Analyzing business and financial cycles using multi-level factor models

Jörg Breitung
University of Cologne and
Deutsche Bundesbank

Sandra Eickmeier
Deutsche Bundesbank and
Centre for Applied Macroeconomic Analysis, ANU

Abstract

This paper compares alternative estimation procedures for multi-level factor models which imply blocks of zero restrictions on the associated matrix of factor loadings. We suggest a sequential least squares algorithm for minimizing the total sum of squared residuals and a two-step approach based on canonical correlations that are much simpler and faster than Bayesian approaches previously employed in the literature. Monte Carlo simulations suggest that the estimators perform well in typical sample sizes encountered in the factor analysis of macroeconomic data sets. We apply the methodologies to study international co-movements of business and financial cycles as well as asymmetries over the business cycle in the US.

Keywords

Factor models, canonical correlations, international business cycles, financial cycles, business cycle asymmetries

JEL Classification

C38, C55

Address for correspondence:

(E) cama.admin@anu.edu.au

Analyzing business and financial cycles using multi-level factor models[*]

Jörg Breitung[†]   Sandra Eickmeier[‡]

May 2014

Abstract

This paper compares alternative estimation procedures for multi-level factor models which imply blocks of zero restrictions on the associated matrix of factor loadings. We suggest a sequential least squares algorithm for minimizing the total sum of squared residuals and a two-step approach based on canonical correlations that are much simpler and faster than Bayesian approaches previously employed in the literature. Monte Carlo simulations suggest that the estimators perform well in typical sample sizes encountered in the factor analysis of macroeconomic data sets. We apply the methodologies to study international comovements of business and financial cycles as well as asymmetries over the business cycle in the US.

JEL classification: C38, C55

Keywords: Factor models, canonical correlations, international business cycles, financial cycles, business cycle asymmetries

Contents

1   Introduction

In recent years (dynamic) factor models have become increasingly popular for macro-economic analysis and forecasting in a data-rich environment.[1] A serious limitation of the standard approximate factor model is that it assumes the common factors to affect all variables of the system. As argued by Boivin and Ng (2006) the efficiency of the Principal Component (PC) estimator may deteriorate substantially if groups of variables are included that do not provide any information about the factors, that is, the corresponding factor loadings of some subgroups of variables are equal to zero. Similarly, if factors are ignored that affect a subset of variables only, the respective idiosyncratic components may be highly correlated, resulting in poor (PC) estimates of those factors which load on all variables.

There are natural examples for models with factors loading on subgroups of variables only. In an international context, for example, factors may represent regional characteristics, and it may be of independent interest to analyze these "regional factors" in addition to "global factors" linking all variables in the model. Alternatively (or in addition), a block structure may represent economic, cultural or other characteristics. A natural way to deal with such block structures is to extract "regional factors" from various subgroups of data (data associated with specific "regions") separately. However, if there exist at the same global factors that affect all regions in the sample, a separate analysis of the regions will mix up regional and global factors which hampers identification of the factors and involves a severe loss of efficiency.

A characterizing feature of such model structures is that the loading matrix of the common factors is subject to blocks of zero restrictions and the technical challenge is to take into account such restrictions when estimating the common factors. Estimating the state space representation of the model employing Bayesian methods is most popular Kose, Otrok and Whiteman (2003), Moench, Ng and Potter (2013), Kaufmann and Schumacher (2012) and Francis, Owyang and Savascin (2012).[2] Other recent papers adapt frequency domain PCs (Hallin and Liska (2011)), a two-step quasi maximum likelihood (ML) estimator (Banbura, Giannone and Reichlin (2010), Cicconi (2012)), two-stage PC approaches (e.g. Beck et al. (2009), Beck, Hubrich

---

[1]For forecasting applications see, e.g., Stock and Watson (2002a), Stock and Watson (2002b), Eickmeier and Ziegler (2008). For structural macro applications see, e.g., Bernanke, Boivin and Eliasz (2005), Eickmeier (2007), Eickmeier and Hofmann (2013), Beck, Hubrich and Marcellino (2009), Kose, Otrok and Whiteman (2003).

[2]The latter two papers assume that the groups of variables are unknown and are determined endogenously in the model. By contrast, the two former papers as well as the present paper determine a priori which variables are associated with which group.

and Marcellino (2011), Aastveit, Bjoernland and Thorsrud (2011)) or a sequential PC approach (Wang (2010)).[3]

In this paper we make several contributions. First, we provide a comprehensive comparison of existing estimation approaches for multi-level factor models and propose two very simple alternative estimation techniques based on sequential least squares (LS) and canonical correlations. The sequential LS algorithm is equivalent to the (quasi) ML estimator assuming Gaussian i.i.d. errors and treating the common factors as unknown parameters. It is closely related to Wang (2010)'s sequential PC approach and the quasi ML approach of Banbura et al. (2010). The estimator based on a canonical correlation analysis (CCA) avoids any iterations and can be computed in two steps. In particular, we employ this computationally convenient and consistent estimator for initializing the LS algorithm in order to ensure that the procedure starts in the neighborhood of the global minimum.

These estimation techniques provide (point) estimates in less than 0.02 seconds (in typical macroeconomic settings) compared to a Bayesian estimator that requires several hours. Moreover, our Monte Carlo simulations suggest that, in some circumstances, the sequential LS and the CCA estimators tend to outperform alternative estimation methods such as the two-step PC estimator and the quasi ML estimator based on the EM algorithm.

The two-step PC estimator involves estimating, in the first step, the global factors as the first PCs of the full dataset. In a second step the global factors are purged of all variables and the regional factors are extracted applying regional-specific PC analyses to the residuals. In Section 2.4.1 we argue that for the consistency of this estimator we need to assume that the number of regions tends to infinity, whereas in empirical practice the number of groups is typically small (often less than 10). In such cases, the largest eigenvalues may correspond to dominating regional factors so that identification of the global factors by the largest eigenvalues breaks down.

We also extend the sequential LS estimation approach to a three-level factor model (with, for example, global factors, regional factors and factors specific to types of variables) with overlapping blocks of factors. Such factor structures are challenging as they cannot be estimated one level after another (which is the rationale for Wang (2010)'s sequential PC approach).

A final contribution are three applications in which we study international co-movements of business and financial cycles as well as asymmetries over the business

---

[3]Other studies estimate small-dimensional multi-level factor models (e.g. Gregory and Head (1999)). In our paper we focus, however, on large-dimensional models and, therefore, do not differences between those papers and ours further.

cycle. The first and third applications use the two-level factor model, while the second application uses the three-level factor model assuming an overlapping factor structure.

In the first application we (basically) replicate the study by Hirata, Kose and Otrok (forthcoming) and apply several estimation methodologies for two-level factor models to an annual real activity dataset of more than 100 countries between 1960 and 2010. We estimate global and regional factors which turn out to be similar across methods. We confirm Hirata et al. (forthcoming)'s main finding that regional (business cycle) factors have become more important and global factors less important over time.

In the second application, we use a large quarterly macro-financial dataset for 24 countries between 1995 and 2011. We estimate a global factor, regional factors, as well as factors specific to types of variables (i.e. macro and financial factors). We find that financial variables strongly comove internationally, to a similar extent as macroeconomic variables. Macroeconomic and financial dynamics share common factors, but financial factors independent from macro factors also matter for financial variables. Finally, the temporal evolution of the estimated financial factors looks plausible.

In the third application, we use a large US dataset comprising monthly macro and financial variables over 1959-2011. We estimate factors which are common to all periods ("symmetric" factors) as well as phase-specific ("asymmetric") factors which drive the variables in recession or expansion phases only. Hence, this application differs from the previous ones where factors common to a subset of variables rather than points in time were considered. We find non-negligible asymmetries over the business cycle. However, the bulk of common dynamics is stable over time. The overall comovement between the variables is higher during recessions than during expansions. The first recession factor is highly correlated with monetary and financial variables whereas the first symmetric and the first expansion factors are related to real activity variables.

The remainder of the paper is organized as follows. We first present the two-level factor model in Section 2.1. In Sections 2.2 and 3 we then suggest a sequential LS estimator for the two-level factor model and, as an extension, the three-level factor model. In Section 2.3 we propose a CCA estimator. We show in Section 2.4.1 that the two-stage PC approach which has been used in the literature works well only under specific conditions. In Sections 2.4.2 and 2.4.3 we compared the sequential LS approach with the sequential PC and the quasi ML approaches. In Section 4

we investigate the relative performance of alternative estimators by means of Monte Carlo simulations. For ease of exposition we assume in the methodological sections that we work with a large international dataset. We label factors associated with all variables as "global factors" and factors associated with specific groups "regional factors" and/or "variable type-specific factors". However, the models are, of course, more general and can be applied to other empirical setups with variables being associated with other groups as well. In Section 5 we present our applications, and we conclude in Section 6.

2 The two-level factor model

2.1 The model

Consider the following two-level factor model

$$y_{r,it} = \gamma'_{r,i} G_t + \lambda'_{r,i} F_{r,t} + u_{r,it} , \tag{1}$$

where $r = 1, ..., R$ indicates the region, the index $i = 1, \ldots, n_r$ denotes the the i'th variable of region r and $t = 1, \ldots, T$ stands for the time period. The vector $G_t = (g_{1,t}, \ldots, g_{m_0,t})'$ comprises m_0 global factors and the $m_r \times 1$ vector $F_{r,t}$ collects the m_r regional factors in region r. The idiosyncratic component is denoted by $u_{r,it}$, where the usual assumptions of an approximate factor model (e.g. Bai (2003)) apply. In vector notation, the factor model for region r is written as

$$
\begin{aligned}
y_{r,\cdot t} &= \Gamma_r G_t + \Lambda_r F_{r,t} + u_{r,\cdot t}, && (2) \\
&= \begin{pmatrix} \Gamma_r & \Lambda_r \end{pmatrix} \begin{pmatrix} G_t \\ F_{r,t} \end{pmatrix} + u_{r,\cdot t} , && (3)
\end{aligned}
$$

where $y_{r,\cdot t} = (y_{r,1t}, \ldots, y_{r,n_r t})'$ and Γ_r, Λ_r and $u_{r,\cdot t}$ are defined conformably. The entire system representing all R regions results as

$$
\begin{pmatrix} y_{1,\cdot t} \\ \vdots \\ y_{R,\cdot t} \end{pmatrix} = \begin{pmatrix} \Gamma_1 & \Lambda_1 & 0 & \cdots & 0 \\ \Gamma_2 & 0 & \Lambda_2 & \cdots & 0 \\ \vdots & & & \ddots & \vdots \\ \Gamma_R & 0 & 0 & \cdots & \Lambda_R \end{pmatrix} \begin{pmatrix} G_t \\ F_{1,t} \\ F_{2,t} \\ \vdots \\ F_{R,t} \end{pmatrix} + \begin{pmatrix} u_{1,\cdot t} \\ u_{2,\cdot t} \\ \vdots \\ u_{R,\cdot t} \end{pmatrix} \tag{4}
$$

$$ y_t = \Lambda^* F_t^* + u_t , \tag{5}$$

where $F_t^* = (G_t', F_{1,t}', \ldots, F_{R,t}')'$. We normalize the factor space by imposing the following assumptions:

(i) $T^{-1} \sum_{t=1}^T G_t G_t' = I_{m_0}$ and $T^{-1} \sum_{t=1}^T F_{r,t} F_{r,t}' = I_{m_r}$ for all r.

(ii) $N^{-1} \Gamma_r' \Gamma_r$ and $N^{-1} \Lambda_r' \Lambda_r$ are diagonal matrices which coincides with the respective assumption of the PC estimator, see e.g. Breitung and Choi (2013).

(iii) $T^{-1} \sum_{t=1}^T F_{r,t} G_t' = 0$ for all r.

As shown by Wang (2010) these restrictions ensure that all parameters are identified. We do not need to assume that the regional factors from different regions are uncorrelated. This assumption is often imposed for a Bayesian analysis of the multi-level factor model (e.g. Kose, Otrok and Whiteman (2003)) and it implies an over-identified model structure.

2.2 The sequential least-squares estimator

Assume that the idiosyncratic components are identically and independent normally distributed (i.i.d.) across i, t and r with $\mathbb{E}(u_{r,it}^2) = \sigma^2$.[4]

Treating the factors and factor loadings as unknown parameters yields the log-likelihood function

$$\mathcal{L}(F^*, \Lambda^*, \sigma^2) = \text{const} - \frac{T}{2} \left(\sum_{r=1}^R n_r \right) \log(\sigma^2) - \frac{1}{2\sigma^2} \sum_{t=1}^T (y_t - \Lambda^* F_t^*)'(y_t - \Lambda^* F_t^*)$$

where $F^* = (F_1^*, \ldots, F_T^*)'$. The maximization of this likelihood function is equivalent to minimizing the sum of squared residuals (RSS)

$$S(F^*, \Lambda^*) = \sum_{t=1}^T (y_t - \Lambda^* F_t^*)'(y_t - \Lambda^* F_t^*)$$
$$= \sum_{r=1}^R \sum_{i=1}^{n_r} \sum_{t=1}^T (y_{r,it} - \gamma_{r,i}' G_t - \lambda_{r,i}' F_{r,t})^2.$$

Assume that we have available suitable initial estimators of the global and regional factors, denoted by $\widehat{G}^{(0)} = (\widehat{G}_1^{(0)}, \ldots, \widehat{G}_T^{(0)})'$ and $\widehat{F}_r^{(0)} = (\widehat{F}_{r,1}^{(0)}, \ldots, \widehat{F}_{r,T}^{(0)})'$. The associated loading coefficients are estimated from $\sum_{r=1}^R n_r$ time series regressions of the form

---

[4] The assumption that the errors are i.i.d. is a simplifying assumption that is used to obtain a simple (quasi) likelihood function. The estimator remains consistent if the errors are heteroskedastic and autocorrelated, cf. Wang (2010).

$$y_{r,it} = \gamma'_{r,i}\widehat{G}_t^{(0)} + \lambda'_{r,i}\widehat{F}_{r,t}^{(0)} + \widetilde{u}_{r,it} \ . \tag{6}$$

Denote the resulting estimates as $\widehat{\gamma}_{r,i}^{(0)}$, $\widehat{\lambda}_{r,i}^{(0)}$ and the respective matrices as $\widehat{\Gamma}_r^{(0)} = (\widehat{\gamma}_{r,1}^{(0)}, \ldots, \widehat{\gamma}_{r,n_r}^{(0)})'$ and $\widehat{\Lambda}_r^{(0)} = (\widehat{\lambda}_{r,1}^{(0)}, \ldots, \widehat{\lambda}_{r,n_r}^{(0)})'$. The loading matrix for the full system is constructed as

$$\widehat{\Lambda}_{(0)}^* = \begin{pmatrix} \widehat{\Gamma}_1^{(0)} & \widehat{\Lambda}_1^{(0)} & 0 & \cdots & 0 \\ \widehat{\Gamma}_2^{(0)} & 0 & \widehat{\Lambda}_2^{(0)} & \cdots & 0 \\ \vdots & & & \ddots & \vdots \\ \widehat{\Gamma}_R^{(0)} & 0 & 0 & \cdots & \widehat{\Lambda}_R^{(0)} \end{pmatrix}.$$

An updated estimator for the vector of factors is obtained from the least-squares regression of y_t on $\widehat{\Lambda}_{(0)}^*$ yielding

$$\widehat{F}_{t,(1)}^* = \begin{pmatrix} \widehat{G}_t^{(1)} \\ \widehat{F}_{1,t}^{(1)} \\ \vdots \\ \widehat{F}_{R,t}^{(1)} \end{pmatrix} = \left(\widehat{\Lambda}_{(0)}^{*\prime}\widehat{\Lambda}_{(0)}^*\right)^{-1}\widehat{\Lambda}_{(0)}^{*\prime}y_t \ , \tag{7}$$

where in each step the factors are normalized to have a unit variance by multiplying the vector of factors with the matrix $\left(T^{-1}\sum_{t=1}^T \widehat{F}_{t,(1)}^* \widehat{F}_{t,(1)}^{*\prime}\right)^{-1/2}$.

Next, the updated factors can be used to obtain the associated loading coefficients from the least-squares regression (6), yielding the updated estimator $\widehat{\Lambda}_{(1)}^*$ which in turn yields the updated factors $\widehat{F}_{t,(2)}^*$. It is easy to see that

$$\mathcal{S}(\widehat{F}_{(0)}^*, \widehat{\Lambda}_{(0)}^*) \ \leq \ \mathcal{S}(\widehat{F}_{(1)}^*, \widehat{\Lambda}_{(0)}^*) \ \leq \ \mathcal{S}(\widehat{F}_{(1)}^*, \widehat{\Lambda}_{(1)}^*) \ \leq \ \cdots$$

since in each step the previous estimators are contained in the parameter space of the subsequent least-squares estimators. Hence the next estimation step cannot yield a larger RSS and, therefore the sequence of least-squares regressions eventually converges to a minimum.

To ensure that the iterative algorithm converges quickly to the global minimum, we initialize the algorithm with suitable starting values for the factors. In our Monte Carlo experiments and in the empirical applications we employ the CCA estimator, which is considered in Section 2.3.

So far we have assumed that the idiosyncratic variances are identical for all variables and regions. Although the resulting LS estimator is consistent in the case of heteroskedastic errors (since the LS estimators are robust against heteroskedastic

errors), the asymptotic efficiency may be improved by using a generalized least-squares (GLS) approach (cf. Breitung and Tenhofen (2011)).

It is important to notice that the proposed algorithm does not impose a particular normalization. Therefore, although the vector of common components $\xi_t = \Lambda^* F_t^*$ is identified and consistently estimated, whereas the factors and loading matrices are estimated consistently up to some arbitrary rotation. In order to impose the normalization proposed in Section 2.1 we first regress the final estimators of the regional factors $\widehat{F}_{r,t}$ ($r = 1, \ldots, R$) on $\widehat{G}_t$. The residuals from these regressions yield the orthogonalized regional factor. In order to adopt the same normalization as in the PC analysis, the normalized global factors can be obtained as the r_g PCs of the estimated common components resulting from the nonzero eigenvalues and the associated eigenvectors of the matrix

$$\widehat{\Gamma} \left(\frac{1}{T} \sum_{t=1}^{T} \widehat{G}_t \widehat{G}_t' \right) \widehat{\Gamma}' .$$

The PC normalization of the regional factors can be imposed in a similar manner by using the covariance matrix of the respective common components.

2.3 The CCA estimator

We start with estimating the $m = m_0 + m_r$ global and regional factors in each region separately by a PC analysis yielding the vector of factors $\widehat{F}_{r,t}^+$ which is a consistent estimator for the factor space of the $m \times 1$ vector of factors $(G_t', F_{r,t}')'$. Since the PCs of two different regions share a common component (the global factor), we apply a CCA to determine the linear combination $\widehat{\mathcal{G}}_{r,t} = \tau_r' \widehat{F}_{r,t}^+$ that is most correlated with the linear combination $\widehat{\mathcal{G}}_{s,t} = \tau_s' \widehat{F}_{s,t}^+$ of some other region s. This problem is equivalent to solving the generalized eigenvalue problem

$$\left| \mu \sum_{t=1}^{T} \widehat{F}_{r,t}^+ \widehat{F}_{r,t}^{+\prime} - \sum_{t=1}^{T} \widehat{F}_{r,t}^+ \widehat{F}_{s,t}^{+\prime} \left(\sum_{t=1}^{T} \widehat{F}_{s,t}^+ \widehat{F}_{s,t}^{+\prime} \right)^{-1} \sum_{t=1}^{T} \widehat{F}_{s,t}^+ \widehat{F}_{r,t}^{+\prime} \right| = 0.$$

The eigenvectors associated with the m_0 largest eigenvalues provide the weights of the linear combination $\widehat{\mathcal{G}}_{r,t} = \tau_r' \widehat{F}_{r,t}^+$ which serves as an estimator of the global factors G_t. As in the appendix of Breitung and Pigorsch (2013) it can be shown that as $N \to \infty$ and $T \to \infty$ the linear combination $\widehat{\mathcal{G}}_{r,t}$ (or $\widehat{\mathcal{G}}_{s,t}$) converges in probability to HG_t, where H is some regular $m_0 \times m_0$ matrix. Hence, $\widehat{\mathcal{G}}_{r,t}$ yields a consistent estimator of the space spanned by G_t.

Obviously, there are $R^2(R-1)/2$ possible pairs $(\widehat{F}_{r,t}^+, \widehat{F}_{s,t}^+)$ that can be employed for a CCA. We suggest to choose the linear combination with the largest canonical correlation (resp. eigenvalue) as the preferred estimate of G_t. In the next step the estimated global factors are purged of all variables and from R region-specific PC analyses the regional factors are extracted.

2.4 Relation to existing (non Bayesian) approaches

2.4.1 Two-step PC estimators

Since the set of regional factors $\{F_{1,t}, \ldots, F_{R,t}\}$ is assumed to be uncorrelated with the vector of global factors G_t, the regional factors may be treated as idiosyncratic components yielding the reduced factor model

$$y_{r,it} = \gamma_{r,i}' G_t + e_{r,it}$$

where $e_{r,it} = \lambda_{r,i}' F_{r,t} + u_{r,it}$. Accordingly, the global factors may be estimated by the first m_0 PCs of the matrix $T^{-1} Y'Y$, where $Y = (Y_1, \ldots, Y_R)$ and $Y_r = (y_{r,it})$ is the $T \times n_r$ data matrix of region r. In a second step, the regional factors may be estimated again with PCs from the covariance matrix of the resulting idiosyncratic components associated with a specific region. We refer to this estimator as the "top-down PC estimator" as this estimator starts from a PC analysis of the entire system. In empirical studies this top-down PC estimator is employed by Beck et al. (2011), Beck et al. (2009), Aastveit et al. (2011) and Thorsrud (2013). A problem with this estimator is that the regional factors give rise to a strong correlation among the regional clusters of idiosyncratic components. Let $\tau_{r,ij} = \max_t \mathbb{E}(|e_{r,it} e_{r,jt}|)$. Since the errors possess a factor structure it follows that $\sum_{i=1}^{n_r} \sum_{j=1}^{n_r} \tau_{r,ij} = O(n_r^2)$ and, therefore,

$$\sum_{r=1}^{R} \sum_{i=1}^{n_r} \sum_{j=1}^{n_r} \tau_{r,ij} = O\left(\sum_{r=1}^{R} n_r^2\right).$$

As shown by Bai (2003) consistent estimation of the factors requires that

$$\frac{1}{\left(\sum_{i=1}^{R} n_r\right)} \sum_{r=1}^{R} \sum_{i=1}^{n_r} \sum_{j=1}^{n_r} \tau_{r,ij} \leq M < \infty$$

and, thus, $\sum_{r=1}^{R} n_r^2 / \sum_{r=1}^{R} n_r$ needs to be bounded. Obviously, this condition is fulfilled if n_r is fixed and $R \to \infty$. In empirical practice, however, n_r is large relative to R so that an asymptotic framework assuming $n_r/R \to 0$ is inappropriate in

8

typical empirical applications.

An identical estimator would be obtained by an alternative two-stage PC estimator. Let $\widetilde{F}_{r,t}$ denote the vector of the first $m_0 + m_r$ PCs of the region-specific covariance matrix $T^{-1}Y_r'Y_r$. The global factor is estimated by a second PC analysis of the covariance matrix of the estimated factors $T^{-1}\sum_{i=1}^{T}\overline{F}_t'\overline{F}_t$ where $\overline{F}_t = \left(\widetilde{F}_{1,t}', \ldots, \widetilde{F}_{R,t}'\right)'$. This estimator may be refereed to as the "bottom-up PC estimator". A problem with the last PC step is that the number of regions is often too small in practice, violating the conditions established by Bai (2003) and Bai and Ng (2002) for consistent estimation of the (global) factors.

For illustration, consider the model with a single global factor G_t. The equivalence to the bottom-up and top-town PC estimators results from the fact that for the eigenvalue problem we have

$$
\begin{aligned}
\max_{v} \frac{v'Y'Yv}{v'v} &= \max_{v} \frac{\sum_{r=1}^{R} v_r'Y_r'Y_r v_r}{\sum_{r=1}^{R} v_r'v_r} \\
&= \max_{a_1,\ldots,a_R} \sum_{r=1}^{R} a_r \max_{v_r} \left. \frac{v_r'X_r'X_r v_r}{v_r'v_r} \right| \text{ subject to } \sum_{r=1}^{R} a_r = 1 \\
&= \max_{a_1,\ldots,a_R} \sum_{r=1}^{R} a_r \widehat{F}_{r,t}^2 \,,
\end{aligned}
\tag{8}
$$

where $v = (v_1, \ldots, v_R)'$ and $a_i = v_i'v_i/\sum_{r=1}^{R} v_r'v_r$. Accordingly, the first PC of the full sample results as a linear combination of the R region-specific PCs and the maximum is obtained as the largest eigenvalue of the sample covariance matrix of the vector $(\widehat{F}_{1,t}', \ldots, \widehat{F}_{R,t}'')'$. It follows that the bottom-up and the top-down PC estimators are equivalent.

2.4.2 The sequential PC approach

Wang (2010) proposes a sequential PC estimator for maximizing the log-likelihood estimator which is also based on the minimization of the RSS of the two-level factor model.

$$
S(F^*, \Lambda^*) \;=\; \sum_{r=1}^{R}\sum_{i=1}^{n_R}\sum_{t=1}^{T}(y_t - \Lambda^*F_t^*)'(y_t - \Lambda^*F_t^*)
$$

with respect to $F^* = (F_1^*, \ldots, F_t^*)'$ and Λ^*.

Assume that we have a suitable initial estimator of the global factors, denoted by $\widehat{G}^{(0)} = (\widehat{G}_1^{(0)}, \ldots, \widehat{G}_T^{(0)})'$. Conditional on these initial estimates it is straightforward to obtain initial estimators of the regional factors in region r. All variables are

purged of the global factor by running regressions of the variables on the estimated global factor. Then regional factors are estimated as the first m_r PCs of the sample covariance matrix

$$\widehat{\Sigma}_r^{(0)} = \frac{1}{T} \sum_{t=1}^{T} Y_r' M_{\widehat{G}^{(0)}} Y_r \ , \tag{9}$$

where $Y_r = (y_{r,it})$ is the $T \times n_r$ matrix of observations from region r and $M_{\widehat{G}^0} = I_T - \widehat{G}^{(0)}(\widehat{G}^{(0)\prime}\widehat{G}^{(0)})^{-1}\widehat{G}^{(0)\prime}$. Denote the $T \times m_r$ matrix of the resulting PCs as $\widehat{F}_r^{(0)}$. To eliminate the regional factors from the sample the following R regressions are performed

$$Y_r = \widehat{F}_r^{(0)} B_r + W_r \quad \text{for } r = 1, 2, \ldots, R. \tag{10}$$

Note that at this stage the assumption is imposed that the regional factors are orthogonal to the global factor (that enters the residual of this regression). Let $\widehat{W} = (\widehat{W}_1, \ldots, \widehat{W}_R)'$ denote the $T \times (\sum_{r=1}^{R} n_r)$ matrix of residuals from the R regressions (10). The updated estimates of the global factors $\widehat{G}_t^{(1)}$ are obtained as the first m_0 PCs obtained from the sample covariance matrix

$$\widehat{\Omega}^{(1)} = \frac{1}{T} \widehat{W}' \widehat{W}.$$

With the updated estimate of the global factors the matrix $\widehat{\Sigma}_r^{(0)}$ can be computed as in (9) but using $M_{\widehat{G}^{(1)}}$ instead of $M_{\widehat{G}^{(0)}}$ in order to obtain the updated estimate $\widehat{F}_r^{(1)}$. These steps are repeated until convergence.

Wang (2010) initializes the algorithm either with global factors obtained with the top-down PC approach considered in the Section 2.4.1 or, alternatively, with a confirmatory factor analysis given the set of admissible rotations of the regional PCs.

Since both sequential LS and PC approaches minimize the sum of squared errors, the fix point is identical, and the approaches should yield the same estimates. The main advantage of the LS estimator is that it can be straightforwardly generalized to more than two factor levels with overlapping factor structures (as we will show in Section 3), whereas the sequential PC estimator is confined to hierarchical factor models. Second, the LS estimator is computationally less demanding and tends to be faster. Third, in models with heteroskedastic or autocorrelated errors, the sequential LS technique can be used to compute the implied ML estimator that is equivalent to minimizing the weighted sum of squared residuals (cf. Breitung and Tenhofen (2011)), which is equivalent to the (pseudo) ML estimator. It is unclear how this could be achieved with the sequential PC approach.

2.4.3 The quasi ML approach

A related estimation procedure based on quasi ML is employed in Banbura et al. (2010). The conceptual difference to the sequential LS approach is that their approach assumes that the factors are normally distributed random variables yielding a log-likelihood function which includes – besides our RSS – an additional expression that is due to the distribution of the vector of factors. To maximize the likelihood function an EM algorithm is adapted that was originally proposed for the standard factor model without block structures. This approach gives rise to a shrinkage estimator for the vector of factors given by

$$\widehat{F}_t^{(1)} = (\widehat{\Lambda}_0^{*\prime}\widehat{\Lambda}_0^* + \bar{\sigma}_0^2 I_n)^{-1}\widehat{\Lambda}_0^{*\prime}y_t \ , \tag{11}$$

where $\bar{\sigma}_0^2 = (NT)^{-1}S(\widehat{F}_{(0)}^*, \widehat{\Lambda}_{(0)}^*)$. As $N \to \infty$ we have $\Lambda'\Lambda/N + (\sigma^2/N)I_m \to \lim_{N\to\infty}\Lambda'\Lambda/N$ and, therefore, the estimators (7) and (11) are asymptotically equivalent.

3 The three-level factor model

The factor model can be extended to include further (overlapping) levels. Assume an international macro-financial panel, where the variables are clustered according to some additional criteria. For example, the variables may be grouped into output-related variables (e.g. production indices, employment), price variables (e.g. consumer prices, producer prices, wages) and financial variables (e.g. interest rates, stock returns). Accordingly, an additional index $k = 1, \ldots, K$ is introduced and the factor model is written as

$$y_{rk,it} = \gamma'_{rk,i}G_t + \lambda'_{rk,i}F_{r,t} + \theta'_{rk,i}H_{k,t} + u_{rk,it}, \tag{12}$$

where $H_{k,t}$ is a $m_k \times 1$ vector of additional factors. The system can be casted (period-wise) as

$$
\begin{pmatrix} y_{11,\cdot t} \\ \vdots \\ y_{R1,\cdot t} \\ y_{12,\cdot t} \\ \vdots \\ y_{R2,\cdot t} \\ \vdots \\ y_{1K,\cdot t} \\ \vdots \\ y_{RK,\cdot t} \end{pmatrix} = \begin{pmatrix} \Gamma_{11} & \Lambda_{11} & 0 & \cdots & 0 & \Theta_{11} & 0 & \cdots & 0 \\ \Gamma_{21} & 0 & \Lambda_{21} & \cdots & 0 & \Theta_{21} & 0 & \cdots & 0 \\ \vdots & & & \ddots & & \vdots & \vdots & & \vdots \\ \Gamma_{R1} & 0 & 0 & \cdots & \Lambda_{R1} & \Theta_{R1} & 0 & \cdots & 0 \\ \Gamma_{12} & \Lambda_{12} & 0 & \cdots & 0 & 0 & \Theta_{12} & \cdots & 0 \\ \Gamma_{22} & 0 & \Lambda_{22} & \cdots & 0 & 0 & \Theta_{22} & \cdots & 0 \\ \vdots & & & \ddots & & \vdots & \vdots & & \vdots \\ \Gamma_{R2} & 0 & 0 & \cdots & \Lambda_{R2} & 0 & \Theta_{R2} & \cdots & 0 \\ \vdots & & & \ddots & & \vdots & \vdots & & \vdots \\ \Gamma_{1K} & \Lambda_{1K} & 0 & \cdots & 0 & 0 & 0 & \cdots & \Theta_{1K} \\ \Gamma_{2K} & 0 & \Lambda_{2K} & \cdots & 0 & 0 & 0 & \cdots & \Theta_{2K} \\ \vdots & & & \ddots & & \vdots & \vdots & & \vdots \\ \Gamma_{RK} & 0 & 0 & \cdots & \Lambda_{RK} & 0 & 0 & \cdots & \Theta_{RK} \end{pmatrix} \begin{pmatrix} G_t \\ F_{1,t} \\ \vdots \\ F_{R,t} \\ H_{1,t} \\ \vdots \\ H_{K,t} \end{pmatrix} + \begin{pmatrix} u_{11,\cdot t} \\ \vdots \\ u_{R1,\cdot t} \\ \vdots \\ u_{1K,\cdot t} \\ \vdots \\ u_{RK,\cdot t} \end{pmatrix}
$$

$$
y_t = \Lambda^{**} F_t^{**} + u_t \ . \tag{13}
$$

To identify the parameters we assume that $\mathbb{E}(H_{k,t} H_{k,t}') = I_{m_k}$ as well as $\mathbb{E}(H_{k,t} G_t') = 0$ and $\mathbb{E}(H_{k,t} F_{r,t}') = 0$. The least-squares principle can be applied to estimate the factors and factor loadings, where the iteration adopts a sequential estimation of the factors G_t, $F_{1,t}, \ldots, F_{R,t}$, and $H_{1,t}, \ldots, H_{K,t}$. In what follows we focus on the sequential LS procedure which is convenient to implement. Consistent starting values can be obtained from a CCA of the relevant subfactors (see below). Let $\widehat{G}_t^{(0)}$, $\widehat{F}_{1,t}^{(0)}, \ldots, \widehat{F}_{R,t}^{(0)}$, and $\widehat{H}_{1,t}^{(0)}, \ldots, \widehat{H}_{K,t}^{(0)}$ denote the initial estimators. The elements of the loading matrices can be estimated by running regressions of $y_{rk,it}$ on the initial factor estimates $\widehat{G}_t^{(0)}$, $\widehat{F}_{1,t}^{(0)}, \ldots, \widehat{F}_{R,t}^{(0)}$, and $\widehat{H}_{1,t}^{(0)}, \ldots, \widehat{H}_{K,t}^{(0)}$. The resulting least-squares estimators for the loading coefficients are organized as in the matrix Λ^{**}, yielding the estimator $\widehat{\Lambda}^{**}$. An update of the factor estimates is obtained by running a regression of y_t on $\widehat{\Lambda}^{**}$ yielding the updated vector of factors, $\widehat{G}_t^{(1)\prime}, \widehat{F}_{1,t}^{(1)\prime}, \ldots, \widehat{F}_{R,t}^{(1)\prime}$, and $\widehat{H}_{1,t}^{(1)}, \ldots, \widehat{H}_{K,t}^{(1)}$. With this updated estimates of the factors we are able to obtain improved estimates of the loading coefficients by running again regressions of $y_{rk,it}$ on the estimated factors. This sequential LS estimation procedure continues until convergence.

The last step involves orthogonalizing the two vectors of factors $\left(\widehat{F}_{1,t}', \ldots, \widehat{F}_{R,t}' \right)$

12

and $\left(\widehat{H}'_{1,t}, \ldots, \widehat{H}'_{K,t}\right)$. Although this orthogonalization step is not necessary for identification of the factors, it enables us to perform a variance decomposition of individual variables with respect to the factors. Orthogonalizing the factors can be achieved by regressing $\left(\widehat{F}'_{1,t}, \ldots, \widehat{F}'_{R,t}\right)$ on $\left(\widehat{H}'_{1,t}, \ldots, \widehat{H}'_{K,t}\right)$ (or *vice versa*) and taking the residuals as new estimates of $\left(F'_{1,t}, \ldots, F'_{R,t}\right)$ (or of $\left(H'_{1,t}, \ldots, H'_{K,t}\right)$). We note that the results may depend on whether we regress $\left(\widehat{F}'_{1,t}, \ldots, \widehat{F}'_{R,t}\right)$ on $\left(\widehat{H}'_{1,t}, \ldots, \widehat{H}'_{K,t}\right)$ or $\left(\widehat{H}'_{1,t}, \ldots, \widehat{H}'_{K,t}\right)$ on $\left(\widehat{F}'_{1,t}, \ldots, \widehat{F}'_{R,t}\right)$.

The initialization for the three-level factor model works as follows. We first estimate the global factor as the first m_0 PCs and the global factors are eliminated from the variables by running least-squares regressions of the variables on the estimated global factors.[5] In the next step the CCA is employed to extract the common component among the $m_r + m_k$ estimated factors from *region r*, *group k* and the estimated vectors from the same *region r* but different *group k'*. This common component is the estimated regional factor. Similarly, the estimated factor $H_{k,t}$ is obtained from a CCA of the factor of region r, group k and a different region r' but the same group. These initial estimates are used to start the sequential LS procedure.

The overall estimation procedure outlined for the three-level factor model with an overlapping factor structure can be generalized straightforwardly to allow for further levels of factors (provided that the number of units in each group is sufficiently large). Furthermore, the levels may be specified as a hierarchical structure (e.g. Moench et al. (2013)), that is, the second level of factors (e.g. regions) is divided into a third level of factors (e.g. countries) such that each third level group is uniquely assigned to one second level group. For such hierarchical structures the CCA can be adapted to yield a consistent initial estimator for a sequential estimation procedure that switches between estimating the factors and (restricted) loadings.

4 Monte Carlo simulations

4.1 Two-level factor model

In this section we first examine the small sample properties of the LS estimation procedure for the two-level factor model (Section 2.2). We compare them to those of

---

[5]Alternatively, a CCA between (i) the variables in region r and group k and (ii) the variables in group r' and k' with $r \neq r'$ and $k \neq k'$ may be employed to extract the common factors. In our experience the two-step top-down estimator used in our simulation performs similarly and has the advantage that the starting values are invariant with respect to a reorganization of the levels (that is interchanging regions and groups).

the simple CCA approach (which provides us with starting values for the sequential LS approach), the two-step PC estimation procedure considered in Section 2.4.1[6] and the quasi ML approach[7]. An advantage of all these approaches is that they do not take long. By contrast, the Bayesian method requires many hours for a single estimation. Therefore, we are not able to include Bayesian methods in our Monte Carlo study, but we will compare the sequential LS estimation and the other procedures to the Bayesian approach in the first empirical application in Section 5.1.

The Monte Carlo setup is as follows. The factors are generated (independently) by a first order autoregressive process, where the autoregressive coefficient is 0.5. The idiosyncratic components are also generated independently and follow an AR(1) process with an autoregressive parameter of 0.1.[8] The innovations of the global factor(s) and the idiosyncratic errors are independently standard normally distributed. The innovations of the regional factors are also independently normally distributed; the standard deviations are set to `stdfacreg` $\in \{0.5, 1, 2\}$ in order to study the effect of the importance of the regional factors (relative to the global factor(s)). All factor loadings are generated as $N(1,1)$, following Boivin and Ng (2006). We finally multiply the idiosyncratic components by a scalar which yields idiosyncratic and common components, which are about equally important. We note that all results improve as the idiosyncratic component gets less important relative to the common component. However, the relative performances of the different methodologies remain unchanged.

We consider $R \in \{2,4\}$ regions with $n_r \in \{20, 50, 80\}$ variables in each region, and one global and one regional factor in each region. The time dimension is $T \in \{50, 200\}$. For each of the experiments we determine the R^2 (or trace R^2) of a regression of the actual on the estimated factors based on 1000 replications of the model.

From the Monte Carlo experiments presented in Table 1 it turns out that the performance of the two-step PC estimator crucially depends on the relative importance of the global and regional factors. Only if the variance of the global factor is large relative to the variance of the regional factors (`stdregfac` $= 0.5$), the two-

---

[6]We have shown in Section 2.4.1 equivalence of the top-down and the bottom-up PC approaches and therefore, only show Monte Carlo results for one of them.

[7]We are grateful to Domenico Giannone for providing us with his Matlab codes.

[8]Note that (apart from the two-step PC estimator which requires the additional assumption that $R \to \infty$) all estimators are consistent if the factors and idiosyncratic components are weakly autocorrelated and heteroscedastic.

step PC estimator yields reliable estimates for the global factors, whereas the global factors are not well estimated if the regional factors dominate (`stdregfrac` $= 2$).

The CCA estimator for the global factor is less sensitive to the relative importance of the global and regional factors and performs reasonably well for all values of `stdfacreg`. This is due to the fact that if the regional factors are more important than the global factor, the largest eigenvalue may correspond to a regional factor instead of the global factor and the two-step PC estimator may confound global with regional factors. In contrast, our two-step estimator identifies the global factors by CCA of the (standardized) factors, which does not depend on the relative importance of the factors.

While the simple CCA approach performs already well, iterations tend to lead to small improvements on average. The sequential LS estimator (which uses CCA-based starting values) produces even more realiable estimates of the global factor in sample sizes typically encountered in macroeconomic datasets. The quasi ML approach, finally, also delivers reliable global factor estimates. The average correlation between true and estimated global factors is never smaller than 0.79 and in general larger than 0.9.

In small samples, the regional factors are less precisely estimated by all methods when they are less important than the global factors. Those estimates tend to improve substantially as the sample size increases. However, as the standard deviation of the regional factors relative to the global factor increases and as the sample size grows, the regional factors are less precisely estimated with the quasi ML approach.

For $n_r = 50$, $T = 200$ and $R = 2$, we also compare the estimated density functions of the R^2 (resp. trace R^2) of the global and regional factors as well as the computing time across methods (on average over the simulations). Figure 1 shows that not only the correspondence between the factors obtained with the two-step PC approach tends to be smaller than the one obtained with the other methods, but also the variance is larger. The sequential procedures yield better factor estimates, but in few cases (for `stdfacreg`$= 2$), the quasi ML approach delivers rather inaccurate solutions.

We have also looked at the average (trace) R^2 (means and distributions) of the sequential PC approach suggested by Wang (2010) and of the sequential LS approach, where we employ the two-step PC approach to generate the starting values for the factor estimates. As expected, we obtain virtually the same results as for the sequential LS approach with CCA-based starting values and, hence, do not show them here.

Among the two methods with no iteration, the two-step PC approach is slightly faster than the CCA approach. It takes, on average over all iterations, between 0.006 and 0.007 seconds (depending on `stdfacreg`) compared to $0.008-0.009$ seconds with the CCA. Notwithstanding, the sequential LS approach with CCA starting values tends to be faster (between 0.016 and 0.021 seconds) than the sequential LS approach starting with the two-step PC approach (between 0.019 and 0.027 seconds). This suggests that, although the starting values do not seem to matter for the precision of the factor estimates, using improved (CCA-based) starting values leads to faster convergence of the algorithm. The sequential PC approach takes longer than both sequential LS approaches, especially as the regional factors become more important (between 0.04 and 0.12 seconds). The quasi ML approach is slower than the other methods. It takes between 0.30 and 1.08 seconds.

4.2 Three-level factor model

We next carry out simulations for the three-level factor model using the sequential LS approach. Third-level factors (e.g. factors specific to certain types of variables) are generated just like the regional factors. We consider `stdfacreg`$\in \{0.5, 1, 2\}$ and `stdfacvar`$\in \{0.5, 1, 2\}$ to study the importance of the regional factors and the factors specific to certain types of variables relative to the global factors, respectively. We further assume that each variable is driven by $m_0 = 1$ global factor, $m_r = 1$ regional factor and $m_k = 1$ variable type-specific factor. We consider $R = 2$ regions with $n_r \in \{20, 50, 80\}$ variables in each region and $N/2$ variables in each of the $K = 2$ groups. The time dimensions are $T \in \{50, 200\}$. Again, we multiply the idiosyncratic components by a scalar so that common components are about equally important.

Overall, our simulation results suggest that in reasonably large samples the LS approach yields very precise estimators of the factors. In small samples, global factors are also quite precisely estimated, whereas the precision of regional and variable type-specific factor estimates depends on the importance of those factors.

5 Applications

In this section, we provide three applications of our methodology to study international business and financial comovements and business cycle asymmetries in the US. The first application serves to compare the methods for estimating a two-level

factor model presented in Section 2 with the Bayesian approach. The second application makes use of the three-level factor model with an overlapping factor structure as outlined in Section 3. The third application uses the sequential LS approach for estimation of the two-level factor model (Section 2.2). But rather than considering factors which are common to specific groups of variables (as in first two applications), we consider factors which are common to specific periods.

5.1 International business cycle comovements

There is an interest reaching far back in describing and understanding the international synchronization of business cycles. Examples for key questions that have been addressed in the literature are: Does increased trade and financial integration lead to more or less synchronization of business cycles (something which is theoretically unclear)?[9] Has there been a decoupling of emerging economies from advanced economies in recent years, for instance due to regional or bilateral integration agreements or similar policies within regions and, hence, emergence of regional cycles?[10]

We basically replicate the analysis conducted by Hirata et al. (forthcoming) using the sequential LS and CCA methodologies, in comparison to their Bayesian approach (and to the two-step PC and the quasi ML approaches). From their dataset of annual consumption, investment and GDP growth for 106 countries[11], we estimate global and regional factors for the entire period 1960-2010 and separately for 1960-1984 and 1985-2010. We initially follow Hirata et al. (forthcoming) and estimate one global factor and one factor for each of seven regions (North America, Europe, Oceania, Latin America and the Carribean, Asia, Sub-Saharan Africa, Middle East and North Africa). Hirata et al. (forthcoming) also estimate country factors. We use a simplified model with no country factors given the small number of series available for each country. Nevertheless, the assumptions on the idiosyncratic components in our model are fairly flexible to account for weak correlation across variables (also within a country).

To apply the LS approach we do not need to make assumptions on the processes for the factors and idiosyncratic components nor do we need to choose priors for the parameters. When adopting the Bayesian approach we specify our model as in Hirata et al. (forthcoming) and refer to their study for details. The regional factors

---

[9]See, e.g., Kose, Otrok and Whiteman (2003), Kose, Otrok and Prasad (2008), Kose, Prasad and Terrones (2003), Kose, Prasad and Terrones (2007).

[10]See, e.g., Hirata et al. (forthcoming), Kose et al. (2008).

[11]We are grateful to Ayhan Kose for kindly sharing his dataset with us.

are normalized to be positively correlated with GDP growth in a large country in each region (here: US, Germany, Australia, Brazil, Japan, South Africa and Morocco), and the global factor is normalized to be positively correlated with US GDP growth.

Figure 2 shows the global and regional factors estimated over the entire period 1960-2010 obtained using the different methodologies. Overall, the sets of factor estimates are similar. The LS approach suggests a somewhat less severe global recession at the end of the sample than the other approaches. All approaches attribute some of the Great Recession to the global factor, but for all regions but Africa and the Middle East another important part is attributed to the regional factors. There are also some minor differences between the levels of the African factors estimated using the Bayesian and the other methodologies over parts of the sample period.

Table 2 reveals that the regional factors estimated based on the LS approach are notably correlated across regions. The highest correlations of above 0.4 in absolute terms are found for the pairs North America with Europe and with Oceania and Africa with Latin America. Most correlations are positive, some are negative, but rather small.[12]

Table 3 shows the variance decomposition of GDP growth estimated based on the sequential LS method on average over all countries in each region for the entire sample period and the two subsamples. We find that the regional factors have become more important over time in almost all regions, and in the second subsample, they are more important than the global factor. Moreover, the importance of the global factor has declined over time in most regions except for the Middle East and Africa. In the latter two regions the shares accounted for by the global factor have broadly doubled (from low shares though).

The shares explained by the common global and regional factors tend to be larger than those estimated by Hirata et al. (forthcoming). A reason might be that Hirata et al. (forthcoming) also estimate country factors while comovements among variables within a country in our approach are only implicitly accounted for by cross-correlated idiosyncratic components.

As a robustness check, we also estimated the model using the sequential LS method by allowing for two global and two regional factors. The overall common-

---

[12]We have also verified correlations between regional factors estimated with the Bayesian approach. Those are correlated to a similar extent (which is not surprising given that factor estimates are similar), although uncorrelated factors are assumed in the underlying model. The explanation is that the Bayesian approach involves overidentifying assumptions (namely that the regional factors are uncorrelated across regions), which are generally not satisfied by the *estimated factors*.

ality rises by 15 percentage points compared to the model with one global factor and one regional factor in both subsamples. A comparison between the two subsamples confirms the main result from the model with one global factor and one regional factor, i.e. that higher business cycle synchronization is due to a greater variance share explained by regional factors.[13] Hence, overall we confirm Hirata et al. (forthcoming)'s main results that the increased business cycle synchronization we have observed in the last decades is due to "regionalization" rather than to "globalization".

5.2 International financial linkages

In the second application, we broadly extend the previous analysis to financial cycles at the global level. We address the following main questions. (i) How strongly do financial variables in different countries comove? (ii) Are macroeconomic and financial dynamics at the global level driven by the same common factor(s)? Or are there (global) financial factors independent of macroeconomic factors? (iii) Is there something like a "financial cycle", i.e. do different groups of financial variables share a common factor, or are there factors specific to individual groups of financial variables? (iv) Are financial factors associated with financial developments in advanced or rather emerging economies or both?

It is far from clear what answers we should expect. While the global financial crisis affected financial markets and economic growth worldwide, other financial crises (such as the Asian crisis in 1997 or the Argentinian crisis in 1999-2002) only mainly affected the neighbouring emerging countries. Financial variables do not only move together during financial busts, but also in boom periods. For example, prior to the latest crisis, many countries experienced simultaneously housing and credit booms. The strong international comovement among financial variables can be explained with financial globalization having led to capital flows, an equalization of asset prices through arbitrage and confidence effects, and cross-border lending and global banks. Moreover, monetary policy has become increasingly similar, at least in advanced countries.[14]

---

[13]One global factor looks almost identical to the one estimated before. The other one seems to match oil price movements fairly well. It has its largest trough around the first oil price shock in 1973/74 and another deep trough around the second oil price shock in 1979/80 (there are no major troughs around the Gulf war and the war with Iraq 1991 and 2003, respectively). Factor plots and variance shares are available upon request.

[14]There has been a general change in the strategy towards inflation targeting. Central banks now tend to react to output growth and inflation which comove internationally. And recently, monetary policy was coordinated explicitly or implicitly to fight the crisis.

We broadly use the dataset built by Eickmeier, Gambacorta and Hofmann (forthcoming). It comprises overall 348 quarterly series from 11 advanced and 13 emerging market economies over 1995-2011. 207 series are financial and 141 macroeconomic series. The macroeconomic block includes, for each country (if available for a sufficiently long time span) price series (consumer prices, producer prices, GDP deflator) and output series (GDP, consumption, investment). The financial block contains stock and house prices, domestic and cross-border credit, interest rates (money market rates, long-term government bond yields), monetary aggregates M0 and M2 as well as implied stock market volatility. All series enter in year-on-year growth rates, except for interest rates and implied stock market volatility which enter in levels. Also, each series is demeaned, and its variance is normalized to one.[15]

We now apply our three-level factor model to the dataset. We estimate a "global factor" G_t, which is common to all variables in our dataset. Moreover, we estimate regional factors F_t, i.e. a factor specific to all variables in advanced countries ("advanced economies' factor") and one specific to all variables in emerging economies ("emerging economies' factor").[16] We consider only two regions because we have less countries in our sample than in the previous application.[17] Finally, we estimate variable type-specific factors H_t. It is unclear a priori how to divide the variables. Hence, we consider several (variable-wise) splits of the data leading to different models[18]:

- real activity series; price series; financial series (all other variables) (model 1)

- real activity series; price series; financial price series (comprising house and stock prices and implied volatility); financial quantities (comprising money and credit aggregates) (model 2)

- real activity series; price series; interest rates; stock prices; house prices; credit;

---

[15]The dataset used originally by Eickmeier et al. (forthcoming) also comprises lots of - less standard - US financial series as well as overnight rates and lending rates for different countries, which are not included here. Overnight and lending rates are not included in order not to give interest rates in our dataset a too large weight. Asset prices are included here, but not in the baseline model of Eickmeier et al. (forthcoming). For more details on the dataset and transformations we refer to their analysis.

[16]Those factors are normalized to be positively correlated with US GDP (global and advanced economies' factors) and GDP of Hong Kong (emerging economies' factor).

[17]Our application is an extension of Eickmeier et al. (forthcoming) who extract factors common to all financial variables and identify them as a global monetary policy factor, a global credit supply factor and a global credit demand factor, but do not consider regional factors.

[18]The variables can certainly be split also in other ways. We leave systematic assessment of the best split to future research.

monetary aggregates; implied stock market volatility.[19] (model 3)

The orthogonalization of variable type-specific and regional factors is achieved by regressing the regional factors on the variable type-specific factors.

One advantage of a finer level of disaggregation is that factors are more easily interpretable. Figure 3, hence, shows the financial factor estimates from model 3 (which are estimates conditional on the global and regional factors). The temporal evolution looks broadly plausible. The financial boom in the mid-2000s is characterized by below average interest rate and implied volatility factors and an above average stock price factor early in the boom, as well as above average credit, money and - to a less clearer extent - house price factors later in the boom. This is consistent with various explanations for the boom and subsequent crisis, including loose monetary policy (in the US and worldwide) (Taylor (2009), Hofmann and Bogdanova (2012)), the "global saving glut" (Bernanke (2005)) (which may have led to lower bond yields), strong credit growth due to deregulation on financial markets (Eickmeier et al. (forthcoming)) and major changes in the housing sector. It is interesting that the housing boom is indeed reflected somewhat in the global housing factor, even though the increase in house prices was not shared by some major emerging and advanced countries (e.g. Thailand, Malaysia, Germany, Japan and Korea) (André (2010), Ferrero (2012)). During the global financial crisis, the implied stock market volatility factor shows the greatest peak, and stock and house price factors display the deepest troughs. At the end of the sample period, we observe that the interest rate factor is still far below average, suggesting a very loose monetary policy stance. The evolution of all factors indicate sharp reversals towards improvements in financial markets, but only conditions in global stock markets seem to have fully recovered after the global financial crisis (at least temporarily).

We are now ready to answer the questions raised at the beginning of this section.

(i) Financial variables worldwide strongly comove, with variance shares explained by common factors of more than 40 percent on average over all financial variables (Table 4). The degree of synchronization among financial variables worldwide is similar to the degree of synchronization among macroeconomic variables. There is, however, a lot of heterogeneity across variables. The commonality is particularly high for fast-moving financial variables such as stock prices and interest rates and

---

[19]The factors were normalized to be positively correlated with US GDP (macro factor and real activity factor), the US GDP deflator (price factor), US stock prices (stock price factor), US house prices (house price factor), US domestic credit (credit factor), US M2 (money factor) and Chinese GDP (emerging factor), US money market rate (interest rate factor), US implied stock market volatility (implied stock market volatility factor).

considerably lower for monetary and credit aggregates as well as house prices. The finding for house prices is not surprising given that houses are not tradable and that regulation and financing in housing markets differ across countries. Interestingly, the commonality is relatively low for stock price volatility. One possible explanation is that the high observed degree of worldwide comovement of financial stress or general uncertainty, which should be reflected in the volatility series, is already captured by other common (global or regional) factors.

(ii) Macroeconomic and financial dynamics are driven by the same (global and regional) factors, which explain together more than 20 percent and roughly 30 percent of the variation in macro and financial variables, respectively. This is in line with Claessens, Kose and Terrones (2012) who illustrate strong linkages between different phases of macro and financial cycles. We find, however, that financial factors independent from macro factors also matter for financial variables, explaining between 10 and 24 percent, depending on the model. Global factors tend to be more important for financial variables than regional factors.

(iii) The overall commonality in the data (all data, but also only financial data) (i.e. the data fit) remains remarkably similar if financial variables are explained by factors specific to individual types of financial variables rather than by one single common financial factor. This is remarkable, given that we would have expected more disaggregated factors to be more highly correlated with individual series and, hence, the explained part to increase with a higher level of disaggregation. (The disaggregated financial factors in model 3 explain indeed a larger share of fluctuations in interest rates and asset prices compared to model 2, but the overall commonality does not increase (because the shares explained by the regional factors are lower in model 3 compared to model 2).) That this is not the case might suggest that it is sufficient to split the data into real activity, prices and financial variables (model 1) or real activity, prices, financial quantities and prices (model 2) (or, put differently, it might suggest existence of a "financial cycle" or a "financial quantity cycle" and a "financial price cycle") and that a finer split may not be necessary. This is useful information for modellers who study the international synchronization of financial variables.

(iv) We also find that the financial factors load highly on variables from many advanced and emerging countries simultaneously with no clear regional pattern (results are not shown, but available upon request). This underlines the global nature of financial market developments.

Our main results are broadly robust once we let the sample end before the global

financial crisis and once we alter the last estimation step and orthogonalize regional and variable type-specific factors by regressing variable type-specific on regional factors rather than regressing regional on variable type-specific factors, as before.

5.3 Asymmetries over the business cycle in the US

There exists ample evidence for asymmetries in economic dynamics over the business cycle. On the empirical side, such asymmetries are typically confirmed by studies using regime switching time series models (e.g. Peersman and Smets (2002), Goodwin (1993), Chauvet and Hamilton (2006), Nason and Tallman (2012)). Also, a general finding in the forecasting literature is that it is particularly hard to predict recessions (and turning points more generally) compared to business cycle movements during expansions[20], which also suggests that there are instabilities over the business cycle. Most recently the difficulty in predicting financial crises (which often coincide with recessions and have unique features (Reinhart and Rogoff (2011)) has been re-emphasized. On the theoretical side, agency problems between lenders and borrowers, capacity constraints or the presence of menu costs are listed as possible explanations for asymmetric dynamics over the business cycle.[21]

In this third application, we apply the two-level factor model to a large monthly US macroeconomic and financial dataset to investigate business cycle asymmetries. The dataset is taken from Stock and Watson (forthcoming). It contains real activity measures, prices, interest rates and spreads, monetary and credit aggregates and exchange rates between January 1959 and August 2011. The original dataset includes 138 variables, but we only use series which are available for the entire period and, hence, work with 108 variables. The series are transformed as in Stock and Watson (forthcoming), i.e. differences or log differences are taken if necessary to make them stationary, and they are standardized to have a zero mean and a unit variance.[22]

We aim to estimate a factor which is common to all variables at each point in time ("symmetric factor") as well as a factor which only exists in recessions

---

[20] Elliott and Timmermann (2013) on p. 297 focus on predictions around turning points, because "downturns and recoveries pose the greatest challenge for economic forecasters".

[21] Capacity constraints or the presence of menu costs (Ball and Mankiw (1994)) give rise to a convex short-run aggregate supply function and, consequently, changes in aggregate demand will have stronger effects on output and weaker effects on inflation in recessions than in expansions. Another reason for asymmetries may be agency problems between lenders and borrowers that are greater in recession (or financial crisis) than in normal times due to binding collateral constraints and increased information asymmetries between lenders and borrowers (Kiyotaki and Moore (1997), Bernanke, Gertler and Gilchrist (1999), Guerrieri and Iacoviello (2012)).

[22] See Stock and Watson (forthcoming) for details on the dataset.

("recession factor") and one which only exists in expansions ("expansion factor"). We use the NBER dating for recessions. The recession (expansion) factor is assumed not to load on the variables in the dataset during expansions (recessions). Hence, unlike in the previous application, there are (phase-specific) factors which are not common across all t rather than all i. This is equivalent to estimating a factor model with factor loadings which vary across different business cycle phases. However, the variables are also driven by a factor, movements of which have a stable impact on the variables. The empirical analysis will attribute movements in the variables to the symmetric and the phase-specific factors. In this way it will inform us about the extent economic dynamics are instable and the characteristics of the instability.

The estimation is performed in several steps. First, we apply the two-level factor model using the sequential LS methodology to the transpose of the $T \times N$ data matrix (where N equals the number of variables in the entire dataset). This yields us loadings of each variable associated with the symmetric factor, the recession factor and the expansion factor. Second, we apply a cross-section regression of our data on the estimated loadings associated with the symmetric factor and the recession (expansion) factor at each point in time during recessions (expansions) to obtain the global factor as well as the recession (expansion) factor. This delivers phase-specific and global factors which are not necessarily orthogonal. In order to carry out a variance decomposition, we (third) regress each phase-specific factor (separately for recession and expansion phases) on the global factor, and the residuals represent the new phase-specific factor estimates (which are now orthogonal to the global factor). Fourth, we re-estimate the loadings. As the loadings associated with the symmetric factor are constant over the sample period, whereas the loadings associated with the phase-specific factors differ across phases, we first regress each variable on the symmetric factor which yields us loadings associated with the symmetric factor. We then regress the difference between each series and the fitted value of that regression on the recession factor during recessions and the expansion factor during expansions. This yields us loadings associated with the phase-specific factors.

We address the following questions. (i) How large is the comovement among variables in recessions compared to expansions? (ii) How important are the phase-specific (recession and expansion) factors compared to the symmetric factor? (iii) Is the recession factor associated with different variables than the expansion factor or the symmetric factor?

(i) To measure the comovement between variables in the two phases, we compute the sum of the shares of the variances explained by the symmetric and the phase-

specific factors. On average over all variables there tends to be higher commonality during recessions than during expansions (Table 5) (see D'Agostino and Giannone (2012) and Liebermann (2012) for a similar finding). During recessions, the common factors explain 24 percent compared to 18 percent during expansions.

(ii) The phase-specific factors seem to matter. They explain 7 and 8 percent of the variation on average over all variables in expansions and recessions, respectively. This indicates that there are indeed non-negligible asymmetries over the business cycle. However, the bulk of the variation is still explained by the symmetric factor which has a time-constant impact on the variables. This holds for both expansions when the symmetric factor explains 11 percent and recessions when it explains 16 percent and suggests that the bulk of the common dynamics is stable over time.

(iii) Table 6 shows the variables with the highest 10 absolute loadings associated with the factors. The symmetric factor is mainly related to labor market variables, and the expansion factor to industrial production as well as credit spreads. The recession factor loads most highly on reserves and monetary aggregates as well as interest rates and spreads. There are several possible interpretations of this latter finding. One is that monetary and financial variables are particularly affected and comove particularly strongly in recessions. Also, financial frictions may be particularly relevant during recessions. Another possible interpretation is that recessions are strongly driven by developments in financial markets.[23] Four of the eight recessions in our sample period coincide with financial crises (following the dating for financial crises by Lopez-Salido and Nelson (2010)). As loadings can be interpreted as both effects factor movements have on certain variables and weights of variables in the factors, we cannot ultimately discriminate between the explanations.

We carry out several robustness checks. First, to assess to what extent our results are driven by the global financial crisis, we re-estimate our model for the pre-crisis period 1959-2007. The overall comovement slightly declines; the sum of the variance shares explained by the phase-specific and the symmetric factors are now at 19 and 16 percent for recessions and expansions, respectively. Otherwise, results remain very similar, and we make them available upon request.

Second, the shares explained by one symmetric factor and one phase-specific factor for recessions and expansions are, overall, quite low. Therefore, we re-estimate the model with two and three symmetric, two and three recession and two and three

---

[23]This would be in line with Del Negro and Schorfheide (2012), Faust, Gilchrist, Wright and Zakrajsek (2012) and Gilchrist and Zakrajsek (forthcoming). They find usefulness of financial frictions (Del Negro and Schorfheide (2012)) and credit spreads (all three) for forecasting output during the global financial crisis period.

expansion factors. For brevity, we only show results on variance shares explained by three (symmetric, recession and expansion) factors in Table 5. The overall commonality rises to 46 percent on average over all variables during recessions and 37 percent during expansions. Hence, the model with more factors confirms the main previous findings: variables comove more in recessions than in expansions; instabilities matter, but the symmetric factors are still more important that the phase-specific factors.

6   Concluding remarks

In this paper we have compared alternative estimation procedures for multi-level factor models which impose blocks of zero restrictions on the associated matrix of factor loadings. For the two-level factor model we have suggested an estimator based on CCA and a simple sequential LS algorithm that minimizes the total sum of squared residuals. The latter estimator is related to Wang (2010)'s sequential PC estimator and to Banbura et al. (2010)'s quasi ML approach, and it is much simpler and faster than Bayesian approaches previously employed in the literature. The sequential LS and CCA estimation approaches can be applied to block structures of two or higher levels of factors (with either overlapping or hierarchical factor structures). Monte Carlo simulations suggest that the estimators perform well (in terms of precision of factor estimates and computing time) in typical sample sizes encountered in the factor analysis of macroeconomic data sets.

We have applied the methodologies to study international comovements of business and financial cycles as well as asymmetries over the business cycle in the US. We first basically replicate the study by Hirata et al. (forthcoming) and also find that regional cycles have become more important and global cycles less important over time. Our factor estimates (based on sequential LS or CCA) and their (Bayesian) factor estimates are similar. We then move on to analyze the comovement of financial variables at the global level. We find that the estimated financial factors plausibly evolve over time. The international synchronization of financial variables is comparable to the comovement of macro variables. Both types of variables share common factors, but independent financial factors also seem to matter. Finally, we demonstrate that the sequential LS approach can be applied to estimate factors that are specific not only to certain groups of variables, but also to certain periods (such as recession and expansions). This idea is adopted in the third application. We find that there are notable asymmetries over the US business cycle, but that the bulk

of common dynamics is stable over time. The comovement of variables is larger in recessions than in expansions. Finally, the recession factor is most highly correlated to monetary and financial variables, where as expansion and symmetric factors are mostly related to real activity variables.

References

Aastveit, K., Bjoernland, H. and Thorsrud, L. (2011), 'The world is not enough! Small open economies and regional dependence', *Norges Bank Working Paper* **2011/16**.

André, C. (2010), 'A bird's eye view of OECD housing markets', *OECD Economics Department Working Paper* **746**.

Bai, J. (2003), 'Inferential theory for factor models of large dimensions', *Econometrica* **71**, 135–171.

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

Ball, L. and Mankiw, N. (1994), 'Asymmetric price adjustment and economic fluctuations', *Economic Journal* **104**, 247–261.

Banbura, M., Giannone, D. and Reichlin, L. (2010), 'Nowcasting', *Oxford Handbook of Economic Forecasting* **Michael P. Clements and David F. Hendry (editors)**.

Beck, G., Hubrich, K. and Marcellino, M. (2009), 'Regional inflation dynamics within and across euro area countries and a comparison with the US', *Economic Policy* **57**, 141–184.

Beck, G., Hubrich, K. and Marcellino, M. (2011), 'On the importance of regional and sectoral shocks for price-setting', *ECB Working Paper* **1134**.

Bernanke, B. (2005), 'The global saving glut and the U.S. current account deficit', *Speech at the Sandridge Lecture, Virginia Association of Economists, Richmond, Virginia* .

Bernanke, B., 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**(1), 387.

Bernanke, B., Gertler, M. and Gilchrist, S. (1999), 'The financial accelerator in a quantitative business cycle framework', *Handbook of Macroeconomics* **1**, 1341–1393.

Boivin, J. and Ng, S. (2006), 'Are more data always better for factor analysis?', *Journal of Econometrics* **132**(1), 169–194.

Breitung, J. and Choi, I. (2013), 'Factor models', *in: N. Hashimzade and M.A. Thornton (eds), Empirical Macroeconomics, Edward Elgar, Chapter 11* .

Breitung, J. and Pigorsch, U. (2013), 'A canonical correlation approach for selecting the number of dynamic factors', *Oxford Bulletin of Economics and Statistics* **75**, 23–36.

Breitung, J. and Tenhofen, J. (2011), 'GLS estimation of dynamic factor models', *Journal of the American Statistical Association* **106**, 1150Ű1166.

Chauvet, M. and Hamilton, J. (2006), 'Dating business cycle turning points', *Nonlinear Time Series Analysis of Business Cycles* **Milas, C., Rothman, P., van Dijk, D., Elsevier**.

Cicconi, C. (2012), 'Essays on macroeconometric short-term forecasting', *Dissertation, Université Libre de Bruxelles* **Chapter 3**.

Claessens, S., Kose, M. and Terrones, M. (2012), 'How do business and financial cycles interact?', *Journal of International Economics* **87**, 178–190.

D'Agostino, A. and Giannone, D. (2012), 'Comparing alternative predictors based on large-panel dynamic factor models', *Oxford Bulletin of Economics and Statistics* **74(2)**, 306–326.

Del Negro, M. and Schorfheide, F. (2012), 'DSGE model-based forecasting', *Handbook of Economic Forecasting* **2**.

Eickmeier, S. (2007), 'Business cycle transmission from the US to Germany–a structural factor approach', *European Economic Review* **51**(3), 521–551.

Eickmeier, S., Gambacorta, L. and Hofmann, B. (forthcoming), 'Understanding global liquidity', *European Economic Review* .

Eickmeier, S. and Hofmann, B. (2013), 'Monetary policy, housing booms and financial (im)balances', *Macroeconomic Dynamics* **17(4)**, 830–860.

Eickmeier, S. and Ziegler, C. (2008), 'How successful are dynamic factor models at forecastin output and inflation? A meta-analytic approach', *Journal of Forecasting* **27**, 237–265.

Elliott, G. and Timmermann, A. (2013), 'Handbook of economic forecasting', **2A, 1st edition, Elsevier, North Holland**.

Faust, J., Gilchrist, S., Wright, J. and Zakrajsek, E. (2012), 'Credit spreads as predictors of real-time economic activity: A Bayesian model-averaging approach', *NBER Working Paper* **16725**.

Ferrero, A. (2012), 'House price booms, current account deficits, and low interest rates', *Federal Reserve Bank of New York Staff Reports* **541**.

Francis, N., Owyang, M. and Savascin, Ö. (2012), 'An endogenously clustered factor approach to international business cycles', *Federal Reserve Bank of St. Louis Working Paper* **2012-014A**.

Gilchrist, S. and Zakrajsek, E. (forthcoming), 'Credit spreads and business cycle fluctuations', *American Economic Review* .

Goodwin, T. (1993), 'Business-cycle analysis with a Markov-Switching model', *Journal of Business and Economic Statistics* **11(3)**, 331–339.

Gregory, A. and Head, A. (1999), 'Common and country-specific fluctuations in productivity, investment, and the current account', *Journal of Monetary Economics* **44(3)**, 423–452.

Guerrieri, L. and Iacoviello, M. (2012), 'Collateral constraints and macroeconomic asymmetries', *Mimeo, Federal Reserve Board* .

Hallin, M. and Liska, R. (2011), 'Dynamic factors in the presence of blocks', *Journal of Econometrics* **163(1)**, 29–41.

Hirata, H., Kose, M. and Otrok, C. (forthcoming), 'Regionalization vs. globalization', *in: Global Interdependence, Decoupling and Recoupling, MIT Press* .

Hofmann, B. and Bogdanova, B. (2012), 'Taylor rules and monetary policy: A global "Great deviation"?', *BIS Quarterly Review* **June**, 37–49.

Kaufmann, S. and Schumacher, C. (2012), 'Finding relevant variables in sparse Bayesian factor models: Economic applications and simulation results', *Bundesbank Discussion Paper* **29/2012**.

Kiyotaki, N. and Moore, J. (1997), 'Credit cycles', *Journal of Political Economy* **105**(2), 211–248.

Kose, A., Otrok, C. and Prasad, E. (2008), 'Global business cycles: Convergence or decoupling?', *IMF Working Paper* .

Kose, M., Otrok, C. and Whiteman, C. (2003), 'International business cycles: World, region, and country-specific factors', *The American Economic Review* **93**(4), 1216–1239.

Kose, M., Prasad, E. and Terrones, M. (2003), 'How does globalization affect the synchronization of business cycles?', *The American Economic Review* **93**(2), 57–62.

Kose, M., Prasad, E. and Terrones, M. (2007), 'How does financial globalization affect risk sharing? Patterns and channels', *IMF Working Paper* **WP/07/238**.

Liebermann, J. (2012), 'Real-time forecasting in a data-rich environment', *MPRA Paper, Munich* **39452**.

Lopez-Salido, D. and Nelson, E. (2010), 'Postwar financial crises and economic recoveries in the United States', *Mimeo* .

Moench, E., Ng, S. and Potter, S. (2013), 'Dynamic hierarchical factor models', *Review of Economics and Statistics* **95(5)**, 1811–1817.

Nason, J. and Tallman, E. W. (2012), 'Business cycles and financial crises: The roles of credit supply and demand shocks', *Federal Reserve Bank of Philadelphia Working Paper* **2012-24**.

Peersman, G. and Smets, F. (2002), 'Are the effects of monetary policy in the euro area greater in recessions than in booms?', *In Mahadeva and Sinclair (eds) Monetary transmission in diverse economies* **Cambridge University Press**, 28–48.

Reinhart, C. and Rogoff, K. (2011), 'This time is different: Eight centuries of financial folly', *Princeton University Press* .

Stock, J. H. and Watson, M. W. (forthcoming), 'Disentangling the channels of the 2007-2009 recession', *Brookings Papers on Economic Activity* .

Stock, J. and Watson, M. (2002*a*), 'Forecasting using principal components from a large number of predictors', *Journal of the American Statistical Association* **97**(460), 1167–1179.

Stock, J. and Watson, M. (2002*b*), 'Macroeconomic forecasting using diffusion indexes', *Journal of Business & Economic Statistics* **20**(2), 147.

Taylor, J. B. (2009), 'Housing and monetary policy', *In: Housing, housing finance, and monetary policy. Proceedings of the Federal Reserve Bank of Kansas City Symposium in Jackson Hole, Wyoming* pp. 463–476.

Thorsrud, L. (2013), 'Global and regional business cycles. Shocks and propagations', *Norges Bank Working Paper* **2013/08**.

Wang, P. (2010), 'Large dimensional factor models with a multi-level factor structure', *Mimeo, Hong Kong University of Science and Technology* .

Table 1: Monte Carlo simulation results: R² (or trace R²) of a regression of actual on estimates factors

(a) Two-level factor model

nr	T	R	stdregfac	2-step PC		CCA		Sequential LS		Quasi ML	
				G	F	G	F	G	F	G	F
20	50	2	0.5	0.92	0.67	0.91	0.64	0.95	0.69	0.95	0.66
20	50	2	1	0.68	0.73	0.85	0.82	0.92	0.85	0.92	0.83
20	50	2	2	0.18	0.69	0.64	0.85	0.79	0.89	0.72	0.83
20	50	4	0.5	0.96	0.73	0.97	0.71	0.98	0.72	0.98	0.71
20	50	4	1	0.86	0.85	0.95	0.86	0.96	0.86	0.96	0.85
20	50	4	2	0.33	0.84	0.84	0.89	0.89	0.90	0.89	0.89
50	50	2	0.5	0.95	0.82	0.96	0.82	0.98	0.84	0.98	0.83
50	50	2	1	0.74	0.82	0.94	0.91	0.97	0.92	0.97	0.90
50	50	2	2	0.18	0.77	0.85	0.93	0.92	0.94	0.87	0.86
50	50	4	0.5	0.98	0.86	0.99	0.85	0.99	0.85	0.99	0.85
50	50	4	1	0.88	0.91	0.98	0.92	0.98	0.93	0.98	0.91
50	50	4	2	0.34	0.89	0.94	0.94	0.96	0.94	0.95	0.90
80	50	2	0.5	0.95	0.87	0.98	0.88	0.99	0.89	0.99	0.87
80	50	2	1	0.74	0.83	0.96	0.93	0.98	0.94	0.98	0.90
80	50	2	2	0.19	0.78	0.90	0.94	0.95	0.95	0.92	0.85
80	50	4	0.5	0.98	0.90	0.99	0.90	0.99	0.90	0.99	0.89
80	50	4	1	0.89	0.92	0.99	0.94	0.99	0.94	0.99	0.92
80	50	4	2	0.35	0.90	0.96	0.95	0.97	0.95	0.97	0.88
20	200	2	0.5	0.93	0.73	0.92	0.70	0.96	0.74	0.96	0.74
20	200	2	1	0.72	0.79	0.87	0.86	0.93	0.88	0.93	0.88
20	200	2	2	0.17	0.74	0.72	0.90	0.84	0.92	0.81	0.90
20	200	4	0.5	0.96	0.77	0.97	0.76	0.98	0.76	0.98	0.77
20	200	4	1	0.88	0.88	0.95	0.88	0.96	0.89	0.97	0.88
20	200	4	2	0.35	0.87	0.88	0.92	0.91	0.92	0.91	0.92
50	200	2	0.5	0.95	0.86	0.97	0.86	0.98	0.88	0.98	0.88
50	200	2	1	0.76	0.86	0.95	0.94	0.97	0.94	0.97	0.93
50	200	2	2	0.17	0.81	0.87	0.96	0.93	0.96	0.92	0.91
50	200	4	0.5	0.98	0.89	0.99	0.89	0.99	0.89	0.99	0.89
50	200	4	1	0.90	0.94	0.98	0.95	0.99	0.95	0.99	0.94
50	200	4	2	0.39	0.92	0.95	0.96	0.97	0.96	0.96	0.94
80	200	2	0.5	0.95	0.90	0.98	0.91	0.99	0.92	0.99	0.91
80	200	2	1	0.77	0.88	0.97	0.96	0.98	0.96	0.98	0.93
80	200	2	2	0.17	0.83	0.92	0.97	0.96	0.97	0.95	0.90
80	200	4	0.5	0.98	0.92	0.99	0.92	0.99	0.93	0.99	0.92
80	200	4	1	0.90	0.95	0.99	0.96	0.99	0.96	0.99	0.95
80	200	4	2	0.40	0.93	0.97	0.97	0.98	0.98	0.98	0.92

Table 1: Monte Carlo simulation results cont.

(b) Three-level factor model (LS method)

nr	T	stdregfac	stdvarfac	G	F	H	nr	T	stdregfac	stdvarfac	G	F	H
20	50	0.5	0.5	0.91	0.51	0.51	20	200	0.5	0.5	0.94	0.66	0.67
20	50	0.5	1	0.86	0.35	0.74	20	200	0.5	1	0.90	0.53	0.84
20	50	0.5	2	0.61	0.13	0.71	20	200	0.5	2	0.76	0.18	0.85
20	50	1	0.5	0.86	0.74	0.34	20	200	1	0.5	0.90	0.84	0.52
20	50	1	1	0.82	0.71	0.70	20	200	1	1	0.87	0.81	0.81
20	50	1	2	0.56	0.42	0.71	20	200	1	2	0.74	0.63	0.87
20	50	2	0.5	0.62	0.72	0.13	20	200	2	0.5	0.76	0.84	0.17
20	50	2	1	0.54	0.70	0.42	20	200	2	1	0.74	0.86	0.62
20	50	2	2	0.41	0.66	0.66	20	200	2	2	0.64	0.82	0.82
20	50	0.5	0.5	0.96	0.60	0.68	20	200	0.5	0.5	0.97	0.71	0.81
20	50	0.5	1	0.91	0.40	0.80	20	200	0.5	1	0.95	0.57	0.91
20	50	0.5	2	0.60	0.17	0.68	20	200	0.5	2	0.78	0.16	0.85
20	50	1	0.5	0.94	0.81	0.53	20	200	1	0.5	0.95	0.86	0.73
20	50	1	1	0.91	0.77	0.83	20	200	1	1	0.94	0.83	0.89
20	50	1	2	0.55	0.46	0.69	20	200	1	2	0.80	0.64	0.89
20	50	2	0.5	0.84	0.85	0.19	20	200	2	0.5	0.89	0.90	0.37
20	50	2	1	0.79	0.85	0.62	20	200	2	1	0.88	0.90	0.81
20	50	2	2	0.54	0.78	0.72	20	200	2	2	0.80	0.86	0.90
50	50	0.5	0.5	0.97	0.77	0.77	50	200	0.5	0.5	0.98	0.84	0.85
50	50	0.5	1	0.95	0.63	0.88	50	200	0.5	1	0.96	0.78	0.93
50	50	0.5	2	0.85	0.26	0.88	50	200	0.5	2	0.91	0.53	0.95
50	50	1	0.5	0.95	0.89	0.63	50	200	1	0.5	0.96	0.93	0.79
50	50	1	1	0.94	0.88	0.87	50	200	1	1	0.95	0.92	0.91
50	50	1	2	0.84	0.74	0.90	50	200	1	2	0.91	0.85	0.95
50	50	2	0.5	0.84	0.87	0.24	50	200	2	0.5	0.91	0.95	0.50
50	50	2	1	0.83	0.90	0.73	50	200	2	1	0.90	0.95	0.85
50	50	2	2	0.71	0.86	0.86	50	200	2	2	0.86	0.93	0.93
50	50	0.5	0.5	0.98	0.80	0.86	50	200	0.5	0.5	0.99	0.86	0.91
50	50	0.5	1	0.97	0.67	0.91	50	200	0.5	1	0.98	0.81	0.96
50	50	0.5	2	0.85	0.29	0.85	50	200	0.5	2	0.95	0.55	0.97
50	50	1	0.5	0.98	0.91	0.77	50	200	1	0.5	0.98	0.94	0.88
50	50	1	1	0.97	0.89	0.92	50	200	1	1	0.98	0.92	0.95
50	50	1	2	0.87	0.77	0.90	50	200	1	2	0.95	0.86	0.97
50	50	2	0.5	0.93	0.92	0.36	50	200	2	0.5	0.96	0.96	0.71
50	50	2	1	0.92	0.93	0.84	50	200	2	1	0.95	0.96	0.92
50	50	2	2	0.79	0.90	0.88	50	200	2	2	0.93	0.94	0.96
80	50	0.5	0.5	0.98	0.84	0.85	80	200	0.5	0.5	0.98	0.90	0.90
80	50	0.5	1	0.97	0.76	0.92	80	200	0.5	1	0.98	0.85	0.95
80	50	0.5	2	0.90	0.37	0.91	80	200	0.5	2	0.95	0.69	0.97
80	50	1	0.5	0.97	0.92	0.76	80	200	1	0.5	0.98	0.96	0.85
80	50	1	1	0.96	0.91	0.91	80	200	1	1	0.97	0.94	0.94
80	50	1	2	0.91	0.84	0.94	80	200	1	2	0.94	0.90	0.97
80	50	2	0.5	0.90	0.91	0.37	80	200	2	0.5	0.94	0.97	0.68
80	50	2	1	0.91	0.93	0.83	80	200	2	1	0.94	0.97	0.90
80	50	2	2	0.84	0.91	0.91	80	200	2	2	0.91	0.96	0.96
80	50	0.5	0.5	0.99	0.87	0.90	80	200	0.5	0.5	0.99	0.91	0.94
80	50	0.5	1	0.98	0.79	0.94	80	200	0.5	1	0.99	0.87	0.97
80	50	0.5	2	0.92	0.41	0.90	80	200	0.5	2	0.97	0.72	0.98
80	50	1	0.5	0.99	0.93	0.85	80	200	1	0.5	0.99	0.96	0.92
80	50	1	1	0.98	0.92	0.94	80	200	1	1	0.98	0.95	0.97
80	50	1	2	0.94	0.86	0.94	80	200	1	2	0.97	0.91	0.98
80	50	2	0.5	0.95	0.94	0.49	80	200	2	0.5	0.97	0.97	0.82
80	50	2	1	0.95	0.95	0.90	80	200	2	1	0.97	0.97	0.94
80	50	2	2	0.90	0.93	0.93	80	200	2	2	0.95	0.96	0.97

Notes: Based on 1000 replications. For details on the simulation design, see the text. G: global factor, F: regional factor, H: variable-specific factor.

Table 2: Application 1 - Correlation between regional factors 1961-2010 (sequential LS)

	(1)	(2)	(3)	(4)	(5)	(6)	(7)
North America (1)	1						
Latin America (2)	0.16	1					
Europe (3)	0.47	0.12	1				
Africa (4)	0.02	0.44	0.22	1			
Asia (5)	-0.01	0.14	0.02	-0.12	1		
Middle East (6)	-0.13	0.15	-0.28	-0.09	-0.12	1	
Oceania (7)	0.46	-0.02	0.33	0.25	0.00	-0.05	1

Table 3: Application 1 - Variance shares of GDP growth explained by global and regional factors in percent (sequential LS)

	1960-2010			1960-1984			1985-2010		
	Glob	Reg	Glob+Reg	Glob	Reg	Glob+Reg	Glob	Reg	Glob+Reg
World	15	16	31	11	19	30	9	28	37
North America	28	58	86	35	47	82	19	73	92
Latin America	12	18	31	13	25	38	7	23	30
Europe	35	19	54	18	31	49	9	52	61
Africa	7	8	15	7	9	16	12	13	24
Asia	18	23	41	13	14	27	7	38	45
Middle East	9	16	25	4	17	21	9	22	31
Oceania	8	44	52	18	43	60	13	54	67

Note: Based on 1 global factor and 1 regional factor.

Table 4: Application 2 - Variance shares explained by the common factors on average over all and over groups of variables in percent for 1995-2011 (sequential LS)

	Global	Regional	Variable-spec.	Idio
Model 1				
All variables	14	13	17	56
Advanced countries	10	12	20	58
Emerging market economies	18	14	13	54
All macro	11	10	26	53
All financial	17	15	10	58
All interest rates	33	17	9	41
All stock prices	11	11	38	40
All house prices	9	12	4	75
All credit	17	16	11	56
All money	11	19	4	66
All implied volatility	12	13	6	69
Model 2				
All variables	15	12	18	56
Advanced countries	12	10	20	58
Emerging market economies	18	13	15	54
All macro	11	9	25	55
All financial	18	13	13	56
All interest rates	38	18	5	39
All stock prices	11	18	26	45
All house prices	5	11	1	83
All credit	15	14	14	57
All money	12	11	18	59
All implied volatility	13	10	12	64
Model 3				
All variables	17	4	20	59
Advanced countries	15	5	22	58
Emerging market economies	19	3	18	60
All macro	19	4	15	62
All financial	16	4	24	57
All interest rates	7	2	40	50
All stock prices	32	1	34	33
All house prices	25	2	31	41
All credit	16	8	13	63
All money	17	4	15	64
All implied volatility	11	5	11	73

Table 5: Application 3 - Variance share explained by the factors on average over all variables in percent (sequential LS)

	Recessions	Expansions
m0 = mr = 1		
Symmetric factor	16	11
Phase-spec. factor	8	7
Symm.+phase-specific	24	18
m0 = mr = 3		
Symmetric factor	30	23
Phase-spec. factor	15	14
Symm.+phase-specific	45	37

Notes: m0: number of symmetric factors, mr: number of phase-specific (i.e. recession and expansion) factors.

Table 6: Application 3 - Variables associated with the factors exhibiting the 10 largest absolute loadings (sequential LS)

| Symmetric factor	Expansion factor	Recession factor
Unemp Rate	IP: Dur Cons. Goods	TotRes
Urate:Age>20 Men	IP: Auto	BAA Bond
BAA-GS10	IP: Consumer goods	Mbase
Urate_ST	Capu Man.	Com Paper
Emp: Private	IP: Total index	TM-6MTH
Emp:Nonfarm	IP: Final products	TB-1YR
Emp:Goods	AAA_GS10	TB-3Mth
Emp: mfg	GS10_tb3m	S&P 500
AAA_GS10	IP: Dur gds materials	AAA Bond
Urate: Age>20 Women	BAA-GS10	DJIA

Figure 1: Smoothed distributions of R² (or trace R²) of a regression of true on estimated factors (or trace R²)
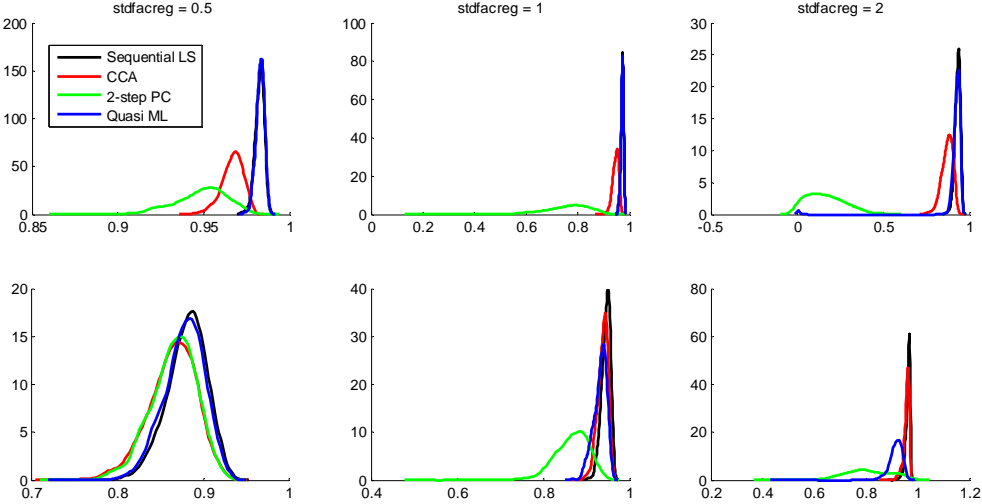
Note: For T = 200, n_g = 50, R = 2. Based on 1000 replications.

Figure 2: Application 1 - Estimates for global and regional factors of international business cycles (black: sequential LS, magenta dashed: CCA, red dotted: Bayesian, blue dashed: quasi ML, green dotted dashed: two-step PC)
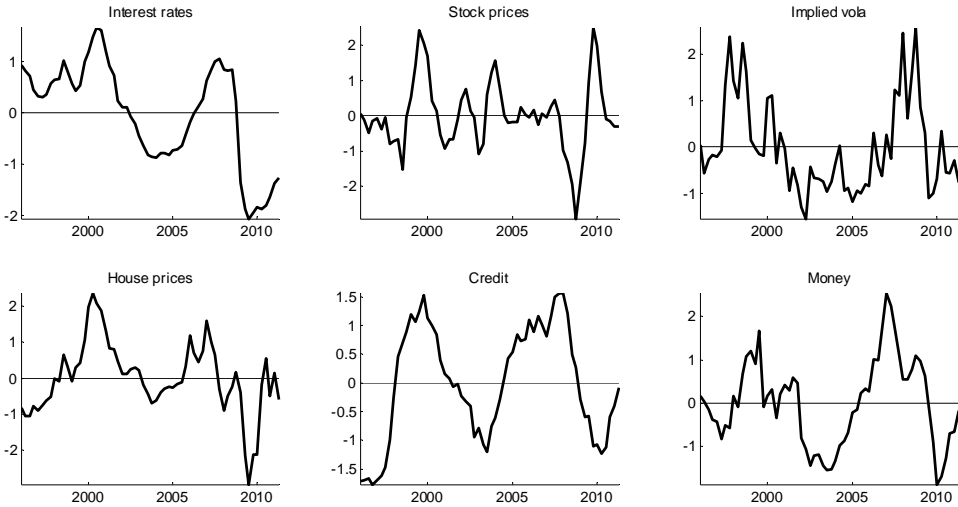
(a) Global factor
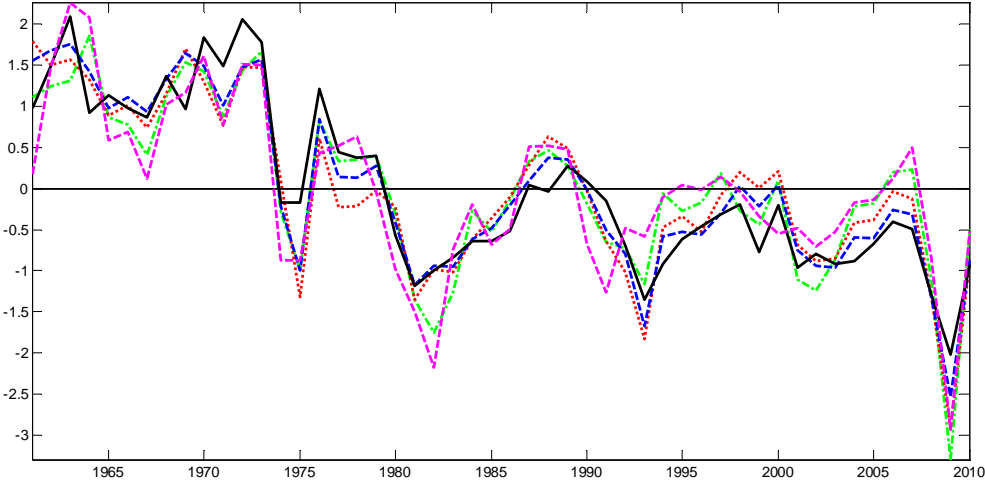
(b) Regional factors
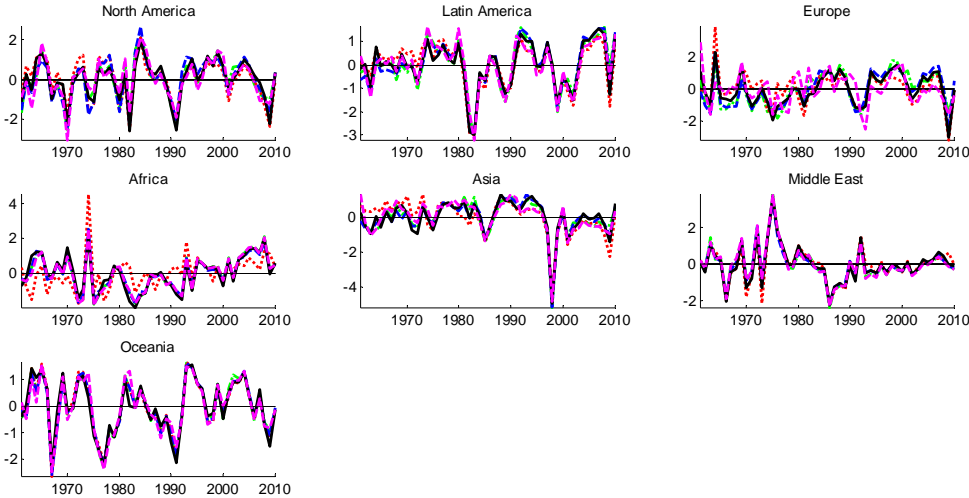
Notes: The Bayesian approach is based on 1,000 burnins and 10,000 draws (which already yields plausible results). We show the posterior means obtained with the Bayesian approach. Model with 1 global factor and 1 regional factor for each country.

Figure 3: Application 2 - Variable-specific factor estimates from model 3 (sequential LS)

Notes: The factors are normalized as described in the main text.