ee}^{-1} \hat{\Lambda})^{-1} \left[\hat{\Lambda}^\intercal \hat{\Sigma}_{ee}^{-1} \hat{\Lambda} + \hat{\Lambda}^\intercal \hat{\Sigma}_{ee}^{-1} (\Lambda - \hat{\Lambda}) \right] \\ &= I_{14} + (\hat{\Lambda}^\intercal \hat{\Sigma}_{ee}^{-1} \hat{\Lambda})^{-1} \hat{\Lambda}^\intercal \hat{\Sigma}_{ee}^{-1} (\Lambda - \hat{\Lambda}) = I_{14} + \hat{H}\hat{\Lambda}^\intercal \hat{\Sigma}_{ee}^{-1} (\Lambda - \hat{\Lambda}) = I_{14} + O_p(T_f^{-1/2}) \end{aligned}$$

where the last equality is due to Lemma C1(a) of Bai and Li (2012). Substituting the preceding display into (A.44), we have

$$\begin{aligned} &\hat{\boldsymbol{\lambda}}_{k,j} - \boldsymbol{\lambda}_{k,j} \\ &= \hat{M}^{-1} M A \boldsymbol{\lambda}_{k,j} + \hat{M}^{-1} \left(\frac{1}{T_f} \sum_{t=1}^{T_f} \boldsymbol{f}_t e_{(k-1)N+j,t} \right) + \hat{M}^{-1} O_p(T_f^{-1/2}) \left(\frac{1}{T_f} \sum_{t=1}^{T_f} \boldsymbol{f}_t e_{(k-1)N+j,t} \right) + o_p(T_f^{-1/2}) \\ &= A \boldsymbol{\lambda}_{k,j} + (\hat{M}^{-1} - M^{-1}) M A \boldsymbol{\lambda}_{k,j} + M^{-1} \left(\frac{1}{T_f} \sum_{t=1}^{T_f} \boldsymbol{f}_t e_{(k-1)N+j,t} \right) \\ &\quad + (\hat{M}^{-1} - M^{-1}) \left(\frac{1}{T_f} \sum_{t=1}^{T_f} \boldsymbol{f}_t e_{(k-1)N+j,t} \right) + \hat{M}^{-1} O_p(T_f^{-1}) + o_p(T_f^{-1/2}). \end{aligned}$$

Given $M^{-1} = O(1)$ and $\hat{M} - M = O_p(T_f^{-1/2})$, we have $\hat{M}^{-1} - M^{-1} = O_p(T_f^{-1/2})$ and $\hat{M}^{-1} = O_p(1)$ (Lemma B4 of Linton and Tang (2021)). Hence the preceding display becomes

$$\hat{\boldsymbol{\lambda}}_{k,j} - \boldsymbol{\lambda}_{k,j} = A \boldsymbol{\lambda}_{k,j} + M^{-1} \left(\frac{1}{T_f} \sum_{t=1}^{T_f} \boldsymbol{f}_t e_{(k-1)N+j,t} \right) + o_p(T_f^{-1/2}).$$

Write in matrix form:

$$\hat{\Lambda}_k - \Lambda_k = A \Lambda_k + F_k + o_p(T_f^{-1/2}), \tag{A.45}$$

where $\hat{\Lambda}_k := (\hat{\lambda}_{k,1}, \ldots, \hat{\lambda}_{k,N})$, $\Lambda_k := (\lambda_{k,1}, \ldots, \lambda_{k,N})$ and

$$F_k := M^{-1} \frac{1}{T_f} \sum_{t=1}^{T_f} \boldsymbol{f}_t \left[e_{(k-1)N+1,t}, e_{(k-1)N+2,t}, \ldots, e_{kN,t} \right]$$

are $14 \times N$ matrices. We are going to solve (A.45) for A in terms of known elements. The idea is exactly the same as that of the proof of Proposition 4.1.

A.8.1 Step I Impose Some Zero Restrictions in $\{\Lambda_k\}_{k=1}^6$

Let $a \subset \{1, 2, \ldots, 14\}$ and $c \subset \{1, \ldots, N\}$ be two vectors of indices, whose identities vary from place to place. From (A.45), we have

$$\begin{aligned} &\left(\hat{\Lambda}_k - \Lambda_k\right)_{a,c} = A_{a,\bullet} \Lambda_{k,\bullet,c} + F_{k,a,c} + o_p(T_f^{-1/2}) = A_{a,b} \Lambda_{k,b,c} + A_{a,-b} \Lambda_{k,-b,c} + F_{k,a,c} + o_p(T_f^{-1/2}) \\ &= A_{a,b} \Lambda_{k,b,c} + F_{k,a,c} + o_p(T_f^{-1/2}) = A_{a,b_1} \Lambda_{k,b_1,c} + A_{a,b_2} \Lambda_{k,b_2,c} + F_{k,a,c} + o_p(T_f^{-1/2}) \quad \text{(A.46)} \end{aligned}$$

where $b \subset \{1, 2, \ldots, 14\}$ is chosen in such a way such that $\Lambda_{k,-b,c} = 0$ for each of the sub-steps of Step I, $-b$ denotes the complement of b, and $b_1 \cup b_2 = b$ with the cardinality of b_2 equal to the cardinality of c and A_{a,b_1} containing solved elements for each of the sub-steps of Step I. Step I is again detailed in Table 11. Imposing $\left(\hat{\Lambda}_k\right)_{a,c} = 0$ and solving (A.46) gives

$$A_{a,b_2} = -\left(A_{a,b_1}\Lambda_{k,b_1,c} + F_{k,a,c}\right)\Lambda^{-1}_{k,b_2,c} + o_p(T_f^{-1/2}). \tag{A.47}$$

A.8.2 Step II Impose $\hat{M}_{j,j} = 1$ for $j = 10, 11, 12, 13$

Consider (A13) of Bai and Li (2012) and write in our notation:

$$\begin{aligned}
\hat{M} - M = &-\hat{H}\hat{\Lambda}^\intercal\hat{\Sigma}_{ee}^{-1}(\hat{\Lambda} - \Lambda)M - M(\hat{\Lambda} - \Lambda)^\intercal\hat{\Sigma}_{ee}^{-1}\hat{\Lambda}\hat{H} \\
&+ \hat{H}\hat{\Lambda}^\intercal\hat{\Sigma}_{ee}^{-1}(\hat{\Lambda} - \Lambda)M(\hat{\Lambda} - \Lambda)^\intercal\hat{\Sigma}_{ee}^{-1}\hat{\Lambda}\hat{H} \\
&+ \hat{H}\hat{\Lambda}^\intercal\hat{\Sigma}_{ee}^{-1}\Lambda\left(\frac{1}{T_f}\sum_{t=1}^{T_f}\boldsymbol{f}_t\boldsymbol{e}_t^\intercal\right)\hat{\Sigma}_{ee}^{-1}\hat{\Lambda}\hat{H} \\
&+ \hat{H}\hat{\Lambda}^\intercal\hat{\Sigma}_{ee}^{-1}\left(\frac{1}{T_f}\sum_{t=1}^{T_f}\boldsymbol{e}_t\boldsymbol{f}_t^\intercal\right)\Lambda^\intercal\hat{\Sigma}_{ee}^{-1}\hat{\Lambda}\hat{H} \\
&+ \hat{H}\left(\sum_{m=1}^{6N}\sum_{\ell=1}^{6N}\frac{1}{\hat{\sigma}_m^2\hat{\sigma}_\ell^2}\hat{\lambda}_m\hat{\lambda}_\ell^\intercal\frac{1}{T_f}\sum_{t=1}^{T_f}\left[e_{m,t}e_{\ell,t} - \mathbb{E}(e_{m,t}e_{\ell,t})\right]\right)\hat{H} \\
&- \hat{H}\left(\sum_{m=1}^{6N}\frac{1}{\hat{\sigma}_m^4}\hat{\lambda}_m\hat{\lambda}_m^\intercal(\hat{\sigma}_m^2 - \sigma_m^2)\right)\hat{H}.
\end{aligned} \tag{A.48}$$

In the paragraph following (A.43), we already showed that the last four terms of the right hand side of the preceding display are $o_p(T_f^{-1/2})$. Since $A = O_p(T_f^{-1/2})$, $A^\intercal MA = o_p(T_f^{-1/2})$. Thus, (A.48) becomes

$$\hat{M} - M + A^\intercal M + MA = o_p(T_f^{-1/2}). \tag{A.49}$$

Imposing $\hat{M}_{10,10} = 1$, the $(10, 10)$th element of the preceding display satisfies

$$o_p(T_f^{-1/2}) = \hat{M}_{10,10} - 1 + M_{10,\bullet}A_{\bullet,10} + \left(A_{\bullet,10}\right)^\intercal M_{\bullet,10} = 2A_{10,10}$$

whence $A_{10,10} = o_p(T_f^{-1/2})$. In the similar way, imposing $\hat{M}_{11,11}, \hat{M}_{12,12}, \hat{M}_{13,13} = 1$, we could deduce that $A_{11,11}, A_{12,12}, A_{13,13} = o_p(T_f^{-1/2})$.

A.8.3 Step III Impose Some Equality Restrictions in $\{\Lambda_k\}_{k=1}^6$

(III.1) Note that (A.45) implies

$$\begin{aligned}
\hat{\Lambda}_{6,9,c} - \Lambda_{6,9,c} &= A_{9,x}\Lambda_{6,x,c} + F_{6,9,c} + o_p(T_f^{-1/2}) \qquad x = \{1, 2, 3, 9\} \\
&= A_{9,[1,9]}\Lambda_{6,[1,9],c} + A_{9,[2,3]}\Lambda_{6,[2,3],c} + F_{6,9,c} + o_p(T_f^{-1/2}) \\
\hat{\Lambda}_{3,12,c} - \Lambda_{3,12,c} &= A_{12,y}\Lambda_{3,y,c} + F_{3,12,c} + o_p(T_f^{-1/2}) \qquad y = \{4, 5, 6, 12\}.
\end{aligned}$$

We then impose $\hat{\Lambda}_{6,9,c} = \hat{\Lambda}_{3,12,c}$ for $c = \{1,2\}$; that is, the loadings of the continent factor on day one and day two are the same for the first two American assets. The preceding display implies

$$A_{9,[1,9]}\Lambda_{6,[1,9],c} = A_{12,y}\Lambda_{3,y,c} + F_{3,12,c} - A_{9,[2,3]}\Lambda_{6,[2,3],c} - F_{6,9,c} + o_p(T_f^{-1/2})$$

whence we have

$$A_{9,[1,9]} = \left(A_{12,y}\Lambda_{3,y,c} + F_{3,12,c} - A_{9,[2,3]}\Lambda_{6,[2,3],c} - F_{6,9,c}\right)\Lambda_{6,[1,9],c}^{-1} + o_p(T_f^{-1/2}).$$

(III.2) Note that (A.45) implies

$$\begin{aligned}
(\hat{\Lambda}_1 - \Lambda_1)_{14,c} &= A_{14,x}\Lambda_{1,x,c} + F_{1,14,c} + o_p(T_f^{-1/2}) \qquad x = \{6,7,8,14\} \\
&= A_{14,[8,14]}\Lambda_{1,[8,14],c} + A_{14,[6,7]}\Lambda_{1,[6,7],c} + F_{1,14,c} + o_p(T_f^{-1/2}) \\
(\hat{\Lambda}_4 - \Lambda_4)_{11,c} &= A_{11,y}\Lambda_{4,y,c} + F_{4,11,c} + o_p(T_f^{-1/2}) \qquad y = \{3,4,5,11\}.
\end{aligned}$$

We then impose $\hat{\Lambda}_{1,14,c} = \hat{\Lambda}_{4,11,c}$ for $c = \{1,2\}$; that is, the loadings of the continent factor on day one and day two are the same for the first two Asian assets. The preceding display implies

$$A_{14,[8,14]}\Lambda_{1,[8,14],c} = A_{11,y}\Lambda_{4,y,c} + F_{4,11,c} - A_{14,[6,7]}\Lambda_{1,[6,7],c} - F_{1,14,c} + o_p(T_f^{-1/2})$$

whence we have

$$A_{14,[8,14]} = \left(A_{11,y}\Lambda_{4,y,c} + F_{4,11,c} - A_{14,[6,7]}\Lambda_{1,[6,7],c} - F_{1,14,c}\right)\Lambda_{1,[8,14],c}^{-1} + o_p(T_f^{-1/2})$$

A.8.4 Step IV Impose Some Restrictions in M

Impose $\hat{M}_{4,12} = \hat{M}_{5,12} = \hat{M}_{6,12} = 0$. Considering the $(4,12)$th, $(5,12)$th and $(6,12)$th elements of the left hand side of (A.49), we have

$$\begin{aligned}
M_{4,\bullet}A_{\bullet,12} + A_{12,4} &= o_p(T_f^{-1/2}) \\
M_{5,\bullet}A_{\bullet,12} + A_{12,5} &= o_p(T_f^{-1/2}) \\
M_{6,\bullet}A_{\bullet,12} + A_{12,6} &= o_p(T_f^{-1/2})
\end{aligned}$$

with the only unknown elements $A_{4,12}, A_{5,12}$ and $A_{6,12}$. The preceding display could be written as

$$M_{a,a}A_{a,12} + M_{a,b}A_{b,12} + (A_{12,a})^\intercal = o_p(T_f^{-1/2}),$$

where $a := \{4,5,6\}, b := \{1,2,3,7,8,9,10,11,12,13,14\}$. Thus, we obtain

$$A_{a,12} = -(M_{a,a})^{-1}\left(M_{a,b}A_{b,12} + (A_{12,a})^\intercal\right)$$

Similarly, imposing $\hat{M}_{2,10} = \hat{M}_{3,10} = \hat{M}_{4,10} = 0$, we could solve $A_{[2,3,4],10}$. Imposing $\hat{M}_{3,11} = \hat{M}_{4,11} = \hat{M}_{5,11} = 0$, we could solve $A_{[3,4,5],11}$. Imposing $\hat{M}_{5,13} = \hat{M}_{6,13} = \hat{M}_{7,13} = 0$, we could solve $A_{[5,6,7],13}$.

A.8.5 Step V Impose Some Restrictions in M

(V.1) Consider the $(4,4)$th and $(6,6)$th elements of (A.49):

$$\hat{M}_{4,4} - 1/(1-\phi^2) + 2M_{4,4}A_{4,4} + 2M_{4,-4}A_{-4,4} = o_p(T_f^{-1/2})$$
$$\hat{M}_{6,6} - 1/(1-\phi^2) + 2M_{6,6}A_{6,6} + 2M_{6,-6}A_{-6,6} = o_p(T_f^{-1/2})$$

Imposing $\hat{M}_{4,4} = \hat{M}_{6,6}$, we could arrange the preceding display into

$$M_{4,4}A_{4,4} + M_{4,-4}A_{-4,4} = M_{6,6}A_{6,6} + M_{6,-6}A_{-6,6} + o_p(T_f^{-1/2}). \tag{A.50}$$

Next, consider the $(4,4)$th and $(6,4)$th elements of (A.49):

$$\hat{M}_{4,4} - 1/(1-\phi^2) + 2M_{4,4}A_{4,4} + 2M_{4,-4}A_{-4,4} = o_p(T_f^{-1/2})$$
$$\hat{M}_{6,4} - \phi^2/(1-\phi^2) + M_{6,\bullet}A_{\bullet,4} + (A_{\bullet,6})^\intercal M_{\bullet,4} = o_p(T_f^{-1/2}).$$

Imposing $\hat{M}_{4,4} - \hat{M}_{6,4} = 1$, we could arrange the preceding display into

$$2M_{4,4}A_{4,4} + 2M_{4,-4}A_{-4,4} = M_{6,\bullet}A_{\bullet,4} + (A_{\bullet,6})^\intercal M_{\bullet,4} + o_p(T_f^{-1/2}). \tag{A.51}$$

Since there are only two unknown elements in (A.50) and (A.51) (i.e., $A_{4,4}$ and $A_{6,6}$), we could thus solve them. Write (A.50) and (A.51) in matrix

$$\begin{pmatrix} M_{4,4} & -M_{6,6} \\ 2M_{4,4} - M_{6,4} & -M_{4,6} \end{pmatrix} \begin{pmatrix} A_{4,4} \\ A_{6,6} \end{pmatrix} = \begin{pmatrix} M_{6,-6}A_{-6,6} - M_{4,-4}A_{-4,4} \\ M_{4,-6}A_{-6,6} - \left(2M_{4,-4} - M_{6,-4}\right)A_{-4,4} \end{pmatrix}.$$

That is,

$$\begin{pmatrix} A_{4,4} \\ A_{6,6} \end{pmatrix} = \begin{pmatrix} M_{4,4} & -M_{6,6} \\ 2M_{4,4} - M_{6,4} & -M_{4,6} \end{pmatrix}^{-1} \begin{pmatrix} M_{6,-6}A_{-6,6} - M_{4,-4}A_{-4,4} \\ M_{4,-6}A_{-6,6} - \left(2M_{4,-4} - M_{6,-4}\right)A_{-4,4} \end{pmatrix}.$$

(V.2) Consider the $(4,4)$th and $(5,5)$th elements of (A.49):

$$\hat{M}_{4,4} - 1/(1-\phi^2) + 2M_{4,\bullet}A_{\bullet,4} = o_p(T_f^{-1/2})$$
$$\hat{M}_{5,5} - 1/(1-\phi^2) + 2M_{5,5}A_{5,5} + 2M_{5,-5}A_{-5,5} = o_p(T_f^{-1/2})$$

Imposing $\hat{M}_{4,4} = \hat{M}_{5,5}$, we could solve the preceding display for $A_{5,5}$:

$$A_{5,5} = \left(M_{4,\bullet}A_{\bullet,4} - M_{5,-5}A_{-5,5}\right)/M_{5,5} + o_p(T_f^{-1/2})$$

A.8.6 Step VI Impose Some Equality Restrictions in $\{\Lambda_k\}_{k=1}^6$

(VI.1) Note that (A.45) implies

$$\begin{aligned}(\hat{\Lambda}_6 - \Lambda_6)_{1,c} &= A_{1,x}\Lambda_{6,x,c} + F_{6,1,c} + o_p(T_f^{-1/2}) \qquad x = \{1,2,3,9\} \\ &= A_{1,[1,9]}\Lambda_{6,[1,9],c} + A_{1,[2,3]}\Lambda_{6,[2,3],c} + F_{6,1,c} + o_p(T_f^{-1/2}) \\ (\hat{\Lambda}_3 - \Lambda_3)_{4,c} &= A_{4,y}\Lambda_{3,y,c} + F_{3,4,c} + o_p(T_f^{-1/2}) \qquad y = \{4,5,6,12\}.\end{aligned}$$

We then impose $\hat{\Lambda}_{6,1,c} = \hat{\Lambda}_{3,4,c}$ for $c = \{1, 2\}$; that is, the loadings of the continent factor on day one and day two are the same for the first two American assets. The preceding display implies

$$A_{1,[1,9]}\Lambda_{6,[1,9],c} = A_{4,y}\Lambda_{3,y,c} + F_{3,4,c} - A_{1,[2,3]}\Lambda_{6,[2,3],c} - F_{6,1,c} + o_p(T_f^{-1/2})$$

whence we have

$$A_{1,[1,9]} = \left(A_{4,y}\Lambda_{3,y,c} + F_{3,4,c} - A_{1,[2,3]}\Lambda_{6,[2,3],c} - F_{6,1,c}\right)\Lambda^{-1}_{6,[1,9],c} + o_p(T_f^{-1/2}).$$

(VI.2) Note that (A.45) implies

$$\begin{aligned}(\hat{\Lambda}_6 - \Lambda_6)_{2,c} &= A_{2,x}\Lambda_{6,x,c} + F_{6,2,c} + o_p(T_f^{-1/2}) \qquad x = \{1, 2, 3, 9\}\\ &= A_{2,[1,2,9]}\Lambda_{6,[1,2,9],c} + A_{2,3}\Lambda_{6,3,c} + F_{6,2,c} + o_p(T_f^{-1/2})\\ (\hat{\Lambda}_3 - \Lambda_3)_{5,c} &= A_{5,y}\Lambda_{3,y,c} + F_{3,5,c} + o_p(T_f^{-1/2}) \qquad y = \{4, 5, 6, 12\}.\end{aligned}$$

We then impose $\hat{\Lambda}_{6,2,c} = \hat{\Lambda}_{3,5,c}$ for $c = \{1, 2, 3\}$; that is, the first lagged loadings of the global factor on day one and day two are the same for the first two American assets. The preceding display implies

$$A_{2,[1,2,9]}\Lambda_{6,[1,2,9],c} = A_{5,y}\Lambda_{3,y,c} + F_{3,5,c} - A_{2,3}\Lambda_{6,3,c} - F_{6,2,c} + o_p(T_f^{-1/2})$$

whence we have

$$A_{2,[1,2,9]} = \left(A_{5,y}\Lambda_{3,y,c} + F_{3,5,c} - A_{2,3}\Lambda_{6,3,c} - F_{6,2,c}\right)\Lambda^{-1}_{6,[1,2,9],c} + o_p(T_f^{-1/2}).$$

(VI.3) Note that (A.45) implies

$$\begin{aligned}(\hat{\Lambda}_6 - \Lambda_6)_{2,c} &= A_{2,x}\Lambda_{6,x,c} + F_{6,2,c} + o_p(T_f^{-1/2}) \qquad x = \{1, 2, 3, 9\}\\ &= A_{2,[1,2,9]}\Lambda_{6,[1,2,9],c} + A_{2,3}\Lambda_{6,3,c} + F_{6,2,c} + o_p(T_f^{-1/2})\\ (\hat{\Lambda}_3 - \Lambda_3)_{5,c} &= A_{5,y}\Lambda_{3,y,c} + F_{3,5,c} + o_p(T_f^{-1/2}) \qquad y = \{4, 5, 6, 12\}.\end{aligned}$$

We then impose $\hat{\Lambda}_{6,2,c} = \hat{\Lambda}_{3,5,c}$ for $c = \{1, 2, 3\}$; that is, the first lagged loadings of the global factor on day one and day two are the same for the first three American assets. The preceding display implies

$$A_{2,[1,2,9]}\Lambda_{6,[1,2,9],c} = A_{5,y}\Lambda_{3,y,c} + F_{3,5,c} - A_{2,3}\Lambda_{6,3,c} - F_{6,2,c} + o_p(T_f^{-1/2})$$

whence we have

$$A_{2,[1,2,9]} = \left(A_{5,y}\Lambda_{3,y,c} + F_{3,5,c} - A_{2,3}\Lambda_{6,3,c} - F_{6,2,c}\right)\Lambda^{-1}_{6,[1,2,9],c} + o_p(T_f^{-1/2}).$$

(VI.4) Note that (A.45) implies

$$\begin{aligned}(\hat{\Lambda}_6 - \Lambda_6)_{3,c} &= A_{3,x}\Lambda_{6,x,c} + F_{6,3,c} + o_p(T_f^{-1/2}) \qquad x = \{1, 2, 3, 9\}\\ (\hat{\Lambda}_3 - \Lambda_3)_{6,c} &= A_{6,y}\Lambda_{3,y,c} + F_{3,6,c} + o_p(T_f^{-1/2}) \qquad y = \{4, 5, 6, 12\}.\end{aligned}$$

We then impose $\hat{\Lambda}_{6,3,c} = \hat{\Lambda}_{3,6,c}$ for $c = \{1, 2, 3, 4\}$; that is, the second lagged loadings of the global factor on day one and day two are the same for the first four American assets. The preceding display implies

$$A_{3,x}\Lambda_{6,x,c} = A_{6,y}\Lambda_{3,y,c} + F_{3,6,c} - F_{6,3,c} + o_p(T_f^{-1/2})$$

whence we have

$$A_{3,x} = \left(A_{6,y}\Lambda_{3,y,c} + F_{3,6,c} - F_{6,3,c}\right)\Lambda_{6,x,c}^{-1} + o_p(T_f^{-1/2})$$

(VI.5) Note that (A.45) implies

$$\begin{aligned}(\hat{\Lambda}_1 - \Lambda_1)_{8,c} &= A_{8,x}\Lambda_{1,x,c} + F_{1,8,c} + o_p(T_f^{-1/2}) \qquad x = \{6, 7, 8, 14\}\\ &= A_{8,[8,14]}\Lambda_{1,[8,14],c} + A_{8,[6,7]}\Lambda_{1,[6,7],c} + F_{1,8,c} + o_p(T_f^{-1/2})\\ (\hat{\Lambda}_4 - \Lambda_4)_{5,c} &= A_{5,y}\Lambda_{4,y,c} + F_{4,5,c} + o_p(T_f^{-1/2}) \qquad y = \{3, 4, 5, 11\}.\end{aligned}$$

We then impose $\hat{\Lambda}_{1,8,c} = \hat{\Lambda}_{4,5,c}$ for $c = \{1, 2\}$; that is, the second lagged loadings of the global factor on day one and day two are the same for the first two Asian assets. The preceding display implies

$$A_{8,[8,14]}\Lambda_{1,[8,14],c} = A_{5,y}\Lambda_{4,y,c} + F_{4,5,c} - A_{8,[6,7]}\Lambda_{1,[6,7],c} - F_{1,8,c} + o_p(T_f^{-1/2})$$

whence we have

$$A_{8,[8,14]} = \left(A_{5,y}\Lambda_{4,y,c} + F_{4,5,c} - A_{8,[6,7]}\Lambda_{1,[6,7],c} - F_{1,8,c}\right)\Lambda_{1,[8,14],c}^{-1} + o_p(T_f^{-1/2}).$$

(VI.6) Note that (A.45) implies

$$\begin{aligned}(\hat{\Lambda}_1 - \Lambda_1)_{7,c} &= A_{7,x}\Lambda_{1,x,c} + F_{1,7,c} + o_p(T_f^{-1/2}) \qquad x = \{6, 7, 8, 14\}\\ &= A_{7,[7,8,14]}\Lambda_{1,[7,8,14],c} + A_{7,6}\Lambda_{1,6,c} + F_{1,7,c} + o_p(T_f^{-1/2})\\ (\hat{\Lambda}_4 - \Lambda_4)_{4,c} &= A_{4,y}\Lambda_{4,y,c} + F_{4,4,c} + o_p(T_f^{-1/2}) \qquad y = \{3, 4, 5, 11\}.\end{aligned}$$

We then impose $\hat{\Lambda}_{1,7,c} = \hat{\Lambda}_{4,4,c}$ for $c = \{1, 2, 3\}$; that is, the first lagged loadings of the global factor on day one and day two are the same for the first three Asian assets. The preceding display implies

$$A_{7,[7,8,14]}\Lambda_{1,[7,8,14],c} = A_{4,y}\Lambda_{4,y,c} + F_{4,4,c} - A_{7,6}\Lambda_{1,6,c} - F_{1,7,c} + o_p(T_f^{-1/2})$$

whence we have

$$A_{7,[7,8,14]} = \left(A_{4,y}\Lambda_{4,y,c} + F_{4,4,c} - A_{7,6}\Lambda_{1,6,c} - F_{1,7,c}\right)\Lambda_{1,[7,8,14],c}^{-1} + o_p(T_f^{-1/2}).$$

(VI.7) Note that (A.45) implies

$$\begin{aligned}(\hat{\Lambda}_1 - \Lambda_1)_{6,c} &= A_{6,x}\Lambda_{1,x,c} + F_{1,6,c} + o_p(T_f^{-1/2}) \qquad x = \{6, 7, 8, 14\}\\ &= A_{6,[7,8,14]}\Lambda_{1,[7,8,14],c} + A_{6,6}\Lambda_{1,6,c} + F_{1,6,c} + o_p(T_f^{-1/2})\\ (\hat{\Lambda}_4 - \Lambda_4)_{3,c} &= A_{3,y}\Lambda_{4,y,c} + F_{4,3,c} + o_p(T_f^{-1/2}) \qquad y = \{3, 4, 5, 11\}.\end{aligned}$$

We then impose $\hat{\Lambda}_{1,6,c} = \hat{\Lambda}_{4,3,c}$ for $c = \{1,2,3\}$; that is, the contemporaneous loadings of the global factor on day one and day two are the same for the first three Asian assets. The preceding display implies

$$A_{6,[7,8,14]}\Lambda_{1,[7,8,14],c} = A_{3,y}\Lambda_{4,y,c} + F_{4,3,c} - A_{6,6}\Lambda_{1,6,c} - F_{1,6,c} + o_p(T_f^{-1/2})$$

whence we have

$$A_{6,[7,8,14]} = \left(A_{3,y}\Lambda_{4,y,c} + F_{4,3,c} - A_{6,6}\Lambda_{1,6,c} - F_{1,6,c}\right)\Lambda^{-1}_{1,[7,8,14],c} + o_p(T_f^{-1/2})$$

A.8.7 To Sum Up

Recall that (A.45) implies that for $k = 1,\dots,6$, $j = 1,\dots,N$,

$$\sqrt{T_f}(\hat{\boldsymbol{\lambda}}_{k,j} - \boldsymbol{\lambda}_{k,j}) = \sqrt{T_f}A\boldsymbol{\lambda}_{k,j} + M^{-1}\frac{1}{\sqrt{T_f}}\sum_{t=1}^{T_f}\boldsymbol{f}_t e_{(k-1)N+j,t} + o_p(1). \tag{A.52}$$

We have just shown that *each* element of the 14×14 matrix A could be expressed into some known (but complicated) linear function involving elements of the following six matrices:

$$M^{-1}\frac{1}{T_f}\sum_{t=1}^{T_f}\boldsymbol{f}_t\left[e_{(k-1)N+1,t}, e_{(k-1)N+2,t}, e_{(k-1)N+3,t}, e_{(k-1)N+4,t}\right]$$

for $k = 1,\dots,6$, plus $o_p(T_f^{-1/2})$. That is, there exists a 196×336 matrix Γ, whose elements are known (but complicated) linear functions of elements of (inverted) submatrices of Λ and M, satisfying

$$\operatorname{vec} A = \Gamma \times \frac{1}{T_f}\sum_{t=1}^{T_f}(\boldsymbol{e}_t^\dagger \otimes \boldsymbol{f}_t) + o_p(T_f^{-1/2}),$$

where $\boldsymbol{e}_t^\dagger$ is a 24×1 vector consisting of $e_{(p-1)N+q,t}$ for $p = 1,\dots,6$ and $q = 1,\dots,4$. Thus (A.52) could be written as

$$\begin{aligned}
&\sqrt{T_f}(\hat{\boldsymbol{\lambda}}_{k,j} - \boldsymbol{\lambda}_{k,j}) \\
&= \sqrt{T_f}(\boldsymbol{\lambda}_{k,j}^\intercal \otimes I_{14})\operatorname{vec} A + M^{-1}\frac{1}{\sqrt{T_f}}\sum_{t=1}^{T_f}\boldsymbol{f}_t e_{(k-1)N+j,t} + o_p(1) \\
&= (\boldsymbol{\lambda}_{k,j}^\intercal \otimes I_{14})\Gamma\frac{1}{\sqrt{T_f}}\sum_{t=1}^{T_f}\left(\boldsymbol{e}_t^\dagger \otimes \boldsymbol{f}_t\right) + M^{-1}\frac{1}{\sqrt{T_f}}\sum_{t=1}^{T_f}\boldsymbol{f}_t e_{(k-1)N+j,t} + o_p(1).
\end{aligned} \tag{A.53}$$

In SM B.5, we show that

$$\begin{aligned}
&\frac{1}{\sqrt{T_f}}\sum_{t=1}^{T_f}\begin{bmatrix} (\boldsymbol{\lambda}_{k,j}^\intercal \otimes I_{14})\Gamma(\boldsymbol{e}_t^\dagger \otimes \boldsymbol{f}_t) \\ M^{-1}\boldsymbol{f}_t e_{(k-1)N+j,t} \end{bmatrix} \xrightarrow{d} \\
&N\left(\begin{bmatrix} \mathbf{0} \\ \mathbf{0} \end{bmatrix}\begin{bmatrix} (\boldsymbol{\lambda}_{k,j}^\intercal \otimes I_{14})\Gamma(\Sigma_{ee}^\dagger \otimes M)\Gamma^\intercal(\boldsymbol{\lambda}_{k,j} \otimes I_{14}) & \operatorname{cov}_{k,j} \\ \operatorname{cov}_{k,j} & M^{-1}\sigma_{k,j}^2 \end{bmatrix}\right)
\end{aligned} \tag{A.54}$$

where $\Sigma_{ee}^{\dagger} := \mathbb{E}[\boldsymbol{e}_t^{\dagger}\boldsymbol{e}_t^{\dagger\intercal}]$ and $\mathrm{cov}_{k,j}$ is an 14×14 matrix defined as

$$\mathrm{cov}_{k,j} := \mathrm{cov}\left((\boldsymbol{\lambda}_{k,j}^{\intercal} \otimes I_{14})\Gamma(\boldsymbol{e}_t^{\dagger} \otimes \boldsymbol{f}_t), e_{(k-1)N+j,t}\boldsymbol{f}_t^{\intercal}M^{-1}\right).$$

By Assumption 2.1, we have $\mathrm{cov}_{k,j} = \mathbf{0}$ for $j > 4$, and

$$\begin{aligned}\mathrm{cov}_{k,j} &= (\boldsymbol{\lambda}_{k,j}^{\intercal} \otimes I_{14})\Gamma\mathbb{E}\left[\boldsymbol{e}_t^{\dagger}e_{(k-1)N+j,t} \otimes \boldsymbol{f}_t\boldsymbol{f}_t^{\intercal}\right]M^{-1} = (\boldsymbol{\lambda}_{k,j}^{\intercal} \otimes I_{14})\Gamma\left[\sigma_{k,j}^2\boldsymbol{\iota}_{k,j} \otimes M\right]M^{-1}\\ &= (\boldsymbol{\lambda}_{k,j}^{\intercal} \otimes I_{14})\Gamma\left[\boldsymbol{\iota}_{k,j} \otimes I_{14}\right]\sigma_{k,j}^2,\end{aligned}$$

for $j \leq 4$, where $\boldsymbol{\iota}_{k,j}$ is a 24×1 zero vector with its $[4(k-1)+j]$th element replaced by one. Thus for $j > 4$, we have

$$\sqrt{T_f}(\hat{\boldsymbol{\lambda}}_{k,j} - \boldsymbol{\lambda}_{k,j}) \xrightarrow{d} N\left(\mathbf{0}, (\boldsymbol{\lambda}_{k,j}^{\intercal} \otimes I_{14})\Gamma(\Sigma_{ee}^{\dagger} \otimes M)\Gamma^{\intercal}(\boldsymbol{\lambda}_{k,j} \otimes I_{14}) + M^{-1}\sigma_{k,j}^2\right).$$

For $j \leq 4$, we have

$$\sqrt{T_f}(\hat{\boldsymbol{\lambda}}_{k,j} - \boldsymbol{\lambda}_{k,j}) \xrightarrow{d} N\left(\mathbf{0}, (\boldsymbol{\lambda}_{k,j}^{\intercal} \otimes I_{14})\Gamma(\Sigma_{ee}^{\dagger} \otimes M)\Gamma^{\intercal}(\boldsymbol{\lambda}_{k,j} \otimes I_{14}) + M^{-1}\sigma_{k,j}^2 + \mathrm{cov}_{k,j} + \mathrm{cov}_{k,j}^{\intercal}\right).$$

A.9 Proof of Theorem 4.4

From (A.49), we have

$$\sqrt{T_f}(\hat{M} - M) = -\sqrt{T_f}\left(A^{\intercal}M + MA\right) + o_p(1).$$

whence we have

$$\begin{aligned}&\sqrt{T_f}\,\mathrm{vech}(\hat{M} - M) = -\sqrt{T_f}\,\mathrm{vech}\left(A^{\intercal}M + MA\right) + o_p(1) = -\sqrt{T_f}D_{14}^{+}\,\mathrm{vec}\left(A^{\intercal}M + MA\right) + o_p(1)\\ &= -\sqrt{T_f}D_{14}^{+}\left[(M \otimes I_{14})\,\mathrm{vec}(A^{\intercal}) + (I_{14} \otimes M)\,\mathrm{vec}\,A\right] + o_p(1)\\ &= -\sqrt{T_f}D_{14}^{+}\left[(M \otimes I_{14})K_{14,14}\,\mathrm{vec}\,A + (I_{14} \otimes M)\,\mathrm{vec}\,A\right] + o_p(1)\\ &= -\sqrt{T_f}D_{14}^{+}\left[K_{14,14}(I_{14} \otimes M)\,\mathrm{vec}\,A + (I_{14} \otimes M)\,\mathrm{vec}\,A\right] + o_p(1)\\ &= -\sqrt{T_f}D_{14}^{+}(K_{14,14} + I_{14^2})(I_{14} \otimes M)\,\mathrm{vec}\,A + o_p(1) = -2\sqrt{T_f}D_{14}^{+}D_{14}D_{14}^{+}(I_{14} \otimes M)\,\mathrm{vec}\,A + o_p(1)\\ &= -2D_{14}^{+}(I_{14} \otimes M)\Gamma\frac{1}{\sqrt{T_f}}\sum_{t=1}^{T_f}\left(\boldsymbol{e}_t^{\dagger} \otimes \boldsymbol{f}_t\right) + o_p(1)\end{aligned}$$

where the second equality is due to symmetry of $A^{\intercal}M + MA$, and the fifth and seventh equalities are due to properties of $K_{14,14}$. Thus we have

$$\sqrt{T_f}\,\mathrm{vech}(\hat{M} - M) \xrightarrow{d} N\left(0, \mathcal{M}\right)$$

where $\mathcal{M}$ is 105×105 and defined

$$\mathcal{M} := 4D_{14}^{+}(I_{14} \otimes M)\Gamma(\Sigma_{ee}^{\dagger} \otimes M)\Gamma^{\intercal}(I_{14} \otimes M)D_{14}^{+\intercal}.$$

A.10 Proof of Theorem 4.5

$$\begin{aligned}\sqrt{N}(\hat{\boldsymbol{f}}_t - \boldsymbol{f}_t) &= -\sqrt{N}\left(\hat{\Lambda}^{\intercal}\hat{\Sigma}_{ee}^{-1}\hat{\Lambda}\right)^{-1}\hat{\Lambda}^{\intercal}\hat{\Sigma}_{ee}^{-1}(\hat{\Lambda} - \Lambda)\boldsymbol{f}_t + \sqrt{N}\left(\hat{\Lambda}^{\intercal}\hat{\Sigma}_{ee}^{-1}\hat{\Lambda}\right)^{-1}\hat{\Lambda}^{\intercal}\hat{\Sigma}_{ee}^{-1}\boldsymbol{e}_t\\ &= -\sqrt{\frac{N}{T_f}}\sqrt{T_f}A^{\intercal}\boldsymbol{f}_t + \sqrt{N}\left(\frac{1}{N}\hat{\Lambda}^{\intercal}\hat{\Sigma}_{ee}^{-1}\hat{\Lambda}\right)^{-1}\frac{1}{N}\hat{\Lambda}^{\intercal}\hat{\Sigma}_{ee}^{-1}\boldsymbol{e}_t. \qquad \text{(A.55)}\end{aligned}$$

Lemma D1 of Bai and Li (2012) still holds in our setting and it reads, in our notation:

$$\frac{1}{N}\hat{\Lambda}^\intercal\hat{\Sigma}_{ee}^{-1}\boldsymbol{e}_t = \frac{1}{N}\Lambda^\intercal\Sigma_{ee}^{-1}\boldsymbol{e}_t + O_p\left(N^{-1/2}T_f^{-1/2}\right) + O_p(T_f^{-1}). \tag{A.56}$$

$$\left(\frac{1}{N}\hat{\Lambda}^\intercal\hat{\Sigma}_{ee}^{-1}\hat{\Lambda}\right)^{-1} = Q^{-1} + o_p(1). \tag{A.57}$$

Substituting (A.56) and (A.57) into (A.55), we have

$$\begin{aligned}
&\sqrt{N}(\hat{\boldsymbol{f}}_t - \boldsymbol{f}_t)\\
&= -\sqrt{\frac{N}{T_f}}\sqrt{T_f}A^\intercal\boldsymbol{f}_t + \sqrt{N}\left(Q^{-1}+o_p(1)\right)\left[\frac{1}{N}\Lambda^\intercal\Sigma_{ee}^{-1}\boldsymbol{e}_t + O_p\left(N^{-1/2}T_f^{-1/2}\right) + O_p(T_f^{-1})\right]\\
&= -\sqrt{\frac{N}{T_f}}\sqrt{T_f}A^\intercal\boldsymbol{f}_t + \left(Q^{-1}+o_p(1)\right)\left[\frac{1}{\sqrt{N}}\Lambda^\intercal\Sigma_{ee}^{-1}\boldsymbol{e}_t + O_p\left(T_f^{-1/2}\right) + O_p(\sqrt{N}T_f^{-1})\right]\\
&= -\sqrt{\Delta}\sqrt{T_f}A^\intercal\boldsymbol{f}_t + \left(Q^{-1}+o_p(1)\right)\left[\frac{1}{\sqrt{N}}\Lambda^\intercal\Sigma_{ee}^{-1}\boldsymbol{e}_t + O_p\left(T_f^{-1/2}\right) + o_p(1)\right]\\
&= -\sqrt{\Delta}\sqrt{T_f}A^\intercal\boldsymbol{f}_t + Q^{-1}\frac{1}{\sqrt{N}}\Lambda^\intercal\Sigma_{ee}^{-1}\boldsymbol{e}_t + o_p(1)\\
&= -\sqrt{\Delta}(\boldsymbol{f}_t^\intercal\otimes I_{14})K_{14,14}\Gamma\frac{1}{\sqrt{T_f}}\sum_{t=1}^{T_f}\left(\boldsymbol{e}_t^\dagger\otimes\boldsymbol{f}_t\right) + Q^{-1}\frac{1}{\sqrt{N}}\Lambda^\intercal\Sigma_{ee}^{-1}\boldsymbol{e}_t + o_p(1)
\end{aligned}\tag{A.58}$$

where the third equality is due to $\sqrt{N}/T_f \to 0$, $N/T_f \to \Delta$ and $\sqrt{T_f}A^\intercal\boldsymbol{f}_t = O_p(1)$, and the fourth equality uses the fact that $Q^{-1} = O_p(1)$ and $N^{-1/2}\Lambda^\intercal\Sigma_{ee}^{-1}\boldsymbol{e}_t = O_p(1)$ by the central limit theorem.

The first two terms on the right side of (A.58), conditioning on $\boldsymbol{f}_t$, are asymptotically normal and asymptotically independent. The former uses the central limit theorem over the time dimension (only depends on the first 4 assets in each continent) and the latter uses the central limit theorem over the cross-sectional dimension. In particular,

$$\left.\begin{pmatrix} -\sqrt{\Delta}(\boldsymbol{f}_t^\intercal\otimes I_{14})K_{14,14}\Gamma\frac{1}{\sqrt{T_f}}\sum_{t=1}^{T_f}\left(\boldsymbol{e}_t^\dagger\otimes\boldsymbol{f}_t\right)\\ Q^{-1}\frac{1}{\sqrt{N}}\Lambda^\intercal\Sigma_{ee}^{-1}\boldsymbol{e}_t\end{pmatrix}\right|\boldsymbol{f}_t \xrightarrow{d} N\left(\begin{bmatrix}\mathbf{0}\\ \mathbf{0}\end{bmatrix}\begin{bmatrix}\maltese & \mathbf{0}\\ \mathbf{0} & Q^{-1}\end{bmatrix}\right)$$

where

$$\maltese := \Delta(\boldsymbol{f}_t^\intercal\otimes I_{14})K_{14,14}\Gamma(\Sigma_{ee}^\dagger\otimes M)\Gamma^\intercal K_{14,14}(\boldsymbol{f}_t\otimes I_{14}).$$

Then the result of the theorem follows.

A.11 Proof of Theorem 4.6

Proof of Theorem 4.6. Recall (3.8)

$$\check{\boldsymbol{\theta}}_m := \arg\min_{\boldsymbol{b}\in\mathbb{R}^{c_2}}\left[\hat{\boldsymbol{h}} - h(\boldsymbol{b})\right]^\intercal W\left[\hat{\boldsymbol{h}} - h(\boldsymbol{b})\right].$$

The minimum distance estimator $\check{\boldsymbol{\theta}}_m$ satisfies the first-order condition:

$$\frac{\partial h(\check{\boldsymbol{\theta}}_m)}{\partial\boldsymbol{\theta}_m^\intercal}W\left[\hat{\boldsymbol{h}} - h(\check{\boldsymbol{\theta}}_m)\right] = 0. \tag{A.59}$$

Do a Taylor expansion

$$h(\check{\boldsymbol{\theta}}_m) = h(\boldsymbol{\theta}_m) + \frac{\partial h(\dot{\boldsymbol{\theta}}_m)}{\partial \boldsymbol{\theta}_m}(\check{\boldsymbol{\theta}}_m - \boldsymbol{\theta}_m),$$

where $\dot{\boldsymbol{\theta}}_m$ is a mid-point between $\check{\boldsymbol{\theta}}_m$ and $\boldsymbol{\theta}_m$. Substituting the preceding display into (A.59), we have

$$\frac{\partial h(\check{\boldsymbol{\theta}}_m)}{\partial \boldsymbol{\theta}_m^\intercal} W \left[\hat{\boldsymbol{h}} - h(\boldsymbol{\theta}_m) - \frac{\partial h(\dot{\boldsymbol{\theta}}_m)}{\partial \boldsymbol{\theta}_m}(\check{\boldsymbol{\theta}}_m - \boldsymbol{\theta}_m)\right] = 0$$

whence we have

$$\begin{aligned}\sqrt{T_f}(\check{\boldsymbol{\theta}}_m - \boldsymbol{\theta}_m) &= \left[\frac{\partial h(\check{\boldsymbol{\theta}}_m)}{\partial \boldsymbol{\theta}_m^\intercal} W \frac{\partial h(\dot{\boldsymbol{\theta}}_m)}{\partial \boldsymbol{\theta}_m}\right]^{-1} \frac{\partial h(\check{\boldsymbol{\theta}}_m)}{\partial \boldsymbol{\theta}_m^\intercal} W \left[\hat{\boldsymbol{h}} - h(\boldsymbol{\theta}_m)\right] \\ &\xrightarrow{d} \left[\frac{\partial h(\boldsymbol{\theta}_m)}{\partial \boldsymbol{\theta}_m^\intercal} W \frac{\partial h(\boldsymbol{\theta}_m)}{\partial \boldsymbol{\theta}_m}\right]^{-1} \frac{\partial h(\boldsymbol{\theta}_m)}{\partial \boldsymbol{\theta}_m^\intercal} W N\left(0, \mathcal{H}\right),\end{aligned}$$

where the convergence in distribution follows from consistency of $\check{\boldsymbol{\theta}}_m$ (i.e., $\check{\boldsymbol{\theta}}_m \xrightarrow{p} \boldsymbol{\theta}_m$). The proof of consistency of $\check{\boldsymbol{\theta}}_m$ is given in SM B.6. Then we have

$$\sqrt{T_f}(\check{\boldsymbol{\theta}}_m - \boldsymbol{\theta}_m) \xrightarrow{d} N\left(0, \mathcal{O}\right).$$

□

A.12 Formulas for $\vec{\mathbb{E}}[\boldsymbol{f}_t \boldsymbol{f}_t^\intercal]$ and $\vec{\mathbb{E}}[\boldsymbol{f}_t \mathring{\boldsymbol{y}}_t^\intercal]$

In this subsection, we give the formulas for $\vec{\mathbb{E}}[\boldsymbol{f}_t \boldsymbol{f}_t^\intercal]$ and $\vec{\mathbb{E}}[\boldsymbol{f}_t \mathring{\boldsymbol{y}}_t^\intercal]$ defined in Section 3.4. We know that

$$\begin{pmatrix} \boldsymbol{f}_t \\ \mathring{\boldsymbol{y}}_t \end{pmatrix} \sim N\left(\begin{pmatrix} \mathbf{0} \\ \mathbf{0} \end{pmatrix} \begin{pmatrix} M & M\Lambda^\intercal \\ \Lambda M & \Sigma_{yy} \end{pmatrix}\right).$$

Recall that in Section 3.4 we treat $\{\boldsymbol{f}_t\}_{t=1}^{T_f}$ as i.i.d. Thus, according to the conditional distribution of the multivariate normal, we have

$$\begin{aligned}&\mathbb{E}[\boldsymbol{f}_t | \{\mathring{\boldsymbol{y}}_t\}_{t=1}^{T_f}; \vec{\boldsymbol{\theta}}^{(i)}] = M\Lambda^\intercal \Sigma_{yy}^{-1} \mathring{\boldsymbol{y}}_t \\ &\mathrm{var}[\boldsymbol{f}_t | \{\mathring{\boldsymbol{y}}_t\}_{t=1}^{T_f}; \vec{\boldsymbol{\theta}}^{(i)}] = M - M\Lambda^\intercal \Sigma_{yy}^{-1} \Lambda M = \mathbb{E}[\boldsymbol{f}_t \boldsymbol{f}_t^\intercal | \{\mathring{\boldsymbol{y}}_t\}_{t=1}^{T_f}; \vec{\boldsymbol{\theta}}^{(i)}] - \mathbb{E}[\boldsymbol{f}_t | \{\mathring{\boldsymbol{y}}_t\}_{t=1}^{T_f}; \vec{\boldsymbol{\theta}}^{(i)}] \mathbb{E}[\boldsymbol{f}_t^\intercal | \{\mathring{\boldsymbol{y}}_t\}_{t=1}^{T_f}; \vec{\boldsymbol{\theta}}^{(i)}].\end{aligned}$$

We then can show that

$$\begin{aligned}&\frac{1}{T_f} \sum_{t=1}^{T_f} \vec{\mathbb{E}}[\mathring{\boldsymbol{y}}_t \boldsymbol{f}_t^\intercal] = S_{yy} \Sigma_{yy}^{-1} \Lambda M \\ &\frac{1}{T_f} \sum_{t=1}^{T_f} \vec{\mathbb{E}}[\boldsymbol{f}_t \boldsymbol{f}_t^\intercal] = M - M\Lambda^\intercal \Sigma_{yy}^{-1} \Lambda M + M\Lambda^\intercal \Sigma_{yy}^{-1} S_{yy} \Sigma_{yy}^{-1} \Lambda M. \end{aligned} \tag{A.60}$$

A.13 Computation of the QMLE

In this subsection, we provide a way to compute the QMLE defined in Section 3.2. Again we will rely on the EM algorithm.

(i) In the E step, calculate $T_f^{-1}\sum_{t=1}^{T_f}\vec{\mathbb{E}}[\mathring{\boldsymbol{y}}_t\boldsymbol{f}_t^\intercal]$ and $T_f^{-1}\sum_{t=1}^{T_f}\vec{\mathbb{E}}[\boldsymbol{f}_t\boldsymbol{f}_t^\intercal]$ as in (A.60).

(ii) In the M step, obtain the factor loading estimates similar to those of the QMLE-res by imposing the factor loading restrictions within the 14^2 restrictions. We only have to change the selection matrices, say, L_1 and L_4 defined in Section 3.4, accordingly.

(iii) Iterate steps (i) and (ii) until the estimates $\hat{\Lambda}, \hat{\Sigma}_{ee}, \hat{M}$ satisfy (3.7) reasonably well.

(iv) Rotate the converged $\hat{\Lambda}$ and $\hat{M}$ so that the rotated $\hat{\Lambda}$ and $\hat{M}$ satisfy all the 14^2 restrictions. This could only be done numerically. In particular, we define a distance function which measures the distance between the restricted elements in Λ and M and the corresponding elements in the rotated $\hat{\Lambda}$ and $\hat{M}$.

B Supplementary Materials

B.1 Missing Because of Continent-Specific Reasons

In this subsection, we discuss how to alter the EM algorithm if we include the scenario of missing observations due to continent-specific reasons such as continent-wide public holidays (e.g., Chinese New Year).

Suppose that continent c last traded at $t = t_1 - 3$, did not open for trading at $t = t_1, t_1 + 3, \ldots, t_1 + 3(\tau - 1)$, and re-opened for trading at $t_1 + 3\tau$, where τ is some integer ≥ 1. There are actually four statuses for this continent at a particular t: *Trade*, *NA*, *Closure* and *Re-open*. Status *Trade* means that the stock market opens normally (e.g., $t = t_1 - 3$), while status *NA* means that it is a time when the stock market closes because of non-synchronised trading (e.g., $t = t_1 - 2$); status *Closure* means that the stock market does not open for trading because of public holidays (e.g., $t = t_1$), while status *Re-open* means the stock market re-opens after public holidays (e.g., $t = t_1 + 3\tau$). Define $\boldsymbol{y}_t^*$ such that

$$\boldsymbol{y}_t^* = \begin{cases} \boldsymbol{y}_t & \text{if Status}_t = \mathit{Trade} \text{ or } \mathit{NA} \\ \boldsymbol{\varepsilon}_t^* & \text{if Status}_t = \mathit{Closure} \\ \sum_{i=0}^{\tau}\boldsymbol{y}_{t_1+3i} & \text{if Status}_t = \mathit{Re\text{-}open} \end{cases} \tag{B.1}$$

where $\boldsymbol{\varepsilon}_t^*$ is to be defined shortly. The idea is that $\boldsymbol{y}_{t_1+3\tau}^*$ is the actual *observed* returns on $t = t_1 + 3\tau$. We now give a concrete example to illustrate this. Suppose that the European continent did not open for trading on $t = 32, 35$ because of some public holiday. Then the values of $\boldsymbol{y}_t^*$ are given in Table 12.

Recall (2.3):

$$\begin{aligned} \boldsymbol{y}_t &= Z_t\boldsymbol{\alpha}_t + \boldsymbol{\varepsilon}_t, & \boldsymbol{\varepsilon}_t &\sim N(0, \Sigma_t) \\ \boldsymbol{\alpha}_{t+1} &= \mathcal{T}\boldsymbol{\alpha}_t + R\boldsymbol{\eta}_t, & \boldsymbol{\eta}_t &\sim N(0, I_2). \end{aligned}$$

	E	U	A	E	U	A	E	U	A	E	U
$t =$	29	30	31	32	33	34	35	36	37	38	39
Status$_t$	*Trade*	*NA*	*NA*	*Closure*	*NA*	*NA*	*Closure*	*NA*	*NA*	*Re-open*	*NA*
$\boldsymbol{y}_t^* =$	$\boldsymbol{y}_{29}$	$\boldsymbol{y}_{30}$	$\boldsymbol{y}_{31}$	$\boldsymbol{\varepsilon}_{32}^*$	$\boldsymbol{y}_{33}$	$\boldsymbol{y}_{34}$	$\boldsymbol{\varepsilon}_{35}^*$	$\boldsymbol{y}_{36}$	$\boldsymbol{y}_{37}$	$\boldsymbol{y}_{32} + \boldsymbol{y}_{35} + \boldsymbol{y}_{38}$	$\boldsymbol{y}_{39}$

Table 12: Letters A, E, U denote the times when we should theoretically observe the closing prices of stocks of in the Asian continent (A), European continent (E), and American continent (U), respectively. Status *Trade* means that the stock market opens normally, while status *NA* means that it is a time when the stock market closes because of non-synchronised trading; status *Closure* means that the stock market does not open for trading because of public holidays, while status *Re-open* means the stock market re-opens after public holidays.

We now write down the sate space model for $\boldsymbol{y}_t^*$.

$$\begin{aligned} \boldsymbol{y}_t^* &= Z_t^* \boldsymbol{\alpha}_t^* + \boldsymbol{\varepsilon}_t^*, \qquad \boldsymbol{\varepsilon}_t \sim N(0, \Sigma_t^*) \\ \boldsymbol{\alpha}_{t+1}^* &= \mathcal{T}_t^* \boldsymbol{\alpha}_t^* + R_t^* \boldsymbol{\eta}_t, \qquad \boldsymbol{\eta}_t \sim N(0, I_2). \end{aligned} \tag{B.2}$$

Parameters $Z_t^*, \boldsymbol{\alpha}_t^*, \Sigma_t^*, \mathcal{T}_t^*, R_t^*$ should be chosen in such a way that (B.2) is consistent with (B.1). In our example, $Z_t^*, \boldsymbol{\alpha}_t^*, \Sigma_t^*, \mathcal{T}_t^*, R_t^*$ take the following values:

$$\boldsymbol{\alpha}_t^* = \boldsymbol{\alpha}_t \qquad t = 29, 30, 31, 32, 39$$

$$\boldsymbol{\alpha}_{33}^* = \begin{bmatrix} \boldsymbol{\alpha}_{33} \\ \mathbf{0}_{4\times1} \\ \boldsymbol{\alpha}_{32} \\ \mathbf{0}_{4\times1} \end{bmatrix}, \boldsymbol{\alpha}_{34}^* = \begin{bmatrix} \boldsymbol{\alpha}_{34} \\ \mathbf{0}_{4\times1} \\ \boldsymbol{\alpha}_{32} \\ \mathbf{0}_{4\times1} \end{bmatrix}, \boldsymbol{\alpha}_{35}^* = \begin{bmatrix} \boldsymbol{\alpha}_{35} \\ \mathbf{0}_{4\times1} \\ \boldsymbol{\alpha}_{32} \\ \mathbf{0}_{4\times1} \end{bmatrix}$$

$$\boldsymbol{\alpha}_{36}^* = \begin{bmatrix} \boldsymbol{\alpha}_{36} \\ \mathbf{0}_{4\times1} \\ \boldsymbol{\alpha}_{32} + \boldsymbol{\alpha}_{35} \\ \mathbf{0}_{4\times1} \end{bmatrix}, \boldsymbol{\alpha}_{37}^* = \begin{bmatrix} \boldsymbol{\alpha}_{37} \\ \mathbf{0}_{4\times1} \\ \boldsymbol{\alpha}_{32} + \boldsymbol{\alpha}_{35} \\ \mathbf{0}_{4\times1} \end{bmatrix}, \boldsymbol{\alpha}_{38}^* = \begin{bmatrix} \boldsymbol{\alpha}_{38} \\ \mathbf{0}_{4\times1} \\ \boldsymbol{\alpha}_{32} + \boldsymbol{\alpha}_{35} \\ \mathbf{0}_{4\times1} \end{bmatrix}.$$

$$Z_t^* = \begin{cases} Z_t & t = 29, 30, 31, 39 \\ \mathbf{0}_{N_E\times4} & t = 32 \\ \mathbf{0}_{N_E\times16} & t = 35 \\ \begin{bmatrix} Z_t & \mathbf{0}_{4\times12} \end{bmatrix} & t = 33, 34, 36, 37 \\ \begin{bmatrix} Z_t & \mathbf{0}_{4\times4} & Z_t & \mathbf{0}_{4\times4} \end{bmatrix} & t = 38 \end{cases}$$

$$\mathcal{T}_{32}^* = \begin{bmatrix} \mathcal{T} \\ \mathbf{0} \\ I_4 \\ \mathbf{0} \end{bmatrix}, \qquad \mathcal{T}_t^* = \begin{bmatrix} \mathcal{T} & \mathbf{0} & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & I_4 & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & I_4 & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \mathbf{0} & I_4 \end{bmatrix}, \quad t = 33, 34, 36, 37$$

$$\mathcal{T}_{35}^* = \begin{bmatrix} \mathcal{T} & \mathbf{0} & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & I_4 & \mathbf{0} & \mathbf{0} \\ I_4 & \mathbf{0} & I_4 & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \mathbf{0} & I_4 \end{bmatrix}, \qquad \mathcal{T}_{38}^* = \begin{bmatrix} \mathcal{T} & \mathbf{0} & \mathbf{0} & \mathbf{0} \end{bmatrix}.$$

$$\Sigma_t^* = \begin{cases} \Sigma_t & t \notin \{32, 35, 38\} \\ 0.01 & t \in \{32, 35\} \\ 3\Sigma_t & t = 38 \end{cases} \qquad R_t^* = \begin{cases} R & t \notin \{32, 33, \ldots, 37\} \\ \begin{bmatrix} R \\ \mathbf{0}_{12\times 4} \end{bmatrix} & t \in \{32, 33, \ldots, 37\} \end{cases}.$$

The main idea of the sate space representation for $\boldsymbol{y}_t^*$ is to extend the state variable from dimension 4 to dimension 16 when needed. The first 4 elements of $\boldsymbol{\alpha}_t^*$ are $\boldsymbol{\alpha}_t$, while the other elements of $\boldsymbol{\alpha}_t^*$ are used to store the states that accumulated during the public holiday.[8]

We are now ready to write down the general formulas for $Z_t^*, \boldsymbol{\alpha}_t^*, \Sigma_t^*, \mathcal{T}_t^*, R_t^*$. Define the dummy variable $\varpi_t = 1$ if $\text{Status}_t = \textit{Closure}$ and $\varpi_t = 0$ if otherwise. In addition, define

$$\boldsymbol{v}_t = \begin{cases} \begin{pmatrix} 1 & 0 & 0 \end{pmatrix} & t \in T_A \\ \begin{pmatrix} 0 & 1 & 0 \end{pmatrix} & t \in T_E \\ \begin{pmatrix} 0 & 0 & 1 \end{pmatrix} & t \in T_U \end{cases}.$$

We shall write $Z_t^* = Z_t A_t$, where

$$A_t = \begin{cases} \begin{bmatrix} \mathbf{0}_{4\times 4} & \mathbf{0}_{4\times 12\sum_{j=1}^{3}\varpi_{t-j}} \end{bmatrix} & \text{if Status}_t = \textit{Closure} \\ \begin{bmatrix} I_4 & \boldsymbol{v}_t \otimes I_4 \end{bmatrix} & \text{if Status}_t = \textit{Re-open} \\ \begin{bmatrix} Z_t & \mathbf{0}_{4\times 12} \end{bmatrix} & \text{if Status}_t = NA \text{ and } \big(\varpi_{t-1} = 1 \text{ or } \varpi_{t-2} = 1\big) \\ I_4 & \text{otherwise} \end{cases}.$$

Similarly, we have

$$\Sigma_t^* = \begin{cases} \Sigma_t & \text{Status}_t \in \{NA, Trade\} \\ 0.01 & \text{Status}_t = \textit{Closure} \\ 3\Sigma_t & \text{Status}_t = \textit{Re-open} \end{cases}$$

$$R_t^* = \begin{cases} \begin{bmatrix} R \\ \mathbf{0}_{12\times 4} \end{bmatrix} & \text{if Status}_t = \textit{Closure} \\ \begin{bmatrix} R \\ \mathbf{0}_{12\times 4} \end{bmatrix} & \text{if Status}_t = NA \text{ and } \big(\varpi_{t-1} = 1 \text{ or } \varpi_{t-2} = 1\big) \\ R & \text{otherwise} \end{cases}$$

$$\mathcal{T}_t^* = \begin{bmatrix} \mathcal{T} & \mathbf{0}_{4\times 12\cdot\sum_{j=1}^{3}\varpi_{t-j}} \\ \varpi_t \cdot \boldsymbol{v}_t^\intercal \otimes I_{4\cdot\sum_{j=0}^{2}\varpi_{t-j}} & I_{4\cdot\sum_{j=0}^{2}\varpi_{t-j}\times 12\cdot\sum_{j=1}^{3}\varpi_{t-j}} \end{bmatrix}.$$

The EM algorithm outlined before could then be applied to the state space model of $\boldsymbol{y}_t^*$. The principle remains unchanged. Again we use the European continent to illustrate

[8] The three continents in our model may have different but overlapping periods of public holidays. One alternative way is to set the dimension of $\boldsymbol{\alpha}_t^*$ to 16 for all t. We could then use the 5th -8th, 9th - 12th, 13th - 16th elements of $\boldsymbol{\alpha}_t^*$ to store the Asian, European, and American accumulated states, respectively for the whole sample. The disadvantage of this treatment is that the KF and KS will be inefficient since the last 12 elements of $\boldsymbol{\alpha}_t^*$ would often be zero. Hence, instead of keeping the dimension of $\boldsymbol{\alpha}_t^*$ to be 16 for all t, we only extend the dimension to 16 when needed.

and suppose that its stock market does not open for $\tau+1$ days because of some public holiday. Equation (3.2) takes the following form:

$$\tilde{\mathbb{E}}\sum_{\substack{t\in T_E\\ \varpi_t=0}}\ell_{1,t}=\sum_{\substack{t\in T_E\\ \varpi_t=0}}\log|\Sigma_t^*|+\sum_{\substack{t\in T_E\\ \varpi_t=0}}\operatorname{tr}\left(\left[\boldsymbol{y}_t^*\boldsymbol{y}_t^{*\intercal}-2Z_t^*\tilde{\mathbb{E}}[\boldsymbol{\alpha}_t^*]\boldsymbol{y}_t^{*\intercal}+Z_t^*\tilde{\mathbb{E}}[\boldsymbol{\alpha}_t^*\boldsymbol{\alpha}_t^{*\intercal}]Z_t^{*\intercal}\right]\Sigma_t^{*-1}\right)$$

$$=\sum_{\substack{t\in T_E\\ \varpi_t=0}}\left\{\log|\Sigma_t(\tau_t+1)|+\operatorname{tr}\left(\left[\boldsymbol{y}_t^*\boldsymbol{y}_t^{*\intercal}-2Z_tA_t\tilde{\mathbb{E}}[\boldsymbol{\alpha}_t^*]\boldsymbol{y}_t^{*\intercal}+Z_tA_t\tilde{\mathbb{E}}[\boldsymbol{\alpha}_t^*\boldsymbol{\alpha}_t^{*\intercal}]A_t^\intercal Z_t^\intercal\right]\frac{\Sigma_t^{-1}}{(\tau_t+1)}\right)\right\}$$

$$=\sum_{t\in T_E}\operatorname{tr}\left(\left[\boldsymbol{y}_t^*\boldsymbol{y}_t^{*\intercal}\frac{(1-\varpi_t)}{(\tau_t+1)}-2Z_tA_t\tilde{\mathbb{E}}[\boldsymbol{\alpha}_t^*]\boldsymbol{y}_t^{*\intercal}\frac{(1-\varpi_t)}{(\tau_t+1)}+Z_tA_t\tilde{\mathbb{E}}[\boldsymbol{\alpha}_t^*\boldsymbol{\alpha}_t^{*\intercal}]A_t^\intercal Z_t^\intercal\frac{(1-\varpi_t)}{(\tau_t+1)}\right]\Sigma_t^{-1}\right)$$

$$+\left(\sum_{t\in T_E}(1-\varpi_t)\log|\Sigma_t|\right)+\text{constant},$$

where $\tau_t=\tau$ if $\text{Status}_t=\textit{Re-open}$ and $\tau_t=0$ if otherwise, and τ is the number of days of the closure due to the public holiday. That is, the terms $\boldsymbol{y}_t\boldsymbol{y}_t^\intercal$, $\tilde{\mathbb{E}}[\boldsymbol{\alpha}_t]\boldsymbol{y}_t^\intercal$, $\tilde{\mathbb{E}}[\boldsymbol{\alpha}_t\boldsymbol{\alpha}_t^\intercal]$, $T/3$ in (3.2) are replaced by $\boldsymbol{y}_t^*\boldsymbol{y}_t^{*\intercal}\frac{(1-\varpi_t)}{(\tau_t+1)}$, $A_t\tilde{\mathbb{E}}[\boldsymbol{\alpha}_t^*]\boldsymbol{y}_t^{*\intercal}\frac{(1-\varpi_t)}{(\tau_t+1)}$, and $A_t\tilde{\mathbb{E}}[\boldsymbol{\alpha}_t^*\boldsymbol{\alpha}_t^{*\intercal}]A_t^\intercal\frac{(1-\varpi_t)}{(\tau_t+1)}$, $\sum_{t\in T_E}(1-\varpi_t)$ respectively. Hence, we have

$$\tilde{Z}^E=\sum_{t\in T_E}\left(\tilde{\mathbb{E}}[\boldsymbol{y}_t^*\boldsymbol{\alpha}_t^{*\intercal}]A_t^\intercal\frac{1-\varpi_t}{\tau_t+1}\right)\left(\sum_{t\in T_E}A_t\tilde{\mathbb{E}}[\boldsymbol{\alpha}_t^*\boldsymbol{\alpha}_t^{*\intercal}]A_t^\intercal\frac{1-\varpi_t}{\tau_t+1}\right)^{-1}$$

$$\tilde{C}_E:=\sum_{t\in T_E}\frac{1-\varpi_t}{\tau_t+1}\left[\boldsymbol{y}_t^*\boldsymbol{y}_t^{*\intercal}-2\tilde{Z}^EA_t\tilde{\mathbb{E}}[\boldsymbol{\alpha}_t^*]\boldsymbol{y}_t^{*\intercal}+\tilde{Z}^EA_t\tilde{\mathbb{E}}[\boldsymbol{\alpha}_t^*\boldsymbol{\alpha}_t^{*\intercal}]A_t^\intercal\tilde{Z}^{E\intercal}\right]$$

$$\tilde{\Sigma}_E=\frac{1}{\sum_{t\in T_E}(1-\varpi_t)}(\tilde{C}_E\circ I_{N_E}).$$

The formula of $\tilde{\phi}$ in (3.4) remains unchanged, since $\tilde{\mathbb{E}}\sum_{t=1}^T\ell_{2,t}$ remains unchanged.

B.2 Motivation of the EM Algorithm

In this subsection, we shall review the motivation of the EM algorithm. The log-likelihood of $Y_{1:T}$ is

$$\ell(Y_{1:T};\boldsymbol{\theta})=\log p(Y_{1:T};\boldsymbol{\theta})=\log\int p(Y_{1:T}|\Xi;\boldsymbol{\theta})p(\Xi;\boldsymbol{\theta})d\Xi.$$

Given $\tilde{\boldsymbol{\theta}}^{(i)}$, we could compute

$$\ell(Y_{1:T};\boldsymbol{\theta})-\ell(Y_{1:T};\tilde{\boldsymbol{\theta}}^{(i)})=\log\int\left(p(\Xi|Y_{1:T},\tilde{\boldsymbol{\theta}}^{(i)})\frac{p(Y_{1:T}|\Xi;\boldsymbol{\theta})p(\Xi;\boldsymbol{\theta})}{p(\Xi|Y_{1:T},\tilde{\boldsymbol{\theta}}^{(i)})}\right)d\Xi-\log p(Y_{1:T};\tilde{\boldsymbol{\theta}}^{(i)})$$

$$=\log\tilde{\mathbb{E}}\left[\frac{p(Y_{1:T}|\Xi;\boldsymbol{\theta})p(\Xi;\boldsymbol{\theta})}{p(\Xi|Y_{1:T},\tilde{\boldsymbol{\theta}}^{(i)})}\right]-\log p(Y_{1:T};\tilde{\boldsymbol{\theta}}^{(i)})=\log\tilde{\mathbb{E}}\left[\frac{p(Y_{1:T}|\Xi;\boldsymbol{\theta})p(\Xi;\boldsymbol{\theta})}{p(\Xi|Y_{1:T},\tilde{\boldsymbol{\theta}}^{(i)})}\right]-\tilde{\mathbb{E}}\left[\log p(Y_{1:T};\tilde{\boldsymbol{\theta}}^{(i)})\right]$$

$$\geq\tilde{\mathbb{E}}\left[\log\frac{p(Y_{1:T}|\Xi;\boldsymbol{\theta})p(\Xi;\boldsymbol{\theta})}{p(\Xi|Y_{1:T},\tilde{\boldsymbol{\theta}}^{(i)})}\right]-\tilde{\mathbb{E}}\left[\log p(Y_{1:T};\tilde{\boldsymbol{\theta}}^{(i)})\right]=\tilde{\mathbb{E}}\left[\log\frac{p(Y_{1:T}|\Xi;\boldsymbol{\theta})p(\Xi;\boldsymbol{\theta})}{p(\Xi|Y_{1:T},\tilde{\boldsymbol{\theta}}^{(i)})p(Y_{1:T};\tilde{\boldsymbol{\theta}}^{(i)})}\right]$$

whence we have

$$\ell(Y_{1:T};\boldsymbol{\theta}) \geq \ell(Y_{1:T};\tilde{\boldsymbol{\theta}}^{(i)}) + \tilde{\mathbb{E}}\left[\log \frac{p(Y_{1:T}|\Xi;\boldsymbol{\theta})p(\Xi;\boldsymbol{\theta})}{p(\Xi|Y_{1:T},\tilde{\boldsymbol{\theta}}^{(i)})p(Y_{1:T};\tilde{\boldsymbol{\theta}}^{(i)})}\right] =: B(\boldsymbol{\theta},\tilde{\boldsymbol{\theta}}^{(i)}).$$

We see that $B(\boldsymbol{\theta},\tilde{\boldsymbol{\theta}}^{(i)})$ is a lower bound for $\ell(Y_{1:T};\boldsymbol{\theta})$, and $\ell(Y_{1:T};\tilde{\boldsymbol{\theta}}^{(i)}) = B(\tilde{\boldsymbol{\theta}}^{(i)},\tilde{\boldsymbol{\theta}}^{(i)})$. Thus we would like to choose $\boldsymbol{\theta}$ to maximise $B(\boldsymbol{\theta},\tilde{\boldsymbol{\theta}}^{(i)})$:

$$\tilde{\boldsymbol{\theta}}^{(i+1)} = \arg\max_{\boldsymbol{\theta}} B(\boldsymbol{\theta},\tilde{\boldsymbol{\theta}}^{(i)}) = \arg\max_{\boldsymbol{\theta}} \tilde{\mathbb{E}}\left[\ell(Y_{1:T},\Xi;\boldsymbol{\theta})\right].$$

B.3 The Expression for $\partial h(\boldsymbol{\theta}_m)/\partial\boldsymbol{\theta}_m$

Let $\mathbf{0}_a$ denote an $a \times 1$ zero vector.

$$\underbrace{\frac{\partial h(\boldsymbol{\theta}_m)}{\partial \boldsymbol{\theta}_m}}_{59\times 10} = \begin{bmatrix}
\mathbf{0}_5 & \mathbf{0}_5 & \mathbf{0}_5 & \mathbf{0}_5 & \mathbf{0}_5 & \mathbf{0}_5 & \mathbf{0}_5 & \mathbf{0}_5 & \mathbf{0}_5 & \mathbf{0}_5 \\
1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
\mathbf{0}_5 & \mathbf{0}_5 & \mathbf{0}_5 & \mathbf{0}_5 & \mathbf{0}_5 & \mathbf{0}_5 & \mathbf{0}_5 & \mathbf{0}_5 & \mathbf{0}_5 & \mathbf{0}_5 \\
0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 \\
\mathbf{0}_2 & \mathbf{0}_2 & \mathbf{0}_2 & \mathbf{0}_2 & \mathbf{0}_2 & \mathbf{0}_2 & \mathbf{0}_2 & \mathbf{0}_2 & \mathbf{0}_2 & \mathbf{0}_2 \\
1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
\mathbf{0}_5 & \mathbf{0}_5 & \mathbf{0}_5 & \mathbf{0}_5 & \mathbf{0}_5 & \mathbf{0}_5 & \mathbf{0}_5 & \mathbf{0}_5 & \mathbf{0}_5 & \mathbf{0}_5 \\
0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 \\
\mathbf{0}_6 & \mathbf{0}_6 & \mathbf{0}_6 & \mathbf{0}_6 & \mathbf{0}_6 & \mathbf{0}_6 & \mathbf{0}_6 & \mathbf{0}_6 & \mathbf{0}_6 & \mathbf{0}_6 \\
0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 \\
\mathbf{0}_5 & \mathbf{0}_5 & \mathbf{0}_5 & \mathbf{0}_5 & \mathbf{0}_5 & \mathbf{0}_5 & \mathbf{0}_5 & \mathbf{0}_5 & \mathbf{0}_5 & \mathbf{0}_5 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 \\
\mathbf{0}_2 & \mathbf{0}_2 & \mathbf{0}_2 & \mathbf{0}_2 & \mathbf{0}_2 & \mathbf{0}_2 & \mathbf{0}_2 & \mathbf{0}_2 & \mathbf{0}_2 & \mathbf{0}_2 \\
0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 \\
\mathbf{0}_5 & \mathbf{0}_5 & \mathbf{0}_5 & \mathbf{0}_5 & \mathbf{0}_5 & \mathbf{0}_5 & \mathbf{0}_5 & \mathbf{0}_5 & \mathbf{0}_5 & \mathbf{0}_5 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 \\
\mathbf{0}_5 & \mathbf{0}_5 & \mathbf{0}_5 & \mathbf{0}_5 & \mathbf{0}_5 & \mathbf{0}_5 & \mathbf{0}_5 & \mathbf{0}_5 & \mathbf{0}_5 & \mathbf{0}_5 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & \frac{2\phi}{(1-\phi^2)^2} & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & \frac{1+\phi^2}{(1-\phi^2)^2} & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1
\end{bmatrix}$$

B.4 Validity of Applying Our Model to Standardised Portfolio Returns

In this subsection, we discuss the validity of applying our model to standardised portfolio returns. The daily value-weighted portfolio return R_t on day t is calculated using the following formula:

$$R_t = \sum_{i=1}^n w_{i,t-1} \frac{P_{i,t} - P_{i,t-1}}{P_{i,t-1}} \approx \sum_{i=1}^n w_{i,t-1} \log \frac{P_{i,t}}{P_{i,t-1}} = \log \left(\frac{\prod_{i=1}^n P_{i,t}^{w_{i,t-1}}}{\prod_{i=1}^n P_{i,t-1}^{w_{i,t-1}}} \right)$$
$$= \log \left(\prod_{i=1}^n P_{i,t}^{w_{i,t-1}} \right) - \log \left(\prod_{i=1}^n P_{i,t-1}^{w_{i,t-1}} \right) \approx \log \left(\prod_{i=1}^n P_{i,t}^{w_{i,t-1}} \right) - \log \left(\prod_{i=1}^n P_{i,t-1}^{w_{i,t-2}} \right)$$

where $P_{i,t}$ denotes the closing price of stock i on day t, $w_{i,t-1}$ denotes the market capitalisation of stock i divided by the market capitalisation of the portfolio on day $t-1$, and n denotes the number of the stocks in the portfolio. The first approximation sign will hold on the assumption that daily returns are often small, and the second approximation sign will hold on the assumption that weights do not change much over a day. The preceding display shows that $\log \left(\prod_{i=1}^n P_{i,t}^{w_{i,t-1}} \right)$ could be interpreted as the log closing price of the portfolio on day t, and we could hence fit R_t with our model:

$$R_t = \mu + \sum_{j=0}^2 z_j f_{g,t-j} + z_3 f_{c,t} + e_t$$

where z_j are scalars, and e_t is a random variable. Let $\bar{R} := T^{-1} \sum_{t=1}^T R_t$ and $\mathrm{se}(R) := \left[T^{-1} \sum_{t=1}^T (R_t - \bar{R})^2 \right]^{1/2}$. Thus

$$\frac{R_t - \bar{R}}{\mathrm{se}(R_t)} \approx \sum_{j=0}^2 \frac{z_j}{\mathrm{se}(R_t)} f_{g,t-j} + \frac{z_3}{\mathrm{se}(R_t)} f_{c,t} + \frac{e_t}{\mathrm{se}(R_t)} =: \sum_{j=0}^2 z_j^* f_{g,t-j} + z_3^* f_{c,t} + e_t^*.$$

We hence see that our model could be applied to standardised portfolio returns in practice by relying a few innocuous approximations.

B.5 Proof of (A.54)

We first state a central limit theorem for the martingale difference array.

Theorem B.1 (McLeish (1974)). *Let $\{X_{n,i}, i = 1, ..., k_n\}$ be a martingale difference array with respect to the triangular array of σ-algebras $\{\mathcal{F}_{n,i}, i = 0, ..., k_n\}$ (i.e., $X_{n,i}$ is $\mathcal{F}_{n,i}$-measurable and $\mathbb{E}[X_{n,i}|\mathcal{F}_{n,i-1}] = 0$ almost surely for all n and i) satisfying $\mathcal{F}_{n,i-1} \subset \mathcal{F}_{n,i}$ for all $n \geq 1$. Assume,*

(i) $\max_{i \leq k_n} |X_{n,i}|$ *is uniformly (in n) bounded in L_2 norm,*

(ii) $\max_{i \leq k_n} |X_{n,i}| \xrightarrow{p} 0$, *and*

(iii) $\sum_{i=1}^{k_n} X_{n,i}^2 \xrightarrow{p} 1$.

Then, $S_n = \sum_{i=1}^{k_n} X_{n,i} \xrightarrow{d} N(0, 1)$ *as $n \to \infty$.*

We now show (A.54). Note that (A.54) is equivalent to

$$\frac{1}{\sqrt{T_f}} \sum_{t=1}^{T_f} \rho^\intercal \begin{bmatrix} (\boldsymbol{\lambda}_{k,j}^\intercal \otimes I_{14}) \Gamma(\boldsymbol{e}_t^\dagger \otimes \boldsymbol{f}_t) \\ M^{-1} \boldsymbol{f}_t e_{(k-1)N+j,t} \end{bmatrix} \xrightarrow{d}$$
$$N\left(0, \rho^\intercal \begin{bmatrix} (\boldsymbol{\lambda}_{k,j}^\intercal \otimes I_{14}) \Gamma(\Sigma_{ee}^\dagger \otimes M) \Gamma^\intercal (\boldsymbol{\lambda}_{k,j} \otimes I_{14}) & \mathrm{cov}_{k,j} \\ \mathrm{cov}_{k,j} & M^{-1} \sigma_{k,j}^2 \end{bmatrix} \rho \right) =: N(0, \rho^\intercal V_a \rho) \quad \text{(B.3)}$$

where ρ is a 24×1 non-zero vector with $\|\rho\|_2 = 1$. Define

$$u_t := T_f^{-1/2} \rho^\intercal \begin{bmatrix} (\boldsymbol{\lambda}_{k,j}^\intercal \otimes I_{14})\Gamma(\boldsymbol{e}_t^\dagger \otimes \boldsymbol{f}_t) \\ M^{-1}\boldsymbol{f}_t e_{(k-1)N+j,t} \end{bmatrix} / \sqrt{\rho^\intercal V_a \rho}.$$

Define the filtration $\mathcal{F}_t := \sigma(\boldsymbol{f}_t, \boldsymbol{e}_t, \boldsymbol{f}_{t-1}, \boldsymbol{e}_{t-1}, \ldots, \boldsymbol{f}_1, \boldsymbol{e}_1)$ and $\mathcal{F}_0 := \{\emptyset, \emptyset^c\}$. It is easy to show that u_t is $\mathcal{F}_t$-measurable. We now show that $\{u_t, \mathcal{F}_t\}$ is a martingale difference sequence (i.e., $\mathbb{E}[u_t|\mathcal{F}_{t-1}] = 0$). It suffices to show $\mathbb{E}[\boldsymbol{e}_t \otimes \boldsymbol{f}_t|\mathcal{F}_{t-1}] = \mathbf{0}$. Write $\boldsymbol{f}_t = \mathcal{T}^\dagger \boldsymbol{f}_{t-1} + \boldsymbol{\eta}_t^\dagger$, where $\mathcal{T}^\dagger$ is a 14×14 matrix consisting of ϕ, and $\boldsymbol{\eta}_t^\dagger$ is a 14×1 vector consisting of $\{\boldsymbol{\eta}_t\}$. Then we have

$$\begin{aligned} &\mathbb{E}[\boldsymbol{e}_t \otimes \boldsymbol{f}_t|\mathcal{F}_{t-1}] = \mathbb{E}\left[\boldsymbol{e}_t \otimes (\mathcal{T}^\dagger \boldsymbol{f}_{t-1} + \boldsymbol{\eta}_t^\dagger)|\mathcal{F}_{t-1}\right] = \mathbf{0} + \mathbb{E}\left[\boldsymbol{e}_t \otimes \boldsymbol{\eta}_t^\dagger|\mathcal{F}_{t-1}\right] \\ &= \mathbb{E}\left[\boldsymbol{e}_t|\mathcal{F}_{t-1}\right] \otimes \mathbb{E}\left[\boldsymbol{\eta}_t^\dagger|\mathcal{F}_{t-1}\right] = \mathbf{0}. \end{aligned}$$

where we have used Assumption 2.1. We now check conditions (i)-(iii) of Theorem B.1. We first investigate at what rate the denominator $\sqrt{\rho^\intercal V_a \rho}$ goes to zero. Since

$$\rho^\intercal V_a \rho \geq \lambda_{\max}(V_a) > 0,$$

where the last inequality is due to Assumption 4.1, we have $1/\sqrt{\rho^\intercal V_a \rho} = O(1)$. Hence, it is easy to show that

$$\begin{aligned} &|u_t| = O\left(\frac{1}{\sqrt{T_f}}\right) \|\boldsymbol{e}_t^\dagger \otimes \boldsymbol{f}_t\|_2 + O\left(\frac{1}{\sqrt{T_f}}\right) \|\boldsymbol{f}_t e_{(k-1)N+j,t}\|_2 = O\left(\frac{1}{\sqrt{T_f}}\right) \|\boldsymbol{e}_t \otimes \boldsymbol{f}_t\|_2 \\ &= O\left(\frac{1}{\sqrt{T_f}}\right) \|\boldsymbol{e}_t \otimes \boldsymbol{f}_t\|_\infty. \end{aligned}$$

We now verify (i) and (ii) of Theorem B.1. We shall use Orlicz norms as defined in van der Vaart and Wellner (1996): Let $\psi : \mathbb{R}^+ \to \mathbb{R}^+$ be a non-decreasing, convex function with $\psi(0) = 0$ and $\lim_{x\to\infty} \psi(x) = \infty$, where $\mathbb{R}^+$ denotes the set of nonnegative real numbers. Then, the Orlicz norm of a random variable X is given by

$$\|X\|_\psi = \inf\left\{C > 0 : \mathbb{E}\psi\left(|X|/C\right) \leq 1\right\},$$

where $\inf \emptyset = \infty$. We shall use Orlicz norms for $\psi(x) = \psi_p(x) = e^{x^p} - 1$ for $p = 1$ in this article. Consider

$$\big\|\|\boldsymbol{e}_t \otimes \boldsymbol{f}_t\|_\infty\big\|_{\psi_1} = \big\|\max_{i,j} |e_{i,t}[\boldsymbol{f}_t]_j|\big\|_{\psi_1} \leq C \max_{i,j} \|e_{i,t}[\boldsymbol{f}_t]_j\|_{\psi_1}$$

where $e_{i,t}$ is the ith element of $\boldsymbol{e}_t$, $[\boldsymbol{f}_t]_j$ is the jth element of $\boldsymbol{f}_t$, and the inequality is due to Lemma 2.2.2 in van der Vaart and Wellner (1996). Since $e_{i,t}, [\boldsymbol{f}_t]_j$ are normal, it follows from Lemma 2.2.1 in van der Vaart and Wellner (1996) that $\|e_{i,t}[\boldsymbol{f}_t]_j\|_{\psi_1} = O(1)$ for all i, j, t. Thus

$$\big\|\max_t |u_t|\big\|_{\psi_1} \leq \log(1 + T_f) \max_t \|u_t\|_{\psi_1} = O\left(\frac{\log T_f}{\sqrt{T_f}}\right) \max_t \big\|\|\boldsymbol{e}_t \otimes \boldsymbol{f}_t\|_\infty\big\|_{\psi_1} = O\left(\frac{\log T_f}{\sqrt{T_f}}\right) = o(1).$$

Since $\|U\|_{L_r} \leq r!\|U\|_{\psi_1}$ for any random variable U (van der Vaart and Wellner (1996), p95), we conclude that (i) and (ii) of Theorem B.1 are satisfied. We now verify condition

(iii) of Theorem B.1. Since we have already shown that $\rho^\intercal V_a \rho$ is bounded away from zero by an absolute constant, it suffices to show

$$\frac{1}{T_f}\sum_{t=1}^{T_f}\left\{\rho^\intercal \left[\begin{array}{c}(\boldsymbol{\lambda}_{k,j}^\intercal \otimes I_{14})\Gamma(\boldsymbol{e}_t^\dagger \otimes \boldsymbol{f}_t)\\ M^{-1}\boldsymbol{f}_t e_{(k-1)N+j,t}\end{array}\right]\right\}^2 - \rho^\intercal V_a \rho = o_p(1).$$

This just follows from Ergodic theorem (Theorem 3.34 of White (2001)) by recognising that $\{u_t\}$ is strictly stationary and ergodic because $\{\boldsymbol{e}_t\}$ and $\{\boldsymbol{\eta}_t\}$ are strictly stationary and ergodic (Theorem 3.35 of White (2001)). Thus, condition (iii) of Theorem B.1 is verified and (B.3) is proved.

B.6 Proof of Consistency of $\check{\boldsymbol{\theta}}_m$

In this subsection, we give a proof for $\check{\boldsymbol{\theta}}_m \xrightarrow{p} \boldsymbol{\theta}_m$. Define

$$\begin{aligned} OB_{T_f}(\boldsymbol{b}) &:= -\left[\hat{\boldsymbol{h}} - h(\boldsymbol{b})\right]^\intercal W \left[\hat{\boldsymbol{h}} - h(\boldsymbol{b})\right] \\ OB(\boldsymbol{b}) &:= -\left[h(\boldsymbol{\theta}_m) - h(\boldsymbol{b})\right]^\intercal W \left[h(\boldsymbol{\theta}_m) - h(\boldsymbol{b})\right]. \end{aligned}$$

It is easy to see that

$$\begin{aligned} \sup_{\boldsymbol{b}} \left| OB_{T_f}(\boldsymbol{b}) - OB(\boldsymbol{b})\right| &= o_p(1) \\ \sup_{\boldsymbol{b}:\|\boldsymbol{b}-\boldsymbol{\theta}_m\|_2 \geq \epsilon} OB(\boldsymbol{b}) &< OB(\boldsymbol{\theta}_m) \end{aligned}$$

for every $\epsilon > 0$. By Theorem 5.7 of van der Vaart (1998), we have $\check{\boldsymbol{\theta}}_m \xrightarrow{p} \boldsymbol{\theta}_m$.

References

Anderson, T. W. (2003). *An Introduction to Multivariate Statistical Analysis*. John Wiley and Sons.

Anderson, T. W. and H. Rubin (1956). Statistical inference in factor analysis. *Proceedings of the Third Berkeley Symposium on Mathematical Statistics and Probability*.

Aruoba, S. B., F. X. Diebold, and C. Scotti (2009). Real-time measurement of business conditions. *Journal of Business & Economic Statistics 27*(4), 417–427.

Bai, J. and K. Li (2012). Statistical analysis of factor models of high dimension. *The Annals of Statistics 40*(1), 436–465.

Bai, J. and K. Li (2016). Maximum likelihood estimation and inference for approximate factor models of high dimension. *The Review of Economics and Statistics 98*(2), 298–309.

Bai, J. and P. Wang (2015). Identification and bayesian estimation of dynamic factor models. *Journal of Business & Economic Statistics 33*(2), 221–240.

Banbura, M., D. Giannone, M. Modugno, and L. Reichlin (2013). Nowcasting and the real-time data flow. In G. Elliott and A. Timmermann (Eds.), *Handbook of Economic Forecasting*, Volume 2, Chapter 4, pp. 195–237. Elsevier, North-Holland.

Barigozzi, M. and M. Luciani (2022). Quasi maximum likelihood estimation and inference of large approximate dynamic factor models via the em algorithm. *arXiv preprint arXiv:1910.03821v3*.

Carter, C. K. and R. Kohn (1994). On gibbs sampling for state space models. *Biometrika 81*(3), 541–553.

Connor, G., L. R. Goldberg, and R. A. Korajczyk (2010). *Portfolio Risk Analysis*. Princeton University Press.

Doz, C., D. Giannone, and L. Reichlin (2012). A quasi-maximum likelihood approach for large, approximate dynamic factor models. *The Review of Economics and Statistics 94*(4), 1014–1024.

Durbin, J. and S. J. Koopman (2012). *Time Series Analysis by State Space Methods*. Oxford.

Fama, E. F. and K. R. French (1993). Common risk factors in the returns on stocks and bonds. *Journal of Financial Economics 33*, 3–56.

Forni, M., M. Hallin, M. Lippi, and L. Reichlin (2000). The generalized dynamic-factor model: Identification and estimation. *The Review of Economics and Statistics 82*(4), 540–554.

Gagnon, L. and G. A. Karolyi (2006). Price and volatility transmission across borders. *Financial Markets, Institutions & Instruments 15*(3), 107–158.

Giannone, D., L. Reichlin, and D. Small (2008). Nowcasting: the real-time informational content of macroeconomic data. *Journal of Monetary Economics 55*, 665–676.

Koch, P. D. and T. W. Koch (1991). Evolution in dynamic linkages across daily national stock indexes. *Journal of International Money and Finance 10*, 231–251.

Linton, O. B. and H. Tang (2021). Estimation of the kronecker covariance model by quadratic form. *Econometric Theory Forthcoming*.

Martens, M. and S.-H. Poon (2001). Returns synchronization and daily correlation dynamics between international stock markets. *Journal of Banking & Finance 25*, 1805–1827.

McLeish, D. L. (1974). Dependent central limit theorems and invariance principles. *The Annals of Probability 2*(4), 620–628.

Pan, W. and J. E. Connett (2002). Selecting the working correlation structure in generalized estimation equations with application to the lung health study. *Statistica Sinica 12*, 475–490.

Pesaran, M. H., T. Schuermann, and S. M. Weiner (2004). Modeling regional interdependencies using a global error-correcting macroeconometric model. *Journal of Business & Economic Statistics 22*(2), 129–162.

Quah, D. and T. J. Sargent (1992). A large dynamic index model for large cross sections. In J. H. Stock and M. W. Watson (Eds.), *Business Cycles, Indicators and Forecasting*, pp. 285 – 310. University of Chicago Press.

Sharma, A. and N. Seth (2012). Literature review of stock market integration: a global perspective. *Qualitative Research in Financial Markets 4* (1), 84–122.

Stock, J. H. and M. W. Watson (2016). Dynamic factor models, factor-augmented vector autoregressions, and structural vector autoregressions in macroeconomics. In *Handbook of Macroeconomics*, Volume 2, Chapter 8, pp. 415–525. Elsevier B.V.

van der Vaart, A. (1998). *Asymptotic Statistics*. Cambridge University Press.

van der Vaart, A. and J. A. Wellner (1996). *Weak Convergence and Empirical Processes*. Springer.

Watson, M. W. and R. F. Engle (1983). Alternative algorithms for the estimation of dynamic factor, mimic and varying coefficient regression models. *Journal of Econometrics 23*, 385–400.

White, H. (2001). *Asymptotic Theory for Econometricians*. Emerald.