TI 2016-088/VI
Tinbergen Institute Discussion Paper

Is there really a Global Business Cycle? A Dynamic Factor Model with Stochastic Factor Selection

Tino Berger[1]
Lorenzo Pozzi[2]

[1] *University of Goettingen, Germany;*
[2] *Erasmus School of Economics, Erasmus University Rotterdam, and Tinbergen Institute, The Netherlands.*

Tinbergen Institute Amsterdam
Gustav Mahlerplein 117
1082 MS Amsterdam
The Netherlands
Tel.: +31(0)20 525 1600

Tinbergen Institute Rotterdam
Burg. Oudlaan 50
3062 PA Rotterdam
The Netherlands
Tel.: +31(0)10 408 8900
Fax: +31(0)10 408 9031

Is there really a global business cycle?
A dynamic factor model with stochastic factor selection

Tino Berger[1] and Lorenzo Pozzi*[2]

[1] *University of Goettingen*

[2] *Erasmus University Rotterdam & Tinbergen Institute*

October 2016

Abstract

We investigate the presence of international business cycles in macroeconomic aggregates (output, consumption, investment) using a panel of 60 countries over the period 1961 − 2014. The paper presents a Bayesian stochastic factor selection approach for dynamic factor models with predetermined factors. The literature has so far ignored model uncertainty in these models as common factors (i.e., global, regional or otherwise) are typically imposed but not tested for. We focus in particular on the existence of a global business cycle as the literature has, in our opinion unjustifiably, taken for granted its existence. In contrast to the literature, we find no evidence to support its presence.

JEL Classification: F44, C52, C32

Keywords: Global business cycle, dynamic factor model, Bayesian, model selection

*Corresponding author: Department of Economics, P.O. Box 1738, 3000 DR Rotterdam, the Netherlands. Email: pozzi@ese.eur.nl. Website: http://people.few.eur.nl/pozzi.

1 Introduction

During the past decades important changes have occurred in the international economic and financial landscape. Global trade in goods and services has steadily increased, supported by a reduction in trade barriers and transportation costs and by an increase in signed trade agreements. Financial flows have rapidly expanded across countries, supported by increased financial development, financial liberalization and financial integration. As such, countries have become more interrelated and affected by international forces. These evolutions have occurred for industrial countries, for emerging markets - which have become major players - and, to a lesser extent, for developing countries.

As a result of these developments, a large literature has investigated the extent of interdependence across countries between macroeconomic aggregates at the business cycle frequency, i.e., the occurrence and the extent of business cycle synchronization across countries. One branch of the literature has focussed on the theoretical and empirical effects of increased trade linkages and financial flows on business cycle synchronization (see e.g., Frankel and Rose, 1998; Kalemli-Ozcan et al., 2001; Kose et al., 2003b; Kose and Yi, 2006). Another branch of the literature has investigated the regional dimension of business cycle synchronization. A comprehensive overview of this literature is provided by Hirata et al. (2013) who provide information on the considered regions, countries and variables, sample period, methodology, objectives and results for about 100 studies.

A number of recent papers consider multiple regions (or other country groups) simultaneously when analyzing business cycle synchronization. Most of these studies employ structured or hierarchical dynamic factor models for analysis that are estimated using Bayesian techniques.[1] Kose et al. (2003a) estimate a factor model with one global and multiple regional common factors using data on output, consumption and investment for 60 countries over the period $1960 - 1990$. They find that the global factor is an important determinant of countries' business cycles, while regional factors play a smaller role. Mumtaz et al. (2011) estimate a factor model with a global factor and regional factors using data on output and inflation for 36 countries over a period of more than 75 years and argue that regional business cycles have become more important over time. Using a large dataset on output, consumption and investment for 106 countries over the period $1960 - 2008$, Kose et al. (2012) find that the importance of the global factor has diminished over time and that country group factors based on the development level of countries - i.e., countries are grouped according to whether they are considered industrialized, emerging or developing - have become more important since the mid 1980s. Francis et al. (2012) estimate a factor model with

[1] In these factor models a hierarchical block structure is typically imposed on the factor loadings. The factors are then identified based on the countries they load on and based on their dynamic specification. As a result, the factors can be interpreted economically as either global, i.e., when they load on all countries, or as belonging to a certain subgroup of countries, i.e., when they load only on countries in that specific subgroup.

a global factor plus a number of endogenously determined country group factors using data on output for 60 countries over the period 1960 − 2007. While they argue that regional proximity matters, their approach identifies clusters of countries that are determined not only by geographical distance but also by other characteristics such as institutions and linguistic similarity. Hirata et al. (2013) estimate a dynamic factor model that contains a global factor and factors for seven geographical regions using a dataset that consists of data on output, consumption and investment for 106 countries over the period 1960 − 2010. They find that regional factors have become more important since the mid 1980s.

A limitation of this literature is that model uncertainty is typically ignored. The common factors that are included in the estimated dynamic factor models (i.e., global, regional or otherwise) are imposed ex ante but no econometric test is conducted to find out whether these factors are actually relevant. The standard practice in dynamic factor models is to use a number of identifying restrictions that impose the common factors a priori, i.e., typically the variance of a common factor innovation is restricted to a positive number while simultaneously at least one factor loading on this factor is restricted to be positive as well. As a result, the common factors cannot possibly drop out of the model when the factor model is estimated. Francis et al. (2012) show that imposing factors in a factor model that are not actually in the data implies a misspecification and may lead to factor estimates that deviate substantially from the true model and to severe reductions in fit.

This paper deals with this limitation and contributes to the literature by proposing the estimation of a hierarchical dynamic factor model combined with a Bayesian stochastic model specification search. The model specification search amounts to a stochastic selection of the pre-specified common factors, i.e., it provides a way to investigate the relevance of the common factors included in the model. To the best of our knowledge, stochastic model selection in structured or hierarchical dynamic factor models with predetermined common factors has not been considered before.[2] The paper deals with the issue that a test for the existence of a common factor via the innovation variance of this factor or via the loadings on this factor is problematic. To this end, our method considers a dynamic factor model with standardized factors instead of scaled factors. These standardized factors have unit variance but enter the factor model multiplied by the factor innovation *standard errors*. The latter can be treated in estimation as regression coefficients, which can potentially be equal to zero in which case factors drop out of the model automatically. We then apply a Bayesian stochastic model selection procedure to our dynamic factor model. The approach used is based on the stochastic variable selection procedure of George and McCulloch (1993) for identifying non-zero regression effects in models with observed

[2]The method differs from the literature that deals with determining the optimal number of common factors in general factor models (see e.g., Hallin and Liska, 2007; Amengual and Watson, 2007; Bai and Ng, 2007) as it specifically tests for the existence of *predetermined* common factors in *hierarchical* factor models.

variables. Frühwirth-Schnatter and Wagner (2010) extend this approach to model selection in models with unobserved variables, i.e., state space models, allowing us to implement the method in the context of dynamic factor models. The approach consists of selecting the relevant common factors in the factor model by assigning binary indicators to each of the common factors in the model. We then sample these binary indicators together with the parameters of the factor model using Markov Chain Monte Carlo methods (MCMC), i.e., the Gibbs sampler. From the sampled binary indicators obtained for the common factors we calculate posterior factor inclusion probabilities - i.e., the probability that a common factor belongs in the model - by averaging the binary indicators over the iterations of the Gibbs sampler. By combining the posterior factor inclusion probabilities we then also calculate posterior model probabilities, i.e., the posterior probability of a particular *combination* of common factors.

The method presented is applied to an investigation of the presence of international common business cycles in international macroeconomic aggregates (i.e., output, consumption and investment), where the paper further contributes to the literature by investigating in particular whether a global worldwide business cycle can be identified. The reason for this focus is that the literature has been concerned with the configuration, presence and evolution of subgroups of countries (e.g., regions) but has generally taken the presence of a global business cycle for granted. For instance, even Francis et al. (2012) who do not impose factors for subgroups of countries ex ante but rather determine subgroups of countries based on an endogenous clustering method do nonetheless simply impose the global common factor in their model. However, the existence of a factor on which countries from all corners of the world load is questionable. The reason is that the commonalities that are typically considered when grouping countries that might potentially experience common business cycles (i.e., geography, institutions, development,...) cannot be invoked when considering a group of countries that potentially consists of all countries in the world. Our considered dynamic factor model hence includes a global common factor to which, in the baseline case, six regional common factors (i.e., for Europe, North America, Oceania, Latin America, Africa, Asia) are added. The factor inclusion probabilities and the factor model are estimated using data on real GDP growth, real private consumption growth and real private investment growth for a sample of 60 countries over the period $1961-2014$. Estimations are also conducted for subsamples, i.e., for the preglobalization period (1961-1985) and the globalization period (1986-2014). We investigate the robustness of the results found for the global business cycle using a battery of robustness checks among which are the use of a larger dataset consisting of 106 countries and the inclusion of common factors for country groups based not on geographical proximity but on other commonalities.

Our model selection results suggest that there is no global business cycle in the data, i.e., the posterior factor inclusion probabilities for the global common factor are negligible both in the baseline estimations and in most of the conducted robustness checks. This implies that we cannot identify a single factor on which countries of all six regions load. This finding contrasts with the literature discussed above where typically a global business cycle is identified. By contrast, our model selection procedure supports the presence of business cycles for subgroups of countries. In particular, we find evidence that supports the presence of regional cycles in the data. The posterior factor inclusion probabilities for Europe, North America, Oceania and Latin America are generally equal or close to one. For Asia, we find strong evidence of a regional business cycle in the globalization period (1986-2014) but no evidence of a regional cycle in the preglobalization period (1961-1985). The evidence to support the existence of an African cycle is mostly weak however. In line with the literature discussed above, our results also point towards increased regionalization in the globalization period, in particular for the regions Europe, North America and - evidently, given the emergence of an Asian cycle during that period - Asia. Finally, we argue that not finding a single global factor on which countries of all six regions load does not imply that regions are independent as we show that the estimated regional cycles are moderately mutually correlated.

The remainder of the paper is organized as follows. Section 2 sets up the factor model, introduces the testing procedure and provides details on estimation. Section 3 presents details on the data used in the estimations. Baseline results and robustness checks are reported in Section 4. Section 5 concludes.

2 A dynamic factor model with stochastic factor selection

We start from a standard dynamic factor model, discuss identification and the problems of testing for common factors within this framework. We then introduce an alternative specification containing standardized common factors. We discuss identification within this alternative framework and introduce the model selection approach that allows to determine which common factors belong in the model. The section then provides details on the Bayesian estimation method and the parameter priors used within this approach. The section ends with a discussion of the variance decompositions that can be applied to our factor model.

2.1 Factor model with scaled factors

2.1.1 Specification

Consider a standard dynamic factor model with multiple observed variables per country and multiple unobserved factors (see e.g., Kose et al., 2003a, 2008, 2012). More specifically, our factor model contains common factors that are common to all variables in a given country (i.e., national common factors) and common factors that are common to all countries in a given group of countries (i.e., international common factors). We denote by N the number of countries, by K the number of observed variables per country and by T the number of periods available for each of the $N \times K$ time series. We then denote by M the number of unobserved international common factors that are common across countries in a given group (and across all variables of the countries in the group). There also are N unobserved national common factors that are common across the variables in every country. As such, $M + N$ is the total number of common factors in our model (i.e., the total of international and national common factors). The dynamic factor model takes the form,

$$y_{it} = \sum_{j=1}^{M+N} \alpha_i^j F_t^j + \mu_{it}, \qquad i = 1, \ldots, K \times N, \quad t = 1, \ldots, T \tag{1}$$

where y_{it} is the observable variable (in deviation from its country-specific mean), α_i^j is the factor loading corresponding to observable i and factor j, F_t^j denotes the jth common factor - which is assumed to have innovation variance $\sigma_{\varepsilon^j}^2$ - and μ_{it} is the purely idiosyncratic "unexplained" component of y_{it}. Note that some of the loadings α_i^j are restricted to be equal to 0 as observed variables do not load on every factor, i.e., observed variables for one country do not load on the national common factor of another country while observed variables for countries that do not belong to a particular country group do not load on the international common factor of that particular country group.

2.1.2 Identification

Factor models of the type given by eq. (1) are not identified without further restrictions as neither the signs nor the scales of the factors and the factor loadings are separately identified, i.e., upon multiplying α_i^j and F_t^j in eq. (1) or α_i^j and σ_{ε^j} in eq. (2) by some constant, their product remains unaffected. The usual approach to obtain scale identification is to normalize the variance of the factor innovations to some positive constant c (where usually $c = 1$), i.e., set $\sigma_{\varepsilon^j} = c$. The usual approach to obtain sign identification is to restrict at least one of the factor loadings α_i^j on each factor j to be positive as well.

2.1.3 Testing issues

To adequately test for the presence of a common factor F_t^j, the set-up must be such that a non-existent factor F_t^j drops out of the model. This is not feasible in this framework. Testing whether all the loadings on factor j are zero is not be feasible since at least one loading is restricted to be nonzero for sign identification. Additionally, as this testing approach involves all the factor loadings α_i^j on F_t^j, testing involves a large number of parameters (i.e., $K \times N$ loadings per factor j) making it a cumbersome and possibly inconclusive method.

Alternatively, testing whether the factor innovation variance $\sigma_{\varepsilon^j}^2$ is zero is not feasible either. By obtaining scale and sign identification through a normalization on the factor loadings (e.g., impose that the average factor loadings across countries are equal to 1 for every factor), it is possible to estimate $\sigma^2_{\varepsilon^j}$. Testing whether $\sigma^2_{\varepsilon^j} = 0$ is problematic however. First, the test is non-regular from a classical point of view as the null hypothesis lies on the boundary of the parameter space.[3] Second, $\sigma^2_{\varepsilon^j}$ will be estimated using a prior typically used for a variance parameter, i.e., an inverse Gamma prior which has no probability mass at 0. Frühwirth-Schnatter and Wagner (2010), however, show that this approach tends to push the posterior distribution of the factor innovation variance away from zero if the true variance is close to or equal to 0. As such, the importance of common factors could be overestimated in this approach.

2.2 Factor model with standardized factors

2.2.1 Specification and testing

To adequately test for the inclusion or exclusion of common factors, we rewrite the factor model in terms of the standardized factors f_t^j instead of the scaled factors F_t^j,

$$y_{it} = \sum_{j=1}^{M+N} \alpha_i^j \sigma_{\varepsilon^j} f_t^j + \mu_{it}, \tag{2}$$

where $f_t^j \equiv \frac{F_t^j}{\sigma_{\varepsilon^j}}$ with σ_{ε^j} the standard deviation of the innovation ε_t^j to factor F_t^j. Note that the standardized factors f_t^j have innovations with standard deviations equal to one. As in this set-up σ_{ε^j} is treated as a regression coefficient, it can be positive, negative or zero as we discuss further below. From eq.(2) we note that a non-existent factor F_t^j for which $\sigma_{\varepsilon^j} = 0$ drops out of the model automatically. Hence, the way we rewrite our factor model provides a natural framework to test for the inclusion or exclusion of common factors.[4]

2.2.2 Identification

In the factor model with standardized factors, estimates of σ_{ε^j} are obtained to determine whether a common factor belongs in the model so that - as we discuss below - cases for which $\sigma_{\varepsilon^j} = 0$ are a possibility. Hence, the normalization $\sigma^2_{\varepsilon^j} = c$ discussed in Section 2.1 is not appropriate. We therefore impose a normalization on the factor loadings instead. In particular, for every factor j, we set the average of the corresponding factor loadings across countries and variables equal to 1, i.e., we set $\overline{\alpha}^j = 1$ (for $j = 1, ..., M+N$) where $\overline{\alpha}^j$ is the average of α_i^j across countries and variables (where loadings α_i^j which

[3] Moreover, even if $\sigma^2_{\varepsilon^j} = 0$ then F_t^j would not necessarily drop out of the model unless F_t^j were an *iid* process which would be too restrictive.

[4] This reformulation is comparable to the non-centered parameterizations of random walk processes used instead of standard random walk processes when testing for time variation in unobserved components or parameters of state space models (see Frühwirth-Schnatter and Wagner, 2010).

are equal to 0 are excluded from the average). This has the additional advantage that, unlike when normalizations are imposed on the factor innovation variances, the signs of the loadings α_i^j and of the scaled common factors F_t^j are now determined as well, i.e., we obtain scale and sign identification using the same restriction.

Importantly, however, the signs of the standardized factors f_t^j and the corresponding standard deviations σ_{ε^j} are *not* identified as it is possible to multiply both by -1 without changing their product F_t^j. As a result of this non-identification, the likelihood function is symmetric around 0 along the σ_{ε^j} dimension. When F_t^j exists ($\sigma^2_{\varepsilon^j} > 0$), the likelihood function is bimodal with modes σ_{ε^j} and $-\sigma_{\varepsilon^j}$. When F_t^j does not exist ($\sigma^2_{\varepsilon^j} = 0$), the likelihood function is unimodal around zero. As such, the non-identification of σ_{ε^j} is convenient as it provides useful information on whether a common factor F_t^j should be included in the model. In our estimation approach - fully detailed in Section 2.3 below - we fully exploit this non-identification by applying a random sign switch to f_t^j and σ_{ε^j} so as to obtain clear-cut bimodal or unimodal posterior distributions for σ_{ε^j}. The bi-or unimodality of these distributions can be considered as preliminary evidence on whether a particular common factor should or should not be included in the model.

2.2.3 Processes for common factors and idiosyncratic components

The dynamic factor model is completed by assuming stochastic laws of motion for f_t^j and μ_{it}. The common factors f_t^j are assumed to follow zero-mean $AR(q)$ processes,

$$f_t^j = \sum_{l=1}^{q} \rho^{j,l} f_{t-l}^j + \varepsilon_t^j, \qquad j = 1, \ldots, M+N \tag{3}$$

where $\varepsilon_t^j \sim iid\mathcal{N}(0,1)$, i.e., the error terms ε_t^j are iid over time and across factors. Their variances equal 1 as the common factors f_t^j are standardized factors. The idiosyncratic components μ_{it} are assumed to follow zero-mean $AR(p)$ processes,

$$\mu_{it} = \sum_{l=1}^{p} \pi_i^l \mu_{i,t-l} + \nu_{it}, \qquad i = 1, \ldots, K \times N \tag{4}$$

where $\nu_{it} \sim iid\mathcal{N}\left(0, \sigma^2_{\nu_i}\right)$, i.e., the error terms ν_{it} are iid over time and across variables and countries. The latter assumption implies that comovements in the data across countries or across variables in a given country are captured by the common factors. Further, we note that the error terms ε_t^j and ν_{it} are assumed to be mutually independent.

2.2.4 Stochastic model selection

To formally test which common factors belong in the factor model, we follow the model selection approach suggested by Frühwirth-Schnatter and Wagner (2010) and use binary indicators δ^j added to each common factor in eq. (2),

$$y_{it} = \sum_{j=1}^{M+N} \alpha_i^j \delta^j \sigma_{\varepsilon^j} f_t^j + \mu_{it}, \tag{5}$$

where δ^j is a binary indicator that is equal to 1 if F_t^j is to be included in the model (with σ_{ε^j} an unconstrained unknown parameter that is estimated from the data) and that is equal to zero if F_t^j is to be excluded from the model (with σ_{ε^j} set qual to zero). The binary indicators δ^j are used to obtain posterior inclusion probabilities for each common factor, i.e., the probability that a common factor belongs in the model. By combining the posterior factor probabilities we can then also calculate posterior model probabilities, i.e., the posterior probability of a particular *combination* of common factors.

2.3 Bayesian estimation

The standard dynamic factor model of eq. (1) could, in principle, be estimated using classical estimation techniques like maximum likelihood. Classical methods however would be extremely difficult to apply given the dimension of the problem where there are a large number of factors and unknown parameters to estimate.[5] However, the factor model presented in eq. (2) expressed in terms of standardized factors and combined with the model selection approach presented in eq. (5) implies a non-regular estimation problem for which classical methods are infeasible. Hence, our dynamic factor model is estimated using Bayesian methods. We use a Gibbs sampling approach which is a Markov Chain Monte Carlo method (MCMC) to simulate draws from the intractable joint and marginal posterior distributions of the unknown parameters and the unobserved common factors using only tractable conditional distributions. The general outline of the Gibbs sampler is presented in Section 2.3.2 below while technical details about the exact implementation of each step of the Gibbs sampler are relegated to Appendix B. First, Section 2.3.1 discusses Bayesian parameter priors however.

[5] Assume that all binary indicators δ^j are set equal to 1 so that no common factors drop out of the model. Then there are $M+N$ common factors to estimate. The number of unknown parameters to estimate equals $M+N$ for σ_{ε^j}, $(M+N)\times q$ for $\rho^{j,l}$, $N \times K$ for $\sigma^2_{\nu_i}$, $N \times K \times p$ for π_i^l, $\sum_{m=1}^{M} N_m \times K \times M$ for the non-zero loadings α_i^m on the international common factors (with $m = 1, ..., M$) and $N \times K$ for the non-zero loadings α_i^n on the national common factors (with $n = 1, ..., N$). With respect to the international common factors, the number of non-zero loadings on these factors depends on the number of countries N_m in every one of the M country groups. With respect to the national common factors, there are only K non-zero loadings for every one of these factors as the K observed variables in a country only load on their own national common factor. As an example consider the case with $N = 20$ and $M = 3$ where the first international common factor is common to all countries ($N_1 = 20$) while the other international common factors are common to two subsets of countries where each country group consists of 10 countries ($N_2 = 10$, $N_3 = 10$). Further assume that $p = q = 2$ and that $K = 3$. Then the total number of estimated parameters equals 669. Of course, if some common factors drop out of the model because $\delta^j = 0$ then the number of parameters σ_{ε^j}, $\rho^{j,l}$ and α_i^j is reduced.

2.3.1 Parameter priors

We first discuss the priors employed for our main parameters of interest, i.e., for the standard deviations of the common factors σ_{ε^j} - which are non-standard priors - and for the binary factor selection indicators δ^j. Next, we discuss the priors used for the other parameters in the factor model.

Gaussian priors centered at zero for σ_{ε^j}

As σ_{ε^j} is a regression coefficient in eqs. (2) and (5), an important advantage of our dynamic factor specification expressed in terms of *standardized* factors f_t^j (rather than scaled factors F_t^j) is that it allows us to use a Gaussian prior centered at zero on σ_{ε^j}. Centering the prior distribution at zero makes sense as, for both $\sigma^2_{\varepsilon^j} = 0$ and $\sigma^2_{\varepsilon^j} > 0$, σ_{ε^j} is symmetric around zero.[6]

We therefore impose a prior distribution for σ_ε^j given by $\mathcal{N}(0, V_0)$ where we choose a prior variance of $V_0 = 10$ which, given the data, is large enough to allow the posterior means of σ_ε^j to deviate substantially from the imposed zero prior means.

Priors for model selection

For the binary indicators δ^j that determine whether a common factor j should or should not be included in the model, we choose a Bernoulli prior distribution where each indicator has a prior probability p_0 of being included in the model, i.e., $p(\delta^j = 1) = p_0$. In our baseline scenario we set $p_0 = 0.5$ but for our main results we also report estimates obtained when assuming $p_0 = 0.25$ and $p_0 = 0.75$.[7]

Other priors

Our Bayesian estimation approach also requires choosing prior distributions for the other parameters in the model. For the factor loadings α_i^j and for the AR parameters $\rho^{j,l}$ and π_i^l we choose Gaussian prior distributions while for the variances of the innovations to the idiosyncratic components, i.e., $\sigma^2_{\nu_i}$, we use inverse Gamma (IG) distributions. In particular, we use Gaussian distributions centered at zero for the factor loadings α_i^j and for the AR parameters $\rho^{j,l}$ and π_i^l, i.e., $\mathcal{N}(0, V_0)$, with the prior variance V_0 chosen such that the prior distribution has support over the range of relevant parameter values. Hence, we set V_0 equal to one for the AR parameters $\rho^{j,l}$ and π_i^l and equal to 10 for the factor loadings α_i^j. The IG distribution for $\sigma^2_{\nu_i}$ is given by $IG(s_0T, s_0m_0T)$ with shape s_0T and scale s_0m_0T where m_0 is the prior

[6] Frühwirth-Schnatter and Wagner (2010) show that - in contrast to a posterior density for $\sigma^2_{\varepsilon^j}$ obtained when imposing a standard Inverse Gamma prior on the variance parameter $\sigma^2_{\varepsilon^j}$ - the posterior density of σ_{ε^j} is not very sensitive to the hyperparameters of the Gaussian distribution and is not pushed away from zero when $\sigma^2_{\varepsilon^j} = 0$.

[7] The reason that we check the robustness of our results to alternative priors for p_0 is that, as noted by Scott and Berger (2010), the prior choice $p_0 = 0.5$ does not provide multiplicity control for the Bayesian variable - in our case, factor - selection. When the number of possible variables is large and each of the binary indicators has a prior probability $p_0 = 0.5$ of being equal to one, the fraction of selected variables will very likely be around 0.5.

belief concerning the value of the variance $\sigma^2_{\nu_i}$ and s_0 is the strength given to this belief expressed as a fraction of the sample size T (see e.g., Bauwens et al., 2000). We set $m_0 = 10$ which is a magnitude that is in accordance with the actual variance of the data and $s_0 = 0.01$ which corresponds to a very loose prior.

2.3.2 Outline of the Gibbs sampler

For notational convenience, define the unknown parameter vector $\phi = (\delta, \sigma_\varepsilon, \alpha, \sigma^2_\nu, \rho, \pi)$ where δ, σ_ε, α, σ^2_ν, ρ and π themselves consist respectively of the stacked country-specific, variable-specific, factor-specific and lag-specific parameters δ^j, σ_{ε^j}, α^j_i, $\sigma^2_{\nu_i}$, $\rho^{j,l}$ and π^l_i (where $i = 1, ..., K \times N$, $j = 1, ..., N + M$ and $l = 1, ..., q$ for ρ or $l = 1, ..., p$ for π). Further, let f denote the stacked common factors f^j_t across j and t and let y denote the stacked data across i and t. The posterior density of interest is then given by $\Lambda(\phi, f|y)$. Given an initial value for f, the Gibbs sampling scheme is as follows:

1. Sample the parameters ϕ from the conditional distribution $\Lambda(\phi|f, y)$.

   (a) Sample the binary indicators δ from $\Lambda(\delta|\alpha, \sigma^2_\nu, \pi, f, y)$ using eq. (5) while marginalizing over the parameters σ_ε for which the factor selection is carried out. The approach follows the stochastic model specification procedure suggested by Frühwirth-Schnatter and Wagner (2010).

   (b) Sample the standard deviations of the common factors σ_ε from $\Lambda(\sigma_\varepsilon|\delta, \alpha, \sigma^2_\nu, \pi, f, y)$ using eq. (5) for those factors j for which $\delta^j = 1$. For the common factors j for which $\delta^j = 0$, set $\sigma_{\varepsilon^j} = 0$.

   (c) Sample the non-zero factor loadings α (i.e., the loadings α^j_i for those observed variables i that load on factors j) and the variances σ^2_ν jointly from $\Lambda(\alpha, \sigma^2_\nu|\delta, \sigma_\varepsilon, \pi, f, y)$ using eq. (5). Impose the normalization condition $\overline{\alpha}^j = 1$ (where the average is over i and excludes cases for which α^j_i is equal to 0).

   (d) Sample the AR parameters ρ from $\Lambda(\rho|f)$ using eq. (3). The approach follows the method of Chib and Greenberg (1994) to deal with AR terms in Bayesian regression models.

   (e) Sample the AR parameters π from $\Lambda(\pi|\delta, \sigma_\varepsilon, \alpha, \sigma^2_\nu, f, y)$ using eq. (4) where δ, σ_ε, α, f and y completely determine the idiosyncratic components μ. Again, the approach follows the method of Chib and Greenberg (1994).

2. Sample the common factors f from the conditional distribution $\Lambda(f|\phi, y)$.

   (a) The common factors that are included in the model (i.e., those for which δ^j=1) can be sampled from $\Lambda(f|\delta, \sigma_\varepsilon, \alpha, \sigma^2_\nu, \rho, \pi, y)$ using eqs.(5) and (3). A state space approach with multimove sampling is followed to estimate the common factors (see e.g., Carter and Kohn, 1994; Kim and Nelson, 1999). The common factors that are excluded from the model (i.e., those for which δ^j=0) are sampled from their prior distribution. Under the prior assumption of zero AR coefficients ρ in the common factors, this distribution is standard Gaussian, i.e., for factors j for which $\delta^j = 0$ we draw $f_t^j \sim iid\mathcal{N}(0,1)$.

(b) Perform a random sign switch on σ_{ε^j} and $\{f_t^j\}_{t=1}^T$ (for $j = 1, ..., M + N$) to exploit the non-identification of the signs of σ_{ε^j} and $\{f_t^j\}_{t=1}^T$, i.e., σ_{ε^j} and $\{f_t^j\}_{t=1}^T$ are left unchanged with probability 0.5 while with the same probability they are replaced by $-\sigma_{\varepsilon^j}$ and $\{-f_t^j\}_{t=1}^T$.

3. Calculate additional quantities such as the scaled common factors $F \equiv \sigma_\varepsilon f$ and variance shares λ obtained from a variance decomposition applied to the estimated factor model (see Section 2.4).

The initial values used for f are taken from the prior distribution which - under the prior assumption of zero AR coefficients ρ in the common factors - is given by $f_t^j \sim iid\mathcal{N}(0,1)$ (for $j = 1, ..., M + N$). Sampling from these steps is iterated D times and, after a sufficiently large number of burn-in draws B, the sequence of draws $(B + 1, ..., D)$ approximates a sample from the virtual posterior distribution $\Lambda(\phi, f|y)$. Technical details on the exact implementation of the steps of the Gibbs algorithm can be found in Appendix B. The results reported below are based on $D = 20000$ iterations with the first $B = 10000$ draws discarded as a burn-in sequence, i.e., the reported results are based on posterior distributions constructed from $D - B = 10000$ draws.

2.4 Variance decompositions

From a variance decomposition applied to eq. (5) we calculate variance shares, i.e., the fraction of the variance in y_{it} explained by factor j is given by,

$$\lambda_i^j = \frac{V(\alpha_i^j \delta^j \sigma_{\varepsilon^j} f_t^j)}{V(y_{it})} \qquad i = 1, \ldots, K \times N, \quad j = 1, \ldots, M + N \tag{6}$$

where $0 \leq \lambda_i^j \leq 1$.

As we conduct a variance analysis separately for every variable k (where $k = 1, ..., K$), we first rewrite the variance shares as $\lambda_n^{k,j}$ where $n = 1, ..., N$. If the global factor (i.e., the factor on which all countries in the sample load) is ordered first in the set of $j = 1, ..., M$ international common factors, then the variance share of the global factor in the variance of variable k in country n is given by $\lambda_n^{k,global} = \lambda_n^{k,1}$. The remaining country group factors are then given by the remaining $j = 2, ..., M$ international common factors. Since the country groups considered in this paper (e.g., regions) are such that every country belongs only to one country group, the variance share of the country group factor in the variance of variable k in country n is given by $\lambda_n^{k,group} = \lambda_n^{k,j}$ with j being the group country n belongs to (with $j \in [2, M]$). Next, the variance share of the national common factor in the variance of variable k in country n is given by $\lambda_n^{k,country} = \lambda_n^{k,M+n}$ which follows from the fact that there is one national common factor per country which for country n is factor $j = M+n$. Finally, the variance share of the idiosyncratic component in the variance of variable k in country n can be calculated as $\lambda_n^{k,idio} = 1 - \lambda_n^{k,global} - \lambda_n^{k,group} - \lambda_n^{k,country}$.

The variance shares $\lambda_n^{k,global}$, $\lambda_n^{k,group}$, $\lambda_n^{k,country}$ and $\lambda_n^{k,idio}$ can then be averaged across different countries so as to obtain the average variance share of the global factor, of the country group factor, of the national country factor and of the idiosyncratic component over all countries over which the averaging takes place. As such, we obtain the average variance shares $\overline{\lambda}_g^{k,global}$, $\overline{\lambda}_g^{k,group}$, $\overline{\lambda}_g^{k,country}$ and $\overline{\lambda}_g^{k,idio}$ where g denotes the group of countries over which we average. In this paper our groups will coincide with the groups of countries for which we define international common factors, i.e., $g = 1, ..., M$ where $g = 1$ is the group containing all countries in the sample (i.e., corresponding to the global international common factor) and where $g = 2, ..., M$ are the remaining country groups that contain countries in the sample that belong to that group (e.g., regional country groups).

3 Country groups and data

Our data are taken from the Penn-World table (PWT) version 9.0 (see Feenstra et al., 2015). Our sample consists of the $N = 60$ countries considered by Kose et al. (2003a) and more recently also by Francis et al. (2012). Following Kose et al. (2003a), the paper focusses mainly on geographical regions as country groups. Hence, we allow for a number of regional factors in addition to the global factor, i.e., Europe, North America, Oceania, Latin America, Africa, Asia. As such, we have $M = 7$. The countries considered and the six geographical regions to which they belong are reported in Appendix A. The dataset consists of annual data over the period $1960-2014$ for three variables ($K = 3$): real GDP (output), real household consumption and real private investment. All per country and per variable time series used for y_{it} are log first-differenced and demeaned (see e.g., Kose et al. (2003a) and many subsequent papers that estimate business cycles using dynamic factor models).[8] The effective sample period is therefore $1961 - 2014$ ($T = 54$). We also conduct our estimations using two separate subsamples, i.e., one for the subperiod $1961 - 1985$ and one for the subperiod $1986 - 2014$. This follows other papers such as Kose et al. (2008), Kose et al. (2012) and Hirata et al. (2013) who assume that there is a demarcation point in the mid 1980s that effectively separates the preglobalization period from the globalization period.

In the robustness checks discussed in Section 4.5 we make a further distinction between developed

[8]Instead of calculating growth rates, an alternative approach to detrend the data would be to use an explicitly filtered measure obtained for instance from a Hodrick-Prescott or bandpass filter. As noted by Canova (1998) however, business cycle facts are not robust to the use of such filtering methods.

and developing Asian countries as in Kose et al. (2003a). Appendix A reports to which of these distinct groups each of our Asian countries belongs. Additionally, in the robustness checks we also use a larger dataset consisting of the 106 countries considered by Kose et al. (2012). We refer to this paper for the exact list of countries included in this larger dataset.

In Section 4.6 alternative country groups are briefly considered as well. First, following Kose et al. (2012), we group countries according to their level of development, i.e., industrial countries, emerging market economies and developing countries. Second, we group countries according to the results obtained by the endogenous clustering method of Francis et al. (2012) who find that the 60 countries in our sample can be placed into three distinct clusters (many industrial countries; former Commonwealth countries; South America, Mexico and a few other countries). Appendix A reports to which of these groups each of our 60 countries belongs.

4 Results

In Section 4.1 we present some preliminary evidence as to which international common factors belong in the model based on the shape of the posterior distributions of the standard deviations of the innovations to these factors. In Section 4.2 we then present the formal factor and model selection results. Section 4.3 presents and discusses the estimated international common factors. Section 4.4 shows the results of applying variance decompositions to our factor model. Section 4.5 then presents a number of robustness checks. While the previous sections discuss the results of a common factor model that includes regional factors, Section 4.6 then briefly considers alternatives to country groups based on geographical proximity, i.e., we group countries based on level of development as in Kose et al. (2012) and we group countries according to the country clusters identified by Francis et al. (2012).

4.1 Preliminary evidence

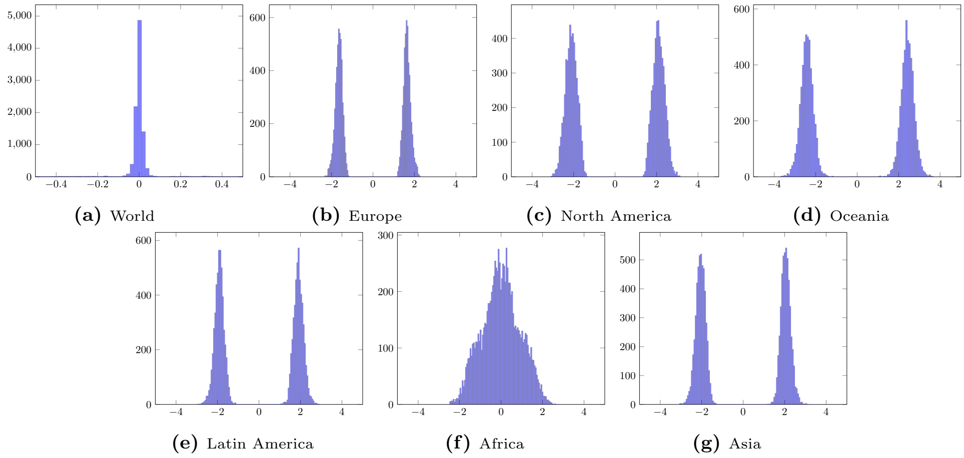
To obtain some preliminary evidence on which international common factors belong in the model, we first estimate the factor model consisting of eq. (2) (or eq. (5) with all binary indicators δ^j - for $j = 1, ..., M + N$ - set equal to 1) and eqs. (3)-(4). We note that all basic estimations are conducted using $AR(1)$ processes for all factors f_t^j and for all idiosyncratic components μ_{it}, i.e., we set $p = q = 1$. In the robustness checks reported below, we show that our conclusions do not change when assuming higher-order AR processes. In Figure 1 we present the plots of the posterior distributions of the standard deviations σ_{ε^j} of the innovations to the $M = 7$ international common factors that we consider, i.e., we have international common factors for the world and 6 geographical regions (Europe, North America, Oceania, Latin America, Africa, Asia).[9] As noted in Section 2.2.2, when the posterior distribution of a particular common factor is bimodal with low or no probability mass at zero this can be considered as evidence that this factor belongs in the model, i.e., it suggests that $\sigma^2_{\varepsilon^j} > 0$. When, on the other hand, the posterior distribution of a particular common factor is unimodal with most of its probability mass around zero this can be considered as evidence that this factor does not belong in the model, i.e., it suggests that $\sigma^2_{\varepsilon^j} = 0$. We present the distributions for the full sample period $1961 - 2014$ as well as for the subperiods $1961 - 1985$ and $1986 - 2014$.

We note, first and foremost, that for every sample period considered the reported posterior distribution of the standard deviation of the world factor innovation is unimodal with all probability mass tightly concentrated around zero. This suggests that there is no global business cycle. As for the regions, we observe clear bimodality in the posterior distributions of σ_{ε^j} for Europe, North America, Oceania, and Latin America over all considered sample periods (even though for the sample period $1986 - 2014$ Oceania has somewhat more probability mass around zero). This suggests that these regions have distinct regional business cycles and that these distinct business cycles have been present during but also before the globalization period. For Asia, the picture is slightly different as the posterior distribution of the standard deviation of its factor innovation is unimodal during the first subperiod and bimodal during the second subperiod. This suggests that Asian economies have become more integrated and synchronized during the globalization era. For Africa, the posterior distributions are unimodal with probability mass around zero for all sample periods considered suggesting that African economies are not sufficiently integrated to share a common business cycle.

[9] As in the paper we focus on international common factors, we do not present graphs for the posterior distributions of the standard deviations of the innovations to the $N = 60$ national common factors but these are available from the authors upon request.

Figure 1: Posterior distributions of the standard deviations σ_{ε^j} of the international common factor innovations for the world and 6 regions

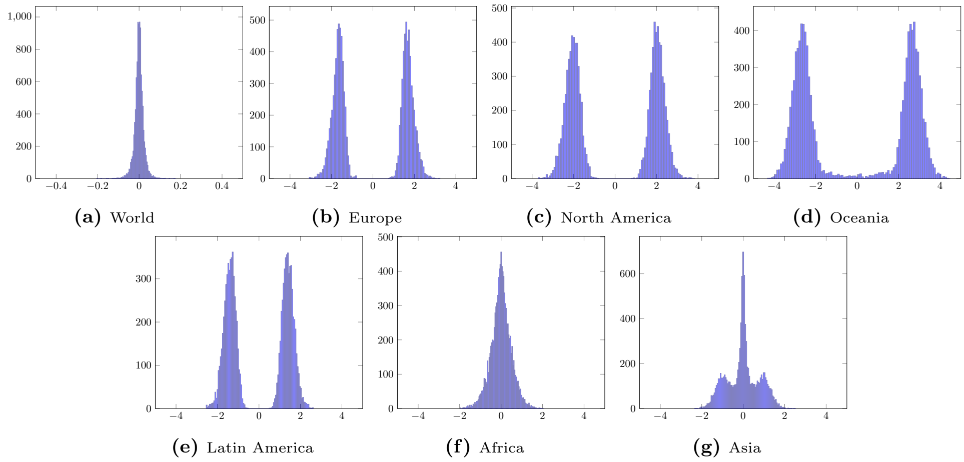
1. Sample period 1961-2014



(a) World (b) Europe (c) North America (d) Oceania

(e) Latin America (f) Africa (g) Asia

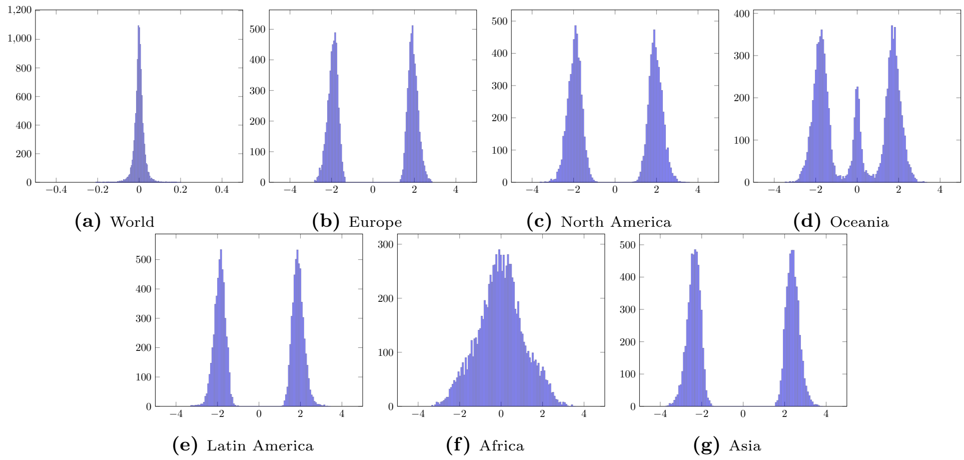
2. Sample period 1961-1985



(a) World (b) Europe (c) North America (d) Oceania

(e) Latin America (f) Africa (g) Asia

3. Sample period 1986-2014



(a) World (b) Europe (c) North America (d) Oceania

(e) Latin America (f) Africa (g) Asia

Note: The posterior distributions are obtained with binary indicators δ^j set to 1 (for $j = 1, ..., M + N$) in eq. (5).

4.2 Model selection

To test more formally which common factors belong in the factor model, we next sample the stochastic binary indicators δ^j in eq. (5) (for $j = 1, ..., M + N$) together with the other parameters in the model.[10] Table 1 presents the posterior inclusion probabilities of the international common factors for the world and the 6 regions that we consider. These probabilities are calculated as the average of the sampled binary indicators δ^j over all iterations of the Gibbs sampler. The probabilities are reported for the full sample period (1961-2014) and for both subsamples (1961-1985 and 1986-2014) as well as for different prior factor inclusion probabilities, i.e., for the baseline case $p_0 = 0.5$ but also for $p_0 = 0.75$ and for $p_0 = 0.25$.

From the table we note that, in line with the results reported in Section 4.1, the inclusion probability of the world factor is very low. It varies between 1% and 9% depending on the assumed prior p_0 but shows no difference between the subperiods $1961 - 1985$ and $1986 - 2014$. As such, our formal testing procedure also suggests that there is no global business cycle in the data, a result that contrasts with the literature - as discussed in Section 1 - where a global factor is typically found to be relevant. In contrast to our approach however, the literature typically imposes a global factor in the model - using identifying restrictions that prevent it to drop out of the model - without testing for it. Not finding a common factor on which countries from all six regions load may not be that surprising since - unlike for subgroups of countries such as regions - for a group consisting of all countries in the world, it is hard to invoke characteristics shared by all countries (such as geographical proximity, level op development, institutions) that might give rise to such a global cycle. The regions Europe, North America and Latin America do clearly command distinct cycles as the posterior inclusion probabilities of their factors equal one over all priors and periods considered. Oceania commands its own cycle even though the evidence in favor of this is slightly weaker in the second subperiod $1986 - 2014$. Asia, on the other hand, has posterior inclusion probabilities equal to one over the full sample period and over the second subperiod $1986 - 2014$ but, again in line with the results reported in Section 4.1, has much lower probabilities of having its own distinct factor in the first subperiod $1961 - 1985$. A regional business cycle seems to have emerged in Asia only during the globalization period. For Africa the evidence is less conclusive. The posterior inclusion probabilities of an African factor are somewhat larger than the corresponding prior probabilities p_0 over the full sample period and in the second subperiod $1986 - 2014$ but this hardly constitutes strong evidence to support the existence of a common African business cycle during the globalization period.

[10] The results do not change if, instead, we only sample the binary indicators for the international common factors, i.e., δ^j for $j = 1, ..., M$, and fix to 1 the binary indicators of the national common factors, i.e., δ^j for $j = M + 1, ..., M + N$.

Table 1: Posterior inclusion probabilities of international common factors (world and 6 regions) over different priors and sample periods

Period	Prior	Posterior factor inclusion probabilities						
		World	Europe	No. Am.	Oceania	Lat. Am.	Africa	Asia
1961-2014	$p_0 = 0.50$	0.03	1.00	1.00	1.00	1.00	0.59	1.00
	$p_0 = 0.25$	0.01	1.00	1.00	1.00	1.00	0.34	1.00
	$p_0 = 0.75$	0.09	1.00	1.00	1.00	1.00	0.86	1.00
1961-1985	$p_0 = 0.50$	0.04	1.00	1.00	0.99	1.00	0.43	0.37
	$p_0 = 0.25$	0.02	1.00	1.00	0.97	1.00	0.21	0.25
	$p_0 = 0.75$	0.09	1.00	1.00	0.98	1.00	0.65	0.66
1986-2014	$p_0 = 0.50$	0.04	1.00	1.00	0.81	1.00	0.61	1.00
	$p_0 = 0.25$	0.01	1.00	1.00	0.34	1.00	0.36	1.00
	$p_0 = 0.75$	0.09	1.00	1.00	0.88	1.00	0.79	1.00

Note: The reported probabilities are calculated as the average of the binary indicators δ^j (for $j = 1, ..., M$) over the 10000 iterations of the Gibbs sampler.

Besides inclusion probabilities of the individual common factors, the model selection search also allows to compute overall model probabilities, i.e., probabilities for combinations of common factors. As there are $M+N = 7+60 = 67$ binary indicators - 1 for each of the common factors (international and national) - there are 2^{67} possible models. In ranking the models from most to least preferred, we do not take into account the national common factors however so that only 2^7 models are effectively ranked. In Table 2 we report the 4 most preferred models out of 2^7 for the full sample period and both subsamples where the ranking is based on results obtained under the prior factor inclusion probability $p_0 = 0.5$. The table also presents the model probabilities of these four models under the alternative priors $p_0 = 0.25$ and $p_0 = 0.75$.

The results in the table show that a factor model without a global factor but with six regional factors included is the preferred model over the full sample period in the baseline case when $p_0 = 0.5$. A model without global factor and without the African factor is ranked second. This ranking holds up also when $p_0 = 0.75$ but is reversed when $p_0 = 0.25$. When looking at the preglobalization period $1961 - 1985$, we notice that the preferred models under $p_0 = 0.5$ and $p_0 = 0.25$ exclude the global factor as well as the African and Asian factors. The absence of African and Asian factors during this period is in accordance with the results presented in Figure 1 and with the low posterior inclusion probabilities for these factors reported in Table 1. For $p_0 = 0.75$, Table 1 reports a posterior probability of 0.65 for the African factor and 0.66 for the Asian factor and this is reflected in a preferred model that includes both these factors. For the globalization period $1986 - 2014$, the models ranked first and second are identical to those reported for the full sample period when $p_0 = 0.5$ and $p_0 = 0.75$. The preferred model is the one with six regional factors and no global factor which is followed by the model with neither a world nor an African factor included. For $p_0 = 0.25$ the preferred model also excludes Oceania. This result corroborates the findings reported above, i.e., the low posterior inclusion probability of the factor for Oceania reported in Table 1 for period $1986-2014$ and prior $p_0 = 0.25$ and a reasonable amount of probability mass concentrated at 0 in the posterior distribution of σ_{ε^j} for Oceania over this period. To summarize, we note that while some regional factors are always unambiguously included (i.e., Europe, North America, Latin America) and the inclusion of other regional factors depends on period and/or prior considered (i.e., Oceania, Africa, Asia) the global factor is in all periods and for all priors convincingly excluded from the model.

Table 2: Posterior model probabilities of the four preferred models over different priors and sample periods

Period	Model							Posterior model probability		
	World	Europe	No. Am.	Oceania	Lat. Am.	Africa	Asia	$p_0 = 0.50$	$p_0 = 0.25$	$p_0 = 0.75$
1961-2014	0	1	1	1	1	1	1	0.57	0.33	0.78
	0	1	1	1	1	0	1	0.40	0.65	0.13
	1	1	1	1	1	1	1	0.02	0.00	0.08
	1	1	1	1	1	0	1	0.01	0.01	0.01
1961-1985	0	1	1	1	1	0	0	0.34	0.56	0.11
	0	1	1	1	1	1	0	0.25	0.15	0.20
	0	1	1	1	1	0	1	0.21	0.19	0.20
	0	1	1	1	1	1	1	0.15	0.05	0.38
1986-2014	0	1	1	1	1	1	1	0.48	0.12	0.63
	0	1	1	1	1	0	1	0.31	0.21	0.17
	0	1	1	0	1	1	1	0.11	0.24	0.09
	0	1	1	0	1	0	1	0.07	0.41	0.02

Note: The ordering of models from most preferred to least preferred is on the basis of posterior model probabilities obtained under prior $p_0 = 0.5$.

4.3 Estimated international common factors

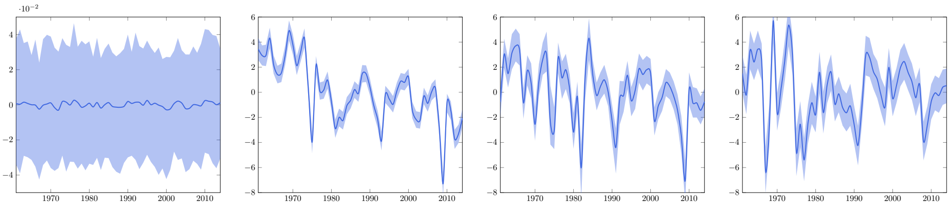
We present the scaled common international factors $F_t^j = \sigma_{\varepsilon^j} f_t^j$ for the world and our six regions in Figure 2. The top panel shows the global and regional factors obtained from estimation of the factor model over the full sample period $1961-2014$. The bottom panel shows the global and regional factors obtained from estimation of the factor model over the the subperiods $1961-1985$ and $1986-2014$ where a vertical bar separates both subperiods. Reported are the means of the posterior distributions of the scaled common factors and the 90% highest posterior density (HPD) intervals, i.e., the 5% and 95% percentiles of these distributions.

Both panels clearly show that, in line with the findings reported in the previous sections, there is no global common factor. For the world factor the HPD is very wide in both panels of the figure with the mean factor itself being almost constant and equal to zero. Further, the evidence to support the existence of an African regional factor is not very strong. The HPD is also rather wide and fluctuations in the mean of the factor are limited.[11] This result for Africa is in line with the inconclusive results reported earlier, i.e., a unimodal posterior distribution for the standard deviation of the African factor innovation (see Section 4.1) and relatively low posterior inclusion probabilities for this factor which, nonetheless, are high enough for models that include the African factor to be ranked favourable in a number of instances (see Section 4.2). The results for the other regions are - again in line with the previous sections - much more convincing. For Asia there is a clear difference between the factor estimates obtained over the full sample period and the factor estimates obtained from both subperiods. The reason is that the evidence reported in Sections 4.1 and 4.2 suggests that the Asian factor only belongs in the model in the globalization period (1986-2014) and not in the preglobalization period (1961-1985). The factor estimates in the top panel are based on an estimated standard deviation for the factor innovation obtained from the full sample period and hence overestimate the true common cyclical fluctuations in the first subperiod. Hence, the bottom panel provides a more reliable representation of the Asian business cycle. The estimated factors for Europe, North America, Oceania, Latin America, and Asia show a number of distinct business cycle episodes but also a number of common business cycle episodes.

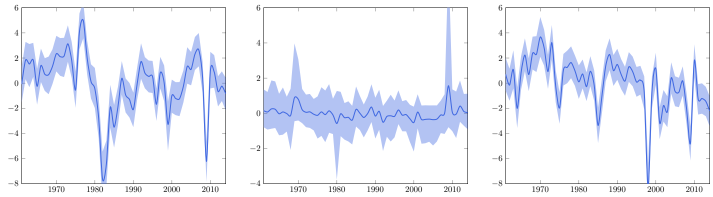
[11] Interestingly, whereas the other regional factors show a big drop during the Great Recession - in particular in the year 2009 - the African regional factor seems to be insulated from the Great Recession (see e.g., Sayeh, 2012) and even shows a modest upward spike in 2009. The latter may be caused by the positive and relatively high average growth rates of - in particular output and consumption - of 6 out of the 7 African countries in our sample that year. Only South Africa experienced a negative growth rate in output and consumption that year while the other countries' growth rates were all - sometimes a lot - higher than 2%.

Figure 2: Scaled international common factors $\sigma_{\varepsilon^j} f_t^j$ for the world and 6 regions

1. Full sample period 1961-2014

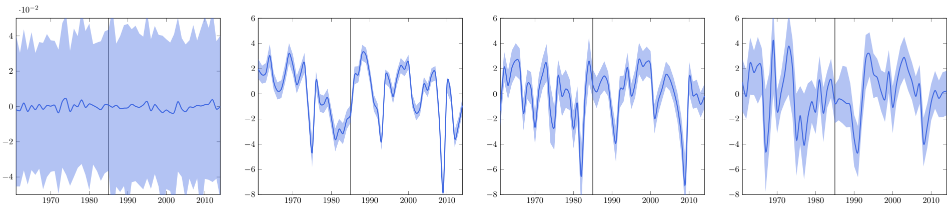


(a) World **(b)** Europe **(c)** North America **(d)** Oceania

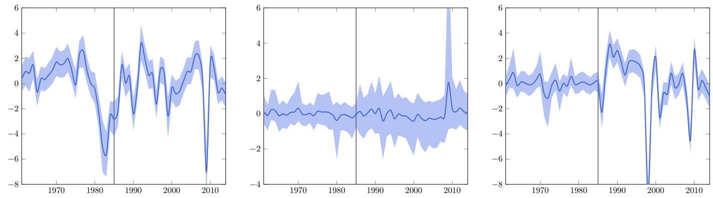


(e) Latin America **(f)** Africa **(g)** Asia

2. Subsamples 1961-1985 and 1986-2014 (separated by a vertical bar)



(a) World **(b)** Europe **(c)** North America **(d)** Oceania



(e) Latin America **(f)** Africa **(g)** Asia

Note: Reported are the mean, the 5th and the 95th percentile of the posterior distribution of the scaled international common factors $F_t^j = \sigma_{\varepsilon^j} f_t^j$ as obtained over the draws of Gibbs sampler. Draws f_t^j and σ_{ε^j} for which $\delta^j = 0$ are excluded from the calculation of the posterior distribution.

With respect to the former, we can observe the Latin American debt crisis of the early 1980s and the Asian financial crisis of the late 1990s. With respect to the latter, we can observe the 1973-74 oil crisis in Europe and North America, the early 1990s recession in Europe, North America and Oceania, the burst of the dot com bubble in 2001 in North America and Europe and the 2007-09 financial crisis and Great Recession in Europe, North America, Oceania, Latin America, and Asia.

As such, not finding one global factor on which countries of all six considered regions load *does not* imply that regions are independent. The reason is that, while the common factors are modeled as independent factors, the estimated common factors are not orthogonal. When considering the regional common factors that are found to be relevant - i.e., those for Europe, North America, Oceania, Latin America and Asia - the highest correlation we measure is 0.57 between the European and North American factors while the lowest correlation is 0.05 between the factors for Oceania and Latin America.

4.4 Variance decompositions

In Table 3 we present the results of applying variance decompositions to our estimated factor model. The table reports the contribution of the world, region, country and idiosyncratic factors to the variance of output, consumption and investment growth for the world as a whole and for the regions considered (i.e., Europe, North America, Oceania, Latin America, Africa, Asia). Results are reported for the full sample period $1961-2014$ and for both subperiods $1961-1985$ and $1986-2014$. We refer to Section 2.4 for the details on the exact calculation of the variance shares. The table presents the medians of the posterior distributions of the variance shares and are expressed in %.

A number of conclusions can be drawn from the table. First, in line with the low posterior inclusion probabilities found for the global factor, the variance shares of the world factor are consistently found to be negligible. Second, with the exception of Africa, the variance shares of the regional factor are always above 10% - and often well above this number - no matter which country group, variable or period that we consider. For Africa the variance share of the regional factor lies between 1% and 5% which is in line with the weak support reported in previous sections for the existence of an African factor in the data. Third, upon comparing both subperiods, there is evidence of increased regionalization, i.e., an increased variance share for the regional factor in the globalization period compared to the preglobalization period. This phenomenon has been documented in the literature (see e.g., Hirata et al., 2013). Regionalization occurs globally as the increased variance share of the regional factor can be observed for the world as a whole. We find that the regions affected are Europe, North America and especially in Asia where - as noted in the previous sections - there is little or no evidence in support of the presence of a regional factor during the preglobalization period whereas the posterior inclusion probabilities of the Asian factor are equal to one for the globalization period. Fourth, country-specific components - i.e., the country-specific common factors and idiosyncratic components - are still responsible for at least 40% (and often a lot more) of the variance of output, consumption and investment growth in all periods and regions considered.

Table 3: Variance decompositions

Period	Variable	Factor	Country group						
			World	Europe	No. Am.	Oceania	Lat. Am.	Africa	Asia
1961-2014	Output	World	0.02	0.01	0.00	0.01	0.02	0.02	0.02
		Region	28.05	48.30	52.99	44.44	15.48	4.12	20.61
		Country	38.61	32.87	30.53	26.57	43.33	39.22	42.37
		Idiosyncratic	33.02	18.61	16.14	28.49	40.94	54.96	36.63
	Consumption	World	0.02	0.02	0.01	0.02	0.02	0.03	0.02
		Region	18.75	26.87	37.64	39.64	12.96	4.17	14.34
		Country	33.56	26.75	32.11	6.74	38.89	34.23	37.94
		Idiosyncratic	47.36	46.08	28.91	51.83	47.89	59.87	47.19
	Investment	World	0.03	0.02	0.01	0.01	0.03	0.04	0.03
		Region	18.31	28.62	43.04	31.11	10.79	2.62	13.98
		Country	27.80	35.43	31.41	41.84	25.40	15.48	23.10
		Idiosyncratic	53.59	35.51	24.58	26.41	63.42	80.78	62.53
1961-1985	Output	World	0.03	0.02	0.01	0.01	0.03	0.04	0.03
		Region	21.79	36.16	47.50	42.20	19.20	1.86	3.74
		Country	35.23	33.15	36.63	27.62	30.33	25.66	49.37
		Idiosyncratic	42.72	30.35	15.89	30.55	50.28	69.85	45.51
	Consumption	World	0.04	0.03	0.01	0.02	0.04	0.05	0.04
		Region	15.88	24.36	29.88	40.20	15.14	1.99	2.68
		Country	31.02	26.40	37.60	6.40	27.13	24.34	45.51
		Idiosyncratic	52.80	48.80	31.41	50.36	57.33	71.34	50.90
	Investment	World	0.04	0.03	0.01	0.01	0.05	0.05	0.05
		Region	14.23	22.41	36.78	32.81	10.33	2.15	3.92
		Country	26.59	34.30	32.53	39.15	20.71	12.92	27.31
		Idiosyncratic	58.92	42.93	30.24	28.54	68.69	83.08	68.18
1986-2014	Output	World	0.02	0.01	0.01	0.01	0.02	0.02	0.02
		Region	30.50	51.52	52.72	34.71	13.27	3.02	33.06
		Country	32.73	23.80	27.24	28.41	41.38	39.39	29.31
		Idiosyncratic	36.60	24.65	19.59	35.07	45.07	55.30	37.29
	Consumption	World	0.03	0.02	0.01	0.02	0.03	0.03	0.03
		Region	20.43	24.35	39.21	31.50	14.59	3.06	24.96
		Country	31.70	28.73	27.72	17.12	37.18	37.99	24.26
		Idiosyncratic	47.67	46.65	32.18	48.54	47.83	56.77	50.33
	Investment	World	0.03	0.02	0.01	0.01	0.03	0.04	0.02
		Region	26.24	42.57	48.78	34.83	14.65	3.42	24.44
		Country	24.72	21.36	28.70	25.02	29.93	14.22	25.61
		Idiosyncratic	48.83	35.87	22.36	39.17	55.07	81.24	49.58

Note: Reported are the medians of the posterior distributions of the variance shares (expressed in %) as discussed in Section 2.4. The posterior distributions are obtained from estimating eq. (2) or eq. (5) with all binary indicators δ^j set to 1 (i.e., for $j = 1, ..., M + N$).

4.5 Robustness checks

In this section, we check the robustness of the results reported in Section 4.2 regarding the inclusion or exclusion of the global factor and the six regional factors (Europe, North America, Oceania, Latin America, Africa, Asia) over the full sample period and over the preglobalization and globalization periods. We consider four checks which are reported in Table 4. The reported results assume a prior factor inclusion probability equal to $p_0 = 0.5$ but they are robust to the use of alternative priors (i.e., $p_0 = 0.25$ or $p_0 = 0.75$). First, we distinguish between developed and developing Asian economies as in Kose et al. (2003a). The results are very similar to those reported in Table 1 for $p_0 = 0.5$. Most importantly, there is no evidence of a world factor in the data. Interestingly, the distinction between developed and developing Asian economies is meaningful as - over all periods considered - only the former group seems to command a distinct regional factor. Second, we assume AR(2) processes for all factors f_t^j and idiosyncratic components μ_{it} instead of AR(1) factors, i.e., $p = q = 2$. When comparing these results to those reported in Table 1, we note that the estimation results obtained with AR(2) processes are almost identical to those obtained with AR(1) processes.[12]

Table 4: Posterior inclusion probabilities of international common factors (world and regions): robustness checks

Check	Period	Posterior factor inclusion probabilities						
		World	Europe	No. Am.	Oceania	Lat. Am.	Africa	Asia
Asia develop.	1961-2014	0.02	1.00	1.00	1.00	1.00	0.67	1.00^a, 0.82^b
	1961-1985	0.04	1.00	1.00	0.97	0.98	0.41	0.85^a, 0.52^b
	1986-2014	0.03	1.00	1.00	0.84	0.96	0.58	1.00^a, 0.16^b
AR(2)	1961-2014	0.03	1.00	1.00	1.00	1.00	0.61	1.00
	1961-1985	0.04	1.00	1.00	1.00	1.00	0.43	0.40
	1986-2014	0.03	1.00	1.00	1.00	1.00	0.62	1.00
106 countries	1961-2014	0.03	1.00	1.00	1.00	1.00	0.18	1.00
	1961-1985	0.04	1.00	1.00	0.93	1.00	0.47	0.16
	1986-2014	0.09	1.00	1.00	0.64	1.00	0.25	1.00
Output only	1961-2014	1.00	0.03	0.99	0.37	1.00	0.26	1.00
	1961-1985	0.03	1.00	0.99	0.46	1.00	0.30	0.20
	1986-2014	0.02	1.00	0.98	0.25	1.00	0.24	1.00

Note: All reported results are based on prior $p_0 = 0.5$ but they are robust to the use of alternative priors $p_0 = 0.25$ or $p_0 = 0.75$. "Asia development" is a check where a distinction is made between developed Asian economies and developing Asian economies resulting in $M = 8$ international common factors instead of 7 with ^{a} denoting the results for the group of developed Asian countries and ^{b} denoting the results for group of developing Asian countries. "AR(2)" is a check where the dynamic factor model is estimated with AR(2) processes assumed for all common factors f_t^j and all idiosyncratic components μ_{it}, i.e., for $p = q = 2$. "106 countries" is a check where the sample consists of $N = 106$ countries that can belong to either one of our 6 regions instead of the 60 countries reported in Appendix A. We refer to Kose et al. (2012) or Hirata et al. (2013) for the list of countries included in this extended sample. "Output only" is a check whether the factor model is estimated using only $K = 1$ variable - i.e., output - instead of 3, i.e., output, consumption, investment.

[12] When inspecting whether the added second lags are different from 0 - in the sense of not finding the value of 0 in the $5\% - 95\%$ interval of their posterior distribution - we find that this is almost never the case.

Third, we consider a larger dataset consisting of 106 countries instead of the 60 countries considered until now. We refer to Kose et al. (2012) for the countries included in this extended dataset. Again, the results reported in Table 1 are largely confirmed. We find no evidence of a global factor. The evidence in favor of an African factor is even weaker in this extended setting however which can be due to the Africa group containing a larger amount of countries compared to the 60 country dataset. Fourth, we estimate the factor model given by eqs. (5), (3) and (4) using only one variable (i.e., output) instead of three (i.e., output, consumption, investment). This setting implies that the national factors and the idiosyncratic components coincide and that detected international cyclical fluctuations are purely output-based. For the full sample period $1961 - 2014$, we find a posterior inclusion probability for the global factor that is equals to one. This result disappears when we look at both subperiods $1961 - 1985$ and $1986 - 2014$ separately however. Since there have been important shifts between subperiods - i.e., the Asian group has no distinct factor during the period $1961 - 1985$ but commands its own factor during the period $1986 - 2014$ - the full sample period results are not very reliable. As such, again, there is no strong evidence in favor of the existence of a global factor. The results for the regional factors are similar to those reported for the three variable case in Table 1 with two exceptions. First, the inclusion probabilities for Oceania are now lower than 0.5 instead of between 0.8 and 1 suggesting that there is no regional business cycle for Oceania if only output is considered. Second, the evidence in favor of an African factor is considerable weaker compared to the three variable case, i.e., there is no more evidence of an African business cycle in the globalization period if only output is considered.

4.6 Alternative country groups

The estimations so far have been based on a dynamic factor model with geographical regions as country groups. In the literature, other country groups have been considered. We discuss two alternative country groups. First, we consider the country groups based on level op development as in Kose et al. (2012) where each country in our 60 country sample belongs to one of three groups: industrial countries (IND), emerging economies (EME) or developing countries (DEV). Second, we consider the country groups endogenously determined by the clustering method of Francis et al. (2012) where each country in our 60 country sample belongs to one of three clusters: many industrialized economies (CL1), the UK and the former Commonwealth countries (CL2) and South America, Mexico and a few other countries (CL3). We refer to Appendix A for the exact composition of the groups IND, EME, DEV, CL1, CL2 and CL3. We investigate whether the international common factors that are considered in these settings - i.e., in particular the global factor - should be included in the model based on our stochastic factor selection approach. The posterior inclusion probabilities of the international common factors in these alternative settings are reported in Table 5. For the factor model containing country groups based on level op development we report, in line with the literature, results for a model containing the three variables considered above (i.e., output, consumption, investment) while for the factor model containing endogenously determined clusters we report both results obtained with one variable (i.e., output only) - which is the framework used by Francis et al. (2012) to determine the clusters - and results obtained with three variables. The results reported in the table, again, suggest that there is no world factor and hence no global business cycle. While there seems to be evidence in favour of a world factor when considering a factor model with development groups based on the full sample period $1961 - 2014$, this result vanishes when looking at both subperiods $1961 - 1985$ and $1986 - 2014$ separately. Since there have been important shifts between subperiods - i.e., the EME group has no distinct factor during the period $1961 - 1985$ but does command its own factor during the period $1986 - 2014$ - the full sample period results are unreliable. With respect to the endogenous clusters, we find that, when considering a factor model for only output over the full sample period, the three clusters identified by Francis et al. (2012) are also included by our approach. This is not surprising as this setting is exactly the setting used by Francis et al. (2012) to determine the clusters. The difference is however that the world factor should not be included in the factor model according to our results whereas it is simply imposed - i.e., not determined endogenously as a distinct cluster - in the set-up of Francis et al. (2012). For both subperiods, again, we find no world factor and, additionally, the posterior inclusion probability for CL2 is now very low suggesting that CL2 does not belong in the model. When looking at the results obtained with all three variables included in the factor model we find - in both subperiods - no evidence in favor of a global factor and we confirm that there is little support for the existence of CL2. The different results that we obtain compared to Francis et al. (2012) when considering different sample periods or more variables than just output suggest that applying the endogenous clustering method in these alternative settings might lead to a different composition of clusters.

Table 5: Posterior inclusion probabilities of international common factors: development levels and clusters

Period	Posterior factor inclusion probabilities											
	Development				Endogenous clusters							
	All variables				Output only				All variables			
	World	IND	EME	DEV	World	CL1	CL2	CL3	World	CL1	CL2	CL3
1961-2014	1.00	0.02	0.23	1.00	0.02	1.00	0.84	1.00	0.78	0.31	0.46	1.00
1961-1985	0.04	1.00	0.15	0.99	0.04	1.00	0.22	1.00	0.04	1.00	0.18	1.00
1986-2014	0.04	1.00	1.00	0.71	0.02	1.00	0.23	1.00	0.04	1.00	0.37	1.00

Note: All reported results are based on prior $p_0 = 0.5$ but they are robust to the use of alternative priors $p_0 = 0.25$ or $p_0 = 0.75$. The columns "All variables" contain results obtained with output, consumption and investment. IND denotes the group of industrial economies, EME denotes the group of emerging economies and DEV denotes the group of developing economies. CL1 denotes cluster 1 which contains many industrialized economies, CL2 denotes cluster 2 which contains the UK and the former Commonwealth countries and CL3 denotes cluster 3 which contains South America, Mexico and a few other countries. We refer to Appendix A for the exact composition of these groups.

5 Conclusions

This paper deals with the identification of international business cycles using international macroeconomic aggregates. To this end, a Bayesian stochastic factor selection approach is developed for hierarchical dynamic factor models with predetermined factors, which are the factor models typically used in the literature on international business cycles. This literature has so far however ignored model uncertainty in these models as common factors are typically imposed but not tested for. This can lead to misleading results. To the best of our knowledge, no comparable factor selection approach for dynamic factor models is available in the literature.

The method is applied to investigate the existence of international business cycles using data for output growth, private consumption growth and private investment growth for 60 countries over the period 1961-2014. The focus lies in particular on whether a global worldwide cycle is present in the data. Our baseline model allows for a global factor and factors based on subgroups of countries, i.e., regions. While our model selection procedure supports the (increased) presence of regional cycles, our results also suggest that there is no global business cycle in the data, i.e., the posterior factor inclusion probabilities for the global common factor are negligible both in the baseline estimations and in most of the conducted robustness checks. We argue that not finding a common factor on which countries from all six regions load may not be that surprising since - unlike for subgroups of countries such as regions - for a group consisting of all countries in the world, it is hard to invoke characteristics shared by all these countries (such as geographical proximity, level op development, institutions) that might give rise to such a global cycle.

References

Amengual, D. and Watson, M. W. (2007). Consistent estimation of the number of dynamic factors in a large N and T panel. *Journal of Business & Economic Statistics*, 25:91–96.

Bai, J. and Ng, S. (2007). Determining the number of primitive shocks in factor models. *Journal of Business & Economic Statistics*, 25:52–60.

Bauwens, L., Lubrano, M., and Richard, J.-F. (2000). *Bayesian inference in dynamic econometric models*. Oxford University Press.

Canova, F. (1998). Detrending and business cycle facts. *Journal of Monetary Economics*, 41(3):475–512.

Carter, C. and Kohn, R. (1994). On Gibbs sampling for state space models. *Biometrika*, 81:541–53.

Chib, S. and Greenberg, E. (1994). Bayes inference in regression models with ARMA(p,q) errors. *Journal of Econometrics*, 64:183–206.

Feenstra, R., Inklaar, R., and Timmer, M. (2015). The next generation of the Penn World Table. *American Economic Review*, 105(10):3150–3182.

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

Frankel, J. A. and Rose, A. K. (1998). The endogeneity of the optimum currency area criteria. *The Economic Journal*, 108(449):1009–1025.

Frühwirth-Schnatter, S. and Wagner, H. (2010). Stochastic model specification search for Gaussian and partial non-Gaussian state space models. *Journal of Econometrics*, 154(1):85–100.

George, E. and McCulloch, R. (1993). Variable selection via Gibbs sampling. *Journal of the American Statistical Association*, 88:881–889.

Hallin, M. and Liska, R. (2007). Determining the number of factors in the general dynamic factor model. *Journal of the American Statistical Association*, 102:603–617.

Hirata, H., Kose, M. A., and Otrok, C. (2013). Regionalization vs. globalization. *IMF Working Paper*, 13/19.

Kalemli-Ozcan, S., Sorensen, B. E., and Yosha, O. (2001). Economic integration, industrial specialization and the asymmetry of macroeconomic fluctuations. *Journal of International Economics*, 55(1):107–137.

Kim, C.-J. and Nelson, C. (1999). *State-space models with regime switching. Classical and Gibbs-sampling approaches with applications.* MIT Press, Cambridge.

Kose, M. A., Otrok, C., and Prasad, E. (2012). Global business cycles: convergence or decoupling? *International Economic Review*, 53(2):511–538.

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

Kose, M. A., Otrok, C., and Whiteman, C. H. (2008). Understanding the evolution of world business cycles. *Journal of International Economics*, 75(1):110–130.

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

Kose, M. A. and Yi, K.-M. (2006). Can the standard international business cycle model explain the relation between trade and comovement? *Journal of International Economics*, 68(2):267–295.

Mumtaz, H., Simonelli, S., and Surico, P. (2011). International comovements, business cycle and inflation: a historical perspective. *Review of Economic Dynamics*, 14(1):176 − 198.

Sayeh, A. (2012). Africa and the Great Recession: changing times. *IMF Direct.*

Scott, J. G. and Berger, J. O. (2010). Bayes and empirical Bayes multiplicity adjustment in the variable-selection problem. *The Annals of Statistics*, 38(5):2587–2619.

Appendix A List of countries

The sample consists of 60 countries which, below, are grouped according to geographical region. Each country belongs to one of 6 regions: Europe, North America, Oceania, Latin America, Africa and Asia. This classification follows Kose et al. (2003a) except for the fact that in the main results of the paper no subdivision is made between developed and developing Asian countries. Between squared brackets the ISO classification of the country is added. Between normal brackets two additional pieces of information are reported that are used in the robustness checks of the paper (see Section 4). First, the development level group to which the country belongs where this classification of countries follows Kose et al. (2012). A country belongs either to the group of industrial economies (IND), to the group of emerging market economies (EME) or to the group of other developing countries (DEV). Second, the endogenously determined cluster to which the country belongs where this classification of countries is based on the results obtained by the endogenous clustering method of Francis et al. (2012). A country belongs either to cluster 1 which contains many industrialized countries (CL1), to cluster 2 which contains the UK and the former Commonwealth countries (CL2) or to cluster 3 which contains South America, Mexico and a few other countries (CL3). In the robustness checks of Section 4 a distinction is further made between developed Asian economies and developing Asian economies where we base this classification on Kose et al. (2003a). Below, the developed Asian economies are denoted with an asterisk next to their names.

Europe (18 countries):

Austria [AUT](IND,CL1), Belgium [BEL](IND,CL1), Denmark [DNK](IND,CL2), Finland [FIN](IND, CL1), France [FRA](IND, CL1), Germany [DEU](IND, CL1), Greece [GRC](IND,CL1), Iceland [ISL](IND,CL1), Ireland [IRL](IND,CL1) Italy [ITA](IND,CL1), Luxembourg [LUX](IND,CL1), Netherlands [NLD](IND,CL1), Norway [NOR](IND,CL1), Portugal [PRT](IND,CL1), Spain [ESP](IND,CL1), Sweden [SWE](IND,CL1), Switzerland [CHE](IND,CL1), United Kingdom [GBR](IND,CL2)

North America (3 countries):

Canada [CAN](IND,CL2), Mexico [MEX](EME,CL3), U.S. [USA](IND,CL2)

Oceania (2 countries):

Australia [AUS](IND,CL2), New Zealand [NZL](IND,CL2)

Latin America (18 countries):

Costa Rica [CRC](DEV,CL3), Dominican Republic [DOM](DEV,CL3), El Salvador [SLV](DEV,CL3), Guatemala [GTM](DEV,CL3), Honduras [HND](DEV,CL3), Jamaica [JAM](DEV,CL3), Panama [PAN](DEV,CL3), Trinidad [TTO](DEV,CL3), Argentina [ARG](EME,CL3), Bolivia [BOL](DEV,CL3), Brazil [BRA](EME,CL1), Chile [CHL](EME,CL3), Columbia [COL](EME,CL3), Ecuador [ECU](DEV,CL3), Paraguay [PRY](DEV,CL3), Peru [PER](EME,CL3), Uruguay [URY](DEV,CL3), Venezuela [VEN](EME,CL3),

Africa (7 countries):

Cameroon [CMR](DEV,CL2), Ivory Coast [CIV](DEV,CL2), Kenya [KEN](DEV,CL2), Morocco [MAR](EME,CL3), Senegal [SEN](DEV,CL3), South Africa [ZAF](EME,CL2), Zimbabwe [ZWE](DEV,CL2),

Asia (12 countries):

Bangladesh [BGD](DEV,CL2), India [IND](EME,CL2), Indonesia [IDN](EME,CL2), Pakistan [PAK](EME,CL1), Philippines [PHL](EME,CL3), Sri Lanka [LKA](DEV,CL2), Hong Kong* [HKG](EME,CL1), Japan* [JAP](IND,CL1), Malaysia* [MYS](EME,CL2), Singapore* [SGP](EME,CL2), South Korea* [KOR](EME,CL1), Thailand* [THA](EME,CL1)

Appendix B Technical details of the the Gibbs sampler

B.1 Sample the parameters ϕ from the conditional distribution $\Lambda(\phi|f, y)$

Subsets of parameters contained in ϕ can be sampled from a standard regression model,

$$x = z^r \beta^r + \omega \tag{B-1}$$

where x is a $n \times 1$ vector containing n stacked observations on the dependent variable, z^r is a $n \times k$ matrix containing n stacked observations of k predictor variables, β^r is the $k \times 1$ parameter vector and ω is the $n \times 1$ vector of error terms for which $\omega \sim iid\mathcal{N}\left(0, \sigma_\omega^2 I_n\right)$. If there are no binary indicators ι in the regression or if all binary indicators in the regression ι are equal to 1, then $z^r = z$ and $\beta^r = \beta$ where z and β are the unrestricted predictor matrix and the corresponding unrestricted coefficient vector. Otherwise, the restricted parameter vector β^r and the corresponding restricted predictor matrix z^r contain only those elements of z and β for which the corresponding binary indicators ι are equal to 1. The prior distribution of β^r is given by $\beta^r \sim \mathcal{N}\left(b_0^r, B_0^r \sigma_\omega^2\right)$ with b_0^r a $k \times 1$ vector and B_0^r a $k \times k$ matrix. The prior distribution of σ_ω^2 is given by $\sigma_\omega^2 \sim \mathcal{IG}\left(c_0, C_0\right)$ with scalars c_0 (shape) and C_0 (scale). The posterior distributions (conditional on x, z^r, and ι) of β^r and σ_ω^2 are then given by $\beta^r \sim \mathcal{N}\left(b^r, B^r \sigma_\omega^2\right)$ and $\sigma_\omega^2 \sim \mathcal{IG}\left(c, C^r\right)$ with,

$$\begin{aligned}
B^r &= \left[(z^r)'z^r + (B_0^r)^{-1}\right]^{-1} \\
b^r &= B^r \left[(z^r)'x + (B_0^r)^{-1} b_0^r\right] \\
c &= c_0 + n/2 \\
C^r &= C_0 + \frac{1}{2}\left[x'x + (b_0^r)'(B_0^r)^{-1} b_0^r - (b^r)'(B^r)^{-1} b^r\right]
\end{aligned} \tag{B-2}$$

To sample the binary indicators ι, a naive implementation of the Gibbs sampler would be to first sample the binary indicators ι from $p(\iota|x, z, \beta, \sigma_\omega^2)$ and next β^r from $p(\beta^r|x, z, \sigma_\omega^2, \iota)$. However, this approach does not result in an irreducible Markov chain as whenever an indicator ι equals 0, the corresponding coefficient in β is also 0 which implies that the chain has absorbing states. Therefore, as in Frühwirth-Schnatter and Wagner (2010), we marginalize over the parameters β when sampling ι and next draw β^r conditional on ι. The posterior distribution of the binary indicators ι is obtained from Bayes' theorem as,

$$p(\iota|x, z, \sigma_\omega^2) \propto p(x|z, \sigma_\omega^2, \iota) p(\iota) \tag{B-3}$$

where $p(\iota)$ is the prior distribution of ι and $p(x|z, \sigma_\omega^2, \iota)$ is the marginal likelihood of regression eq. (B-1) where the effect of the parameters β has been integrated out. We refer to Frühwirth-Schnatter and Wagner (2010) (their eq.(25)) for the closed form expression of the marginal likelihood for the general regression model of eq. (B-1). For our purposes, a simpler marginal likelihood expression suffices as we calculate the marginal likelihood under the restriction $\sigma_\omega^2 = 1$ as will be detailed in Section B.1.1 below. We then have,

$$p(x|z, \sigma_\omega^2, \iota) \propto \frac{|B^r|^{0.5}}{|B_0^r|^{0.5}} \exp\left[-\frac{x'x + (b_0^r)'(B_0^r)^{-1}b_0^r - (b^r)'(B^r)^{-1}b^r}{2}\right] \quad \text{(B-4)}$$

where b_0^r, B_0^r, b^r and B^r are as reported above.[13]

B.1.1 Sample the binary indicators δ from $\Lambda(\delta|\alpha, \sigma_\nu^2, \pi, f, y)$

We draw the binary indicators $\iota = \delta$ one-by-one per factor j (with $j = 1, ..., M+N$). To this end, we first rewrite eq. (5) as,

$$y_{it}^j = \delta^j \sigma_{\varepsilon^j} f_{it}^{*j} + \mu_{it} \quad \text{(B-5)}$$

where $y_{it}^j \equiv y_{it} - \sum_{m \neq j} \alpha_i^m \delta^m \sigma_{\varepsilon^m} f_t^m$ and $f_{it}^{*j} \equiv \alpha_i^j f_t^j$. This equation cannot immediately be cast into the standard regression eq.(B-1) above as μ_{it} is not *iid* but follows an AR(p) process (see eq. (4)). Using the parameters $\pi_i = \left[\pi_i^1, ..., \pi_i^p\right]'$ we can define the lag polynomial $\Pi_i(L) = 1 - \pi_i^1 L - \pi_i^2 L^2 - ... - \pi_i^p L^p$ and premultiply both sides of eq. (B-5) with it to obtain,

$$\tilde{y}_{it}^j = \delta^j \sigma_{\varepsilon^j} \tilde{f}_{it}^{*j} + \nu_{it} \quad \text{(B-6)}$$

where $\tilde{y}_{it}^j = \Pi_i(L) y_{it}^j$, $\tilde{f}_{it}^{*j} = \Pi_i(L) f_{it}^{*j}$ and where we use the result, from eq. (4), that $\nu_{it} = \Pi_i(L)\mu_{it}$.

As such, we have observations for $\tilde{y}_{it}^j$ and $\tilde{f}_{it}^{*j}$ for $t = p+1, ..., T$. The first p observations for $\tilde{y}_{it}^j$ and $\tilde{f}_{it}^{*j}$ are calculated from the first p observations of y_{it}^j and f_{it}^{*j} using π_i and following the approach of Chib and Greenberg (1994). In particular, stack the first p observations for y_{it}^j and f_{it}^{*j} into the $p \times 1$ vectors $y_{i0}^j = \left[y_{i1}^j, ..., y_{ip}^j\right]'$ and $f_{i0}^{*j} = \left[f_{i1}^{*j}, ..., f_{ip}^{*j}\right]'$. Then calculate the $p \times 1$ vectors $\tilde{y}_{i0}^j = \left[\tilde{y}_{i1}^j, ..., \tilde{y}_{ip}^j\right]'$ and $\tilde{f}_{i0}^{*j} = \left[\tilde{f}_{i1}^{*j}, ..., \tilde{f}_{ip}^{*j}\right]'$ from $\tilde{y}_{i0}^j = Q^{-1} y_{i0}^j$ and $\tilde{f}_{i0}^{*j} = Q^{-1} f_{i0}^{*j}$ with the $p \times p$ transformation matrix Q. The matrix Q satisfies $QQ' = \Sigma_p$ where the $p \times p$ matrix Σ_p is defined from $vec(\Sigma_p) = (I_{p^2} - \Phi \otimes \Phi)^{-1} vec(ee')$ with $e = (1, 0, ..., 0)'$ the $p \times 1$ unit vector and with the $p \times p$ matrix $\Phi = \begin{bmatrix} & \pi_i' \\ I_{p-1} & 0_{(p-1)\times 1} \end{bmatrix}$.

To sample δ^j from eq. (B-6) we need to stack observations over both t and i so that the error term ν_{it} is still not *iid*, i.e., it is heteroskedastic across i since its variance $\sigma_{\nu_i}^2$ is different for every i. We therefore

[13] Eq. (B-4) can be obtained from Frühwirth-Schnatter and Wagner's eq. (45) by replacing their variance matrix Σ by 1 (after adjusting differences in notation).

follow a GLS approach and divide both sides of eq. (B-6) by σ_{ν_i} to obtain,

$$\frac{\tilde{y}^j_{it}}{\sigma_{\nu_i}} = \delta^j \sigma_{\varepsilon^j} \frac{\tilde{f}^{*j}_{it}}{\sigma_{\nu_i}} + \frac{\nu_{it}}{\sigma_{\nu_i}} \tag{B-7}$$

Eq. (B-7) fits in the framework of eq. (B-1) both when $\delta^j = 1$ and when $\delta^j = 0$. If $\delta^j = 1$ then $x = \frac{\tilde{y}^j}{\sigma_\nu}$ is a $KNT \times 1$ vector (i.e., $n = KNT$) of stacked values of $\frac{\tilde{y}^j_{it}}{\sigma_{\nu_i}}$, $z^r = z = \frac{\tilde{f}^{*j}}{\sigma_\nu}$ is the $KNT \times 1$ predictor matrix (i.e., $k = 1$) containing the stacked values of $\frac{\tilde{f}^{*j}_{it}}{\sigma_{\nu_i}}$ with corresponding parameter $\beta^r = \beta = \sigma_{\varepsilon^j}$ and $\omega = \frac{\nu}{\sigma_\nu}$ is a $KNT \times 1$ vector containing stacked values of $\frac{\nu_{it}}{\sigma_{\nu_i}}$. We note that $\sigma^2_\omega = 1$. When $\delta^j = 0$ we have the same values for x, ω and σ^2_ω but z is a $KNT \times 1$ vector of zeros and $\beta = 0$.

As $\sigma^2_\omega = 1$ both when $\delta^j = 1$ and $\delta^j = 0$, we calculate the marginal likelihoods $p(x|\delta^j = 1, z, \sigma^2_\omega, \delta^{-j})$ and $p(x|\delta^j = 0, z, \sigma^2_\omega, \delta^{-j})$ using eq. (B-4) above with the priors for $\beta = \sigma_{\varepsilon^j}$ discussed in Section 2.3.1, i.e., $b^r_0 = 0$ and $B^r_0 = 10$. Upon combining the marginal likelihoods with the Bernoulli prior distributions of the binary indicators $p(\delta^j = 1) = p_0$ and $p(\delta^j = 0) = 1 - p_0$, the posterior distributions $p(\delta^j = 1|x, z, \sigma^2_\omega, \delta^{-j})$ and $p(\delta^j = 0|x, z, \sigma^2_\omega, \delta^{-j})$ are obtained from which the probability $prob(\delta^j = 1|x, z, \sigma^2_\omega, \delta^{-j}) = \frac{p(\delta^j=1|x,z,\sigma^2_\omega,\delta^{-j})}{p(\delta^j=0|x,z,\sigma^2_\omega,\delta^{-j})+p(\delta^j=1|x,z,\sigma^2_\omega,\delta^{-j})}$ is calculated which is used to sample δ^j.

B.1.2 Sample the standard deviations of the common factors σ_ε from $\Lambda(\sigma_\varepsilon|\delta, \alpha, \sigma^2_\nu, \pi, f, y)$

Given δ, α, σ^2_ν, π, f and y we can sample σ_ε simultaneously for all factors j (with $j = 1, ..., M + N$). We rewrite eq. (5) as,

$$y_{it} = \sum_{j=1}^{M+N} \sigma_{\varepsilon^j} f^{*j}_{it} + \mu_{it} \tag{B-8}$$

where $f^{*j}_{it} \equiv \delta^j \alpha^j_i f^j_t$. This equation cannot immediately be cast into the standard regression eq. (B-1) above as μ_{it} is not *iid* but follows an AR(p) process (see eq. (4)). Using the parameters $\pi_i = \left[\pi^1_i, ..., \pi^p_i\right]'$ we can define the lag polynomial $\Pi_i(L) = 1 - \pi^1_i L - \pi^2_i L^2 - ... - \pi^p_i L^p$ and premultiply both sides of eq. (B-8) with it to obtain,

$$\tilde{y}_{it} = \sum_{j=1}^{M+N} \sigma_{\varepsilon^j} \tilde{f}^{*j}_{it} + \nu_{it} \tag{B-9}$$

where $\tilde{y}_{it} = \Pi_i(L)y_{it}$, $\tilde{f}^{*j}_{it} = \Pi_i(L)f^{*j}_{it}$ and where we use the result, from eq. (4), that $\nu_{it} = \Pi_i(L)\mu_{it}$.

As such, we have observations for $\tilde{y}_{it}$ and $\tilde{f}^{*j}_{it}$ for $t = p + 1, ..., T$. The first p observations for $\tilde{y}_{it}$ and $\tilde{f}^{*j}_{it}$ are calculated from the first p observations of y_{it} and f^{*j}_{it} using π_i following the approach of Chib and Greenberg (1994) as discussed in Section B.1.1.

To sample σ_{ε^j} from eq. (B-9) we need to stack observations over both t and i so that the error term ν_{it} is still not *iid*, i.e., it is heteroskedastic across i since its variance $\sigma^2_{\nu_i}$ is different for every i. We therefore follow a GLS approach and divide both sides of eq. (B-9) by σ_{ν_i} to obtain,

$$\frac{\tilde{y}_{it}}{\sigma_{\nu_i}} = \sum_{j=1}^{M+N} \sigma_{\varepsilon^j} \frac{\tilde{f}_{it}^{*j}}{\sigma_{\nu_i}} + \frac{\nu_{it}}{\sigma_{\nu_i}} \tag{B-10}$$

Eq. (B-10) fits in the framework of eq. (B-1) with $x = \frac{\tilde{y}}{\sigma_\nu}$ the $KNT \times 1$ vector of stacked (across t and i) values of $\frac{\tilde{y}_{it}}{\sigma_{\nu_i}}$ (i.e., $n = KNT$). The $n \times k$ predictor matrix z^r is given by $z^r = \left[..., \frac{\tilde{f}^{*j}}{\sigma_\nu}, ...\right]$ where $\frac{\tilde{f}^{*j}}{\sigma_\nu}$ contains the stacked (across t and i) values $\frac{\tilde{f}_{it}^{*j}}{\sigma_{\nu_i}}$ and where a regressor $\frac{\tilde{f}^{*j}}{\sigma_\nu}$ is included in z^r if it has $\delta^j = 1$, i.e., for the number of columns k of z^r we have $k \leq M + N$. The corresponding $k \times 1$ parameter vector is given by $\beta^r = [..., \sigma_{\varepsilon^j}, ...]'$. The vector $\omega = \frac{\nu}{\sigma_\nu}$ is the $KNT \times 1$ vector of error terms $\frac{\nu_{it}}{\sigma_{\nu_i}}$ for which $\sigma_\omega^2 = 1$. The priors used when sampling σ_{ε^j} are discussed in Section 2.3.1, i.e., we use prior means and variances for σ_{ε^j} of respectively 0 and 10. As such, we have $\beta^r \sim \mathcal{N}\left(b_0^r, B_0^r \sigma_\omega^2\right)$ with $k \times 1$ vector $b_0^r = [0, ..., 0]'$ and $k \times k$ matrix $B_0^r = 10 I_k$.

Given these priors, we sample $\beta^r = [..., \sigma_{\varepsilon^j}, ...]'$ from $\mathcal{N}\left(b^r, B^r \sigma_\omega^2\right)$ where, from eq. (B-2), we have $b^r = B^r \left[(z^r)' x\right]$ and $B^r = \left[(z^r)' z^r + (B_0^r)^{-1}\right]^{-1}$. The standard deviations σ_{ε^j} corresponding to factors j for which $\delta^j = 0$ are set to 0.

B.1.3 Sample the factor loadings α and the variances σ_ν^2 jointly from $\Lambda(\alpha, \sigma_\nu^2 | \delta, \sigma_\varepsilon, \pi, f, y)$

Given δ, σ_ε, π, f and y we can sample the non-zero factor loadings α and error variances σ_ν^2 per variable/country i (with $i = 1, ..., K \times N$). We rewrite eq. (5) as,

$$y_{it} = \sum_{j=1}^{M+N} \alpha_i^j f_t^{*j} + \mu_{it} \tag{B-11}$$

where $f_t^{*j} \equiv \delta^j \sigma_{\varepsilon^j} f_t^j$. This equation cannot immediately be cast into the standard regression eq. (B-1) above as μ_{it} is not *iid* but follows an AR(p) process (see eq. (4)). Using the parameters $\pi_i = \left[\pi_i^1, ..., \pi_i^p\right]'$ we can define the lag polynomial $\Pi_i(L) = 1 - \pi_i^1 L - \pi_i^2 L^2 - ... - \pi_i^p L^p$ and premultiply both sides of eq.(B-11) with it to obtain,

$$\tilde{y}_{it} = \sum_{j=1}^{M+N} \alpha_i^j \tilde{f}_t^{*j} + \nu_{it} \tag{B-12}$$

where $\tilde{y}_{it} = \Pi_i(L) y_{it}$, $\tilde{f}_t^{*j} = \Pi_i(L) f_t^{*j}$ and where we use the result, from eq. (4), that $\nu_{it} = \Pi_i(L)\mu_{it}$.

As such, we have observations for $\tilde{y}_{it}$ and $\tilde{f}_t^{*j}$ for $t = p + 1, ..., T$. The first p observations for $\tilde{y}_{it}$ and $\tilde{f}_t^{*j}$ are calculated from the first p observations of y_{it} and f_t^{*j} using π_i following the approach of Chib and Greenberg (1994) as discussed in Section B.1.1.

Eq. (B-12) fits in the framework of eq. (B-1) with $x = \tilde{y}_i$ the $T \times 1$ vector of stacked values of $\tilde{y}_{it}$ (i.e., $n = T$). The $n \times k$ predictor matrix z^r is given by $z^r = \left[..., \tilde{f}^{*j}, ...\right]$ where $\tilde{f}^{*j}$ contains the stacked values $\tilde{f}_t^{*j}$ and where a regressor $\tilde{f}^{*j}$ is included in z^r if it has $\delta^j = 1$, i.e., for the number of columns k of z^r we have $k \leq M + N$. The corresponding $k \times 1$ parameter vector is given by $\beta^r = \left[..., \alpha_i^j, ...\right]'$. The vector $\omega = \nu_i$ is the $T \times 1$ vector of error terms ν_{it} where from eq. (4) we note that $\sigma_\omega^2 = \sigma_{\nu_i}^2$. The priors used when sampling α_i and $\sigma_{\nu_i}^2$ are discussed in Section 2.3.1, i.e., we use prior means and variances for α_i^j of respectively 0 and 10 and prior beliefs and strengths for $\sigma_{\nu_i}^2$ of respectively 10 and 0.01. As such, we have $\beta^r \sim \mathcal{N}\left(b_0^r, B_0^r \sigma_\omega^2\right)$ with $k \times 1$ vector $b_0^r = [0, ..., 0]'$ a $k \times 1$ and $k \times k$ matrix $B_0^r = (10/\sigma_\omega^2) \times I_k$ and we have $\sigma_\omega^2 \sim \mathcal{IG}\left(c_0, C_0\right)$ with $c_0 = 0.01 \times T$ and $C_0 = 0.01 \times 10 \times T$.

Given these priors, we sample $\sigma_\omega^2 = \sigma_{\nu_i}^2$ from $\mathcal{IG}\left(c, C^r\right)$ and, conditional on σ_ω^2, we sample $\beta^r = \left[..., \alpha_i^j, ...\right]'$ from $\mathcal{N}\left(b^r, B^r \sigma_\omega^2\right)$ where, from eq. (B-2), we have $c = c_0 + T/2$, $C^r = C_0 + \frac{1}{2}\left[x'x - (b^r)'(B^r)^{-1} b^r\right]$, $b^r = B^r\left[(z^r)'x\right]$ and $B^r = \left[(z^r)'z^r + (B_0^r)^{-1}\right]^{-1}$. Then, we impose the normalization condition $\overline{\alpha}^j = 1$ (where the average is over i and excludes cases for which α_i^j is equal to 0). The loadings α_i^j corresponding to factors j for which $\delta^j = 0$ are set to 0.

B.1.4 Sample the AR parameters ρ from $\Lambda(\rho|f)$

Given the common factors f^j (with $j = 1, ..., M + N$), the AR parameters $\rho^{j,l}$ (with $l = 1, ..., q$) can be sampled per factor j from eq. (B-1) where $x = f^j$ is the $(T - q) \times 1$ vector of stacked values of f_t^j (i.e., $n = T - q$), $z^r = z = \left[f_{-1}^j, ..., f_{-q}^j\right]$ is the $(T - q) \times q$ matrix of stacked lagged values of f_t^j (i.e., $k = q$), $\beta^r = \beta = \rho^j = \left[\rho^{j,1}, ..., \rho^{j,q}\right]'$ is the $q \times 1$ vector of parameters $\rho^{j,l}$ (for $l = 1, ..., q$), $\omega = \varepsilon^j$ is the $(T - q) \times 1$ vector of error terms ε_t^j where from eq. (3) we note that $\sigma_\omega^2 = 1$ (for all j). The priors used when sampling $\rho^{j,l}$ are discussed in Section 2.3.1, i.e., we use prior means and variances for $\rho^{j,l}$ of respectively 0 and 1. As such, we have $\beta \sim \mathcal{N}\left(b_0, B_0\right)$ with $b_0 = [0, ..., 0]'$ a $q \times 1$ vector and $B_0 = I_q$.

Given these priors, we follow Chib and Greenberg (1994) and first simulate a candidate draw β^* from $\beta \sim \mathcal{N}\left(b, B\right)$ where, from eq. (B-2), $B = \left[z'z + (I_q)^{-1}\right]^{-1}$ and $b = B(z'x)$. If the candidate draw β^* does not satisfy the stationarity condition, it is discarded and the previous draw β' is withheld.[14] If the candidate draw β^* satisfies the stationarity condition, we conduct a Metropolis-Hastings step and accept it as the next sample value with probability $min\left[(\Psi(\beta^*)/\Psi(\beta')), 1\right]$ where $\Psi(\beta) = |\Sigma_q|^{-1/2} exp\left(-\frac{1}{2} x_0' \Sigma_q^{-1} x_0\right)$ with $x_0 = [x_1, ..., x_q]' = \left[f_1^j, ..., f_q^j\right]'$ the first q data points assumed to be drawn from the stationary distribution $x_0|\beta \sim \mathcal{N}\left(0, \Sigma_q\right)$. The $q \times q$ matrix Σ_q is defined from $vec(\Sigma_q) = (I_{q^2} - \Phi \otimes \Phi)^{-1} vec(ee')$ with $e = (1, 0, ..., 0)'$ the $q \times 1$ unit vector and with the $q \times q$ matrix $\Phi = \begin{bmatrix} & \rho^{j'} \\ I_{q-1} & 0_{(q-1)\times 1} \end{bmatrix}$. If the candidate draw β^* is rejected, the previous draw β' is withheld.

[14] The stationarity condition is fulfilled if, if for some variable m, the moduli of the roots of the polynomial $\rho^j(m) = 1 - \rho^{j,1}m - \rho^{j,2}m^2 - ... - \rho^{j,q}m^q$ lie outside of the unit circle.

B.1.5 Sample the AR parameters π from $\Lambda(\pi|\delta, \sigma_\varepsilon, \alpha, \sigma_\nu^2, f, y)$

Given δ, σ_ε, α, f and y we can calculate the idiosyncratic components μ from eq. (5). Given the idiosyncratic components μ_i (with $i = 1, ..., K \times N$), the AR parameters π_i^l (with $l = 1, ..., p$) can be sampled per country/variable i from eq. (B-1) where $x = \mu_i$ is the $(T-p) \times 1$ vector of stacked values of μ_{it} (i.e., $n = T - p$), $z^r = z = [\mu_{i,-1}, ..., \mu_{i,-p}]$ is the $(T-p) \times p$ matrix of stacked lagged values of μ_{it} (i.e., $k = p$), $\beta^r = \beta = \pi_i = \left[\pi_i^1, ..., \pi_i^p\right]'$ is the $p \times 1$ vector of parameters π_i^l (for $l = 1, ..., p$), $\omega = \nu_i$ is the $(T-p) \times 1$ vector of error terms ν_{it} where from eq. (4) we note that $\sigma_\omega^2 = \sigma_{\nu_i}^2$ (for all i). We note that when sampling π_i we do not sample $\sigma_{\nu_i}^2$ but take it as given as $\sigma_{\nu_i}^2$ is already sampled in Section B.1.3 above. The priors used when sampling π_i^l are discussed in Section 2.3.1, i.e., we use prior means and variances for π_i^l of respectively 0 and 1. As such, we have $\beta \sim \mathcal{N}\left(b_0, B_0\sigma_\omega^2\right)$ with $b_0 = [0, ..., 0]'$ a $p \times 1$ vector and $B_0 = I_p$.

Given these priors, we follow Chib and Greenberg (1994) and first simulate a candidate draw β^* from $\beta \sim \mathcal{N}\left(b, B\sigma_\omega^2\right)$ where, from eq. (B-2), $B = \left[z'z + (I_p)^{-1}\right]^{-1}$ and $b = B(z'x)$. If the candidate draw β^* does not satisfy the stationarity condition, it is discarded and the previous draw β' is withheld. If the candidate draw β^* satisfies the stationarity condition, we conduct a Metropolis-Hastings step and accept it as the next sample value with probability $min\left[(\Psi(\beta^*)/\Psi(\beta')), 1\right]$ where $\Psi(\beta) = |\Sigma_p|^{-1/2} exp\left(-\frac{1}{2\sigma_\omega^2} x_0' \Sigma_p^{-1} x_0\right)$ with $x_0 = [x_1, ..., x_p]' = [\mu_{i1}, ..., \mu_{ip}]'$ the first p data points assumed to be drawn from the stationary distribution $x_0|\beta \sim \mathcal{N}\left(0, \Sigma_p \sigma_\omega^2\right)$. The $p \times p$ matrix Σ_p is defined from $vec(\Sigma_p) = (I_{p^2} - \Phi \otimes \Phi)^{-1} vec(ee')$ with $e = (1, 0, ..., 0)'$ the $p \times 1$ unit vector and with the $p \times p$ matrix $\Phi = \begin{bmatrix} & \pi_i' \\ I_{p-1} & 0_{(p-1)\times 1} \end{bmatrix}$. If the candidate draw β^* is rejected, the previous draw β' is withheld.

B.2 Sample the common factors f from the conditional distribution $\Lambda(f|\phi, y)$

The common factors that are included in the model - i.e., those for which $\delta^j = 1$ - are sampled conditional on the parameters ϕ and data y using a state space approach where the unobserved states are the common factors f. In particular, we use the forward-filtering backward-sampling approach discussed in detail in Kim and Nelson (1999) to sample the unobserved states. The general form of the state space model is given by,

$$Y_t = Z_t S_t + V_t, \qquad V_t \sim iid\mathcal{N}\left(0, H_t\right), \qquad \text{(B-13)}$$

$$S_t = T_t S_{t-1} + K_t E_t, \qquad E_t \sim iid\mathcal{N}\left(0, Q_t\right), \qquad t = 1, \ldots, T \qquad \text{(B-14)}$$

$$S_1 \sim iid\mathcal{N}\left(s_1, P_1\right), \qquad \text{(B-15)}$$

where Y_t is a $n \times 1$ vector of observations and S_t an unobserved $n^s \times 1$ state vector. The matrices Z_t, T_t, K_t, H_t, Q_t and the mean s_1 and variance P_1 of the initial state vector S_1 are assumed to be known (conditioned upon) and the error terms V_t and E_t are assumed to be serially uncorrelated and independent of each other at all points in time. Note that E_t is a $n^{ss} \times 1$ matrix (where $n^{ss} \leq n^s$). As eqs. (B-13)-(B-15) constitute a linear Gaussian state space model, the unknown state variables in S_t can be filtered using the standard Kalman filter. Sampling $S = [S_1, \ldots, S_T]$ from its conditional distribution can then be done using the multimove Gibbs sampler of Carter and Kohn (1994).

The state space approach and forward-filtering backward-sampling algorithm is only used for common factors that are effectively included in the model, i.e., factors for which $\delta^j = 1$. Assuming there are $n^f \leq M + N$ of such factors (where n^f can be different in every Gibbs iteration), we rewrite eq. (5) in the main text as,

$$y_{it} = \sum_{j=1}^{n^f} \alpha_i^j \sigma_{\varepsilon^j} f_t^j + \mu_{it} \tag{B-16}$$

Eq. (B-16) is not an observation equation of the form of eq. (B-13) as the component μ_{it} is not *iid* but follows an AR(p) process (see eq. (4)). Using the parameters $\pi_i = \left[\pi_i^1, ..., \pi_i^p\right]'$ we can define the lag polynomial $\Pi_i(L) = 1 - \pi_i^1 L - \pi_i^2 L^2 - ... - \pi_i^p L^p$ and premultiply both sides of eq. (B-16) with it to obtain,

$$\tilde{y}_{it} = \sum_{j=1}^{n^f} \alpha_i^j \sigma_{\varepsilon^j} \Pi_i(L) f_t^j + \nu_{it} \tag{B-17}$$

where $\tilde{y}_{it} = \Pi_i(L) y_{it}$ and where we use the result, from eq. (4), that $\nu_{it} = \Pi_i(L)\mu_{it}$. As such, we have observations for $\tilde{y}_{it}$ for $t = p+1, ..., T$. The first p observations for $\tilde{y}_{it}$ are calculated from the first p observations of y_{it} using π_i following the approach of Chib and Greenberg (1994) as discussed in Section B.1.1.

Eq. (B-17) and the n^f equations containing the law of motion for the factors f_t^j as given by eq. (3) in the text (now for $j = 1, ..., n^f$) can be put in the state space form of eqs.(B-13)-(B-15). The dimensions are $n = K \times N$, $n^s = max\left[(p+1) \times n^f, q \times n^f\right]$ and $n^{ss} = n^f$. The system matrices are $Y_t = \left[\begin{array}{ccc} \tilde{y}_{1t} & ... & \tilde{y}_{nt} \end{array}\right]'$, $V_t = \left[\begin{array}{ccc} \nu_{1t} & ... & \nu_{nt} \end{array}\right]'$, $H_t = \left[\begin{array}{ccc} \sigma_{\nu_1}^2 & & 0 \\ & \ddots & \\ 0 & & \sigma_{\nu_n}^2 \end{array}\right]$. The matrix Z_t takes the form $Z_t = \left[\begin{array}{ccc} Z_{1t} & ... & Z_{nt} \end{array}\right]'$ with $Z_{it} = \left[\begin{array}{ccc} Z_{it}^1 & ... & Z_{it}^{n^f} \end{array}\right]$ (for $i = 1, ..., K \times N$). When $q \leq (p+1)$ we have $Z_{it}^j = \left[\begin{array}{cccc} \alpha_i^j \sigma_{\varepsilon^j} & -\alpha_i^j \sigma_{\varepsilon^j} \pi_i^1 & ... & -\alpha_i^j \sigma_{\varepsilon^j} \pi_i^p \end{array}\right]$ (for $j = 1, ..., n^f$) where Z_{it}^j is a $1 \times (p+1)$ matrix. When $q > (p+1)$ we have $Z_{it}^j = \left[\begin{array}{cccccc} \alpha_i^j \sigma_{\varepsilon^j} & -\alpha_i^j \sigma_{\varepsilon^j} \pi_i^1 & ... & -\alpha_i^j \sigma_{\varepsilon^j} \pi_i^p & 0 & ... & 0 \end{array}\right]$ (for $j = 1, ..., n^f$) where Z_{it}^j is a $1 \times q$ matrix. The state vector S_t then takes the form $S_t = \left[\begin{array}{ccc} S_t^1 & ... & S_t^{n^f} \end{array}\right]'$ where for $q \leq (p+1)$ we have $S_t^j = \left[\begin{array}{cccc} f_t^j & f_{t-1}^j & ... & f_{t-p}^j \end{array}\right]'$ a $(p+1) \times 1$ matrix (for $j = 1, ..., n^f$) and where for $q > (p+1)$

we have $S_t^j = \begin{bmatrix} f_t^j & f_{t-1}^j & ... & f_{t-p}^j & f_{t-p-1}^j & ... & f_{t-q+1}^j \end{bmatrix}'$ a $q \times 1$ matrix (for $j = 1, ..., n^f$). The matrix T_t is given by $T_t = \begin{bmatrix} T_t^1 & & 0 \\ & \ddots & \\ 0 & & T_t^{n^f} \end{bmatrix}$. When $q \leq (p+1)$ we have $T_t^j = \begin{bmatrix} \rho^{j'} & 0_{1\times(p+1-q)} \\ I_p & 0_{p\times 1} \end{bmatrix}$ a $(p+1) \times (p+1)$ matrix (for $j = 1, ..., n^f$) with ρ^j the $q \times 1$ vector containing the AR parameters ρ_l^j (for $l = 1, ..., q$). When $q > (p+1)$ we have $T_t^j = \begin{bmatrix} & \rho^{j'} \\ I_{q-1} & 0_{(q-1)\times 1} \end{bmatrix}$ a $q \times q$ matrix (for $j = 1, ..., n^f$). For the matrix Q_t we have $Q_t = I_{n^f}$. The matrix K_t is given by $K_t = \begin{bmatrix} K_t^1 & ... & K_t^{n^f} \end{bmatrix}'$. Each matrix K_t^j (for $j = 1, ..., n^f$) is a $(p+1) \times n^f$ matrix when $q \leq (p+1)$ or a $q \times n^f$ matrix when $q > (p+1)$ and consists entirely of zeros except for row 1 and column j where a "1" is placed. For the the mean s_1 of the initial state vector we have $s_1 = 0_{(p+1)n^f \times 1}$ when $q \leq (p+1)$ or $s_1 = 0_{qn^f \times 1}$ when $q > (p+1)$. The variance P_1 of the initial state vector is given by $P_1 = \begin{bmatrix} P_1^1 & & 0 \\ & \ddots & \\ 0 & & P_1^{n^f} \end{bmatrix}$ where the element P_1^j is calculated for $j = 1, ..., n^f$. When $q \leq (p+1)$ then P_1^j is a $(p+1) \times (p+1)$ matrix calculated from $vec(P_1^j) = (I_{(p+1)^2} - T_t^j \otimes T_t^j)^{-1} vec(ee')$ with T_t^j the $(p+1) \times (p+1)$ matrix defined above and $e = (1, 0, ..., 0)'$ the $(p+1) \times 1$ unit vector. When $q > (p+1)$ then P_1^j is a $q \times q$ matrix calculated from $vec(P_1^j) = (I_{q^2} - T_t^j \otimes T_t^j)^{-1} vec(ee')$ with T_t^j the $q \times q$ matrix defined above and $e = (1, 0, ..., 0)'$ the $q \times 1$ unit vector.