GLS Estimation of Dynamic Factor Models

Jörg BREITUNG and Jörn TENHOFEN

In this article a simple two-step estimation procedure of the dynamic factor model is proposed. The estimator allows for heteroscedastic and serially correlated errors. It turns out that the feasible two-step estimator has the same limiting distribution as the generalized least squares (GLS) estimator assuming that the covariance parameters are known. In a Monte Carlo study of the small sample properties, we find that the GLS estimators may be substantially more efficient than the usual estimator based on principal components. Furthermore, it turns out that the iterated version of the estimator may feature considerably improved properties in sample sizes usually encountered in practice.

KEY WORDS: Factor models; Generalized least squares; Heteroscedasticity; Principal components; Serial correlation.

1. INTRODUCTION

Since the influential work of Forni et al. (2000), Stock and Watson (2002a, 2002b), Bai and Ng (2002), and Bai (2003), dynamic factor models have become an important tool in macroeconomic forecasting (e.g., Watson 2003; Eickmeier and Ziegler 2008) and structural analysis (e.g., Giannone, Reichlin, and Sala 2002; Bernanke, Boivin, and Eliasz 2005; Eickmeier 2007). Under the weak assumptions of an approximate factor model (Chamberlain and Rothschild 1983), the parameters can be consistently estimated by applying the traditional principal component (PC) estimator (Stock and Watson 2002b; Bai 2003) or—in the frequency domain—by using the dynamic principal component estimator (Forni et al. 2000). Assuming Gaussian iid errors, the PC estimator is equivalent to the maximum likelihood (ML) estimator. It is well known that a generalized least squares (GLS)-type criterion function yields a more efficient estimator than the ordinary least squares (OLS)-based PC estimator if the errors are heteroscedastic (e.g., Boivin and Ng 2006; Doz, Giannone, and Reichlin 2006; Choi 2011). It is less clear how the estimator can be improved in the case of serially correlated errors. Stock and Watson (2005) suggest a GLS transformation similar to the one that is used to correct for autocorrelation in the linear regression model. However, this transformation affects the static representation of the factor model.

In this article, we consider the Gaussian (pseudo) ML estimator in models, where the errors are assumed to be heteroscedastic and autocorrelated. We derive the first-order conditions for a local maximum of the (approximate) log-likelihood function and show that the resulting system of equations can be solved by running a sequence of GLS regressions. Specifically, the factors can be estimated by taking into account possible heteroscedasticity of the errors, whereas the factor loadings are estimated by using the usual GLS transformation for autocorrelated errors. We show that the feasible two-step GLS estimation procedure is asymptotically equivalent to the estimator that locally maximizes the approximate likelihood function. In small samples, however, our Monte Carlo simulations suggest that the iterated PC–GLS estimator can be substantially more efficient than the simpler two-step estimator. Jungbacker and Koopman (2008) consider the state space representation of the factor model, where the number of variables (N) is fixed and the vector of common factors has a VARMA representation. As we will argue below, as $N \to \infty$ their (exact) ML estimator converges to the approximate ML estimator suggested in this article. Thus, the two-step GLS estimator can be seen as a simplification of the exact ML approach proposed by Jungbacker and Koopman (2008) as N gets large. Furthermore, we do not specify a particular parametric model for the vector of common factors, as the data-generating process of the factors becomes irrelevant as $N \to \infty$. Accordingly, our approach sidesteps the problem of choosing an appropriate lag length for the VARMA representation of the factors.

It may be argued that in practice the efficiency gain from taking into account serial correlation and heteroscedasticity may be small if the variances of the idiosyncratic components are similar and their autocorrelations are small. To assess the potential of the suggested estimator, we therefore consider the distribution of the variances and first-order autocorrelations estimated from the widely used dataset provided by Stock and Watson (2005). This dataset contains 132 monthly United States series including measures of real economic activity, prices, interest rates, money and credit aggregates, stock prices, and exchange rates. The sampling period runs from 1960 to 2003.[1] As usual, the time series are differenced if unit root tests are not able to reject the null hypothesis of nonstationarity. Applying the information criteria of Bai and Ng (2002) suggests that the number of common factors is $r = 7$. The idiosyncratic component is obtained by subtracting the estimated common component from the standardized series. The resulting histograms with respect to sample variances and first-order autocorrelations of the idiosyncratic components are presented in Figures 1 and 2. Since the variables are standardized, the variances of the idiosyncratic components are identical to $1 - c_i$, where c_i is the "commonality" of variable i. A value of c_i close to zero implies that the factors do not contribute to the variance of the time series. In our example, 13% of the variables have a commonality less than 0.05. Furthermore, the variances do not seem to be concentrated around some common value. Accordingly, ignoring the heteroscedasticity in the data will lead to a severe loss

Jörg Breitung is Professor, Institute of Econometrics, University of Bonn, Adenauerallee 24-42, 53113 Bonn, Germany (E-mail: *breitung@uni-bonn.de*). Jörn Tenhofen is Economist, Department of Economics, Deutsche Bundesbank, Wilhelm-Epstein-Straße 14, 60431 Frankfurt am Main, Germany (E-mail: *joern.tenhofen@bundesbank.de*). We thank In Choi, Domenico Giannone, Franz Palm, Hashem Pesaran, and Yongcheol Shin for helpful comments and suggestions. Furthermore, we are grateful to two anonymous referees for their valuable and constructive comments.

Journal of the American Statistical Association
September 2011, Vol. 106, No. 495, Theory and Methods
DOI: 10.1198/jasa.2011.tm09693

[1] The original dataset is provided for the years 1959 to 2003. Some observations are, however, missing in 1959. We therefore decided to use a balanced dataset starting in 1960.

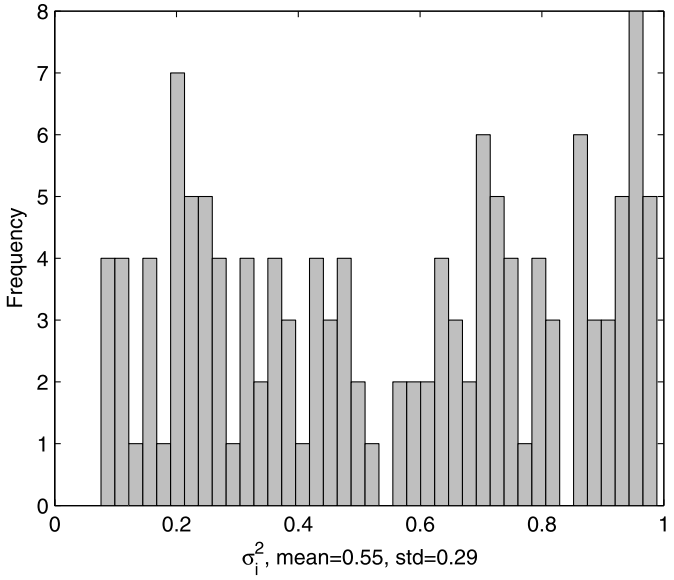

Figure 1. Histogram of the sample variances.

of efficiency. A similar picture emerges for the autocorrelations of the idiosyncratic errors. Most of the estimates are far away from zero. Moreover, there is substantial heterogeneity among the estimates suggesting that the model should allow for individual specific autocorrelations. In order to investigate the impact of those features of the data and to illustrate the potential of our suggested estimators, one of the Monte Carlo experiments presented in the latter part of the article is based on Stock and Watson's (2005) dataset.

The rest of the article is organized as follows. In Section 2, we consider some prerequisites of the dynamic factor model. Section 3 introduces the PC–GLS estimator and Section 4 studies the asymptotic properties of the two-step estimator. The relative asymptotic efficiency of the standard PC and PC–GLS estimators is investigated in Section 5. The small sample properties of different estimators are compared by means of Monte Carlo simulations in Section 6. Finally, Section 7 concludes.

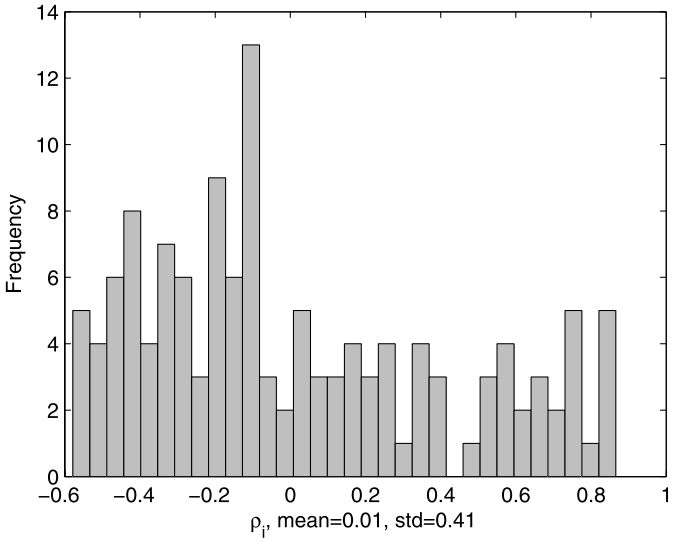

Figure 2. Histogram of the sample autocorrelations.

2. THE DYNAMIC FACTOR MODEL

Let x_{it} be the ith variable ($i = 1, \ldots, N$) observed in period t ($t = 1, \ldots, T$). The factor model is given by

$$x_{it} = \boldsymbol{\lambda}_i' \mathbf{F}_t + e_{it}, \tag{1}$$

where $\mathbf{F}_t = [f_{1t}, \ldots, f_{rt}]'$ is an r-dimensional vector of common factors, $\boldsymbol{\lambda}_i$ is the corresponding vector of r factor loadings, and e_{it} denotes the idiosyncratic component. The components $\mathbf{F}_t$ and e_{it} are generated by dynamic processes which will be characterized below. As usual in this literature, we ignore possible deterministic terms and assume $E(x_{it}) = E(e_{it}) = 0$. In matrix notation, the model is written as

$$\mathbf{X} = \mathbf{F}\boldsymbol{\Lambda}' + \mathbf{e}, \tag{2}$$

where $\mathbf{X} = [\mathbf{X}_1, \ldots, \mathbf{X}_T]'$ is the $T \times N$ matrix of observations with rows $\mathbf{X}_t' = [x_{1t}, \ldots, x_{Nt}]$, $\mathbf{e} = [\mathbf{e}_1, \ldots, \mathbf{e}_T]'$ is a $T \times N$ matrix of idiosyncratic errors, and $\mathbf{e}_t' = [e_{1t}, \ldots, e_{Nt}]$. Furthermore, $\mathbf{F} = [\mathbf{F}_1, \ldots, \mathbf{F}_T]'$ and $\boldsymbol{\Lambda} = [\boldsymbol{\lambda}_1, \ldots, \boldsymbol{\lambda}_N]'$. Under fairly weak assumptions, the factors and factor loadings can be estimated consistently as $N \to \infty$ and $T \to \infty$ by the PC estimator that minimizes the total sum of squares

$$S(\mathbf{F}, \boldsymbol{\Lambda}) = \operatorname{tr}[(\mathbf{X} - \mathbf{F}\boldsymbol{\Lambda}')'(\mathbf{X} - \mathbf{F}\boldsymbol{\Lambda}')],$$

subject to the constraint $T^{-1}\mathbf{F}'\mathbf{F} = \mathbf{I}_r$ (Bai and Ng 2002; Stock and Watson 2002b). The estimators of $\mathbf{F}$ and $\boldsymbol{\Lambda}$ result as $\widehat{\mathbf{F}} = \sqrt{T}\widehat{\mathbf{V}}_r$ and $\widehat{\boldsymbol{\Lambda}} = T^{-1/2}\mathbf{X}'\widehat{\mathbf{V}}_r$, respectively, where $\widehat{\mathbf{V}}_r$ is the matrix of the r orthonormal eigenvectors corresponding to the r largest eigenvalues of the matrix $\mathbf{X}\mathbf{X}'$ (e.g., Stock and Watson 2002b). These estimators of $\boldsymbol{\Lambda}$ and $\mathbf{F}$ will be called the PC–OLS estimators.

If the idiosyncratic errors are heteroscedastic or autocorrelated, the PC–OLS estimator is not efficient. For the heteroscedastic case Choi (2011) suggests GLS-type estimators that minimize the weighted sum of squares

$$S(\mathbf{F}, \boldsymbol{\Lambda}, \boldsymbol{\Omega}) = \operatorname{tr}[\boldsymbol{\Omega}^{-1}(\mathbf{X} - \mathbf{F}\boldsymbol{\Lambda}')'(\mathbf{X} - \mathbf{F}\boldsymbol{\Lambda}')],$$

where $\boldsymbol{\Omega} = \operatorname{diag}[E(e_{1t}^2), \ldots, E(e_{Nt}^2)]$ for all t. Forni et al. (2005) and Choi (2011) consider the case of an arbitrary covariance matrix $\boldsymbol{\Omega}$. It should be noted, however, that the factors are not identified without additional assumptions on the matrix $\boldsymbol{\Omega}$. To see this, consider the spectral decomposition $\boldsymbol{\Omega} = \sum_{i=1}^N \mu_i \mathbf{v}_i \mathbf{v}_i'$, where μ_i and $\mathbf{v}_i$ denote the ith eigenvalue and corresponding eigenvector, respectively. The matrix $\boldsymbol{\Omega}$ may be decomposed in form of a factor model yielding $\boldsymbol{\Omega} = \boldsymbol{\Gamma}\boldsymbol{\Gamma}' + \boldsymbol{\Omega}^*$ where, for example, $\boldsymbol{\Gamma}\boldsymbol{\Gamma}' = \sum_{i=1}^k w_i \mu_i \mathbf{v}_i \mathbf{v}_i'$, $k \le N$, $0 < w_i < 1$ for all i, and

$$\boldsymbol{\Omega}^* = \sum_{i=1}^{k} (1 - w_i)\mu_i \mathbf{v}_i \mathbf{v}_i' + \sum_{i=k+1}^{N} \mu_i \mathbf{v}_i \mathbf{v}_i'$$

is a symmetric positive definite matrix. Thus, $E(\mathbf{X}_t \mathbf{X}_t') = \boldsymbol{\Lambda}\boldsymbol{\Lambda}' + \boldsymbol{\Omega} = \boldsymbol{\Lambda}^* \boldsymbol{\Lambda}^{*\prime} + \boldsymbol{\Omega}^*$, where $\boldsymbol{\Lambda}^* = [\boldsymbol{\Lambda}, \boldsymbol{\Gamma}]$. In order to distinguish the common factors from the idiosyncratic components, the covariance matrix $\boldsymbol{\Omega}$ has to be restricted in such a way that the idiosyncratic errors cannot mimic the pervasive correlation due to the common factors. This is usually ensured by assuming that all eigenvalues of $\boldsymbol{\Omega}$ are bounded as $N \to \infty$. One possibility is to specify $\boldsymbol{\Omega}$ as a diagonal matrix. An alternative is to allow for some spatial correlation of the form $\boldsymbol{\Omega} = \sigma^2(\mathbf{I}_N - \varrho \mathbf{W}_N)^{-1}(\mathbf{I}_N - \varrho \mathbf{W}_N')^{-1}$, where all eigenvalues

of the spatial weight matrix $\mathbf{W}_N$ are smaller than one and $0 \le \varrho < 1$ (e.g., Chudik, Pesaran, and Tosetti 2010). Another problem is that in a model with arbitrary covariance matrix $\mathbf{\Omega}$, the number of parameters increases with the square of N, i.e., the model implies a large number of additional parameters that may even exceed the number of observations. Finally, the estimator $\widehat{\mathbf{\Omega}} = T^{-1}(\mathbf{X} - \widehat{\mathbf{F}}\widehat{\mathbf{\Lambda}}')'(\mathbf{X} - \widehat{\mathbf{F}}\widehat{\mathbf{\Lambda}}')$ is singular and, hence, the inverse does not exist (see also Boivin and Ng 2006). As a result, when deriving our estimator we start from an approximate likelihood function featuring mutually uncorrelated idiosyncratic components, thereby following the traditional factor framework. Our main results concerning this estimator, however, are obtained under much weaker assumptions. In particular, the idiosyncratic components are assumed to be weakly correlated in the sense of Stock and Watson (2002a) and Bai and Ng (2002).

In the following section, we propose a GLS-type estimator which, in contrast to earlier work focusing on either heteroscedastic (Forni et al. 2005; Doz, Giannone, and Reichlin 2006; Choi 2011) *or* autocorrelated errors (Stock and Watson 2005) accommodates both features.

3. THE PC–GLS ESTIMATOR

In this section, we follow Stock and Watson (2005) and assume that the idiosyncratic components have a stationary heterogeneous autoregressive representation of the form

$$\rho_i(L)e_{it} = \varepsilon_{it}, \tag{3}$$

where $\rho_i(L) = 1 - \rho_{1,i}L - \cdots - \rho_{p_i,i}L^{p_i}$. It is important to note that (3) is employed as an auxiliary model to capture the main features of the idiosyncratic dynamics. Our asymptotic analysis allows for misspecification of the dynamic process that gives rise to some remaining autocorrelation in ε_{it}.

We do not need to make specific assumptions about the dynamic properties of the vector of common factors, $\mathbf{F}_t$. Apart from some minor regularity conditions the only consequential assumption that we impose for the derivation of the limiting distribution is that the factors are weakly serially correlated (Assumption 1 in Section 4).

Consider the approximate Gaussian log-likelihood function

$$S(\boldsymbol{\theta}) = -\sum_{i=1}^{N} \frac{T - p_i}{2} \log \sigma_i^2 - \sum_{i=1}^{N} \sum_{t=p_i+1}^{T} \frac{(e_{it} - \rho_{1,i}e_{i,t-1} - \cdots - \rho_{p_i,i}e_{i,t-p_i})^2}{2\sigma_i^2}, \tag{4}$$

where $\boldsymbol{\theta} = [\mathrm{vec}(\mathbf{F})', \mathrm{vec}(\mathbf{\Lambda})', \sigma_1^2, \ldots, \sigma_N^2, \rho_{1,1}, \ldots, \rho_{p_1,1}, \ldots, \rho_{1,N}, \ldots, \rho_{p_N,N}]'$ and $\sigma_i^2 = E(\varepsilon_{it}^2)$. Note that this likelihood function results from conditioning on p_i initial values. If x_{it} is normally distributed and $N \to \infty$, then the PC–GLS estimator is asymptotically equivalent to the ML estimator. This can be seen by writing the log-likelihood function as $\mathcal{L}(\mathbf{X}) = \mathcal{L}(\mathbf{X}|\mathbf{F}) + \mathcal{L}(\mathbf{F})$, where $\mathcal{L}(\mathbf{X}|\mathbf{F})$ denotes the log-likelihood function of $x_{11}, \ldots, x_{NT}$ conditional on the factors $\mathbf{F}$ and $\mathcal{L}(\mathbf{F})$ is the log-likelihood of $(\mathbf{F}_1', \ldots, \mathbf{F}_T')$. Since $\mathcal{L}(\mathbf{X}|\mathbf{F})$ is $O_p(NT)$ and $\mathcal{L}(\mathbf{F})$ is $O_p(T)$, it follows that as $N \to \infty$ maximizing $\mathcal{L}(\mathbf{X}|\mathbf{F})$ is equivalent to maximizing the full log-likelihood function $\mathcal{L}(\mathbf{X})$.

An important problem with maximizing this likelihood function is that the likelihood function is unbounded in general (see, e.g., Anderson 1984, p. 570). To see this, consider a factor model with a single factor (i.e., $r = 1$). If $\widehat{F}_t = x_{it}/\sqrt{T^{-1}\sum_{t=1}^{T} x_{it}^2}$ and $\widehat{\lambda}_i = 1$ for some i and $t = 1, \ldots, T$, then $\widehat{\sigma}_i^2 = 0$ and, therefore, the likelihood tends to infinity. This problem is well known also in other fields of statistics (e.g., the estimation of mixture densities) and we adapt techniques for obtaining the maximum likelihood estimator that were developed to cope with this problem. Specifically, we are focusing on the estimator $\widehat{\boldsymbol{\theta}}$ that attains a *local* maximum of the likelihood function. Redner and Walker (1984) provide two conditions under which the local maximum in a neighborhood of the true values $\boldsymbol{\theta}^0$ yields a consistent and asymptotically normally distributed estimator. These two conditions ensure that the likelihood function is concave in a neighborhood of the true values.

The derivatives of the log-likelihood function are given by

$$\mathbf{g}_{\boldsymbol{\lambda}_i}(\cdot) = \frac{\partial S(\cdot)}{\partial \boldsymbol{\lambda}_i} = \frac{1}{\sigma_i^2}\left\{\sum_{t=p_i+1}^{T} \varepsilon_{it}[\rho_i(L)\mathbf{F}_t]\right\}, \tag{5}$$

$$\begin{aligned}\mathbf{g}_{\mathbf{F}_t}(\cdot) &= \frac{\partial S(\cdot)}{\partial \mathbf{F}_t} \\ &= \sum_{i=1}^{N} \frac{1}{\sigma_i^2}\left(\varepsilon_{it}\boldsymbol{\lambda}_i - \rho_{1,i}\varepsilon_{i,t+1}\boldsymbol{\lambda}_i - \cdots - \rho_{p_i,i}\varepsilon_{i,t+p_i}\boldsymbol{\lambda}_i\right) \\ &= \sum_{i=1}^{N} \frac{1}{\sigma_i^2}[\rho_i(L^{-1})\varepsilon_{it}]\boldsymbol{\lambda}_i,\end{aligned} \tag{6}$$

$$g_{\rho_{k,i}}(\cdot) = \frac{\partial S(\cdot)}{\partial \rho_{k,i}} = \frac{1}{\sigma_i^2}\sum_{t=p_i+1}^{T} \varepsilon_{it}(x_{i,t-k} - \boldsymbol{\lambda}_i'\mathbf{F}_{t-k}), \tag{7}$$

$$g_{\sigma_i^2}(\cdot) = \frac{\partial S(\cdot)}{\partial \sigma_i^2} = \frac{\sum_{t=p_i+1}^{T} \varepsilon_{it}^2}{2\sigma_i^4} - \frac{T - p_i}{2\sigma_i^2}, \tag{8}$$

where $\varepsilon_{is} = 0$ for $s > T$. It is not difficult to verify that condition 1 of Redner and Walker (1984) related to the derivatives of the likelihood function is satisfied. Furthermore, the Fisher information matrix is well defined and positive definite at the true value $\boldsymbol{\theta}^0$ (condition 2 of Redner and Walker 1984). It follows that the ML estimator that locally maximizes the log-likelihood function is consistent and asymptotically normally distributed. Our proposed estimator maximizes the likelihood in the neighborhood of the PC estimator. Since the latter is consistent for a particular normalization of the parameters, the local maximizer of the log-likelihood function in the neighborhood of the PC estimator is consistent and asymptotically normally distributed.

A practical problem is the large dimension of the system consisting of $2Nr + N + \sum p_i$ equations. Accordingly, in many practical situations it is very demanding to compute the inverse of the Hessian matrix that is required to construct an iterative minimization algorithm. We therefore suggest a simple two-step estimator that is asymptotically equivalent to locally maximizing the Gaussian likelihood function.

Let us first assume that the covariance parameters $\boldsymbol{\rho}$ and $\mathbf{\Sigma}$ are known, where $\boldsymbol{\rho} = [\rho_{1,1}, \ldots, \rho_{p_1,1}, \ldots, \rho_{1,N}, \ldots, \rho_{p_N,N}]'$ and $\mathbf{\Sigma} = \mathrm{diag}(\sigma_1^2, \ldots, \sigma_N^2)$. The (infeasible) two-step estimators

$\widetilde{\mathbf{F}}_t$ $(t = 1, \ldots, T)$ and $\widetilde{\boldsymbol{\lambda}}_i$ $(i = 1, \ldots, N)$ that result from using the PC estimators as first-stage estimators are obtained by solving the following sets of equations:

$$\mathbf{g}_{\boldsymbol{\lambda}_i}(\widetilde{\boldsymbol{\lambda}}_i, \widehat{\mathbf{F}}, \boldsymbol{\rho}, \boldsymbol{\Sigma}) = \mathbf{0}, \tag{9}$$

$$\mathbf{g}_{\mathbf{F}_t}(\widehat{\boldsymbol{\Lambda}}, \widetilde{\mathbf{F}}_t, \boldsymbol{\rho}, \boldsymbol{\Sigma}) = \mathbf{0}, \tag{10}$$

where $\widehat{\mathbf{F}} = [\widehat{\mathbf{F}}_1, \ldots, \widehat{\mathbf{F}}_T]'$ and $\widehat{\boldsymbol{\Lambda}} = [\widehat{\boldsymbol{\lambda}}_1, \ldots, \widehat{\boldsymbol{\lambda}}_N]'$ are the ordinary PC–OLS estimators of $\mathbf{F}$ and $\boldsymbol{\Lambda}$.

It is not difficult to see that the two-step estimator of $\boldsymbol{\lambda}_i$ resulting from (9) is equivalent to the least-squares estimator of $\boldsymbol{\lambda}_i$ in the regression:

$$(\rho_i(L)x_{it}) = (\rho_i(L)\widehat{\mathbf{F}}_t)'\boldsymbol{\lambda}_i + \varepsilon^*_{it} \qquad (t = p_i + 1, \ldots, T), \tag{11}$$

where $\varepsilon^*_{it} = \varepsilon_{it} + \rho_i(L)(\mathbf{F}_t - \widehat{\mathbf{F}}_t)'\boldsymbol{\lambda}_i$.

The two-step estimator of $\mathbf{F}_t$ (given $\widehat{\boldsymbol{\Lambda}}$) is more difficult to motivate. Condition (10) gives rise to the two-way GLS transformation that accounts for both serial correlation and heteroscedasticity:

$$\frac{1}{\sigma_i}\rho_i(L)x_{it} = \frac{1}{\sigma_i}\widehat{\boldsymbol{\lambda}}_i'[\rho_i(L)\mathbf{F}_t] + \frac{1}{\sigma_i}\varepsilon_{it}, \tag{12}$$

where for notational convenience we assume $p_i = p$ for all i. Furthermore, our notation ignores the estimation error that results from replacing $\boldsymbol{\lambda}_i$ by $\widehat{\boldsymbol{\lambda}}_i$.[2] It is not difficult to see[3] that the estimator of $\mathbf{F}_t$ based on (12) is equivalent to the least-squares estimator of $\mathbf{F}_t$ in the regression

$$\frac{1}{\omega_i}x_{it} = \left(\frac{1}{\omega_i}\widehat{\boldsymbol{\lambda}}_i'\right)\mathbf{F}_t + u^*_{it} \qquad (i = 1, \ldots, N), \tag{13}$$

where $u^*_{it} = \omega_i^{-1}[e_{it} + (\boldsymbol{\lambda}_i - \widehat{\boldsymbol{\lambda}}_i)'\mathbf{F}_t]$ and $\omega_i^2 = E(e_{it}^2)$, that is, ignoring the GLS transformation with respect to autocorrelation. In what follows, we focus on this simplified version of the two-step estimation approach as its properties are equivalent to those of the full two-way GLS estimation procedure.

4. ASYMPTOTIC DISTRIBUTION OF THE TWO–STEP PC–GLS ESTIMATOR

Our analysis is based on a similar set of assumptions as in Bai (2003), which is restated here for completeness.

Assumption 1. There exists a positive constant $M < \infty$ such that for all N and T: (i) $E\|\mathbf{F}_t\|^4 \le M$ for all t and $T^{-1}\sum_{t=1}^T \mathbf{F}_t\mathbf{F}_t' \xrightarrow{p} \boldsymbol{\Psi}_F$ (p.d.). (ii) $\|\boldsymbol{\lambda}_i\| \le \overline{\lambda} < \infty$ for all i and $N^{-1}\boldsymbol{\Lambda}'\boldsymbol{\Lambda} \to \boldsymbol{\Psi}_\Lambda$ (p.d.). (iii) For the idiosyncratic components it is assumed that $E(e_{it}) = 0$, $E|e_{it}|^8 \le M$, $0 < |\gamma_N(s,s)| \le M$, $T^{-1}\sum_{s=1}^T\sum_{t=1}^T |\gamma_N(s,t)| \le M$, where $\gamma_N(s,t) = E(N^{-1} \times \sum_{i=1}^N e_{is}e_{it})$. Furthermore, $N^{-1}\sum_{i=1}^N\sum_{j=1}^N \tau_{ij} \le M$, $\sum_{i=1}^N \tau_{ij} \le M$, where $\tau_{ij} = \sup_t\{|E(e_{it}e_{jt})|\}$,

$$\frac{1}{NT}\sum_{i=1}^N\sum_{j=1}^N\sum_{t=1}^T\sum_{s=1}^T |E(e_{it}e_{js})| \le M,$$

$$E\left|\frac{1}{\sqrt{N}}\sum_{i=1}^N [e_{is}e_{it} - E(e_{is}e_{it})]\right|^4 \le M.$$

(iv) $E(N^{-1}\sum_{i=1}^N \|T^{-1/2}\sum_{t=1}^T \mathbf{F}_{t-k}e_{it}\|^2) \le M$ for all N, T, and k.

(v) For all t, k, N, and T:

$$E\left\|\frac{1}{\sqrt{NT}}\sum_{s=1}^T\sum_{i=1}^N \mathbf{F}_{s-k}[e_{is}e_{it} - E(e_{is}e_{it})]\right\|^2 \le M,$$

$$E\left\|\frac{1}{\sqrt{NT}}\sum_{s=1}^T\sum_{i=1}^N \mathbf{F}_{s-k}\boldsymbol{\lambda}_i'e_{is}\right\|^2 \le M,$$

$$\frac{1}{\sqrt{N}}\sum_{i=1}^N \boldsymbol{\lambda}_i e_{it} \xrightarrow{d} \mathcal{N}\big(\mathbf{0}, \mathbf{V}^{(t)}_{\boldsymbol{\lambda}\mathbf{e}}\big),$$

where $\mathbf{V}^{(t)}_{\boldsymbol{\lambda}\mathbf{e}} = \lim_{N\to\infty} N^{-1}\sum_{i=1}^N\sum_{j=1}^N \boldsymbol{\lambda}_i\boldsymbol{\lambda}_j'E(e_{it}e_{jt})$ and for each i

$$\frac{1}{\sqrt{T}}\sum_{t=p_i+1}^{T-p_i} \mathbf{F}_t e_{i,t+k} \xrightarrow{d} \mathcal{N}\big(\mathbf{0}, \mathbf{V}^{(i)}_{\mathbf{Fe}}\big) \quad \text{for } -p_i \le k \le p_i,$$

where $\mathbf{V}^{(i)}_{\mathbf{Fe}} = \lim_{T\to\infty} T^{-1}\sum_{s=1}^T\sum_{t=1}^T E(\mathbf{F}_t\mathbf{F}_s'e_{i,s-k}e_{i,t-k})$.

For a thorough discussion of these assumptions, see Bai and Ng (2002) and Bai (2003). It is well known (e.g., Bai and Ng 2002) that for the asymptotic analysis of the estimators, the factors have to be normalized such that in the limit the common factors obey the same normalization as the estimated factors. Following Bai and Ng (2002), this is achieved by normalizing the factors as

$$\begin{aligned}\mathbf{F}\boldsymbol{\Lambda}' &= (\mathbf{F}\mathbf{H})(\mathbf{H}^{-1}\boldsymbol{\Lambda}')\\ &= \mathbf{F}_*\boldsymbol{\Lambda}_*',\end{aligned}$$

where

$$\mathbf{H} = T\boldsymbol{\Lambda}'\boldsymbol{\Lambda}\mathbf{F}'\widehat{\mathbf{F}}(\widehat{\mathbf{F}}'\mathbf{X}\mathbf{X}'\widehat{\mathbf{F}})^{-1}.$$

It can be shown that $T^{-1}\mathbf{F}_*'\mathbf{F}_* \xrightarrow{p} \mathbf{I}_r$ and, therefore, $\mathbf{F}_*$ has asymptotically the same normalization as $\widehat{\mathbf{F}}$.

As we do not impose the assumptions of a strict factor model with stationary idiosyncratic errors, we define the following "pseudo-true" values of the autoregressive and variance parameters:

$$\omega_i^2 = \lim_{T\to\infty} T^{-1}\sum_{t=1}^T E(e_{it}^2),$$

$$\big[\rho_{1,i}, \ldots, \rho_{p_i,i}\big]' = \boldsymbol{\Gamma}_{i,11}^{-1}\boldsymbol{\Gamma}_{i,10},$$

[2] The complete error term is given by $\sigma_i^{-1}[\varepsilon_{it} + (\boldsymbol{\lambda}_i - \widehat{\boldsymbol{\lambda}}_i)'\rho_i(L)\mathbf{F}_t]$. However, as will be shown below, the estimation error in $\widehat{\boldsymbol{\lambda}}_i$ does not affect the asymptotic properties of the estimator.

[3] See the working article version of this article (Breitung and Tenhofen 2009). The equivalence results from the fact that the Equations (12) for $i = 1, \ldots, N$ can be written in form of a system of seemingly unrelated regressions (SUR) with identical regressors. In such systems the GLS estimator, taking into account the autocorrelation of the errors, is identical to the OLS estimator.

where

$$\mathbf{\Gamma}_i = \lim_{T\to\infty} E\left(\frac{1}{T}\sum_{t=p_i+1}^{T}\begin{bmatrix} e_{i,t-1} \\ \vdots \\ e_{i,t-p_i}\end{bmatrix}[\, e_{it} \quad \cdots \quad e_{i,t-p_i}\,]\right)$$
$$= [\, \mathbf{\Gamma}_{i,10} \quad \mathbf{\Gamma}_{i,11}\,],$$

$\mathbf{\Gamma}_{i,10}$ is a $p_i \times 1$ vector, and $\mathbf{\Gamma}_{i,11}$ is a $p_i \times p_i$ matrix.

For the asymptotic analysis we need to impose the following assumption.

Assumption 2. (i) There exists a positive constant $C < \infty$, such that for all i: $\frac{1}{C} < \omega_i^2 < C$. (ii) The matrix $\mathbf{\Gamma}_{i,11}$ is positive definite.

In practice, the covariance parameters are usually unknown and must be replaced by consistent estimates. The feasible two-step PC–GLS estimators $\widetilde{\boldsymbol{\lambda}}_{i,\widehat{\rho}}$ and $\widetilde{\mathbf{F}}_{t,\widehat{\omega}}$ solve the first-order conditions

$$\widetilde{\mathbf{g}}_{\boldsymbol{\lambda}_i}\big(\widetilde{\boldsymbol{\lambda}}_{i,\widehat{\rho}}, \widehat{\mathbf{F}}, \widehat{\boldsymbol{\rho}}^{(i)}\big) = \sum_{t=p_i+1}^{T}[\widehat{\rho}_i(L)(x_{it} - \widetilde{\boldsymbol{\lambda}}'_{i,\widehat{\rho}}\widehat{\mathbf{F}}_t)][\widehat{\rho}_i(L)\widehat{\mathbf{F}}_t]$$
$$= \mathbf{0}, \tag{14}$$

$$\widetilde{\mathbf{g}}_{\mathbf{F}_t}\big(\widehat{\mathbf{\Lambda}}, \widetilde{\mathbf{F}}_{t,\widehat{\omega}}, \widehat{\mathbf{\Omega}}\big) = \sum_{i=1}^{N}\frac{1}{\widehat{\omega}_i^2}(x_{it} - \widehat{\boldsymbol{\lambda}}_i'\widetilde{\mathbf{F}}_{t,\widehat{\omega}})\widehat{\boldsymbol{\lambda}}_i = \mathbf{0}, \tag{15}$$

where

$$\widehat{\omega}_i^2 = \frac{1}{T}\sum_{t=1}^{T}\widehat{e}_{it}^2 \tag{16}$$

and $\widehat{e}_{it} = x_{it} - \widehat{\boldsymbol{\lambda}}_i'\widehat{\mathbf{F}}_t$. Furthermore, $\widehat{\boldsymbol{\rho}}^{(i)} = [\widehat{\rho}_{1,i}, \ldots, \widehat{\rho}_{p_i,i}]'$ is the least-squares estimator from the regression

$$\widehat{e}_{it} = \widehat{\rho}_{1,i}\widehat{e}_{i,t-1} + \cdots + \widehat{\rho}_{p_i,i}\widehat{e}_{i,t-p_i} + \widehat{\varepsilon}_{it}. \tag{17}$$

To study the limiting distribution of the feasible two-step PC–GLS estimator, the following lemma is used.

Lemma 1. Let $\widehat{\boldsymbol{\rho}}^{(i)} = [\widehat{\rho}_{1,i}, \ldots, \widehat{\rho}_{p_i,i}]'$ denote the least-squares estimates from (17) and $\widehat{\omega}_i^2$ is the estimator defined in (16). Under Assumption 1 we have as $(N, T \to \infty)$

$$\widehat{\boldsymbol{\rho}}^{(i)} = \boldsymbol{\rho}^{(i)} + O_p\big(T^{-1/2}\big) + O_p(\delta_{NT}^{-2}) \qquad \text{and}$$
$$\widehat{\omega}_i^2 = \omega_i^2 + O_p\big(T^{-1/2}\big) + O_p(\delta_{NT}^{-2}),$$

where $\delta_{NT} = \min(\sqrt{N}, \sqrt{T})$.

The following theorem presents the limiting distribution of the feasible two-step PC–GLS estimator.

Theorem 1. (i) Under Assumptions 1–2 and if $(N, T \to \infty)$ and $\sqrt{T}/N \to 0$, then for each i,

$$\sqrt{T}(\widetilde{\boldsymbol{\lambda}}_{i,\widehat{\rho}} - \mathbf{H}^{-1}\lambda_i) \xrightarrow{d} \mathcal{N}\big(\mathbf{0}, \widetilde{\mathbf{\Psi}}_F^{(i)^{-1}}\widetilde{V}_{\mathbf{Fe}}^{(i)}\widetilde{\Psi}_F^{(i)^{-1}}\big),$$

where

$$\widetilde{\mathbf{V}}_{\mathbf{Fe}}^{(i)} = \lim_{T\to\infty}\frac{1}{T}\sum_{s=p_i+1}^{T}\sum_{t=p_i+1}^{T}E[\rho_i(L)\mathbf{H}'\mathbf{F}_t\rho_i(L)\mathbf{F}_s'\mathbf{H}\varepsilon_{is}\varepsilon_{it}],$$

$$\varepsilon_{it} = \rho_i(L)e_{it},$$

$$\widetilde{\mathbf{\Psi}}_F^{(i)} = \lim_{T\to\infty}\frac{1}{T}\sum_{t=p_i+1}^{T}E\big\{[\rho_i(L)\mathbf{H}'\mathbf{F}_t][\rho_i(L)\mathbf{H}'\mathbf{F}_t]'\big\}.$$

(ii) If $(N, T \to \infty)$ and $\sqrt{N}/T \to 0$, then for each t,

$$\sqrt{N}(\widetilde{\mathbf{F}}_{t,\widehat{\omega}} - \mathbf{H}'\mathbf{F}_t) \xrightarrow{d} \mathcal{N}\big(\mathbf{0}, \widetilde{\mathbf{\Psi}}_{\mathbf{\Lambda}}^{-1}\widetilde{\mathbf{V}}_{\lambda\mathbf{e}}^{(t)}\widetilde{\mathbf{\Psi}}_{\mathbf{\Lambda}}^{-1}\big),$$

where

$$\widetilde{\mathbf{V}}_{\lambda\mathbf{e}}^{(t)} = \lim_{N\to\infty}N^{-1}\sum_{i=1}^{N}\sum_{j=1}^{N}\frac{1}{\omega_i^2\omega_j^2}\mathbf{H}^{-1}\boldsymbol{\lambda}_i\boldsymbol{\lambda}_j'\mathbf{H}'^{-1}E(e_{it}e_{jt}),$$

$\widetilde{\mathbf{\Psi}}_{\mathbf{\Lambda}} = \lim_{N\to\infty}N^{-1}\mathbf{H}^{-1}\mathbf{\Lambda}'\mathbf{\Omega}^{-1}\mathbf{\Lambda}\mathbf{H}'^{-1}$, and $\mathbf{\Omega} = \operatorname{diag}(\omega_1^2, \ldots, \omega_N^2)$.

Remark 1. From (i) it follows that the asymptotic distribution remains the same if the estimate $\widehat{\rho}_i(L)\widehat{\mathbf{F}}_t$ in (11) is replaced by $\rho_i(L)\mathbf{H}'\mathbf{F}_t$. This suggests that the estimation error in $\widehat{\mathbf{F}}_t$ and $\widehat{\rho}_i(L)$ does not affect the asymptotic properties of the estimator $\widetilde{\boldsymbol{\lambda}}_{i,\widehat{\rho}}$. A similar result holds for the regressor $\widehat{\omega}_i^{-1}\widehat{\boldsymbol{\lambda}}_i$. In other words, the additional assumptions on the relative rates of N and T ensure that the estimates of the regressors in Equations (11) and (13) can be treated as "super-consistent."

Remark 2. The assumptions on the relative rates of N and T may appear to be in conflict with each other. However, the two parts of Theorem 1 are fulfilled if $N = cT^{\delta}$ where $1/2 < \delta < 2$. Therefore, the limiting distribution should give a reliable guidance if both dimensions N and T are of comparable magnitude.

Remark 3. It is interesting to compare the result of Theorem 1 with the asymptotic distribution obtained by Choi (2011). In the latter article it is assumed that $E(\mathbf{e}_t\mathbf{e}_t') = \mathbf{\Omega}$ for all t, where $\mathbf{e}_t = [e_{1t}, \ldots, e_{Nt}]'$, that is, the idiosyncratic components are assumed to be stationary. In this case, the model can be transformed as $\mathbf{X}^* = \mathbf{F}^*\mathbf{\Lambda}^{*\prime} + \mathbf{e}^*$, where $\mathbf{X}^* = \mathbf{X}\mathbf{\Omega}^{-1/2}$, $\mathbf{F}^* = \mathbf{F}\mathbf{J}$, $\mathbf{\Lambda}^* = \mathbf{\Omega}^{-1/2}\mathbf{\Lambda}(\mathbf{J}')^{-1}$, $\mathbf{e}^* = \mathbf{e}\mathbf{\Omega}^{-1/2}$, and

$$\mathbf{J} = T\mathbf{\Lambda}'\mathbf{\Omega}^{-1}\mathbf{\Lambda}\mathbf{F}'\widetilde{\mathbf{F}}(\widetilde{\mathbf{F}}'\mathbf{X}\mathbf{\Omega}^{-1}\mathbf{X}'\widetilde{\mathbf{F}})^{-1},$$

such that the covariance matrix of $\mathbf{e}^*$ is identical the identity matrix. Note that the matrix normalizing the factors in Choi (2011), $\mathbf{J}$, is different from the one employed for the PC–OLS estimator (and for our PC–GLS estimator), $\mathbf{H}$. Imposing the former normalization, the asymptotic covariance matrix of Choi's (2011) GLS estimator $\widetilde{\mathbf{F}}$ reduces to a diagonal matrix.

Remark 4. If the errors are serially uncorrelated, our PC–GLS estimator of $\mathbf{F}_t$ is related to the estimator suggested by Forni et al. (2005). Analogous to Choi (2011), an important distinguishing feature is the different normalization of the factors. Forni et al. (2005) propose an estimator of $\mathbf{\Omega}$ that is obtained from integrating the estimated spectral density matrix of the idiosyncratic errors. The factors are obtained from solving the generalized eigenvalue problem $|\nu\widetilde{\mathbf{\Omega}} - T^{-1}\mathbf{X}'\mathbf{X}| = 0$, where $\widetilde{\mathbf{\Omega}}$ denotes the estimated covariance matrix of the idiosyncratic errors. Note that the matrix of eigenvectors, $\widetilde{\mathbf{V}}$, of the generalized eigenvalue problem obeys the normalization $\widetilde{\mathbf{V}}'\widetilde{\mathbf{\Omega}}\widetilde{\mathbf{V}} = \mathbf{I}$.

Remark 5. The two-step approach can also be applied to an unbalanced dataset with different numbers of time periods for the variables. Stock and Watson (2002b) suggest an EM algorithm, where the missing values are replaced by an estimate of the common component. Let $\hat{x}_{it} = \widehat{\boldsymbol{\lambda}}_i'\widehat{\mathbf{F}}_t$ denote the estimated observation based on the balanced dataset ignoring all time periods with missing observations. The updated estimates of the

common factors and factor loadings are obtained by applying the PC–OLS estimator to the enhanced dataset, where the missing values are replaced by the estimates $\hat{x}_{it}$. Employing the updated estimates of $\mathbf{F}_t$ and $\boldsymbol{\lambda}_i$, results in improved estimates of the missing values that can in turn be used to yield new PC–OLS estimates of the common factors and factor loadings. This estimation procedure can be iterated until convergence. Similarly, the two-step estimation procedure can be initialized by using the reduced (balanced) dataset to obtain the PC–OLS estimates $\widehat{\mathbf{F}}_t$ and $\widehat{\boldsymbol{\lambda}}_i$. In the second step, the vector of common factors is estimated from regression (13). As the T cross-section regressions may employ different numbers of observations, missing values do not raise any problems. Similarly, the N time-series regressions (11) may be based on different sample sizes. As in the EM algorithm, this estimation procedure can be iterated until convergence.

Remark 6. The two-step estimators can be iterated using the resulting estimates $\widetilde{\boldsymbol{\lambda}}_{i,\widehat{\rho}}$ and $\widetilde{\mathbf{F}}_{t,\widehat{\omega}}$ instead of the estimates $\widehat{\mathbf{F}}_t$ and $\widehat{\boldsymbol{\lambda}}_i$ in regressions (11) and (13). Similarly, improved estimators of the covariance parameters can be obtained from the second-step residuals $\widetilde{e}_{it} = x_{it} - \widetilde{\boldsymbol{\lambda}}'_{i,\widehat{\rho}}\widetilde{\mathbf{F}}_{t,\widehat{\omega}}$. However, since the estimation error of the factors, factor loadings, and covariance parameters does not affect the limiting distribution of the two-step estimators, additional iteration steps do not improve their *asymptotic* properties. Nevertheless, further iterations may improve the performance in *small samples*.

5. ASYMPTOTIC EFFICIENCY

In this section, we study the relative asymptotic efficiency of the estimators. The following theorem shows that the PC–GLS estimator is generally more efficient than the PC–OLS estimator if the temporal and contemporaneous variance and covariance functions of the error, e_{it}, are correctly specified.

Theorem 2. Under Assumptions 1–2 and if $E(\mathbf{e}_t\mathbf{e}_t') = \boldsymbol{\Omega} = \operatorname{diag}\{\omega_1^2, \ldots, \omega_N^2\}$, $E(\boldsymbol{\varepsilon}_i\boldsymbol{\varepsilon}_i') = \sigma_i^2\mathbf{I}_{T-p_i}$, and $\mathbf{F}_t$ is independent of e_{it} for all i and t, then the PC–GLS estimators of $\boldsymbol{\lambda}_i$ and $\mathbf{F}_t$ are asymptotically more efficient than the respective PC–OLS estimator in the sense that the difference of the respective asymptotic covariance matrices is positive semidefinite.

Admittedly, this result is of limited practical use as we typically cannot assume that all variance and covariance functions are correctly specified. In Section 3 we have argued that the PC–GLS estimator can be seen as a pseudo-ML estimator, which provides us with simple and generally more efficient estimators even if the variance and covariance functions are misspecified. Indeed, our Monte Carlo experiments (some of which are presented in Section 6) suggest that the PC–GLS estimator tends to outperform the PC–OLS estimator even if the covariance functions are misspecified. This finding may suggest that the PC–GLS estimator is always more efficient as it takes into account at least some form of heteroscedasticity and autocorrelation, whereas the PC–OLS estimator simply ignores the possibility of individual specific variances and serial correlation. Unfortunately, this is not true as it is possible to construct special cases characterized by misspecification of the variance or covariance function, where the PC–OLS estimator is asymptotically more efficient than the PC–GLS estimator.[4] One possibility to cope with this problem is to minimize the misspecification by carefully determining the autoregressive lag order employing, for instance, the Akaike or Schwarz criterion.

In what follows, we propose an alternative approach that combines the two aforementioned estimators yielding a "hybrid estimator" which is at least as efficient as each of the two estimators. To construct such an estimator, we combine the moment conditions of PC–OLS and PC–GLS such that the respective generalized method of moments (GMM) estimator based on two sets of moment conditions is more efficient than any estimator based on a subset of moment conditions. The respective moments for estimating the common factors are

$$\text{PC–OLS:}\quad \mathbf{m}_1(\mathbf{F}_t) = \sum_{i=1}^{N} \widehat{\boldsymbol{\lambda}}_i(x_{it} - \widehat{\boldsymbol{\lambda}}_i'\mathbf{F}_t), \tag{18}$$

$$\text{PC–GLS:}\quad \mathbf{m}_2(\mathbf{F}_t) = \sum_{i=1}^{N} \frac{1}{\widehat{\omega}_i^2}\widehat{\boldsymbol{\lambda}}_i(x_{it} - \widehat{\boldsymbol{\lambda}}_i'\mathbf{F}_t). \tag{19}$$

Define $\mathbf{z}_i = [\widehat{\boldsymbol{\lambda}}_i', \widehat{\boldsymbol{\lambda}}_i'/\widehat{\omega}_i^2]'$ as the vector of instruments and $\mathbf{Z} = [\mathbf{z}_1, \ldots, \mathbf{z}_N]'$ is an $N \times (2r)$ matrix. The GMM estimator $\widehat{\mathbf{F}}_t^{\text{gmm}}$ is given by (e.g., Hayashi 2000, chap. 3)

$$\widehat{\mathbf{F}}_t^{\text{gmm}} = (\widehat{\boldsymbol{\Lambda}}'\mathbf{Z}\mathbf{W}_t\mathbf{Z}'\widehat{\boldsymbol{\Lambda}})^{-1}\widehat{\boldsymbol{\Lambda}}'\mathbf{Z}\mathbf{W}_t\mathbf{Z}'\mathbf{X}_t. \tag{20}$$

The optimal weight matrix $\mathbf{W}_t$ results as

$$\mathbf{W}_t = [E(\mathbf{Z}'\mathbf{e}_t\mathbf{e}_t'\mathbf{Z})]^{-1}.$$

If $\mathbf{e}_t$ is independent of $\mathbf{Z}$ and $E(\mathbf{e}_t\mathbf{e}_t') = \boldsymbol{\Omega}$ for all $t = 1, \ldots, T$, we can invoke the law of iterated expectations yielding $\mathbf{W}_t = \mathbf{W} = [E(\mathbf{Z}'\boldsymbol{\Omega}\mathbf{Z})]^{-1}$. This suggests to estimate the weight matrix as $\widehat{\mathbf{W}} = (\mathbf{Z}'\widehat{\boldsymbol{\Omega}}\mathbf{Z})^{-1}$, where $\widehat{\boldsymbol{\Omega}} = T^{-1}\sum_{t=1}^{T}\widehat{\mathbf{e}}_t\widehat{\mathbf{e}}_t'$, and $\widehat{\mathbf{e}}_t$ denotes the residual vector from the PC–OLS or PC–GLS estimator, respectively. Unfortunately, this yields a singular weight matrix since the residual vector is orthogonal to the respective columns of the instrument matrix $\mathbf{Z}$. We therefore employ the estimator for the covariance matrix of the idiosyncratic components suggested by Forni et al. (2000, 2005). This estimator (denoted as $\widetilde{\boldsymbol{\Omega}}$) is obtained from the dynamic principal component method by integrating the estimated spectral density matrix of the vector of idiosyncratic components. Estimating the weight matrix as $\widetilde{\mathbf{W}} = (\mathbf{Z}'\widetilde{\boldsymbol{\Omega}}\mathbf{Z})^{-1}$ yields the GMM estimator

$$\begin{aligned}\widehat{\mathbf{F}}^{\text{gmm}} &= [\widehat{\mathbf{F}}_1^{\text{gmm}}, \ldots, \widehat{\mathbf{F}}_T^{\text{gmm}}]' \\ &= \mathbf{X}\mathbf{Z}\widetilde{\mathbf{W}}\mathbf{Z}'\widehat{\boldsymbol{\Lambda}}(\widehat{\boldsymbol{\Lambda}}'\mathbf{Z}\widetilde{\mathbf{W}}\mathbf{Z}'\widehat{\boldsymbol{\Lambda}})^{-1}.\end{aligned} \tag{21}$$

A similar approach can be employed to combine the moment conditions of the PC–OLS and PC–GLS estimators of the factor

[4] For example, the PC–GLS estimator of $\boldsymbol{\lambda}_i$ is less efficient than the PC–OLS estimator if we fit an AR(1) model to the idiosyncratic errors, which are in fact generated by the MA(2) model $e_{it} = \varepsilon_{it} + 0.7\varepsilon_{i,t-1} - 0.7\varepsilon_{i,t-2}$. Note that in this case the fitted AR(1) model implies a positive second-order autocorrelation, whereas the actual second-order autocorrelation of the errors is negative.

loadings. First, consider the moments of the PC–GLS estimator based on AR(1) errors:

$$\begin{aligned}\mathbf{m}^*(\boldsymbol{\lambda}_i) &= \sum_{t=2}^{T}(\widehat{\mathbf{F}}_t - \widehat{\rho}_i\widehat{\mathbf{F}}_{t-1})[x_{it} - \widehat{\rho}_i x_{i,t-1} - \boldsymbol{\lambda}_i'(\widehat{\mathbf{F}}_t - \widehat{\rho}_i\widehat{\mathbf{F}}_{t-1})] \\ &= (1+\widehat{\rho}_i^2)\sum_{t=1}^{T}\widehat{\mathbf{F}}_t(x_{it} - \boldsymbol{\lambda}_i'\widehat{\mathbf{F}}_t) \\ &\quad - \widehat{\rho}_i\sum_{t=2}^{T-1}(\widehat{\mathbf{F}}_{t+1} + \widehat{\mathbf{F}}_{t-1})(x_{it} - \boldsymbol{\lambda}_i'\widehat{\mathbf{F}}_t) + O_p(1).\end{aligned}$$

Therefore, if T is large, these moments are equivalent to a linear combination of the moments

$$\mathbf{m}_3(\boldsymbol{\lambda}_i) = \sum_{t=1}^{T}\widehat{\mathbf{F}}_t(x_{it} - \boldsymbol{\lambda}_i'\widehat{\mathbf{F}}_t), \tag{22}$$

$$\mathbf{m}_4(\boldsymbol{\lambda}_i) = \sum_{t=2}^{T-1}(\widehat{\mathbf{F}}_{t+1} + \widehat{\mathbf{F}}_{t-1})(x_{it} - \boldsymbol{\lambda}_i'\widehat{\mathbf{F}}_t). \tag{23}$$

Since $\mathbf{m}_3(\boldsymbol{\lambda}_i)$ is the moment of the PC–OLS estimator, the "hybrid estimator" is obtained by employing the vector of instruments $\boldsymbol{\xi}_t = [\widehat{\mathbf{F}}_t', \widehat{\mathbf{F}}_{t+1}' + \widehat{\mathbf{F}}_{t-1}']'$ for $t = 2, \ldots, T-1$, $\boldsymbol{\xi}_1 = [\widehat{\mathbf{F}}_1, \mathbf{0}]$, and $\boldsymbol{\xi}_T = [\widehat{\mathbf{F}}_T, \mathbf{0}]$. Furthermore, define the matrix $\boldsymbol{\Xi} = [\boldsymbol{\xi}_1, \ldots, \boldsymbol{\xi}_T]'$. The GMM estimator for $\boldsymbol{\lambda}_i$ results as

$$\widehat{\boldsymbol{\lambda}}_i^{\text{gmm}} = (\widehat{\mathbf{F}}'\boldsymbol{\Xi}\mathbf{W}_i\boldsymbol{\Xi}'\widehat{\mathbf{F}})^{-1}\widehat{\mathbf{F}}'\boldsymbol{\Xi}\mathbf{W}_i\boldsymbol{\Xi}'\mathbf{X}_i. \tag{24}$$

To estimate the weight matrix $\mathbf{W}_i$, we employ the usual heteroscedasticity and autocorrelation consistent covariance (HAC) estimator suggested by Hansen (1982). Using $\widehat{e}_{it} = x_{it} - \widehat{\boldsymbol{\lambda}}_i'\widehat{\mathbf{F}}_t$ and

$$\widehat{\boldsymbol{\Gamma}}_i(k) = \frac{1}{T}\sum_{t=k+1}^{T}\widehat{e}_{it}\widehat{e}_{i,t-k}\boldsymbol{\xi}_t\boldsymbol{\xi}_{t-k}',$$

the weight matrix is estimated as

$$\widehat{\mathbf{W}}_i = \left[\widehat{\boldsymbol{\Gamma}}_i(0) + \sum_{j=1}^{\ell}\tau_j(\widehat{\boldsymbol{\Gamma}}_i(j) + \widehat{\boldsymbol{\Gamma}}_i'(j))\right]^{-1},$$

where $\tau_j = (\ell - j + 1)/(\ell + 1)$ is the weight function and ℓ denotes the truncation lag obeying $\ell/T \to 0$.

Since the hybrid estimators for $\boldsymbol{\lambda}_i$ and $\mathbf{F}_t$ employ r additional moments, the small sample properties may deteriorate if r is large relative to T. Thus, although the GMM estimators are asymptotically more efficient than the PC–OLS or PC–GLS estimators, the hybrid estimators may perform worse in small samples, in particular, when the weight matrices are poorly estimated. Some improvement of the small sample properties may be achieved by using the continuously updated GMM estimator proposed by Hansen, Heaton, and Yaron (1996).

6. SMALL SAMPLE PROPERTIES

In order to investigate the small sample properties of the proposed estimators, we perform a Monte Carlo study. In particular, we calculate a measure of efficiency and compare the performance of five different approaches: the standard PC estimator, the two-step and iterated PC–GLS estimators as described above, the hybrid estimator introduced in the previous section, and the quasi-maximum likelihood (QML) estimator of Doz, Giannone, and Reichlin (2006). The latter authors suggest maximum likelihood estimation of the approximate dynamic factor model via the Kalman filter employing the EM algorithm. In order to make the standard estimation approach of traditional factor analysis feasible in the large approximate dynamic factor environment, Doz, Giannone, and Reichlin (2006, 2007) abstract from possible cross-sectional correlation of the idiosyncratic component. However, their estimation procedure does take into account factor dynamics as well as heteroscedasticity of the idiosyncratic errors.[5]

Furthermore, we consider two sets of simulation experiments. First, we study a setup featuring a single factor, where the parameters governing the data-generating process only exhibit a relatively loose relation to those obtained from real-life data. This allows us to get a transparent overview of the relative merits of the respective estimator in various—perfectly controlled—environments concerning autocorrelation, heteroscedasticity, as well as cross-sectional correlation. In our second experiment, on the other hand, we generate data based on the widely used dataset of Stock and Watson (2005), where we consider the case of multiple factors. This puts us in a position to study the performance of the various estimators when applied to "more realistic" datasets, that is, datasets representative for the ones typically encountered in practice.

6.1 Simulation in a Controlled Environment

The data-generating process of our first Monte Carlo experiment is the following:

$$x_{it} = \lambda_i F_t + e_{it},$$

where

$$\begin{aligned}&F_t = \gamma F_{t-1} + u_t, \qquad u_t \overset{\text{iid}}{\sim} \mathcal{N}(0, (1-\gamma^2)),\\ &e_{it} = \rho_i e_{i,t-1} + \varepsilon_{it},\\ &\boldsymbol{\varepsilon}_t \overset{\text{iid}}{\sim} \mathcal{N}(\mathbf{0}, \mathbf{R}\boldsymbol{\Gamma}\mathbf{R}), \qquad \boldsymbol{\varepsilon}_t = [\varepsilon_{1t}, \ldots, \varepsilon_{Nt}]',\\ &\boldsymbol{\Gamma} = \boldsymbol{\Sigma}\boldsymbol{\Omega}\boldsymbol{\Sigma},\\ &\mathbf{R} = \operatorname{diag}\big(\sqrt{1-\rho_1^2}, \ldots, \sqrt{1-\rho_N^2}\big),\\ &\boldsymbol{\Sigma} = \operatorname{diag}(\sigma_1, \ldots, \sigma_N),\\ &\rho_i \overset{\text{iid}}{\sim} \mathcal{U}[a, b],\\ &\lambda_i \overset{\text{iid}}{\sim} \mathcal{U}[0, 1],\end{aligned} \tag{25}$$

where $\mathcal{U}[a, b]$ denotes a random variable uniformly distributed on the interval $[a, b]$. As mentioned above, in these baseline simulations, we set the number of static and dynamic factors equal to one and, therefore, F_t is a scalar.

[5] Even though their actual implementation of the estimator does not allow for serial correlation of the idiosyncratic components, Doz, Giannone, and Reichlin (2006) point out that, in principle, it is possible to take into account this feature in the estimation approach. However, the resulting estimator is computationally demanding as it implies N additional transition equations for the idiosyncratic components.

Table 1. Efficiency: one factor, autocorrelated errors

	Loadings ($\boldsymbol{\lambda}_i$)				Factors ($\mathbf{F}_t$)			
	PC	Two-step	Iterated	QML	PC	Two-step	Iterated	QML
$T = 50$								
$N = 50$	0.287	0.525	0.622	0.267	0.735	0.730	0.848	0.640
$N = 100$	0.294	0.559	0.646	0.287	0.812	0.809	0.924	0.752
$N = 200$	0.298	0.576	0.653	0.300	0.858	0.855	0.959	0.809
$N = 300$	0.304	0.591	0.660	0.313	0.884	0.884	0.974	0.846
$T = 100$								
$N = 50$	0.487	0.756	0.774	0.438	0.837	0.833	0.871	0.761
$N = 100$	0.511	0.781	0.793	0.492	0.908	0.906	0.935	0.875
$N = 200$	0.525	0.792	0.801	0.519	0.945	0.943	0.967	0.928
$N = 300$	0.525	0.794	0.802	0.522	0.957	0.956	0.978	0.945
$T = 200$								
$N = 50$	0.685	0.875	0.876	0.645	0.872	0.870	0.881	0.830
$N = 100$	0.700	0.886	0.888	0.683	0.932	0.930	0.939	0.915
$N = 200$	0.708	0.890	0.891	0.701	0.963	0.962	0.969	0.956
$N = 300$	0.711	0.892	0.893	0.707	0.973	0.973	0.979	0.969

NOTE: Entries are the R^2 of a regression of the true factors or loadings on the corresponding estimate and a constant. *PC* is the ordinary principal component estimator, *Two-step* and *Iterated* indicate the two-step PC–GLS and iterated PC–GLS estimators, respectively, introduced in Section 3, and *QML* is the quasi-maximum likelihood estimator of Doz, Giannone, and Reichlin (2006). The following parameter values are used: $\gamma = 0.7$, $\rho_i \overset{\text{iid}}{\sim} \mathcal{U}[0.5, 0.9]$, $\sigma_i^2 = 2$ for all i.

Four different scenarios are considered. In the first two, we abstract from cross-sectional correlation, that is, $\boldsymbol{\Omega} = \mathbf{I}$, and concentrate on either autocorrelation or heteroscedasticity. In the autocorrelation case, we focus on the dynamic aspects and set $\gamma = 0.7$, $\rho_i \overset{\text{iid}}{\sim} \mathcal{U}[0.5, 0.9]$, as well as $\sigma_i^2 = 2$ for all i. In the case focusing on heteroscedasticity, we set $\gamma = 0$, $\rho_i = 0$ for all i, and $\sigma_i \overset{\text{iid}}{\sim} |\mathcal{N}(\sqrt{2}, 0.25)|$.

In the other two scenarios, we allow for nonzero cross-sectional correlation, so that $\boldsymbol{\Omega}$ is not the identity matrix. In constructing $\boldsymbol{\Omega}$ we follow Chang (2002), in order to ensure that only weak cross-sectional correlation is present. In particular, the covariance matrix is generated using the spectral decomposition with $\boldsymbol{\Omega} = \mathbf{HVH}'$, where $\mathbf{H}$ is a matrix consisting of orthonormal column vectors and $\mathbf{V}$ is a diagonal matrix. $\mathbf{H}$ is constructed as $\mathbf{H} = \mathbf{M}(\mathbf{M}'\mathbf{M})^{-1/2}$, where the elements of $\mathbf{M}$ are drawn from $\mathcal{U}[0, 1]$. To obtain $\mathbf{V}$, a set of eigenvalues $v_i, i = 2, \ldots, N-1$, is generated by drawing from $\mathcal{U}[w, 1]$, where $w > 0$. Furthermore, to control the ratio of the minimum to the maximum eigenvalue via w, $v_1 = w$ and $v_N = 1$. In line with Chang (2002), we choose $w = 0.1$ in all simulations. By construction, in the scenarios featuring cross-sectional correlation, heteroscedasticity is always present, where we also set $\sigma_i \overset{\text{iid}}{\sim} |\mathcal{N}(\sqrt{2}, 0.25)|$. As a result, we distinguish only the cases when autocorrelation is present or not. In the former scenario we set $\gamma = 0.7$ and $\rho_i \overset{\text{iid}}{\sim} \mathcal{U}[0.5, 0.9]$, whereas in the latter $\gamma = 0$ and $\rho_i = 0$ for all i.

We generate 1000 replications for different sample sizes. In particular, we set $N = 50, 100, 200, 300$ and $T = 50, 100, 200$. In order to assess how precisely we can estimate the true factors or loadings, we compute a measures of efficiency. Concerning the common factors, it is simply $R^2(\mathbf{F}, \widehat{\mathbf{F}})$, that is, the coefficient of determination of a regression of $\mathbf{F}$ (the true factor) on $\widehat{\mathbf{F}}$ (the estimator under consideration) and a constant. Obviously, an analogous measure can also be defined for the factor loadings. Employing the coefficient of determination has the advantage that our measure of efficiency is invariant to the normalization of the factors (or loadings).[6]

6.1.1 Autocorrelation and Heteroscedasticity. Table 1 reports the results for the case of autocorrelated errors when abstracting from cross-sectional correlation. Apparently, the PC and QML estimators perform poorly, in particular with respect to the factor loadings, where the R^2's are of comparable magnitude. The low accuracy can be explained by the fact that both estimators fail to take into account serial correlation of the idiosyncratic component. The QML procedure takes into account the dynamics of the common factors. However, as has been argued in Section 3, the dynamic properties of the factors are irrelevant for the asymptotic properties as $N \to \infty$. In contrast, for the factor loadings the R^2's for both the two-step and the iterated PC–GLS estimator are considerably larger than the ones for the PC and QML estimators. In particular, with larger T the two PC–GLS estimators become increasingly accurate and show a similar performance as expected from Theorem 1. This picture changes somewhat when examining the results for the factors. Using the two-step estimator basically leads to the same R^2's as using standard PC. In this respect, note that the two-step regression for the common factors is not affected by possible autocorrelation of the errors but exploits possible heteroscedasticity. Interestingly, iterating the PC–GLS estimator until convergence, on the other hand, leads to a substantial increase in accuracy.[7] This is due to the fact that the loadings are estimated more precisely by taking into account the autocorrelation of the errors. Thus, in the second step, the regressors have a smaller

[6] Doz, Giannone, and Reichlin (2006) also employ the (trace) R^2 as their performance measure.

[7] The number of iterations is limited to a maximum of five. First, this reduces the computational burden and we find no further improvement if the number of iterations is increased.

Table 2. Efficiency: one factor, heteroscedastic errors

	Loadings ($\boldsymbol{\lambda}_i$)				Factors ($\mathbf{F}_t$)			
	PC	Two-step	Iterated	QML	PC	Two-step	Iterated	QML
$T = 50$								
$N = 50$	0.569	0.559	0.618	0.629	0.833	0.917	0.929	0.932
$N = 100$	0.605	0.596	0.623	0.632	0.915	0.964	0.969	0.970
$N = 200$	0.630	0.618	0.630	0.640	0.958	0.984	0.987	0.987
$N = 300$	0.630	0.618	0.625	0.635	0.972	0.991	0.992	0.992
$T = 100$								
$N = 50$	0.714	0.710	0.770	0.774	0.849	0.929	0.935	0.939
$N = 100$	0.756	0.751	0.774	0.778	0.924	0.968	0.970	0.972
$N = 200$	0.772	0.767	0.777	0.781	0.962	0.986	0.988	0.988
$N = 300$	0.777	0.772	0.778	0.782	0.975	0.992	0.993	0.993
$T = 200$								
$N = 50$	0.821	0.820	0.871	0.872	0.857	0.931	0.934	0.938
$N = 100$	0.858	0.857	0.875	0.876	0.929	0.970	0.972	0.973
$N = 200$	0.872	0.870	0.878	0.878	0.964	0.987	0.988	0.988
$N = 300$	0.875	0.874	0.878	0.879	0.976	0.992	0.993	0.993

NOTE: Entries are the R^2 of a regression of the true factors or loadings on the corresponding estimate and a constant. The following parameter values are used: $\gamma = 0$, $\rho_i = 0$ for all i, $\sigma_i \overset{\text{iid}}{\sim} |\mathcal{N}(\sqrt{2}, 0.25)|$. For further information see Table 1.

error and this improves the efficiency of the factor estimates. This increase in accuracy is mainly noticeable for small T. For larger sample sizes, iteration still leads to more precise estimates, but in absolute terms the improvement is marginal and all estimators show a similar performance characterized by high accuracy.

The results for heteroscedastic errors and without cross-sectional correlation are presented in Table 2. Considering the results for the factor loadings, standard PC and two-step PC–GLS estimation show a similar performance, where both estimators become increasingly accurate with sample size. These findings are not surprising, since the two-step PC–GLS estimator of the factor loadings has the same asymptotic properties as the ordinary PC estimator if the errors are serially uncorrelated. A slight efficiency improvement with respect to the loadings is attainable by employing the iterated PC–GLS estimator, in particular if N is small compared to T. Analogous to the case with autocorrelated errors, the efficiency gain is due to the fact that by estimating the factors more precisely via incorporating heteroscedasticity, in the second step, the regressors have a smaller error, thus improving the accuracy of the estimated factor loadings. However, in line with Theorem 1, for larger samples the two PC–GLS estimators perform similarly. There are two reasons for the limited gain in efficiency via iteration compared to standard PC and two-step PC–GLS estimation. First, as explained in Section 3, relevant for the efficient estimation of the factor loadings is to allow for autocorrelation of the errors. Not surprisingly, as there is no serial correlation in this scenario, the accuracy of PC (and two-step PC–GLS) is relatively high. Second, and in contrast to the autocorrelation case, the reduction in the error of the regressors is not that large as indicated by the rather small absolute gain in efficiency with respect to the common factors by taking into account heteroscedasticity. Overall, the results concerning the common factors imply that heteroscedasticity of the errors does not seem to be that severe of a problem, as the R^2's of the four estimators under consideration basically all indicate high accuracy. While taking into account heteroscedasticity of the errors does indeed lead to an increase in the R^2's compared to standard PC, the difference is really noticeable only in small samples. Finally, the QML estimates of the factors as well as the factor loadings show a strong performance, even slightly better than the iterated PC–GLS estimator. This is due to the fact, that in this scenario the approximating model coincides with the true model and the QML estimator is equivalent to the exact ML estimator.

6.1.2 Introducing Cross-Sectional Correlation. Another typical feature of many datasets is nonzero cross-sectional correlation. Consequently, in the next simulations, we check whether the superior performance of the two PC–GLS estimators still holds under such a correlation structure. First, consider the case of autocorrelated idiosyncratic components as presented in Table 3.[8] The general conclusions from the autocorrelation case presented above carry over to this scenario with added cross-correlation, even though with some modifications due to the presence of heteroscedasticity. With respect to the loadings, the gain in efficiency of using the two-step and iterated PC–GLS estimators compared to standard PC is considerable, where the two PC–GLS estimators show a similar performance in large samples. The improvement when iterating the PC–GLS estimator is even more noticeable. This is due to the presence of heteroscedasticity in addition to autocorrelation. Consequently, there is not only a direct beneficial effect in terms of taking into account dynamic aspects with respect to the factor loadings, but also an indirect effect in terms of a

[8] As mentioned above, heteroscedasticity is always present by construction in this set of simulations. Moreover, due to the computational burden, we do not present results for the QML estimator for these simulations. We checked, however, whether the overall findings of the previous scenarios carry over to the cross-correlation case by running a subset of the simulations including the QML estimator. Indeed, we do not find a substantial change in results.

Table 3. Efficiency: one factor, cross-sectional correlation, autocorrelated errors

	Loadings ($\boldsymbol{\lambda}_i$)					Factors ($\mathbf{F}_t$)				
	PC	Two-step	Iterated	$\text{GMM}_{\mathbf{e}}$	$\text{GMM}_{\hat{\mathbf{e}}}$	PC	Two-step	Iterated	$\text{GMM}_{\mathbf{W}}$	$\text{GMM}_{\widehat{\mathbf{W}}}$
$T = 50$										
$N = 50$	0.419	0.649	0.751	0.700	0.507	0.828	0.893	0.955	0.894	0.829
$N = 100$	0.442	0.702	0.762	0.753	0.541	0.903	0.949	0.982	0.948	0.907
$N = 200$	0.459	0.721	0.761	0.770	0.560	0.941	0.973	0.993	0.973	0.950
$N = 300$	0.464	0.727	0.761	0.776	0.568	0.952	0.980	0.995	0.979	0.962
$T = 100$										
$N = 50$	0.611	0.816	0.862	0.835	0.740	0.890	0.944	0.960	0.947	0.889
$N = 100$	0.644	0.851	0.869	0.867	0.781	0.944	0.976	0.983	0.977	0.944
$N = 200$	0.654	0.860	0.870	0.876	0.793	0.970	0.989	0.993	0.990	0.971
$N = 300$	0.659	0.863	0.870	0.878	0.797	0.977	0.993	0.996	0.993	0.979
$T = 200$										
$N = 50$	0.767	0.901	0.927	0.906	0.877	0.912	0.957	0.962	0.960	0.909
$N = 100$	0.794	0.922	0.931	0.927	0.902	0.957	0.982	0.984	0.982	0.956
$N = 200$	0.801	0.928	0.932	0.932	0.909	0.978	0.992	0.993	0.992	0.978
$N = 300$	0.803	0.929	0.932	0.933	0.910	0.984	0.995	0.996	0.995	0.984

NOTE: Entries are the R^2 of a regression of the true factors or loadings on the corresponding estimate and a constant. $\text{GMM}_{\mathbf{e}}$ is the hybrid estimator for the factor loadings as suggested in Section 5 using the true idiosyncratic errors to compute the weighting matrix, whereas $\text{GMM}_{\hat{\mathbf{e}}}$ is the corresponding estimator using the *estimated* idiosyncratic components. $\text{GMM}_{\mathbf{W}}$ is the hybrid estimator for the common factors using the true covariance matrix of the idiosyncratic components to compute the optimal weighting matrix, whereas $\text{GMM}_{\widehat{\mathbf{W}}}$ employs the estimator for the covariance matrix suggested by Forni et al. (2000, 2005). The following parameter values are used: $v_i \overset{\text{iid}}{\sim} \mathcal{U}[0.1, 1]$, $\gamma = 0.7$, $\rho_i \overset{\text{iid}}{\sim} \mathcal{U}[0.5, 0.9]$, $\sigma_i \overset{\text{iid}}{\sim} |\mathcal{N}(\sqrt{2}, 0.25)|$. For further information see Table 1.

reduction of the error in the regressors, that is, the common factors, by taking into account heteroscedasticity. Concerning the common factors, standard PC, two-step PC–GLS, and iterated PC–GLS all show a strong performance. Still, employing the two PC–GLS estimators leads to more efficient estimates than PC, where the gain is most noticeable in small samples. Due to the presence of heteroscedasticity, the performance of two-step PC–GLS relative to standard PC is more comparable to the heteroscedasticity case presented above. Again, the benefit from iterating the PC–GLS estimator is even more noticeable than in the autocorrelation scenario abstracting from cross-sectional correlation. Analogous to the factor loadings, this is the result of the presence of both autocorrelation and heteroscedasticity. Consequently, not only the factors are estimated more precisely by taking into account heteroscedasticity, but there is also the indirect effect of reducing the error in the regressors by taking into account autocorrelation in the idiosyncratic component. This leads to R^2's very close to one.

Second, consider the scenario without autocorrelation presented in Table 4. The overall findings are very similar to the ones with heteroscedasticity but abstracting from cross-sectional correlation, so that the introduction of the latter does not really seem to affect the performance of the suggested estimators. With respect to the factor loadings, the precision of standard PC and two-step PC–GLS is again very similar due to the absence of autocorrelation of the idiosyncratic errors. Iterating the PC–GLS estimator leads to slight efficiency improvements for the reasons stated above, while for large samples the two PC–GLS estimators perform similarly, in line with Theorem 1. Moreover, the results concerning the common factors again indicate high accuracy for all estimators, that is, the R^2's tend to be very close to one, where the increase in the R^2's of using the two PC–GLS estimators is really noticeable only in small samples.

6.1.3 The Hybrid Estimator. As a final investigation within our first Monte Carlo setup, we consider the hybrid estimator suggested in Section 5. As it is constructed to circumvent the problem that in some special cases, characterized by misspecification of the variance or covariance function, standard PC is asymptotically more efficient than the PC–GLS estimator, we compare our new estimator in particular to PC–OLS and the two-step PC–GLS estimator.[9] In particular, we consider the data-generating process with nonzero cross-sectional correlation and heteroscedasticity as presented in Tables 3 and 4, where setups with and without autocorrelation are distinguished.

Moreover, while asymptotically the GMM estimator is more efficient than both the PC–OLS and PC–GLS estimators, this does not need to hold in finite samples. In particular, it turns out that having a reliable estimate of the optimal weighting matrix greatly affects the performance of the hybrid estimator. Consequently, we present results for two versions of this estimator. First, a variant which actually estimates the optimal weighting matrix as suggested in Section 5. In the second variant, we employ improved estimates of this matrix. For the estimator of the common factors, it is computed using the true covariance matrix of the idiosyncratic components resulting from our simulation setup. Concerning the estimator of the factor loadings, we replace the estimated idiosyncratic components by the true errors generated in each simulation run to obtain a better estimate of the corresponding weighting matrix. The second variant, while

[9] It is not possible to actually construct such a pathological case in our standard simulation setup. But to be consistent with our results discussed above, we stay within this setting. The simulation experiment nevertheless allows to draw comparisons between the different estimators and gives a comprehensive picture of the hybrid estimator's performance in finite samples.

Table 4. Efficiency: one factor, cross-sectional correlation, no autocorrelation in the errors

	Loadings ($\boldsymbol{\lambda}_i$)					Factors ($\mathbf{F}_t$)				
	PC	Two-step	Iterated	$\text{GMM}_{\mathbf{e}}$	$\text{GMM}_{\hat{\mathbf{e}}}$	PC	Two-step	Iterated	$\text{GMM}_{\mathbf{W}}$	$\text{GMM}_{\widehat{\mathbf{W}}}$
$T = 50$										
$N = 50$	0.726	0.719	0.749	0.772	0.712	0.910	0.956	0.960	0.958	0.909
$N = 100$	0.748	0.740	0.754	0.795	0.734	0.956	0.982	0.984	0.983	0.956
$N = 200$	0.757	0.748	0.753	0.802	0.740	0.979	0.992	0.993	0.992	0.978
$N = 300$	0.759	0.750	0.753	0.804	0.742	0.986	0.995	0.996	0.995	0.985
$T = 100$										
$N = 50$	0.836	0.833	0.859	0.854	0.828	0.917	0.960	0.962	0.962	0.917
$N = 100$	0.851	0.848	0.860	0.869	0.843	0.959	0.983	0.983	0.983	0.958
$N = 200$	0.863	0.860	0.864	0.879	0.855	0.980	0.993	0.993	0.993	0.980
$N = 300$	0.865	0.862	0.864	0.881	0.856	0.987	0.996	0.996	0.996	0.987
$T = 200$										
$N = 50$	0.903	0.902	0.924	0.910	0.900	0.919	0.961	0.962	0.963	0.919
$N = 100$	0.922	0.921	0.928	0.928	0.919	0.961	0.984	0.985	0.984	0.961
$N = 200$	0.926	0.926	0.929	0.931	0.923	0.981	0.993	0.993	0.993	0.981
$N = 300$	0.928	0.927	0.929	0.933	0.925	0.987	0.996	0.996	0.996	0.987

NOTE: Entries are the R^2 of a regression of the true factors or loadings on the corresponding estimate and a constant. $\text{GMM}_{\mathbf{e}}$ is the hybrid estimator for the factor loadings as suggested in Section 5 using the true idiosyncratic errors to compute the weighting matrix, whereas $\text{GMM}_{\hat{\mathbf{e}}}$ is the corresponding estimator using the *estimated* idiosyncratic components. $\text{GMM}_{\mathbf{W}}$ is the hybrid estimator for the common factors using the true covariance matrix of the idiosyncratic components to compute the optimal weighting matrix, whereas $\text{GMM}_{\widehat{\mathbf{W}}}$ employs the estimator for the covariance matrix suggested by Forni et al. (2000, 2005). The following parameter values are used: $v_i \overset{\text{iid}}{\sim} \mathcal{U}[0.1, 1]$, $\gamma = 0$, $\rho_i = 0$ for all i, $\sigma_i \overset{\text{iid}}{\sim} |\mathcal{N}(\sqrt{2}, 0.25)|$. For further information see Table 1.

not feasible in practice, clearly illustrates the effect of using a (possibly poor) estimator of the optimal weighting matrix.

Consider first the case of autocorrelated errors as presented in Table 3. The hybrid estimator with an improved estimate of the optimal weighting matrix already delivers the results in finite samples, which we expect the standard version to attain asymptotically. It is at least as efficient as both the standard PC and two-step PC–GLS estimator for the common factors and factor loadings. In most cases, in particular for the loadings, this variant of the hybrid estimator yields considerably more efficient estimates than both reference estimators. The gain in efficiency, however, vanishes in larger samples. The hybrid estimator employing an estimated weight matrix, on the other hand, does not yield as strong a performance. With respect to the factor loadings, the corresponding R^2's lie in between those of the standard PC and two-step PC–GLS estimators. The loss in efficiency in finite samples compared to the (more efficient) two-step estimator, however, becomes less severe as N and T get larger. This does not result for the common factors, where the efficiency of the hybrid estimator with estimated weight matrix is similar to that of the less efficient PC–OLS estimator for all sample sizes considered.

Analogous findings are obtained in the setup without autocorrelation in the idiosyncratic components (Table 4). The only difference is that the hybrid estimator for the factor loadings with estimated weight matrix performs about as well as the two reference estimators, where only a slight loss in efficiency can be observed in small samples. This is, of course, due to the absence of autocorrelation in the errors so that there is basically no difference in the performance of the standard PC and two-step PC–GLS estimators. The hybrid estimator with an improved estimate of the optimal weighting matrix is again at least as efficient as the two reference estimators, with a particularly strong performance for the factor loadings. The version with estimated weight matrix, moreover, estimates the common factors about as well as the less efficient standard PC estimator. In sum, these results clearly illustrate the important role played by the estimation of the optimal weighting matrix.

6.2 Simulation Based on Stock and Watson's (2005) Dataset

In our second Monte Carlo experiment, we study the performance of the different estimators when applied to more realistic datasets. The starting point is the well-known set of time series provided by Stock and Watson (2005). Since we do not want to impose a specific ARMA-structure on the simulated common factors or idiosyncratic components, which would constitute an advantage for the ARMA-based PC–GLS estimators, we employ the circular block bootstrap of Politis and Romano (1992). In particular, motivated by its superior performance in the first simulation experiment, we apply the iterated PC–GLS estimator to the aforementioned dataset. In line with Stock and Watson's (2005) findings, we set the number of factors equal to seven. Thus, we obtain estimates for the common factors, $\mathbf{F}$, the factor loadings, $\boldsymbol{\Lambda}$, and thus for the common component, $\mathbf{F}\boldsymbol{\Lambda}'$, as well as for the idiosyncratic errors, $\mathbf{e}$. In each simulation run, we resample overlapping blocks of a given length from the estimated factors and idiosyncratic errors to obtain a new set of factors and idiosyncratic components. Those series are then combined with the estimated factor loadings to obtain an individual sample for our simulation experiment. The block length differs between the factors and idiosyncratic errors, but is the same over all individual factors and errors, respectively. It is chosen optimally, based on Politis and White (2004) and including the correction of Patton, Politis, and White (2009).[10] In

[10] We use the MATLAB-implementation provided by Andrew Patton on his web page (*http://econ.duke.edu/~ap172/code.html*). As it only delivers the op-

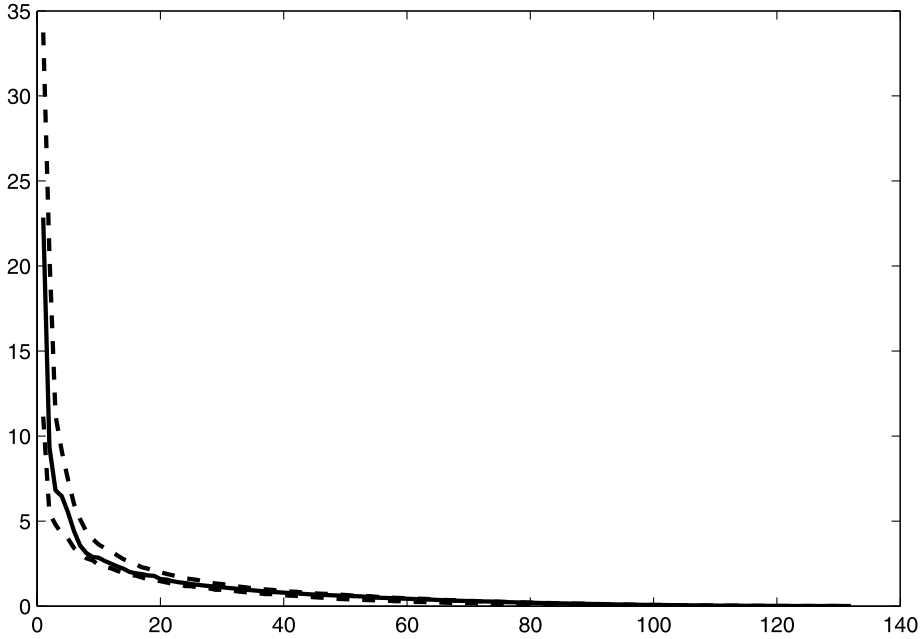

Figure 3. Scree plots. *Note*: This figure shows the scree plot of Stock and Watson's (2005) original dataset (solid line) as well as the respective smallest and largest eigenvalue over 1000 simulation runs from the simulated series (dashed lines) ($N = 132$, $T = 526$).

order to preserve the structure of the cross-sectional correlation, we choose the same permutation for all individual series, that is, factors and errors, respectively, within one simulation run. The scree plots presented in Figure 3 suggest that our simulated series are in fact representative for the real-life dataset of Stock and Watson (2005). This figure shows the scree plot of Stock and Watson's (2005) original dataset as well as the respective smallest and largest eigenvalue over 1000 simulation runs from the simulated series. As can be seen from the graph, while there is of course some variation over the different simulations, the scree plots are close so that the basic dependence structure is preserved in the simulations.

Again, we generate 1000 replications for different sample sizes. Since this set of simulations is based on Stock and Watson's (2005) dataset, we use their number of variables in all simulations, that is, $N = 132$. We consider different time series dimensions, however, and set $T = 100, 200, 400, 526, 800$, where $T = 526$ is the number of time periods in Stock and Watson's (2005) dataset. Varying the times-series dimension is achieved by resampling the respective T observations *with replacement* from the estimated factors and idiosyncratic components. Such a bootstrap resampling scheme allows us, in particular, to obtain the necessary time series with a dimension that is larger than that of Stock and Watson's (2005) original dataset. To assess the performance of the different estimators, the same measure of efficiency as introduced above is used. It has to be generalized, however, to make it applicable to the multifactor case. In particular, now the *trace* R^2 of a regression of the true factors or loadings on the respective estimated factors or loadings and a constant is used.[11]

The results of this final Monte Carlo experiment are presented in Table 5. With respect to the factor loadings, an efficiency gain compared to standard PC is observed for both the two-step and iterated PC–GLS estimators. While the increase in accuracy of the two-step estimator is not that large, it is more pronounced for the iterated version. The latter again stems from the reduction in the error of the regressors, which is supported by the particular relation between N and T observed in this simulation. An increasing T relative to N positively affects the accurate estimation of the common factors and adversely affects the precision of the estimated loadings for the PC–OLS and two-step PC–GLS estimators. Consequently, this leaves more

Table 5. Efficiency using Stock and Watson's (2005) dataset: circular block bootstrap

	Loadings ($\boldsymbol{\lambda}_i$)			Factors ($\mathbf{F}_t$)		
	PC	Two-step	Iterated	PC	Two-step	Iterated
$N = 132$						
$T = 100$	0.579	0.602	0.679	0.818	0.856	0.884
$T = 200$	0.683	0.706	0.770	0.829	0.870	0.892
$T = 400$	0.758	0.777	0.826	0.836	0.878	0.897
$T = 526$	0.782	0.798	0.837	0.839	0.881	0.896
$T = 800$	0.810	0.823	0.857	0.843	0.884	0.897

NOTE: Entries are the trace R^2 of a regression of the true factors or loadings on the corresponding estimate and a constant. Simulation based on circular block bootstrap of factors and idiosyncratic component as estimated from Stock and Watson's (2005) dataset, where $r = 7$. For further information see Table 1.

timal block length of the individual series, we use the mean of the individual block lengths as our optimal value. As a robustness check, we also used larger and smaller block lengths than suggested by the procedure. This does not, however, change our results.

[11] We do not present results for the QML estimator in the second simulation experiment. A setup with seven factors increases the computational burden considerably, making it infeasible in practice to compute the required 1000 estimations in a reasonable amount of time.

room for improvement with respect to the factor loadings and, furthermore, the regressors for the additional steps of the iterated estimator are estimated quite precisely, which additionally boosts the performance of that estimator.

Similar findings arise for the estimated common factors, where the efficiency of standard PC is larger than for the factor loadings, however. Nevertheless, employing two-step PC–GLS and, in particular, the iterated PC–GLS estimator leads to more precise estimates compared to PC–OLS. The gain in efficiency when using two-step PC–GLS compared to standard PC is already quite noticeable, whereas the subsequent increase when using iterated PC–GLS is not as large as for the factor loadings. This also stems from the particular relation between N and T present in this simulation. As noted above, the increasing T relative to N positively affects the two-step estimates of the common factors, so that there is less room for improvement via iterating the PC–GLS estimator. Furthermore, since the relation between N and T adversely affects the accuracy of the estimated factor loadings, the reduction in the error of the regressors in the subsequent iteration steps is not that large. Overall, this table clearly illustrates the advantages of employing the two-step PC–GLS estimator or its iterated version, when confronted with a real-life dataset.

7. CONCLUSION

In this article we propose a GLS-type estimation procedure that allows for heteroscedastic and autocorrelated errors. Since the estimation of the covariance parameters does not affect the limiting distribution of the estimators, the feasible two-step PC–GLS estimator is asymptotically as efficient as the infeasible GLS-estimator (assuming that the covariance parameters are known) and the iterated version that solves the first-order conditions of the (approximate) ML estimator. Furthermore, we show that the PC–GLS estimator is generally more efficient than the PC–OLS estimator provided the variance and covariance functions are correctly specified. We also propose a GMM estimator that combines the moments of the PC–OLS and PC–GLS estimators to yield an estimator that is asymptotically at least as efficient as each of the two estimators if the second moments are misspecified. Notwithstanding these asymptotic results, our Monte Carlo experiments suggest that in small samples the hybrid estimator suffers from the poor properties of the estimated weight matrix. We therefore recommend the (iterated) PC–GLS estimator for datasets of moderate sample size.

If one is willing to accept the framework of a strict factor model (that is a model with cross-sectionally uncorrelated factors and idiosyncratic errors), then our approach can also be employed for inference. For example, recent work by Breitung and Eickmeier (2011) shows that a Chow-type test for structural breaks can be derived using the iterated PC–GLS estimator. Other possible applications are LR tests for the number of common factors or tests of hypotheses on the factor space.

APPENDIX

The following lemma plays a central role in the proof of the following theorems:

Lemma A.1. It holds for all $k \le p_i$ that

$$\text{(i)} \quad T^{-1}\sum_{t=p_i+1}^{T}(\widehat{\mathbf{F}}_t - \mathbf{H}'\mathbf{F}_t)\mathbf{F}'_{t-k} = O_p(\delta_{NT}^{-2}),$$

$$T^{-1}\sum_{t=p_i+1}^{T}(\widehat{\mathbf{F}}_t - \mathbf{H}'\mathbf{F}_t)\widehat{\mathbf{F}}'_{t-k} = O_p(\delta_{NT}^{-2}),$$

$$\text{(ii)} \quad T^{-1}\sum_{t=p_i+1}^{T}\widehat{\mathbf{F}}_t\widehat{\mathbf{F}}'_{t-k} = T^{-1}\sum_{t=p_i+1}^{T}\mathbf{H}'\mathbf{F}_t\mathbf{F}'_{t-k}\mathbf{H} + O_p(\delta_{NT}^{-2}),$$

$$\text{(iii)} \quad T^{-1}\sum_{t=p_i+1}^{T}(\widehat{\mathbf{F}}_t - \mathbf{H}'\mathbf{F}_t)e_{i,t-k} = O_p(\delta_{NT}^{-2}),$$

$$\text{(iv)} \quad N^{-1}\sum_{i=1}^{N}\frac{1}{\omega_i^2}(\widehat{\boldsymbol{\lambda}}_i - \mathbf{H}^{-1}\boldsymbol{\lambda}_i)\boldsymbol{\lambda}_i' = O_p(\delta_{NT}^{-2}),$$

$$N^{-1}\sum_{i=1}^{N}\frac{1}{\omega_i^2}(\widehat{\boldsymbol{\lambda}}_i - \mathbf{H}^{-1}\boldsymbol{\lambda}_i)\widehat{\boldsymbol{\lambda}}_i' = O_p(\delta_{NT}^{-2}),$$

$$\text{(v)} \quad N^{-1}\sum_{i=1}^{N}\frac{1}{\omega_i^2}(\widehat{\boldsymbol{\lambda}}_i - \mathbf{H}^{-1}\boldsymbol{\lambda}_i)e_{it} = O_p(\delta_{NT}^{-2}).$$

Proof. (i) The proof follows closely the proof for $k = 0$ provided by Bai (2003, lemmas B.2 and B.3). We therefore present only the main steps.

We start from the representation

$$\widehat{\mathbf{F}}_t - \mathbf{H}'\mathbf{F}_t = \frac{1}{NT}\mathbf{V}_{NT}^{-1}(\widehat{\mathbf{F}}'\mathbf{F}\boldsymbol{\Lambda}'\mathbf{e}_t + \widehat{\mathbf{F}}'\mathbf{e}\boldsymbol{\Lambda}\mathbf{F}_t + \widehat{\mathbf{F}}'\mathbf{e}\mathbf{e}_t),$$

where $\mathbf{e}_t = [e_{1t}, \ldots, e_{Nt}]'$, $\mathbf{e} = [e_1, \ldots, e_T]'$, and $\mathbf{V}_{NT}$ is an $r \times r$ diagonal matrix of the r largest eigenvalues of $(NT)^{-1}\mathbf{X}\mathbf{X}'$ (Bai 2003, theorem 1). Consider

$$\begin{aligned}
&\frac{1}{T}\sum_{t=p_i+1}^{T}(\widehat{\mathbf{F}}_t - \mathbf{H}'\mathbf{F}_t)\mathbf{F}'_{t-k} \\
&\quad= \frac{1}{NT^2}\mathbf{V}_{NT}^{-1}\Bigg(\widehat{\mathbf{F}}'\mathbf{F}\boldsymbol{\Lambda}'\sum_{t=p_i+1}^{T}\mathbf{e}_t\mathbf{F}'_{t-k} \\
&\qquad + \widehat{\mathbf{F}}'\mathbf{e}\boldsymbol{\Lambda}\sum_{t=p_i+1}^{T}\mathbf{F}_t\mathbf{F}'_{t-k} + \widehat{\mathbf{F}}'\mathbf{e}\sum_{t=p_i+1}^{T}\mathbf{e}_t\mathbf{F}'_{t-k}\Bigg) \\
&\quad= I + II + III.
\end{aligned}$$

From Assumption 1(v) it follows that

$$\boldsymbol{\Lambda}'\sum_{t=p_i+1}^{T}\mathbf{e}_t\mathbf{F}'_{t-k} = \sum_{i=1}^{N}\sum_{t=p_i+1}^{T}e_{it}\boldsymbol{\lambda}_i\mathbf{F}'_{t-k} = O_p(\sqrt{NT}),$$

and using lemma B.2 of Bai (2003) it follows that $T^{-1}\widehat{\mathbf{F}}'\mathbf{F} = T^{-1}\mathbf{H}' \times \mathbf{F}'\mathbf{F} + T^{-1}(\widehat{\mathbf{F}} - \mathbf{F}\mathbf{H})'\mathbf{F} = T^{-1}\mathbf{H}'\mathbf{F}'\mathbf{F} + O_p(\delta_{NT}^{-2})$. Thus, we obtain

$$I = \mathbf{V}_{NT}^{-1}(T^{-1}\widehat{\mathbf{F}}'\mathbf{F})\left(\frac{1}{\sqrt{NT}}\boldsymbol{\Lambda}'\sum_{t=p_i+1}^{T}\mathbf{e}_t\mathbf{F}'_{t-k}\right)\frac{1}{\sqrt{NT}} = O_p\left(\frac{1}{\sqrt{NT}}\right).$$

Next, we consider

$$\boldsymbol{\Lambda}'\mathbf{e}'\widehat{\mathbf{F}} = \boldsymbol{\Lambda}'\sum_{t=1}^{T}\mathbf{e}_t\mathbf{F}_t'\mathbf{H} + \boldsymbol{\Lambda}'\sum_{t=1}^{T}\mathbf{e}_t(\widehat{\mathbf{F}}_t - \mathbf{H}'\mathbf{F}_t)'.$$

Following Bai (2003, p. 160), we have

$$\frac{1}{NT}\boldsymbol{\Lambda}'\sum_{t=1}^{T}\mathbf{e}_t\mathbf{F}_t'\mathbf{H} = O_p\left(\frac{1}{\sqrt{NT}}\right),$$

$$\frac{1}{NT}\boldsymbol{\Lambda}'\sum_{t=1}^{T}\mathbf{e}_t(\widehat{\mathbf{F}}_t - \mathbf{H}'\mathbf{F}_t)' = O_p\left(\frac{1}{\delta_{NT}\sqrt{N}}\right).$$

Using $T^{-1}\sum_{t=p_i+1}^{T}\mathbf{F}_t'\mathbf{F}_{t-k} = O_p(1)$, we obtain

$$II = \mathbf{V}_{NT}^{-1}\left(\frac{1}{NT}\widehat{\mathbf{F}}'\mathbf{e}\boldsymbol{\Lambda}\right)\left(\frac{1}{T}\sum_{t=p_i+1}^{T}\mathbf{F}_t\mathbf{F}_{t-k}'\right)$$

$$= \left[O_p\left(\frac{1}{\sqrt{NT}}\right) + O_p\left(\frac{1}{\delta_{NT}\sqrt{N}}\right)\right]O_p(1).$$

For the remaining term, we obtain

$$\frac{1}{NT^2}\widehat{\mathbf{F}}'\mathbf{e}\sum_{t=p_i+1}^{T}\mathbf{e}_t\mathbf{F}_{t-k}' = \frac{1}{NT^2}\sum_{s=1}^{T}\sum_{t=p_i+1}^{T}\mathbf{e}_s'\mathbf{e}_t\widehat{\mathbf{F}}_s\mathbf{F}_{t-k}'$$

$$= \frac{1}{T^2}\sum_{s=1}^{T}\sum_{t=p_i+1}^{T}\widehat{\mathbf{F}}_s\mathbf{F}_{t-k}'\zeta_{NT}(s,t) + \frac{1}{T^2}\sum_{s=1}^{T}\sum_{t=p_i+1}^{T}\widehat{\mathbf{F}}_s\mathbf{F}_{t-k}'\gamma_N(s,t),$$

where

$$\zeta_{NT}(s,t) = \mathbf{e}_s'\mathbf{e}_t/N - \gamma_N(s,t),$$

$$\gamma_N(s,t) = E(\mathbf{e}_s'\mathbf{e}_t/N).$$

As in Bai (2003, p. 164f), we obtain

$$III = \mathbf{V}_{NT}^{-1}\left[O_p\left(\frac{1}{\delta_{NT}\sqrt{T}}\right) + O_p\left(\frac{1}{\delta_{NT}\sqrt{N}}\right)\right].$$

Collecting these results, we obtain

$$I + II + III = O_p\left(\frac{1}{\sqrt{NT}}\right) + O_p\left(\frac{1}{\sqrt{T}\delta_{NT}}\right) + O_p\left(\frac{1}{\sqrt{N}\delta_{NT}}\right)$$

$$= O_p\left(\frac{1}{\delta_{NT}^2}\right).$$

The proof of the second result in (i) is a similar modification of lemma A.1 in Bai (2003) and is therefore omitted.

(ii) Consider

$$T^{-1}\sum_{t=p_i+1}^{T}\widehat{\mathbf{F}}_t\widehat{\mathbf{F}}_{t-k}'$$

$$= T^{-1}\sum_{t=p_i+1}^{T}[\mathbf{H}'\mathbf{F}_t + (\widehat{\mathbf{F}}_t - \mathbf{H}'\mathbf{F}_t)][\mathbf{H}'\mathbf{F}_{t-k} + (\widehat{\mathbf{F}}_{t-k} - \mathbf{H}'\mathbf{F}_{t-k})]'$$

$$= T^{-1}\Bigg(\sum_{t=p_i+1}^{T}\mathbf{H}'\mathbf{F}_t\mathbf{F}_{t-k}'\mathbf{H} + (\widehat{\mathbf{F}}_t - \mathbf{H}'\mathbf{F}_t)\mathbf{F}_{t-k}'\mathbf{H} + \mathbf{H}'\mathbf{F}_t(\widehat{\mathbf{F}}_{t-k}' - \mathbf{F}_{t-k}'\mathbf{H}) + (\widehat{\mathbf{F}}_t - \mathbf{H}'\mathbf{F}_t)(\widehat{\mathbf{F}}_{t-k}' - \mathbf{F}_{t-k}'\mathbf{H})\Bigg)$$

$$= T^{-1}\Bigg(\sum_{t=p_i+1}^{T}\mathbf{H}'\mathbf{F}_t\mathbf{F}_{t-k}'\mathbf{H} + \underbrace{\mathbf{H}'\mathbf{F}_t(\widehat{\mathbf{F}}_{t-k}' - \mathbf{F}_{t-k}'\mathbf{H})}_{Ta} + \underbrace{(\widehat{\mathbf{F}}_t - \mathbf{H}'\mathbf{F}_t)\widehat{\mathbf{F}}_{t-k}'}_{Tb}\Bigg)$$

$$= \left(T^{-1}\sum_{t=p_i+1}^{T}\mathbf{H}'\mathbf{F}_t\mathbf{F}_{t-k}'\mathbf{H}\right) + a + b.$$

Using (i) the terms a and b can be shown to be $O_p(\delta_{NT}^{-2})$.

(iii) The proof of $k = 0$ is given in Bai (2003, lemma B.1). It is not difficult to see that the result remains unchanged if $k \neq 0$.

(iv) Following Bai (2003, p. 165), we have

$$\widehat{\boldsymbol{\lambda}}_i - \mathbf{H}^{-1}\boldsymbol{\lambda}_i = T^{-1}\mathbf{H}'\mathbf{F}'\mathbf{e}_i + T^{-1}\widehat{\mathbf{F}}'(\mathbf{F} - \widehat{\mathbf{F}}\mathbf{H}^{-1})\boldsymbol{\lambda}_i + T^{-1}(\widehat{\mathbf{F}} - \mathbf{F}\mathbf{H})'\mathbf{e}_i, \quad \text{(A.1)}$$

where $\mathbf{e}_i = [e_{i1}, \ldots, e_{iT}]'$. Postmultiplying by $\omega_i^{-2}\boldsymbol{\lambda}_i'$ and averaging yields

$$N^{-1}\sum_{i=1}^{N}\frac{1}{\omega_i^2}(\widehat{\boldsymbol{\lambda}}_i - \mathbf{H}^{-1}\boldsymbol{\lambda}_i)\boldsymbol{\lambda}_i'$$

$$= T^{-1}\mathbf{H}'\mathbf{F}'\left(N^{-1}\sum_{i=1}^{N}\frac{1}{\omega_i^2}\mathbf{e}_i\boldsymbol{\lambda}_i'\right) + T^{-1}\widehat{\mathbf{F}}'(\mathbf{F} - \widehat{\mathbf{F}}\mathbf{H}^{-1})\left(N^{-1}\sum_{i=1}^{N}\frac{1}{\omega_i^2}\boldsymbol{\lambda}_i\boldsymbol{\lambda}_i'\right) + T^{-1}(\widehat{\mathbf{F}} - \mathbf{F}\mathbf{H})'\left(N^{-1}\sum_{i=1}^{N}\frac{1}{\omega_i^2}\mathbf{e}_i\boldsymbol{\lambda}_i'\right).$$

From Bai (2003, p. 165) it follows that the last two terms are $O_p(\delta_{NT}^{-2})$. From Assumptions 1(v) and 2(i) it follows that

$$\left\|\sum_{t=1}^{T}\frac{1}{\omega_i^2}\mathbf{H}'\mathbf{F}_t\boldsymbol{\lambda}_i'\mathbf{e}_t\right\| \leq \frac{1}{\omega_{\min}^2}\left\|\sum_{t=1}^{T}\mathbf{H}'\mathbf{F}_t\boldsymbol{\lambda}_i'e_{it}\right\| = O_p(1/\sqrt{T}),$$

where $\omega_{\min} = \min(\omega_1, \ldots, \omega_N)$. Thus, the first part of (iv) is $O_p(\delta_{NT}^{-2})$. The second equation can be shown by using the first part and Lemma A.1(v).

(v) From (A.1) it follows that

$$N^{-1}\sum_{i=1}^{N}(\widehat{\boldsymbol{\lambda}}_i - \mathbf{H}^{-1}\boldsymbol{\lambda}_i)e_{it}$$

$$= N^{-1}T^{-1}\sum_{s=1}^{T}\sum_{i=1}^{N}\widehat{\mathbf{F}}_s e_{is}e_{it} + N^{-1}T^{-1}\sum_{s=1}^{T}\sum_{i=1}^{N}\widehat{\mathbf{F}}_s(\mathbf{F}_s - \mathbf{H}'^{-1}\widehat{\mathbf{F}}_s)'\boldsymbol{\lambda}_i e_{it}$$

$$= a + b.$$

For expression a we write

$$N^{-1}T^{-1}\sum_{s=1}^{T}\widehat{\mathbf{F}}_s\sum_{i=1}^{N}e_{is}e_{it} = T^{-1}\sum_{s=1}^{T}\widehat{\mathbf{F}}_s\left[N^{-1}\sum_{i=1}^{N}e_{is}e_{it} - E(e_{is}e_{it})\right] + T^{-1}\sum_{s=1}^{T}\widehat{\mathbf{F}}_s\gamma_N(s,t).$$

From lemma A.2(a) and (b) of Bai (2003), it follows that the first term on the rhs is $O_p(N^{-1/2}\delta_{NT}^{-1})$, whereas the second term is $O_p(T^{-1/2}\delta_{NT}^{-1})$.

To analyze b we note that by Lemma A.1(i) and Assumption 1(v)

$$\left[T^{-1}\sum_{s=1}^{T}\widehat{\mathbf{F}}_s(\mathbf{F}_s-\mathbf{H}'^{-1}\widehat{\mathbf{F}}_s)'\right]\left[N^{-1}\sum_{i=1}^{N}\boldsymbol{\lambda}_i e_{it}\right] = O_p(\delta_{NT}^{-2})O_p\big(N^{-1/2}\big).$$

Collecting these results, it follows that

$$\begin{aligned}&\left\|N^{-1}\sum_{i=1}^{N}\frac{1}{\omega_i^2}(\widehat{\boldsymbol{\lambda}}_i-\mathbf{H}^{-1}\boldsymbol{\lambda}_i)e_{it}\right\|\\&\quad\le\frac{1}{\omega_{\min}^2}\left\|N^{-1}\sum_{i=1}^{N}(\widehat{\boldsymbol{\lambda}}_i-\mathbf{H}^{-1}\boldsymbol{\lambda}_i)e_{it}\right\|\\&\quad=O_p\big(T^{-1/2}\delta_{NT}^{-1}\big)+O_p\big(N^{-1/2}\delta_{NT}^{-1}\big)+O_p\big(N^{-1/2}\delta_{NT}^{-2}\big)\\&\quad=O_p(\delta_{NT}^{-2}).\end{aligned}$$

Proof of Lemma 1

Let

$$\mathbf{z}_t=\begin{bmatrix}e_{it}\\\vdots\\e_{i,t-p_i+1}\end{bmatrix}\quad\text{and}\quad\widehat{\mathbf{z}}_t=\begin{bmatrix}x_{it}-\widehat{\boldsymbol{\lambda}}_i'\widehat{\mathbf{F}}_t\\\vdots\\x_{i,t-p_i+1}-\widehat{\boldsymbol{\lambda}}_i'\widehat{\mathbf{F}}_{t-p_i+1}\end{bmatrix}.$$

Using the same arguments as in lemma 4 of Bai and Ng (2002), it can be shown that

$$T^{-1}\sum_{t=p_i+1}^{T}\widehat{e}_{it}\widehat{\mathbf{z}}_{t-1}-T^{-1}\sum_{t=p_i+1}^{T}e_{it}\mathbf{z}_{t-1}=O_p(\delta_{NT}^{-2})$$

and $T^{-1}\sum_{t=p_i+1}^{T}(\widehat{\mathbf{z}}_{t-1}\widehat{\mathbf{z}}_{t-1}'-\mathbf{z}_{t-1}\mathbf{z}_{t-1}')=O_p(\delta_{NT}^{-2})$. Therefore, we obtain for the least-squares estimator of $\boldsymbol{\rho}^{(i)}$

$$\begin{aligned}\widehat{\boldsymbol{\rho}}^{(i)}&=\boldsymbol{\rho}^{(i)}+\left(\sum_{t=p_i+1}^{T}\mathbf{z}_{t-1}\mathbf{z}_{t-1}'\right)^{-1}\sum_{t=p_i+1}^{T}\mathbf{z}_{t-1}\varepsilon_{it}+O_p(\delta_{NT}^{-2})\\&=\boldsymbol{\rho}^{(i)}+O_p\big(T^{-1/2}\big)+O_p(\delta_{NT}^{-2})\end{aligned}$$

and, similarly, for the least-square estimator of ω_i^2:

$$\begin{aligned}\widehat{\omega}_i^2&=\omega_i^2+\left(T^{-1}\sum_{t=p_i+1}^{T}e_{it}^2-\omega_i^2\right)+\left(T^{-1}\sum_{t=p_i+1}^{T}(\widehat{e}_{it}^2-e_{it}^2)\right)\\&=\omega_i^2+O_p\big(T^{-1/2}\big)+O_p(\delta_{NT}^{-2}).\end{aligned}$$

Proof of Theorem 1

The autoregressive structure of the idiosyncratic component can be represented in matrix format by defining the $(T-p_i)\times T$ matrix

$$\mathbf{R}\big(\boldsymbol{\rho}^{(i)}\big)=\begin{bmatrix}-\rho_{p_i,i}&-\rho_{p_i-1,i}&-\rho_{p_i-2,i}&\cdots&1&0&0&\cdots\\0&-\rho_{p_i,i}&-\rho_{p_i-1,i}&\cdots&-\rho_{1,i}&1&0&\cdots\\\vdots&\ddots&\ddots&&\ddots&\ddots&\ddots&\ddots\end{bmatrix}.$$

Thus, the autoregressive representation (3) is written in matrix form as

$$\mathbf{R}\big(\boldsymbol{\rho}^{(i)}\big)\mathbf{e}_i=\boldsymbol{\varepsilon}_i,$$

where $\boldsymbol{\varepsilon}_i=[\varepsilon_{i,p_i+1},\ldots,\varepsilon_{iT}]'$ and $\mathbf{e}_i=[e_{i1},\ldots,e_{iT}]'$.

The feasible two-step estimator of $\boldsymbol{\lambda}_i$ is obtained as

$$\begin{aligned}\widetilde{\boldsymbol{\lambda}}_{i,\widehat{\rho}}&=\big[\widehat{\mathbf{F}}'\mathbf{R}\big(\widehat{\boldsymbol{\rho}}^{(i)}\big)'\mathbf{R}\big(\widehat{\boldsymbol{\rho}}^{(i)}\big)\widehat{\mathbf{F}}\big]^{-1}\widehat{\mathbf{F}}'\mathbf{R}\big(\widehat{\boldsymbol{\rho}}^{(i)}\big)'\mathbf{R}\big(\widehat{\boldsymbol{\rho}}^{(i)}\big)\mathbf{X}_i\\&=\big[\widehat{\mathbf{F}}'\mathbf{R}\big(\widehat{\boldsymbol{\rho}}^{(i)}\big)'\mathbf{R}\big(\widehat{\boldsymbol{\rho}}^{(i)}\big)\widehat{\mathbf{F}}\big]^{-1}\widehat{\mathbf{F}}'\mathbf{R}\big(\widehat{\boldsymbol{\rho}}^{(i)}\big)'\mathbf{R}\big(\widehat{\boldsymbol{\rho}}^{(i)}\big)\\&\quad\times(\mathbf{F}\boldsymbol{\lambda}_i+\mathbf{e}_i)\\&=\big[\widehat{\mathbf{F}}'\mathbf{R}\big(\widehat{\boldsymbol{\rho}}^{(i)}\big)'\mathbf{R}\big(\widehat{\boldsymbol{\rho}}^{(i)}\big)\widehat{\mathbf{F}}\big]^{-1}\widehat{\mathbf{F}}'\mathbf{R}\big(\widehat{\boldsymbol{\rho}}^{(i)}\big)'\mathbf{R}\big(\widehat{\boldsymbol{\rho}}^{(i)}\big)\\&\quad\times\big\{[\widehat{\mathbf{F}}+(\mathbf{F}\mathbf{H}-\widehat{\mathbf{F}})]\mathbf{H}^{-1}\boldsymbol{\lambda}_i+\mathbf{e}_i\big\},\\\widetilde{\boldsymbol{\lambda}}_{i,\widehat{\rho}}-\mathbf{H}^{-1}\boldsymbol{\lambda}_i&=\big[\widehat{\mathbf{F}}'\mathbf{R}\big(\widehat{\boldsymbol{\rho}}^{(i)}\big)'\mathbf{R}\big(\widehat{\boldsymbol{\rho}}^{(i)}\big)\widehat{\mathbf{F}}\big]^{-1}\widehat{\mathbf{F}}'\mathbf{R}\big(\widehat{\boldsymbol{\rho}}^{(i)}\big)'\mathbf{R}\big(\widehat{\boldsymbol{\rho}}^{(i)}\big)\\&\quad\times[(\mathbf{F}\mathbf{H}-\widehat{\mathbf{F}})\mathbf{H}^{-1}\boldsymbol{\lambda}_i+\mathbf{e}_i].\end{aligned}$$

Using Lemmas 1 and A.1(ii) we obtain

$$\begin{aligned}&\frac{1}{T}\widehat{\mathbf{F}}'\mathbf{R}\big(\widehat{\boldsymbol{\rho}}^{(i)}\big)'\mathbf{R}\big(\widehat{\boldsymbol{\rho}}^{(i)}\big)\widehat{\mathbf{F}}\\&\quad=\frac{1}{T}\sum_{t=p_i+1}^{T}\big(\widehat{\mathbf{F}}_t-\widehat{\rho}_{1,i}\widehat{\mathbf{F}}_{t-1}-\cdots-\widehat{\rho}_{p_i,i}\widehat{\mathbf{F}}_{t-p_i}\big)\\&\qquad\times\big(\widehat{\mathbf{F}}_t-\widehat{\rho}_{1,i}\widehat{\mathbf{F}}_{t-1}-\cdots-\widehat{\rho}_{p_i,i}\widehat{\mathbf{F}}_{t-p_i}\big)'\\&\quad=\frac{1}{T}\widehat{\mathbf{F}}'\mathbf{R}\big(\boldsymbol{\rho}^{(i)}\big)'\mathbf{R}\big(\boldsymbol{\rho}^{(i)}\big)\widehat{\mathbf{F}}+O_p\big(T^{-1/2}\big)+O_p(\delta_{NT}^{-2})\\&\qquad\text{(by Lemma 1)}\\&\quad=\frac{1}{T}\mathbf{H}'\mathbf{F}'\mathbf{R}\big(\boldsymbol{\rho}^{(i)}\big)'\mathbf{R}\big(\boldsymbol{\rho}^{(i)}\big)\mathbf{F}\mathbf{H}+O_p\big(T^{-1/2}\big)+O_p(\delta_{NT}^{-2})\\&\qquad\text{[by Lemma A.1(ii)].}\end{aligned}$$

Lemma A.1(i) yields $T^{-1}\sum_{t=p_i+1}^{T}\widehat{\mathbf{F}}_{t-k}(\widehat{\mathbf{F}}_{t-k}'-\mathbf{F}_{t-k}'\mathbf{H})=O_p(\delta_{NT}^{-2})$ and, by using Lemma 1

$$T^{-1}\widehat{\mathbf{F}}'\mathbf{R}\big(\widehat{\boldsymbol{\rho}}^{(i)}\big)'\mathbf{R}\big(\widehat{\boldsymbol{\rho}}^{(i)}\big)(\widehat{\mathbf{F}}-\mathbf{F}\mathbf{H})\mathbf{H}^{-1}\boldsymbol{\lambda}_i=O_p(\delta_{NT}^{-2})+O_p\big(\delta_{NT}^{-2}T^{-1/2}\big).$$

Next, we consider

$$\begin{aligned}&T^{-1/2}\sum_{t=p_i+1}^{T}[\widehat{\rho}_i(L)\widehat{\mathbf{F}}_t][\widehat{\rho}_i(L)e_{it}]\\&\quad=T^{-1/2}\sum_{t=p_i+1}^{T}\widehat{\rho}_i(L)[\mathbf{H}'\mathbf{F}_t+(\widehat{\mathbf{F}}_t-\mathbf{H}'\mathbf{F}_t)]\widehat{\rho}_i(L)e_{it}\\&\quad=T^{-1/2}\sum_{t=p_i+1}^{T}\rho_i(L)\mathbf{H}'\mathbf{F}_t[\rho_i(L)e_{it}]+O_p(\sqrt{T}/\delta_{NT}^2)\\&\qquad+O_p\big(T^{-1/2}\big),\end{aligned}$$

where Lemmas A.1(iii) and 1 are invoked. Hence, we find

$$\begin{aligned}\sqrt{T}(\widetilde{\boldsymbol{\lambda}}_{i,\widehat{\rho}}-\mathbf{H}^{-1}\boldsymbol{\lambda}_i)&=\big[T^{-1}\mathbf{H}'\mathbf{F}'\mathbf{R}\big(\boldsymbol{\rho}^{(i)}\big)'\mathbf{R}\big(\boldsymbol{\rho}^{(i)}\big)\mathbf{F}\mathbf{H}\big]^{-1}\\&\quad\times T^{-1/2}\mathbf{H}'\mathbf{F}'\mathbf{R}\big(\boldsymbol{\rho}^{(i)}\big)'\mathbf{R}\big(\boldsymbol{\rho}^{(i)}\big)\mathbf{e}_i\\&\quad+O_p(\sqrt{T}/\delta_{NT}^2)+O_p\big(T^{-1/2}\big),\end{aligned}$$

where $\sqrt{T}/\delta_{NT}^2\to0$ if $\sqrt{T}/N\to0$. Finally, Assumption 1(v) implies

$$T^{-1/2}\mathbf{H}'\mathbf{F}'\mathbf{R}\big(\boldsymbol{\rho}^{(i)}\big)'\mathbf{R}\big(\boldsymbol{\rho}^{(i)}\big)\mathbf{e}_i\xrightarrow{d}\mathcal{N}\big(\mathbf{0},\widetilde{\mathbf{V}}_{\mathbf{Fe}}^{(i)}\big),$$

where $\widetilde{\mathbf{V}}_{\mathbf{Fe}}^{(i)}$ is defined in Theorem 1. With these results, part (i) of the theorem follows.

The proof of part (ii) is similar. We therefore present the main steps only. The feasible two-step estimator of the common factors is given

by

$$\begin{aligned}\widetilde{\mathbf{F}}_{t,\widehat{\omega}} &= (\widehat{\boldsymbol{\Lambda}}'\widehat{\boldsymbol{\Omega}}^{-1}\widehat{\boldsymbol{\Lambda}})^{-1}\widehat{\boldsymbol{\Lambda}}'\widehat{\boldsymbol{\Omega}}^{-1}\mathbf{X}_t \\ &= (\widehat{\boldsymbol{\Lambda}}'\widehat{\boldsymbol{\Omega}}^{-1}\widehat{\boldsymbol{\Lambda}})^{-1}\widehat{\boldsymbol{\Lambda}}'\widehat{\boldsymbol{\Omega}}^{-1}[(\widehat{\boldsymbol{\Lambda}}-\widehat{\boldsymbol{\Lambda}}+\boldsymbol{\Lambda}\mathbf{H}'^{-1})\mathbf{H}'\mathbf{F}_t+\mathbf{e}_t], \\ \widetilde{\mathbf{F}}_{t,\widehat{\omega}}-\mathbf{H}'\mathbf{F}_t &= (\widehat{\boldsymbol{\Lambda}}'\widehat{\boldsymbol{\Omega}}^{-1}\widehat{\boldsymbol{\Lambda}})^{-1}\widehat{\boldsymbol{\Lambda}}'\widehat{\boldsymbol{\Omega}}^{-1}[(\boldsymbol{\Lambda}\mathbf{H}'^{-1}-\widehat{\boldsymbol{\Lambda}})\mathbf{H}'\mathbf{F}_t+\mathbf{e}_t],\end{aligned}$$

where $\mathbf{e}_t=[e_{1t},\ldots,e_{Nt}]'$. Under Lemma 1 the (diagonal) elements of $(\widehat{\boldsymbol{\Omega}}^{-1}-\boldsymbol{\Omega}^{-1})$ are $O_p(T^{-1/2})+O_p(\delta_{NT}^{-2})$. Following Bai (2003) and using Lemma A.1(iv) and (v), we obtain

$$\begin{aligned}N^{-1}\widehat{\boldsymbol{\Lambda}}'\widehat{\boldsymbol{\Omega}}^{-1}\widehat{\boldsymbol{\Lambda}} &= N^{-1}\mathbf{H}^{-1}\boldsymbol{\Lambda}'\boldsymbol{\Omega}^{-1}\boldsymbol{\Lambda}\mathbf{H}'^{-1}+O_p\big(T^{-1/2}\big)+O_p(\delta_{NT}^{-2}) \\ &\xrightarrow{p}\widetilde{\boldsymbol{\Psi}}_{\boldsymbol{\Lambda}}, \\ N^{-1}\widehat{\boldsymbol{\Lambda}}'\widehat{\boldsymbol{\Omega}}^{-1}(\widehat{\boldsymbol{\Lambda}}-\boldsymbol{\Lambda}\mathbf{H}'^{-1}) &\le N^{-1}\widehat{\boldsymbol{\Lambda}}'\boldsymbol{\Omega}^{-1}(\widehat{\boldsymbol{\Lambda}}-\boldsymbol{\Lambda}\mathbf{H}'^{-1}) \\ &\quad+\left(\frac{1}{N}\sum_{i=1}^{N}|\widehat{\omega}_i^{-2}-\omega_i^{-2}|^2\right)^{1/2}\left(\frac{1}{N}\|(\widehat{\boldsymbol{\Lambda}}'-\mathbf{H}^{-1}\boldsymbol{\Lambda}')\widehat{\boldsymbol{\Lambda}}\|^2\right)^{1/2} \\ &= O_p(\delta_{NT}^{-2})+O_p\big(T^{-1/2}/\delta_{NT}^2\big), \\ N^{-1}(\widehat{\boldsymbol{\Lambda}}-\boldsymbol{\Lambda}\mathbf{H}'^{-1})'\widehat{\boldsymbol{\Omega}}^{-1}\mathbf{e}_t &= N^{-1}(\widehat{\boldsymbol{\Lambda}}-\boldsymbol{\Lambda}\mathbf{H}'^{-1})'\boldsymbol{\Omega}^{-1}\mathbf{e}_t \\ &\quad+N^{-1}\sum_{i=1}^{N}\left(\frac{1}{\widehat{\omega}_i^2}-\frac{1}{\omega_i^2}\right)e_{it}(\widehat{\boldsymbol{\lambda}}_i-\mathbf{H}^{-1}\boldsymbol{\lambda}_i) \\ &= O_p(\delta_{NT}^{-2})+O_p\big(T^{-1/2}/\delta_{NT}^2\big), \\ N^{-1/2}\mathbf{H}^{-1}\boldsymbol{\Lambda}'\widehat{\boldsymbol{\Omega}}^{-1}\mathbf{e}_t &= N^{-1/2}\mathbf{H}^{-1}\boldsymbol{\Lambda}'\boldsymbol{\Omega}^{-1}\mathbf{e}_t+O_p\big(T^{-1/2}\big)+O_p(\delta_{NT}^{-2}) \\ &\xrightarrow{d}\mathcal{N}\big(\mathbf{0},\widetilde{\mathbf{V}}_{\lambda\mathbf{e}}^{(t)}\big), \\ \widetilde{\mathbf{V}}_{\lambda\mathbf{e}}^{(t)} &= E\Big(\lim_{N\to\infty}N^{-1}\mathbf{H}^{-1}\boldsymbol{\Lambda}'\boldsymbol{\Omega}^{-1}\mathbf{e}_t\mathbf{e}_t'\boldsymbol{\Omega}^{-1}\boldsymbol{\Lambda}\mathbf{H}'^{-1}\Big) \\ &= \lim_{N\to\infty}N^{-1}\sum_{i=1}^{N}\sum_{j=1}^{N}\frac{1}{\omega_i^2\omega_j^2}\mathbf{H}^{-1}\boldsymbol{\lambda}_i\boldsymbol{\lambda}_j'\mathbf{H}'^{-1}E(e_{it}e_{jt}).\end{aligned}$$

From these results the limit distribution stated in Theorem 1(ii) follows.

Proof of Theorem 2

First, we compare the asymptotic covariance matrices of the PC–OLS estimator $\widehat{\mathbf{F}}_t$ and the PC–GLS estimator $\widetilde{\mathbf{F}}_t$ (where for notational convenience the dependence on $\widehat{\omega}$ is suppressed). Using the results presented in Theorem 1(ii), the asymptotic covariance matrix of $\widetilde{\mathbf{F}}_t$ can be written as

$$\lim_{N,T\to\infty}N(\boldsymbol{\Lambda}_0'\boldsymbol{\Omega}^{-1}\boldsymbol{\Lambda}_0)^{-1}\boldsymbol{\Lambda}_0'\boldsymbol{\Omega}^{-1}E(\mathbf{e}_t\mathbf{e}_t')\boldsymbol{\Omega}^{-1}\boldsymbol{\Lambda}_0(\boldsymbol{\Lambda}_0'\boldsymbol{\Omega}^{-1}\boldsymbol{\Lambda}_0)^{-1},$$

where $\boldsymbol{\Lambda}_0=\boldsymbol{\Lambda}\mathbf{H}'^{-1}$. If the covariance structure is correctly specified, then $E(\mathbf{e}_t\mathbf{e}_t')=\boldsymbol{\Omega}$ and the asymptotic covariance matrix reduces to

$$\left(\lim_{N,T\to\infty}\frac{1}{N}\boldsymbol{\Lambda}_0'\boldsymbol{\Omega}^{-1}\boldsymbol{\Lambda}_0\right)^{-1}.$$

The asymptotic covariance matrix of the PC–OLS estimator is (Bai 2003)

$$\begin{aligned}&\lim_{N,T\to\infty}N(\boldsymbol{\Lambda}_0'\boldsymbol{\Lambda}_0)^{-1}\boldsymbol{\Lambda}_0'E(\mathbf{e}_t\mathbf{e}_t')\boldsymbol{\Lambda}_0(\boldsymbol{\Lambda}_0'\boldsymbol{\Lambda}_0)^{-1} \\ &\qquad=\lim_{N,T\to\infty}N(\boldsymbol{\Lambda}_0'\boldsymbol{\Lambda}_0)^{-1}\boldsymbol{\Lambda}_0'\boldsymbol{\Omega}\boldsymbol{\Lambda}_0(\boldsymbol{\Lambda}_0'\boldsymbol{\Lambda}_0)^{-1},\end{aligned}$$

if the covariance matrix $\boldsymbol{\Omega}$ is correctly specified.

Let $\widehat{\mathbf{F}}_t=\widetilde{\mathbf{F}}_t+\boldsymbol{\delta}_t$ where $\boldsymbol{\delta}_t=\widehat{\mathbf{F}}_t-\widetilde{\mathbf{F}}_t$. From

$$\operatorname{var}(\widehat{\mathbf{F}}_t)=\operatorname{var}(\widetilde{\mathbf{F}}_t)+\operatorname{var}(\boldsymbol{\delta}_t)+\operatorname{cov}(\widetilde{\mathbf{F}}_t,\boldsymbol{\delta}_t)+\operatorname{cov}(\boldsymbol{\delta}_t,\widetilde{\mathbf{F}}_t)$$

it follows that $\widetilde{\mathbf{F}}_t$ is asymptotically more efficient than $\widehat{\mathbf{F}}_t$ if $N\times\operatorname{cov}(\widetilde{\mathbf{F}}_t,\boldsymbol{\delta}_t)\to 0$ or

$$\begin{aligned}&\lim_{N,T\to\infty}NE[(\widetilde{\mathbf{F}}_t-\mathbf{H}'\mathbf{F}_t)(\widehat{\mathbf{F}}_t-\mathbf{H}'\mathbf{F}_t)'] \\ &\qquad=\lim_{N,T\to\infty}NE[(\widetilde{\mathbf{F}}_t-\mathbf{H}'\mathbf{F}_t)(\widetilde{\mathbf{F}}_t-\mathbf{H}'\mathbf{F}_t)'].\end{aligned}$$

Since

$$\begin{aligned}&\lim_{N,T\to\infty}NE[(\widetilde{\mathbf{F}}_t-\mathbf{H}'\mathbf{F}_t)(\widehat{\mathbf{F}}_t-\mathbf{H}'\mathbf{F}_t)'] \\ &\quad=\lim_{N,T\to\infty}N(\boldsymbol{\Lambda}_0'\boldsymbol{\Omega}^{-1}\boldsymbol{\Lambda}_0)^{-1}\boldsymbol{\Lambda}_0'\boldsymbol{\Omega}^{-1}E(\mathbf{e}_t\mathbf{e}_t')\boldsymbol{\Lambda}_0(\boldsymbol{\Lambda}_0'\boldsymbol{\Lambda}_0)^{-1} \\ &\quad=\lim_{N,T\to\infty}\left(\frac{1}{N}\boldsymbol{\Lambda}_0'\boldsymbol{\Omega}^{-1}\boldsymbol{\Lambda}_0\right)^{-1},\end{aligned}$$

it follows that the difference of the asymptotic covariance matrices of the PC–GLS and PC–OLS estimators is positive semidefinite.

In a similar manner, it can be shown that the PC–GLS estimator of $\boldsymbol{\lambda}_i$ is asymptotically more efficient than the PC–OLS estimator. Let $\mathbf{F}_0=\mathbf{F}\mathbf{H}$ and $\mathbf{R}(\boldsymbol{\rho}^{(i)})$ as defined in the proof of Theorem 1. The asymptotic distribution of the PC–GLS estimator $\widetilde{\boldsymbol{\lambda}}_i$ presented in Theorem 1(i) can be written as

$$\begin{aligned}&\left[\lim_{N,T\to\infty}E\left(\frac{1}{T}\mathbf{F}_0'\mathbf{R}(\boldsymbol{\rho}^{(i)})'\mathbf{R}(\boldsymbol{\rho}^{(i)})\mathbf{F}_0\right)\right]^{-1} \\ &\quad\times\lim_{N,T\to\infty}E\left(\frac{1}{T}\mathbf{F}_0'\mathbf{R}(\boldsymbol{\rho}^{(i)})'\mathbf{R}(\boldsymbol{\rho}^{(i)})\mathbf{e}_i\mathbf{e}_i'\mathbf{R}(\boldsymbol{\rho}^{(i)})'\mathbf{R}(\boldsymbol{\rho}^{(i)})\mathbf{F}_0\right) \\ &\quad\times\left[\lim_{N,T\to\infty}E\left(\frac{1}{T}\mathbf{F}_0'\mathbf{R}(\boldsymbol{\rho}^{(i)})'\mathbf{R}(\boldsymbol{\rho}^{(i)})\mathbf{F}_0\right)\right]^{-1}.\end{aligned}$$

If the autoregressive model for the idiosyncratic errors is correctly specified, we have $E(\boldsymbol{\varepsilon}_i\boldsymbol{\varepsilon}_i')=\sigma_i^2\mathbf{I}_{T-p_i}$, where $\boldsymbol{\varepsilon}_i=\mathbf{R}(\boldsymbol{\rho}^{(i)})\mathbf{e}_i$. If $\mathbf{F}_0$ is independent of $\boldsymbol{\varepsilon}_i$, it follows from the law of iterated expectations that the asymptotic covariance matrix of the PC–GLS estimator $\widetilde{\boldsymbol{\lambda}}_i$ reduces to

$$\sigma_i^2\left[\lim_{N,T\to\infty}E\left(\frac{1}{T}\mathbf{F}_0'\mathbf{R}(\boldsymbol{\rho}^{(i)})'\mathbf{R}(\boldsymbol{\rho}^{(i)})\mathbf{F}_0\right)\right]^{-1}.$$

Consider

$$\begin{aligned}&\lim_{N,T\to\infty}TE[(\widetilde{\boldsymbol{\lambda}}_i-\mathbf{H}^{-1}\boldsymbol{\lambda}_i)(\widehat{\boldsymbol{\lambda}}_i-\mathbf{H}^{-1}\boldsymbol{\lambda}_i)'] \\ &\quad=\left[\lim_{N,T\to\infty}E\left(\frac{1}{T}\mathbf{F}_0'\mathbf{R}(\boldsymbol{\rho}^{(i)})'\mathbf{R}(\boldsymbol{\rho}^{(i)})\mathbf{F}_0\right)\right]^{-1} \\ &\qquad\times\lim_{N,T\to\infty}E\left(\frac{1}{T}\mathbf{F}_0'\mathbf{R}(\boldsymbol{\rho}^{(i)})'\mathbf{R}(\boldsymbol{\rho}^{(i)})\mathbf{e}_i\mathbf{e}_i'\mathbf{F}_0\right) \\ &\qquad\times\left[\lim_{N,T\to\infty}E\left(\frac{1}{T}\mathbf{F}_0'\mathbf{F}_0\right)\right]^{-1} \\ &\quad=\left[\lim_{N,T\to\infty}E\left(\frac{1}{T}\mathbf{F}_0'\mathbf{R}(\boldsymbol{\rho}^{(i)})'\mathbf{R}(\boldsymbol{\rho}^{(i)})\mathbf{F}_0\right)\right]^{-1} \\ &\qquad\times\lim_{N,T\to\infty}E_{\mathbf{F}_0}\left\{E\left[\frac{1}{T}\mathbf{F}_0'\mathbf{R}(\boldsymbol{\rho}^{(i)})'\mathbf{R}(\boldsymbol{\rho}^{(i)})\mathbf{e}_i\mathbf{e}_i'\mathbf{R}(\boldsymbol{\rho}^{(i)})'\mathbf{R}(\boldsymbol{\rho}^{(i)})\right.\right. \\ &\qquad\left.\left.\times\left[\mathbf{R}(\boldsymbol{\rho}^{(i)})'\mathbf{R}(\boldsymbol{\rho}^{(i)})\right]^{-1}\mathbf{F}_0\middle|\mathbf{F}_0\right]\right\} \\ &\qquad\times\left[\lim_{N,T\to\infty}E\left(\frac{1}{T}\mathbf{F}_0'\mathbf{F}_0\right)\right]^{-1}\end{aligned}$$

$$= \sigma_i^2 \left[\lim_{N,T\to\infty} E\left(\frac{1}{T} \mathbf{F}_0' \mathbf{R}(\boldsymbol{\rho}^{(i)})' \mathbf{R}(\boldsymbol{\rho}^{(i)}) \mathbf{F}_0 \right) \right]^{-1}$$

and, therefore, the asymptotic covariance between $\widetilde{\boldsymbol{\lambda}}_i$ and $\widehat{\boldsymbol{\lambda}}_i - \widetilde{\boldsymbol{\lambda}}_i$ tends to zero. It follows that the PC–GLS estimator $\widetilde{\boldsymbol{\lambda}}_i$ is asymptotically more efficient than the PC–OLS estimator $\widehat{\boldsymbol{\lambda}}_i$.

[Received November 2009. Revised January 2011.]

REFERENCES

Anderson, T. W. (1984), *An Introduction to Multivariate Statistical Analysis* (2nd ed.), Hoboken, NJ: Wiley. [1152]

Bai, J. (2003), "Inferential Theory for Factor Models of Large Dimensions," *Econometrica*, 71, 135–172. [1150,1153,1162,1163,1165]

Bai, J., and Ng, S. (2002), "Determining the Number of Factors in Approximate Factor Models," *Econometrica*, 70, 191–221. [1150-1153,1164]

Bernanke, B. S., Boivin, J., and Eliasz, P. (2005), "Measuring the Effects of Monetary Policy: A Factor-Augmented Vector Autoregressive (FAVAR) Approach," *The Quarterly Journal of Economics*, 120, 387–422. [1150]

Boivin, J., and Ng, S. (2006), "Are More Data Always Better for Factor Analysis?" *Journal of Econometrics*, 132, 169–194. [1150,1152]

Breitung, J., and Eickmeier, S. (2011), "Testing for Structural Breaks in Dynamic Factor Models," *Journal of Econometrics*, 163, 71–84. [1162]

Breitung, J., and Tenhofen, J. (2009), "GLS Estimation of Dynamic Factor Models," mimeo, University of Bonn, available at *http://www.ect.uni-bonn.de/mitarbeiter/joerg-breitung/factor-gls*. [1153]

Chamberlain, G., and Rothschild, M. (1983), "Arbitrage, Factor Structure, and Mean–Variance Analysis on Large Asset Markets," *Econometrica*, 51, 1281–1304. [1150]

Chang, Y. (2002), "Nonlinear IV Unit Root Tests in Panels With Cross-Sectional Dependency," *Journal of Econometrics*, 110, 261–292. [1157]

Choi, I. (2011), "Efficient Estimation of Factor Models," *Econometric Theory*, to appear. [1150-1152,1154]

Chudik, A., Pesaran, M. H., and Tosetti, E. (2010), "Weak and Strong Cross Section Dependence and Estimation of Large Panels," *The Econometrics Journal*, 14, C45–C90. [1152]

Doz, C., Giannone, D., and Reichlin, L. (2006), "A Quasi Maximum Likelihood Approach for Large Approximate Dynamic Factor Models," *Reviews of Economics and Statistics*, to appear. [1150,1152,1156,1157]

——— (2007), "A Two-Step Estimator for Large Approximate Dynamic Factor Models Based on Kalman Filtering," *Journal of Econometrics*, 164, 188–205. [1156]

Eickmeier, S. (2007), "Business Cycle Transmission From the US to Germany—A Structural Factor Approach," *European Economic Review*, 51, 521–551. [1150]

Eickmeier, S., and Ziegler, C. (2008), "How Successful Are Dynamic Factor Models at Forecasting Output and Inflation? A Meta-Analytic Approach," *Journal of Forecasting*, 27, 237–265. [1150]

Forni, M., Hallin, M., Lippi, M., and Reichlin, L. (2000), "The Generalized Dynamic Factor Model: Identification and Estimation," *The Review of Economics and Statistics*, 82, 540–554. [1150,1155,1159,1160]

——— (2005), "The Generalized Dynamic Factor Model: One-Sided Estimation and Forecasting," *Journal of the American Statistical Association*, 100, 830–840. [1151,1152,1154,1155,1159,1160]

Giannone, D., Reichlin, L., and Sala, L. (2002), "Tracking Greenspan: Systematic and Unsystematic Monetary Policy Revisited," Discussion Paper 3550, Centre for Economic Policy Research, London, U.K. [1150]

Hansen, L. P. (1982), "Large Sample Properties of Generalized Method of Moments Estimators," *Econometrica*, 50, 1029–1054. [1156]

Hansen, L. P., Heaton, J., and Yaron, A. (1996), "Finite Sample Properties of Some Alternative GMM Estimators," *Journal of Business & Economic Statistics*, 14, 262–280. [1156]

Hayashi, F. (2000), *Econometrics*, Princeton, NJ: Princeton University Press. [1155]

Jungbacker, B., and Koopman, S. J. (2008), "Likelihood-Based Analysis for Dynamic Factor Models," mimeo, VU University Amsterdam. [1150]

Patton, A., Politis, D. N., and White, H. (2009), "Correction to "Automatic Block-Length Selection for the Dependent Bootstrap" by D. Politis and H. White," *Econometric Reviews*, 28, 372–375. [1160]

Politis, D. N., and Romano, J. P. (1992), "A Circular Block-Resampling Procedure for Stationary Data," in *Exploring the Limits of Bootstrap*, eds. R. LePage and L. Billard, New York: Wiley, pp. 263–270. [1160]

Politis, D. N., and White, H. (2004), "Automatic Block-Length Selection for the Dependent Bootstrap," *Econometric Reviews*, 23, 53–70. [1160]

Redner, R. A., and Walker, H. F. (1984), "Mixture Densities, Maximum Likelihood and the EM Algorithm," *SIAM Review*, 26, 195–239. [1152]

Stock, J. H., and Watson, M. W. (2002a), "Macroeconomic Forecasting Using Diffusion Indexes," *Journal of Business & Economic Statistics*, 20, 147–162. [1150,1152]

——— (2002b), "Forecasting Using Principal Components From a Large Number of Predictors," *Journal of the American Statistical Association*, 97, 1167–1179. [1150,1151,1154]

——— (2005), "Implications of Dynamic Factor Models for VAR Analysis," Working Paper 11467, National Bureau of Economic Research, Cambridge, MA. [1150-1152,1156,1160,1161]

Watson, M. W. (2003), "Macroeconomic Forecasting Using Many Predictors," in *Advances in Economics and Econometrics: Theory and Applications*, Vol. III, eds. M. Dewatripont, L. P. Hansen, and S. J. Turnovsky, Cambridge, U.K.: Cambridge University Press, pp. 87–114. [1150]