OXFORD BULLETIN OF ECONOMICS AND STATISTICS, 71, 5 (2009) 0305-9049
doi: 10.1111/j.1468-0084.2009.00556.x

# Multivariate Business Cycle Synchronization in Small Samples*

BERTRAND CANDELON†, JAN PIPLACK‡ and STEFAN STRAETMANS§

†*Department of Economics, University of Maastricht, Maastricht, The Netherlands*
*(e-mail: b.candelon@algec.unimaas.nl)*
‡*Rabobank and University of Utrecht, Utrecht, The Netherlands*
*(e-mail: jan.piplack@rabobank.com)*
§*Department of Finance, University of Maastricht, Maastricht, The Netherlands*
*(e-mail: s.straetmans@finance.unimaas.nl)*

## Abstract

In this paper, we study the degree of business cycle synchronization by means of a small sample version of the Harding and Pagan's [*Journal of Econometrics* (2006) Vol. 132, pp. 59–79] Generalized Method of Moment test. We show that the asymptotic version of the test gets increasingly distorted in small samples when the number of countries grows large. However, a block bootstrapped version of the test can remedy the size distortion when the time series length divided by the number of countries *T/n* is sufficiently large. Applying the technique to a number of business cycle proxies of developed economies, we are unable to reject the null hypothesis of a non-zero common multivariate synchronization index for certain economically meaningful subsets of these countries.

## I. Introduction

An accurate measurement of the degree of business cycle synchronization enables policy-makers to assess the optimality and the survival probability of a monetary union and its accompanying common monetary policy.[1] Measuring the degree of

*The authors thank two anonymous referees as well as Michel Beine, Stefano Fachin, Stephan Smeekes and Jean-Pierre Urbain for comments on the paper. They have also benefited from discussions with the participants of the 'Frontiers in Time Series Analysis' Congress in Olbia, the Schumpeter Seminar at Humboldt Universität zu Berlin, the Money and Banking Research Day in Maastricht, the INFER Congress at the European Commission and the Congress on Business Cycles at Eurostat. The usual disclaimers apply.
JEL Classification numbers: C15, F15.

[1] See the recent surveys of Artis (2003) and de Haan, Inklaar and Jong-a-Pina (2005) for a more in-depth discussion on the policy implications of business cycle synchronization.

business cycle synchronization is a two-stage procedure because it requires the determination of business cycle phases prior to estimating the degree of cyclical synchronization. How to determine business cycle phases constitutes a quickly growing area in the empirical business cycle literature but this will not be the focus of the current paper.[2] We rather prefer to focus on the synchronization measurement issue using the periods of boom and recession as inputs. A myriad of alternative synchronization definitions and accompanying estimation procedures have been recently proposed. Croux, Forni and Reichlin (2001) study business cycle synchronization within the frequency domain framework and define synchronization as the ‘coherence’ within a particular frequency range. Comparable approaches to measuring business cycle synchronization are provided by Hughes-Hallet and Richter (2004) or Breitung and Candelon (2001). Another strand of literature defines synchronization as the phase shift between the stochastic cycles within a state space framework; see, e.g. Koopman and Azevedo (2003).

More recently, Harding and Pagan (2006) measured the amount of bivariate business cycle synchronization by Pearson-type correlations on binary variables representing business cycle busts and booms. Next, they proposed an asymptotic procedure for testing whether all bivariate correlations are equal to each other. The considered null hypothesis in their paper either takes the form of ‘perfect synchronization’ (all correlations equal to 1) or ‘perfect non-synchronization’ (all correlations equal to zero). The present paper builds further on their framework. First, we test the null hypothesis of perfect synchronization. If that null hypothesis is rejected, we test the null hypothesis of ‘imperfect synchronization’. This basically amounts to testing the cross-equality of all bivariate binary correlations for all possible restricted synchronization values strictly smaller than 1 (including 0). This seems a more realistic null hypothesis than testing against either a value of 0 or 1 only. To this aim, we can use the same test statistic that Harding and Pagan proposed for testing the null hypothesis of ‘perfect non-synchronization’: the limiting distribution stays the same as long as the value of the restricted synchronization index under the null hypothesis stays strictly below 1.

Moreover, and provided that the null hypothesis of imperfect synchronization is not always rejected, the test for imperfect synchronization produces an estimator for the multivariate synchronization index. First, the test renders the range of Generalized Method of Moments (GMM)-restricted synchronization estimates that do not lead to rejection of the cross-equality hypothesis for the bivariate correlations. Next, one can select the GMM estimate that minimizes the test statistic as ‘best attainable’ estimate (i.e. that leads to the strongest non-rejection) for the ‘imperfect multivariate synchronization index’.

Also, we show by Monte Carlo simulation that the asymptotic version of the synchronization test is biased towards rejection and that the size distortion increases with the number of countries considered. A bootstrap procedure for the small sample

[2] See Artis, Krolzig and Toro (2004) for a recent survey.

critical values (CV) of the imperfect synchronization test removes the size distortion nearly entirely provided the number of countries is not too big relative to the time series length of the business cycle indicators.

Anticipating our results, we find support for (imperfectly) synchronized business cycles for economically meaningful subsets of European and Anglosaxon countries. The results are found to be robust to changing the dating algorithms for the business cycle. Also, the point estimates for imperfect synchronization are relatively stable across subsamples.

The remainder of the paper is organized as follows. In section II, the Harding–Pagan framework for measuring and testing business cycle synchronization is briefly revisited. Section III contains a Monte Carlo investigation of the strong multivariate non-synchronization (SMNS) test for varying numbers of countries and time series lengths. A bootstrap algorithm for size-corrected CV is proposed in the same section. Empirical evidence on imperfect multivariate business cycle synchronization is provided in section IV. Conclusions are drawn in section V.

## II. Test of imperfect multivariate synchronization

We are interested in identifying the cross-country co-movements between phases of the business cycle. Consider $n$ countries with time series $X_{it}$ $(i=1,\dots,n; t=1,\dots,T)$ proxying time-varying economic activity. Business cycle booms or busts are dated using the marginal transform (or 'filter') $F(\cdot)$ such that $F(X_{it})=S_{it}$ $(\forall i)$ and where $S_{it}$ is 0 or 1 in case of bust or boom, respectively.[3] Examples of such dating algorithms will be discussed in the next section.

Sample means of the business cycle dummies and pairwise correlations can now be easily defined:

$$\begin{aligned} E(S_{it}) &= \mu_i, \\ \rho^{ij} &= \frac{E(S_{it}S_{jt})-\mu_i\mu_j}{\sqrt{\mu_i(1-\mu_i)\mu_j(1-\mu_j)}} \\ t&=1,\dots,T; \quad (i\times j)\in(1,\dots,n)^2. \end{aligned} \tag{1}$$

The sample means reflect the likelihoods that each of the countries' business cycles is in a boom state. The pairwise correlation $\rho^{ij}$ measures business cycle synchronization for countries $i$ and $j$. Finally, the expectation of the cross-product $E(S_{it}S_{jt})$ in the numerator of equation (1) reflects the likelihood of a simultaneous boom in both countries' economies.

Harding and Pagan (2006) propose a GMM-based procedure to test whether all bivariate correlations in equation (1) are equal to either 0 or 1. Let $\theta'=[\mu_1,\dots,\mu_n, \rho^{12},\dots,\rho^{n(n-1)}]$ represent the row vector of parameters to be estimated and $S_t$ be the

[3]For the sake of convenience, we allow for two regimes but the discussed techniques can be easily generalized.

$(n \times 1)$ vector of time $t$ business cycle dummies. Harding and Pagan start from the following set of $n(n+1)/2$ moment conditions in order to test for SMNS:

$$E[h_t(\theta, S_t)] = 0, \tag{2}$$

with

$$h_t(\theta, S_t) = \begin{bmatrix} S_{1t} - \mu_1 \\ \vdots \\ S_{1n} - \mu_n \\ \frac{(S_{1t}-\mu_1)'(S_{2t}-\mu_2)}{\sqrt{\mu_1(1-\mu_1)\mu_2(1-\mu_2)}} - \rho^{12} \\ \vdots \\ \frac{(S_{(n-1)t}-\mu_{n-1})'(S_{nt}-\mu_n)}{\sqrt{\mu_{n-1}(1-\mu_{n-1})\mu_n(1-\mu_n)}} - \rho^{(n-1)n} \end{bmatrix}. \tag{3}$$

The first subset of $n$ moment conditions in equation (2) defines the sample means of the cycle dummies whereas the second subset of $n(n-1)/2$ moment conditions characterizes estimates of all bivariate cycle correlations. The deviation from the moment conditions in equation (2) can now be easily determined by calculating the time series average of equation (3):

$$g(\theta, \{S\}_{t=1}^{T}) = \frac{1}{T}\sum_{t=1}^{T} h_t(\theta, S_t).$$

Let $\hat{\theta}' = [\hat{\mu}_1, \ldots, \hat{\mu}_n, \hat{\rho}^{12}, \ldots, \hat{\rho}^{n(n-1)}]$ be the vector of unrestricted parameter estimates and let $\theta_0' = [\mu_1, \ldots, \mu_n, 0, \ldots, 0]$ be the restricted vector under the null hypothesis of SMNS. A test statistic for the moment conditions in equation (2) boils down to:

$$W = \sqrt{T}g(\theta_0, \{S\}_{t=1}^{T})'\hat{V}^{-1}\sqrt{T}g(\theta_0, \{S\}_{t=1}^{T}) \xrightarrow{d} \chi^2_{(n-1)n/2} \tag{4}$$

(see Harding and Pagan, 2006, p. 70). The estimated covariance matrix $\hat{V}$ is a heteroskedasticity and autocorrelation consistent estimator of the covariance matrix of $\sqrt{T}g(\theta_0, S_{t=1}^{T})$; see, e.g. Newey and West (1987).[4]

On the other side of the spectrum, the moment conditions for testing strong perfect positive synchronization (SPPS) boil down to:

$$\begin{cases} ES_i = \mu_0 \\ E\left[\frac{(S_{jt}-\mu_j)'(S_{it}-\mu_i)}{\sqrt{\mu_j(1-\mu_j)\mu_i(1-\mu_i)}} - 1\right] = 0 \quad t = 1, \ldots, T; \quad (i \times j) \in (1, \ldots, n)^2. \end{cases} \tag{5}$$

Given that the cycle dummies' sample means can be interpreted as the country likelihoods for being in an expansionary phase, equality of these sample means constitutes a necessary condition for SPPS. This null hypothesis can be tested using

[4]The statistic is defined as a quadratic form in the penalty vector $g(\cdot, \cdot)$. The latter defines deviations from the moment conditions; the stronger the deviations from the moment conditions the more likely a rejection of the null hypothesis of SMNS becomes. The number of degrees of freedom of the limiting distribution in equation (4) equals the number of pairwise correlation restrictions to be tested.

equation (4) with $n-1$ degrees of freedom for the limiting distribution. The second subset of moment conditions reflects that the cross-country cycle correlations should all be equal to 1 under the null hypothesis of SPPS. It is now tempting to use the $W$-statistic [equation (4)] to test the joint hypothesis $\rho^{12}=\cdots=\rho^{n(n-1)/2}=1$. However, the asymptotic distribution turns out to be a weighted average of Chi-squared distributions where the weights have to be determined by simulation (see Gourieroux, Holly and Monfort, 1982).

We implement the SPPS test as a 'pretest' prior to identifying multivariate 'imperfect' synchronization. In the case where SPPS is rejected (nearly always as will be illustrated in the empirical section), we propose to test the 'weaker' null hypothesis that $\rho^{12}=\cdots=\rho^{n(n-1)}=\rho_0$ with $-1\leq\rho_0<1$. We call this null hypothesis strong multivariate synchronization of order $\rho_0$ (SMS($\rho_0$)) in the rest of the paper. The non-rejection of SMS($\rho_0$) provides evidence for the presence of a common index of synchronization.[5] Testing this null hypothesis is straightforward because the moment conditions, test statistic $W$ and limiting behaviour under SMS($\rho_0$) are analogous to the SMNS case.

In order to solve the problem of testing against an unknown value of $\rho_0$, we calculate the GMM test statistic in equation (4) for different values of $\rho_0$ $(-1\leq\rho_0<1)$. This grid search renders the interval of GMM estimates $[\rho_-,\rho_+]\subset]-1,1[$ that do not lead to rejection of the null hypothesis SMS($\rho_0$) at a prespecified nominal size. From this interval estimate, we select the GMM point estimate $\hat{\rho}_0$ that minimizes the test statistic $W$.[6] Or in formula form:

$$\hat{\rho}_0=\operatorname{argmin}_{\rho\in[\rho_-,\rho_+]}\sqrt{T}g(\theta_0,\{S\}_{t=1}^{T})'\hat{V}^{-1}\sqrt{T}g(\theta_0,\{S\}_{t=1}^{T}).$$

Before putting the estimator of the common synchronization index cum testing procedure to work in an empirical application we evaluate the small sample behaviour of the test. It may be that the asymptotic distribution of the $W$-test only poorly approximates the test's small sample behaviour when the number of countries grows large.

## III. Small sample properties of the SMS($\rho_0$) test

Previous papers (see, e.g. Christiano and den Haan, 1996; Koenker and Machado, 1999) already argued that existing asymptotic theory for GMM estimators may break down in small samples. In this section, we investigate whether the GMM-based asymptotic test for SMNS in equation (4) also suffers from small sample problems. We suspect that this problem might be more severe when the number of countries (cross-sectional dimension) grows large relative to the length of the economic time series.

[5]The existence of a common synchronization index does not necessarily imply the existence of a 'common business cycle'. Identification of the latter typically presupposes estimating a full-fledged structural model of business cycle synchronization whereas our approach has a reduced form character.

[6]More sophisticated optimization algorithms like, e.g. the Newton–Raphson technique could also be applied in order to determine the minimum point of the test statistic.

We set up a Monte Carlo experiment for four partly encompassing data-generating processes (DGP) that we consider sufficiently representative for the current business cycle literature (but that nevertheless differ in their sense of reality). Under the simplest DGP, we draw $(n \times 1)$ vectors $X_t$ $(t = 1, \ldots, T)$ from a multivariate standard normal distribution with unit marginal variance and equal pairwise correlation $\rho$. The second DGP models the business cycle as a random walk without drift, $X_t = X_{t-1} + u_t$, with disturbance vector $u_t$ drawn from a multivariate normal distribution.

For $DGP_3$, we choose an ARI(1) model with drift. Prior to simulation, the parameters are estimated using the growth rate $(\Delta x_t = \Delta \ln(X_t))$ of US industrial production:

$$\Delta x_t = 0.0042 + 0.435 \Delta x_{t-1} + u_t. \tag{6}$$

As for our final DGP ($DGP_4$), we use an IMA(1) process with drift[7]:

$$\Delta x_t = 0.0074 + \eta_t + 0.596 \eta_{t-1}. \tag{7}$$

For simulation purposes, we draw $u_t$ and $\eta_t$ from $N(0, 0.016)$.[8] Clearly the differing stationarity properties constitute the main distinguishing feature of the four DGP. The first DGP renders stationary draws without prior detrending (standard normal draws are stationary around a zero trend) whereas the three remaining DGP only become stationary after proper detrending.[9]

Next to generating data in different ways, we also consider two algorithms for dating the business cycle phases. First, the 'calculus rule' attaches a one to positive random draws and a zero otherwise. It is only meaningful to apply this dating rule to stationary (properly detrended) data, i.e. to $DGP_1$ or to the series in first differences $\Delta x_t$ generated under the remaining DGP. The reason is that we hardly get any negative values in $X_t$ for upward trending processes.

The second filter is the Bry and Boschan (BBQ) dating algorithm (1971) that can be applied to both the 'raw' and detrended data. This algorithm discriminates between periods of generalized upward and downward trends (identified as expansions and contractions, respectively). More specifically, the criterion locates turning points (peaks and troughs) that correspond to local maxima and minima of the series. The sequencing of peaks and troughs determines whether one goes from recession into expansion or vice versa. Loosely speaking, a peak/trough in the series $X$ for country $i$ $(i = 1, \ldots, n)$ occurs at time $t$ $(t = 1, \ldots, T)$ when $X_{it}$ reaches a local maximum (minimum) in a window of width of five quarters.[10] Or in formula form:

[7]We thank two anonymous referees for suggesting the two latter DGP.

[8]The disturbances' variance in the Monte Carlo experiments is chosen so that it lies close to the empirical residual variance in equations (6) and (7).

[9]Conform with the business cycle literature, the cycles defined by these DGP correspond to a 'classical' (prior to detrending) or a 'growth' (after detrending) business cycle.

[10]The choice of the window width is exogenously imposed and constitutes an educated guess of the complete cycle duration. The peak/trough dating may be slightly sensitive to this choice; however, the used window width is in accordance with previous business cycle literature; see, e.g. Watson (1994) or Harding and Pagan (2002).

$$\begin{cases} \text{peak at } t & \text{if } X_{i,t-2} < X_{i,t}, X_{i,t-1} < X_{i,t};\ X_{i,t} > X_{i,t+1}, X_{i,t} > X_{i,t+2}, \\ \text{trough at } t & \text{if } X_{i,t-2} > X_{i,t}, X_{i,t-1} > X_{i,t};\ X_{i,t} < X_{i,t+1}, X_{i,t} < X_{i,t+2}. \end{cases}$$

Such a rule actually detects changes in the slope of the process. It also ensures that phases (expansions and recessions) of the cycles have a minimum duration of two periods whereas the completed cycles have a minimum duration of five periods.[11]

The Monte Carlo investigation combines different DGP with different dating algorithms. First, the series in levels generated by the four DGP can all be dated with BBQ. As $DGP_1$ is stationary by definition, it can be coupled with the calculus rule. Finally, both the calculus rule as well as BBQ dating can be applied to the series in first differences generated by $DGP_3$ and $DGP_4$, i.e. after removal of stochastic and deterministic trends. This renders in total nine meaningful scenarios for choosing a DGP and an accompanying dating algorithm.

Without loss of generality we limit ourselves to analysing the size distortion of equation (4) under the null hypothesis of SMNS, i.e. $\rho^{ij} = 0$, for $i \neq j$.[12]

Figure 1 shows the small sample size (nominal size equal to 5%) of the asymptotic test for SMNS ($\rho = 0$) as a function of $n$ (the number of countries) and $T$ (the length of the time series). The horizontal axes allow for an upper bound of 10 countries and 1,000 time series observations. First, the outcomes do not seem to differ greatly across different DGP or dating algorithms. More importantly, however, the rejection rates reveal that size distortion grows rapidly with the number of countries and is only negligible in the bivariate case. In the worst case scenario of ten countries, the asymptotic GMM test nearly always rejects the null hypothesis of absence of synchronization, even with time series of 1,000 observations. To better grasp the intuition behind this outcome, notice that the number of moment conditions to be estimated in equation (3) using time series of fixed length $T$ grows more rapidly than the number of countries $n$, e.g. $(10 \times 9)/2 = 45$ moment conditions have to be estimated for a panel of $n = 10$ countries.[13]

Lengthening the time series removes at least part of the problem. Indeed, Figure 1 shows that the size distortion somewhat decreases for larger values of $T$ but this reduction is far from sufficient to talk about a 'size corrected' test. Additional Monte Carlo simulations for the ten countries' 'worst case' scenario revealed that one would need time series of at least 10,000 data points in order to have a non-distorted test statistic. However, these sample lengths are non-realistic in business cycle research:

[11] Hamilton's (1989) switching regime approach constitutes an important parametric alternative to the BBQ dating approach. However, combining the Hamilton dating approach with our synchronization framework did not radically alter the dating. Moreover, given the non-parametric bootstrap procedure that we implement to remedy the size distortion of the synchronization test, we decided to also date the business cycle phases in a non-parametric way for reasons of theoretical consistency. The results using the Markov switching filter are therefore not included in the paper but available upon request.

[12] Similar conclusions on small sample behaviour hold when simulating under the null hypothesis SMS($\rho_0$), with $\rho_0 \neq 0$. The latter simulations are therefore omitted for sake of space considerations but are available upon request.

[13] We suspect that the speed of convergence to the limiting $\chi^2_{(n-1)n/2}$ distribution may also be negatively influenced when $n$ increases because the variance of the limiting distribution rises with $n$, i.e. $\sigma^2(\chi^2_{n(n-1)/2}) = n(n-1)$.

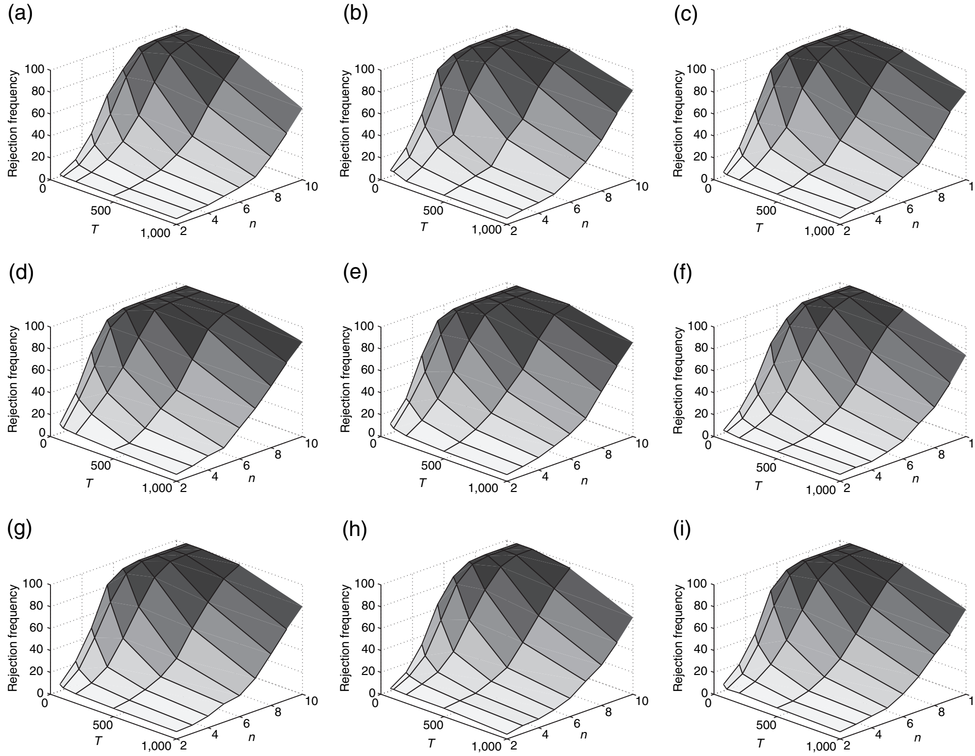


Figure 1. Size of the asymptotic SMS($\rho_0$) test. (a) $DGP_1$, calculus rule; (b) $DGP_1$, BBQ; (c) $DGP_2$, BBQ; (d) $DGP_3$, BBQ; (e) $DGP_4$, BBQ; (f) $DGP_3$, calculus rule, detrended; (g) $DGP_3$, BBQ, detrended; (h) $DGP_4$, calculus rule, detrended; (i) $DGP_4$, BBQ, detrended

*Notes*: $Z$-axis shows rejection frequency of the asymptotic version of SMS($\rho_0$) test under several DGP and dating rules for business cycles. $T$- and $n$-axes stand for the sample size $T$ and the number of countries $n$, respectively. SMS, strong multivariate synchronization; DGP, data-generating process; BBQ, Bry and Boschan algorithm

even switching to monthly data – if available – would not render sufficient time series information over the considered time period. We therefore propose to bootstrap the small sample distribution of the GMM test. The asymptotic pivotality of the $W$-test in equation (4) and its continuity in the parameter vector $\theta_0$ ensure proper convergence of the bootstrapped distribution function to the true asymptotic distribution; see, e.g. Horowitz (2001).

Apart from proper asymptotic convergence properties, we also want to know whether the bootstrapped CV are sufficiently size-corrected in small samples. Analogous to the preceding size study for the asymptotic test, we study the performance of the bootstrap under the null hypothesis of SMNS, i.e. $H_0 : \rho_0 = 0$.

The bootstrap algorithm we propose is non-parametric in nature (only the block length parameter has to be determined). Obviously, opting for either a parametric or a non-parametric bootstrap is to some extent a matter of taste and it is impossible to claim that one approach is strictly better than the other. Our viewpoint is that it is rather unrealistic to assume that researchers know the underlying DGP of the business cycle. Moreover, not imposing parametric business cycle models prevents the risk of misspecification. Also, popular dating algorithms like the BBQ algorithm are non-parametric (the window width being the only parameter to be chosen). Finally, instances like the National Bureau of Economic Research (NBER) or the Economic Cycle Research Institute (ECRI) typically publish data on business cycle phases without making public how they dated the cycles. Thus, if one decides to use NBER or ECRI data, it seems natural to perform non-parametric bootstraps because one can only obtain info on the 0-1 variables determining business cycle phases (we neither know the raw data originally used nor the dating algorithm).

The bootstrap is performed in blocks to account for the temporal persistence in 0-1 dummies describing the business cycle phases. There is no tailor-made solution to the determination of the optimal block length $w$. Hall and Sen (1999) suggest choosing $w = c.T^{1/4}$. However, the parameter $c$ is only known for certain specific parametric specifications and we just argued that we do not want to impose parametric business cycle models and dating procedures. We therefore opt for a heuristic approach (see step BO2 next): we iteratively determine the block length such that the average cycle duration in the original sample is preserved in the bootstrapped samples.[14]

In order to evaluate the bootstrap's small sample performance we implement the following multi-stage Monte Carlo cum bootstrap procedure:

1. *Step MC1*: Consider a particular DGP and a dating algorithm $F(\cdot)$. Draw a $T \times n$ matrix of standard normally distributed residuals $U$ with characteristic element $u_{i,t}^{MC}$. Use the residuals in order to build $n$ time series $X_{i,t}^{MC}$, with $t = 1, \dots, T$ and $i = 1, \dots, n$. This renders the $T \times n$ matrix $X$.
2. *Step MC2*: Use the dating algorithms to build a $T \times n$ binary variable matrix $S$ with characteristic element $F(X_{i,t}^{MC}) = S_{i,t}^{MC}$. Calculate the cross-sectional average number of business cycle phases $bc$.
3. *Step MC3*: Compute the simulated $W$-statistic for $\rho_0 = 0$, i.e. $W^{MC}(\rho_0 = 0)$ using $S$ as input.
4. *Step BO1*: Generate a $T \times n$ matrix of bootstrap replications $S^B$ by randomly drawing 25 consecutive $S_t$ with replacement.
5. *Step BO2*: Calculate the average number of business cycle phases in the bootstrapped sample $bc_{\text{boo}}$. If $bc_{\text{boo}} > 1.05bc$ (resp. $bc_{\text{boo}} < 0.95bc$), increase (resp. decrease) by 5 the consecutive $S_t$, which are drawn in BO1. Repeat BO1 and BO2, until $bc_{\text{boo}} \in [0.95bc, 1.05bc]$.[15]
6. *Step BO3*: Compute the bootstrapped $W$-statistic for $\rho_0 = 0$, i.e. $W^B(\rho_0 = 0)$ using $S^B$as input.
7. Repeat the bootstrap steps 4–6 a 'sufficient' number of times ($M$). $M$ is endogenously determined using the three-step method of Andrews and Buchinsky

[14]We are grateful to one of the referees for suggesting this improvement.
[15]Robustness checks with $bc \in [0.975bc, 1.025bc]$ have been performed and provide similar results.

(2000).[16] A CV is then obtained as the $\alpha$-percent quantile, say $W^B_{\text{crit}}$, from the empirical distribution of the bootstrap test statistic. The nominal size is set to $\alpha = 5\%$. The null hypothesis of SMS($\rho_0 = 0$) is rejected if $W^{MC}(\rho_0 = 0) > W^B_{\text{crit}}$.

8. Repeat steps 1–7 several times to obtain the rejection frequencies of the test $W(\rho_0 = 0)$.

Figure 2 shows the rejection rates of the bootstrap-based version of the test for SMNS ($\rho = 0$) as a function of $n$ (the number of countries) and $T$ (the length of the time series). We consider the same combinations of DGP and dating algorithm as in the previous figure and the nominal size is again set equal to 5%.

It turns out that the size distortion is greatly reduced as compared with the asymptotic version of the test for most combinations of $T$ and $n$. However, even the bootstrap no longer seems to be a valid remedy for the size distortion in case $T$ is small relative to $n$. For example, in the case of a sample size $T = 75$, considering a number of countries $n$ larger than six would still lead to over-rejection of the null hypothesis of SMNS ($\rho = 0$). Thus, this highlights the limits of our procedure when information is restricted, i.e. $T$ is relatively small with respect to $n$.

Figure 2 also reveals that size distortions are negligible when business cycles are dated using the calculus rule. Our intuition is that the magnitude of the size distortion also depends on the average business cycle length ($bc$).[17] More specifically, we expect that when the average length of the cycles is low (i.e. more numerous business cycle phases and thus more cyclical information in the data), the test's small sample distribution will converge more quickly to its asymptotic counterpart. To investigate this hypothesis, Figure 3 plots the size distortions for the asymptotic and bootstrap-based test against the average lengths of the business cycle for all nine experiments and for the cases ($n = 5, T = 200$) and ($n = 10, T = 75$).[18]

Figure 3 reveals that calculus rule dating typically generates shorter business cycles than the BBQ algorithm (using the calculus rule business cycles exhibit an average length of 6.36 quarters vs. 10.63 quarters with BBQ). Most importantly, however, we find a very strong positive relationship between the business cycle length and the asymptotic size distortion. The relationship disappears for the bootstrap with ($n = 5, T = 200$), but is still present for the bootstrap with ($n = 10, T = 75$). This illustrates once more that the bootstrap has its limits in removing the size distortion, especially when $T/n$ is relatively small.

Finally, we also investigate the small sample power of the bootstrap-based $W$-test. The power is size-corrected in the sense that we used the bootstrapped CV obtained from our algorithm. Power results are summarized in Figures 4 and 5 for two different alternative hypotheses.

The power is calculated under the null hypothesis of SMNS against the alternatives, $\rho = 0.10$ and $\rho = 0.25$. We find that the power is of acceptable magnitude, even

[16] The endogenous number of bootstrap replications $M$ was found to vary between 200 and 2,000.

[17] We thank an anonymous referee for providing us with this suggestion.

[18] For sake of space considerations, we only report these two cases but the positive link between asymptotic size distortion and business cycle length is found to hold for other combinations of $T$ and $n$ as well.

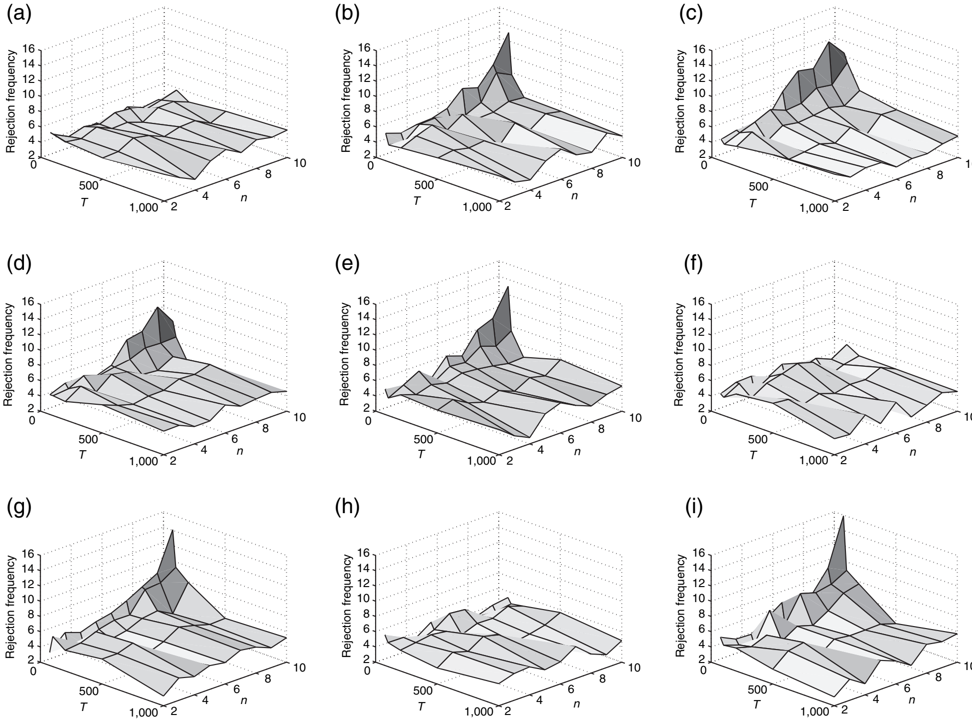


Figure 2. Size of the bootstrap SMS($\rho_0$) test. (a) $DGP_1$, calculus rule; (b) $DGP_1$, BBQ; (c) $DGP_2$, BBQ; (d) $DGP_3$, BBQ; (e) $DGP_4$, BBQ; (f) $DGP_3$, calculus rule, detrended; (g) $DGP_3$, BBQ, detrended; (h) $DGP_4$, calculus rule, detrended; (i) $DGP_4$, BBQ, detrended

*Notes*: $Z$-axis shows rejection frequency of bootstrap-based version of SMS($\rho_0$) under several DGP and dating rules for business cycles. $T$- and $n$-axes correspond with the sample size $T$ and the number of countries $n$, respectively. SMS, strong multivariate synchronization; DGP, data-generating process; BBQ, Bry and Boschan algorithm

for low values of $T$ and $n$. Moreover, the power does not differ much across different DGP and dating algorithms. As expected, we find the power to rise with the sample size $T$.[19]

## IV. Empirical application

We apply the business cycle synchronization framework to nine developed economies (seven European and two North American). In the spirit of Canova, Ciccarelli and Ortega (2007), we allow for possible cross-Atlantic differences by identifying the degree of business cycle synchronization for Europe and North America separately.[20]

[19]The power also rises when $T$ is small relative to $n$, which seems rather counterintuitive. However, given the earlier observation that the bootstrap is unable to mitigate the size distortion in this situation, the size-corrected power results where $T$ is small relative to $n$ cannot be trusted.

[20]Notice, however, that Canova *et al.* (2007) try to relate co-cyclicality to explanatory variables by means of a structural VAR analysis of business cycle synchronization. We limit ourselves to estimating reduced form estimators for business cycle synchronization.

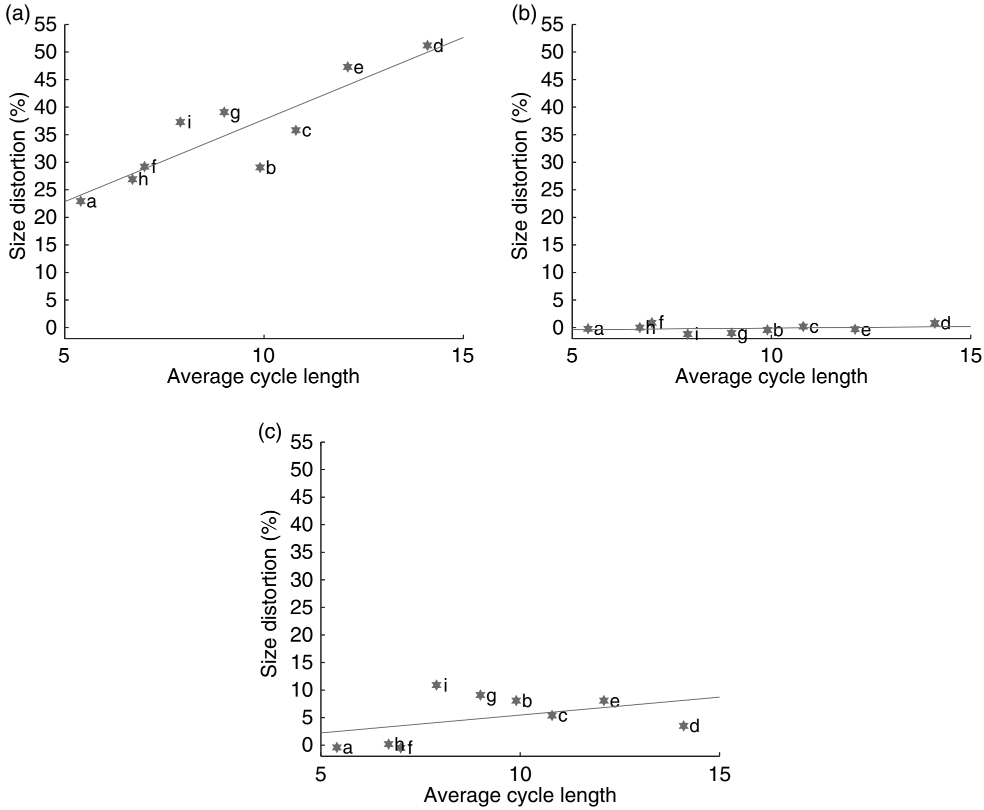


Figure 3. Size distortion and average business cycle length. (a) Asymptotic with $n=5$, $T=200$; (b) bootstrap with $n=5$, $T=200$; (c) bootstrap with $n=10$, $T=75$

*Notes*: The vertical (horizontal) axis reflects the size distortion (in per cent) and average cycle length, respectively. Stars and letters (a–i) identify the different data-generating process and dating rules. Letters are the same as in Figures 1 and 2. The figures also contain the ordinary least square regression lines relating (asymptotic, bootstrapped) size distortion and average business cycle length

The process of European economic integration (i.e. the creation of the 'internal market') and the growing importance of the euro suggests that a Eurocycle may exist that potentially differs from the rest of the world. On the other hand, it could as well be argued that more European integration, harmonization and the introduction of the single currency may have induced more transatlantic similarity.

Next to comparing business cycle synchronization across countries, country groups or continents, one may also question the business cycle synchronization measure's stability over time. Full sample and subsample results will therefore be reported to identify possible structural shifts in synchronization due to, e.g. the gradual process of economic globalization. Estimation results are complemented with a battery of tests. First, we apply the Harding–Pagan test for SPPS. Next, provided the former null hypothesis is rejected, we test the 'weaker' null hypothesis of SMS of order $\rho_0$ (SMS($\rho_0$)). We earlier argued that even the bootstrapped version of the SMS($\rho_0$) test

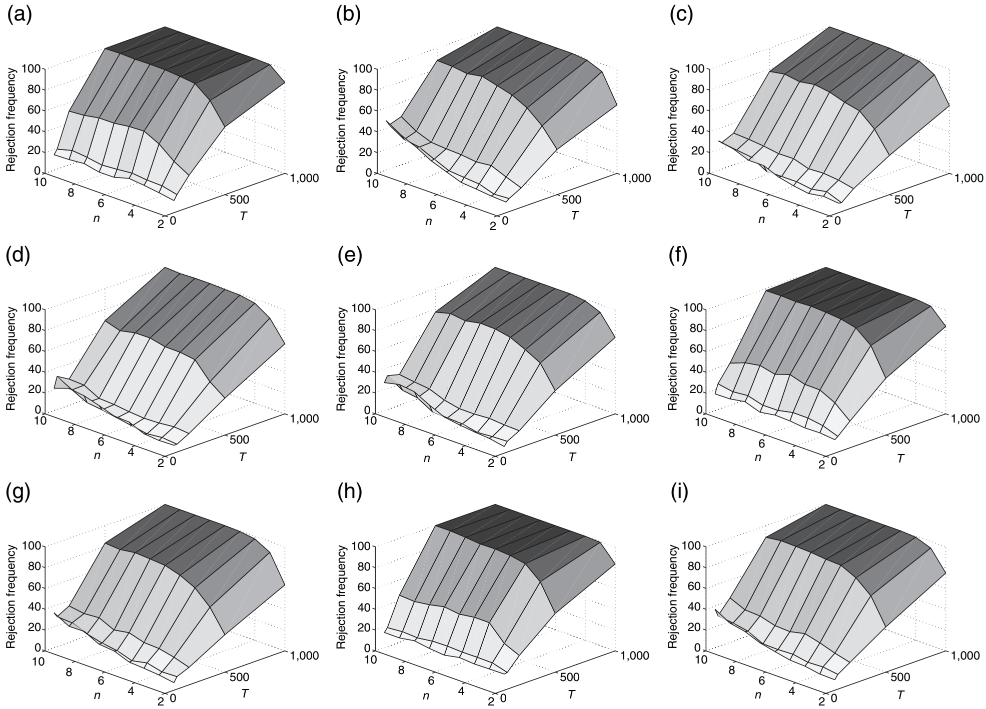


Figure 4. Power of the bootstrap SMS($\rho_0$) test against $\rho = 0.10$. (a) $DGP_1$, calculus rule; (b) $DGP_1$, BBQ; (c) $DGP_2$, BBQ; (d) $DGP_3$, BBQ; (e) $DGP_4$, BBQ; (f) $DGP_3$, calculus rule, detrended; (g) $DGP_3$, BBQ, detrended; (h) $DGP_4$, calculus rule, detrended; (i) $DGP_4$, BBQ, detrended

*Notes*: $Z$-axis shows rejection frequency of bootstrap version of SMS($\rho_0$) test under different DGP and dating rules. We simulate small sample power under the null hypothesis $\rho = 0$ of SMS($\rho_0$) when the true value of $\rho$ is 0.1. $T$- and $n$-axes show sample size $T$ and number of countries $n$, respectively. SMS, strong multivariate synchronization; DGP, data-generating process; BBQ, Bry and Boschan algorithm

gets increasingly size distorted for sets of countries bigger than six. Thus, it seems more prudent to test the null hypothesis SMS($\rho_0$) for 'economically meaningful' country groups of medium size. Indeed, even if the common synchronization hypothesis SMS($\rho_0$) is rejected for certain sets of countries, it may still hold for narrower subsets.

Business cycle data were downloaded from the ECRI and run from the first quarter of 1970 until the last quarter of 2007. This amounts to $T = 152$ quarterly observations.[21] Comparable with what the NBER publishes on US business cycles, the ECRI reports binary (0-1) business cycle information without specifying either the raw data used or the dating algorithm for determining business cycle peaks and troughs. In order to perform some sensitivity analysis, we also apply the synchronization framework to industrial production and (un)employment rates of the same countries. Deseasonalized

[21]The data are available on their website at http://www.businesscycle.com except for Belgium and the Netherlands.

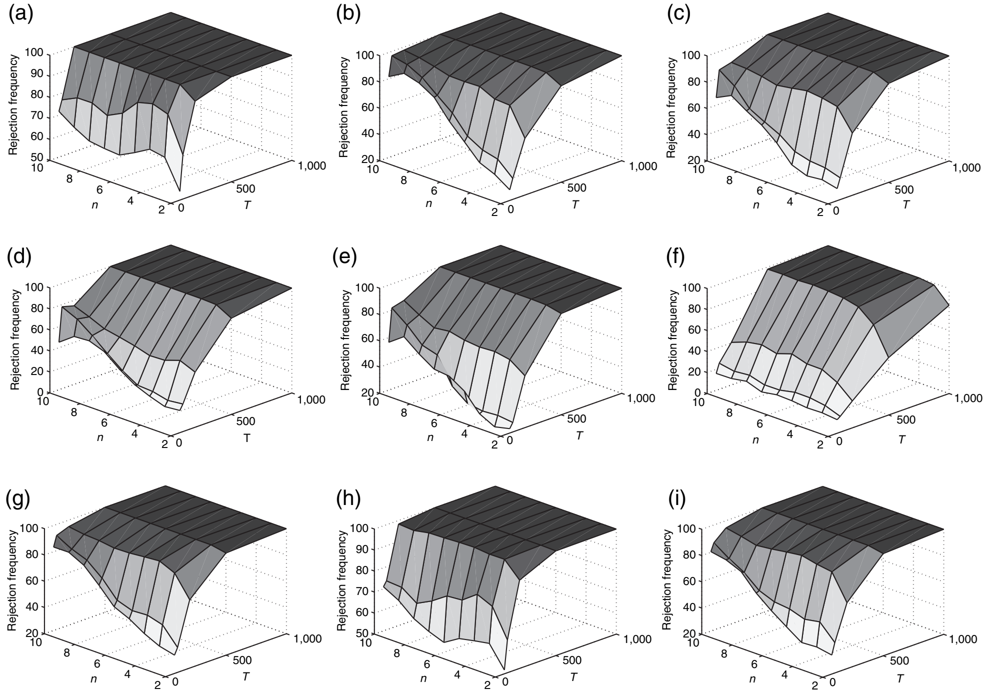

Figure 5. Power of the bootstrap SMS($\rho_0$) test against $\rho = 0.25$. (a) $DGP_1$, calculus rule; (b) $DGP_1$, BBQ; (c) $DGP_2$, BBQ; (d) $DGP_3$, BBQ; (e) $DGP_4$, BBQ; (f) $DGP_3$, calculus rule, detrended; (g) $DGP_3$, BBQ, detrended; (h) $DGP_4$, calculus rule, detrended; (i) $DGP_4$, BBQ, detrended

*Notes*: $Z$-axis shows rejection frequency of bootstrap version of SMS($\rho_0$) test under different DGP and dating rules. We simulate small sample power under the null hypothesis $\rho = 0$ of SMS($\rho_0$) when the true value of $\rho$ is 0.25. $T$- and $n$-axes show sample size $T$ and number of countries $n$, respectively. SMS, strong multivariate synchronization; DGP, data-generating process; BBQ, Bry and Boschan algorithm

industrial production, employment and unemployment figures are extracted on a quarterly basis from the IMF International Financial Statistics database over the period 1970–2007. Finally, we also downloaded US dollar-denominated and dividend-adjusted monthly stock market indices for the considered countries and for the same sample period.

The cycle dummies $S_{i,t}$ ($i = 1, \ldots, n; t = 1, \ldots, T$) are obtained either via the ECRI data set or by applying the BBQ dating algorithm to the raw time series.[22] We apply the BBQ algorithm on the series in levels (no detrending). Given the ongoing controversy on the nature and specification of trends, detrending might do more harm than good. Moreover, this ‘classical’ business cycle approach is in line with the companion paper by Harding and Pagan (2006).

[22]The exact dates of the estimated peaks and troughs for each series are not reported in a separate table but are available upon request.

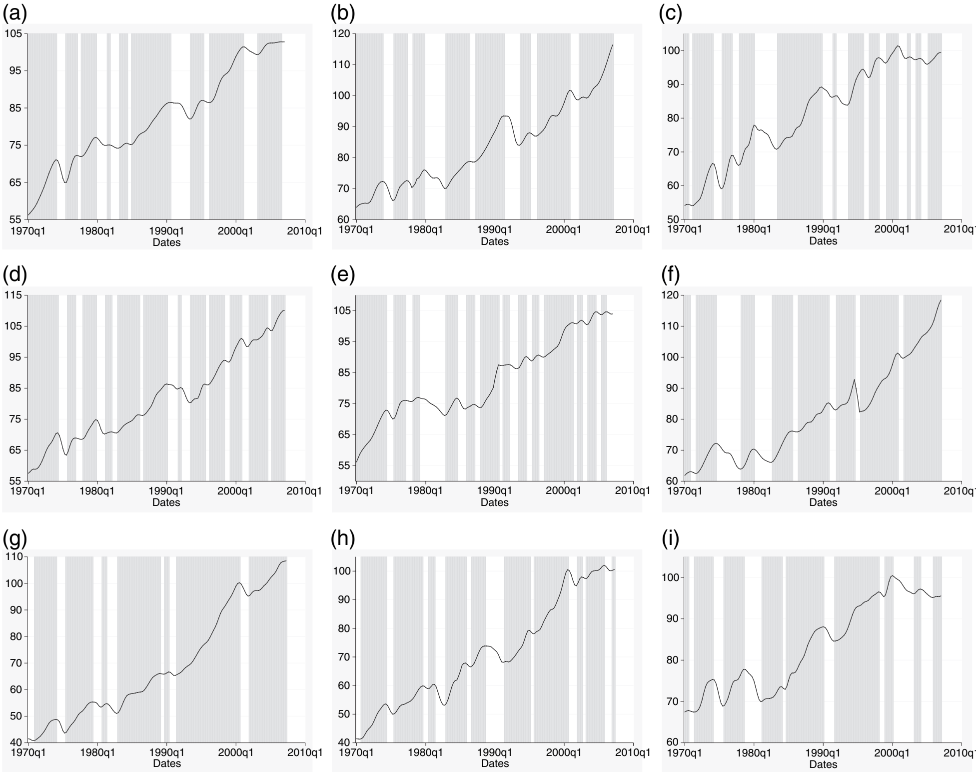


Figure 6. Dating business cycles for industrial production indexes. (a) France; (b) Germany; (c) Italy; (d) Belgium; (e) The Netherlands; (f) Sweden; (g) US; (h) Canada; (i) UK

*Notes*: Shaded areas correspond to expansionary phases. Underlying time series have been seasonally adjusted with Demetra 2.1 using the Tramo/Seats method

Figure 6 contains the evolution of the (log) industrial production for each of the countries, where the boom periods have been shaded to facilitate visual inspection.[23]

The figure provides some casual evidence for cross-country co-movement or 'synchronization' between boom and bust periods. The BBQ dating algorithm is able to detect the 'textbook' recessions of both oil crises in the seventies, the 90-91 recession and the negative real effects in the aftermath of the dotcom bubble burst and 9/11 (2000–2). However, differences in synchronization between North American and European countries are not visible with the naked eye. In order to assess the degree of synchronization and whether it differs across country sets (or whether it changes over time), we have to resort to the more advanced statistical tools introduced in the previous sections.

Table 1 reports bivariate correlations for ECRI business cycle data (below the diagonal) and industrial production indices or IPI (above the diagonal).[24]

[23]Graphs of stock prices, employment and unemployment series are omitted for sake of space considerations.

[24]Correlations for employment, unemployment and stock market data are not included in the paper for sake of space considerations.

TABLE 1

*Correlations for IPI and ECRI (1970–2007)*

| | *FRA* | *GER* | *ITA* | *BEL* | *NL* | *SWE* | *US* | *CAN* | *UK* |
|---|---|---|---|---|---|---|---|---|---|
| FRA | 1 | 0.35 | 0.41 | | | | 0.15 | 0.02 | 0.24 |
| GER | 0.61 | 1 | 0.54 | | | | 0.43 | 0.36 | 0.36 |
| ITA | 0.50 | 0.47 | 1 | | | | 0.32 | 0.28 | 0.21 |
| BEL | 0.52 | 0.52 | 0.61 | 1 | | | | | |
| NL | 0.41 | 0.35 | 0.34 | 0.33 | 1 | | | | |
| SWE | 0.24 | 0.36 | 0.25 | 0.34 | 0.25 | 1 | | | |
| US | 0.32 | 0.34 | 0.35 | 0.35 | 0.25 | 0.25 | 1 | 0.40 | 0.26 |
| CAN | 0.24 | 0.23 | 0.21 | 0.27 | 0.21 | 0.18 | 0.66 | 1 | 0.44 |
| UK | 0.26 | 0.06 | 0.33 | 0.33 | 0.06 | 0.11 | 0.34 | 0.14 | 1 |

*Notes*: Lower triangular values correspond to bivariate correlations based on the IPI. Upper triangular values correspond to the bivariate correlations based on the data from the ECRI. IPI, industrial production indexes; ECRI, Economic Cycle Research Institute; FRA, France; GER, Germany; ITA, Italy; BEL, Belgium; NL, The Netherlands; SWE, Sweden; US, United States; CAN, Canada; UK, United Kingdom.

ECRI correlations are lower and exhibit wider confidence intervals (not reported in table) than their IPI counterparts in a majority of the cases. The wider ECRI confidence intervals should not surprise given the fact that the ECRI data contain much less cyclical information than the IPI series (the ECRI cycle duration is found to be twice as high as the IPI cycle duration). We already showed in the simulation section that cycle duration also influences the size distortion of our asymptotic GMM test. Most importantly, however, the table shows that European and North American business cycle correlations tend to exceed cross-Atlantic business cycle correlations. This suggests that one should identify business cycle synchronization for separate groups of European and North American countries.

Turning to our multivariate framework, Table 2 reports different types of tests for multivariate synchronization. We include results for two business cycle proxies (ECRI and IPI), stock market prices, employment and unemployment series. We consider two subsets of European countries: the smallest subset ($E3$) solely consisting of France, Germany, and Italy whereas an extended country set ($E6$) consisting of $E3$ plus Sweden, Netherlands and Belgium. We do not include the UK in the European country sets because Table 1 suggests that the UK business cycle tends to be more strongly co-moving with the North American business cycle. As in Canova *et al.* (2007), we consider business cycle synchronization for a group of North American countries (Canada and the US or $A2$) possibly augmented with the UK ($A3$).[25]

The columns SPPS(i) and SPPS(ii) contain the test statistics of Harding and Pagan's (2006) test of SPPS. First, one tests whether the likelihoods of being in expansionary phases are equal across countries (SPPS(i)). Provided SPPS(i) is not rejected,

[25] It is less obvious that one should partition the countries in the same way for the other macroeconomic series; for sake of comparison, however, we keep the same country subsets for all macroeconomic series.

TABLE 2

*Multivariate synchronization: estimation/test results (1970–2007)*

| | SPPS | | | SMS($\rho_0$) | | | |
|---|---|---|---|---|---|---|---|
| *Countries* | *SPPS(i)* | *SPPS(ii)* | $W(\hat{\rho}_0)$ | $CV_B$ | $\hat{\rho}_0$ | $[\hat{\rho}_-; \hat{\rho}_+]$ | *PC (%)* |
| *Panel A: ECRI* | | | | | | | |
| *E*3 | 3.00 | 38.74*** | 1.73 | 63.27 | 0.47 | [−0.19; 0.99] | — |
| *A*2 | 0.91 | 6.34*** | — | 8.55 | 0.40 | [−0.01; 0.80] | — |
| *A*3 | 0.95 | 37.04*** | 7.41 | 61.77 | 0.44 | [−0.04; 0.91] | — |
| *Panel B: Industrial production* | | | | | | | |
| *E*3 | 2.79 | 48.68*** | 1.95 | 23.99 | 0.54 | [0.29; 0.80] | 80.98 |
| *E*6 | 6.85 | 402.22*** | 57.36 | 260.45 | 0.56 | [0.17; 0.95] | 63.31 |
| *A*2 | 3.09 | 5.23*** | — | 11.15 | 0.66 | [0.47; 0.84] | 88.31 |
| *A*3 | 5.89 | 72.00*** | 26.26 | 34.45 | 0.71 | [0.60; 0.82] | 76.16 |
| *Panel C: Employment* | | | | | | | |
| *E*3 | 13.06*** | — | 6.61 | 38.77 | 0.21 | [−0.13; 0.54] | 55.95 |
| *E*6 | 15.24*** | — | 79.37 | 656.2 | 0.21 | [−0.42; 0.84] | 49.07 |
| *A*2 | 1.09 | 8.43*** | — | 11.29 | 0.37 | [−0.14; 0.88] | 78.11 |
| *A*3 | 6.63*** | — | 0.60 | 70.54 | 0.47 | [0.10; 0.85] | 65.54 |
| *Panel D: Unemployment* | | | | | | | |
| *E*3 | 3.37 | 98.07*** | 7.56 | 23.04 | 0.39 | [0.07; 0.70] | 66.93 |
| *E*6 | 9.19 | 452.92*** | 74.89 | 371.92 | 0.55 | [−0.15; 0.99] | 59.38 |
| *A*2 | 0.24 | 20.03*** | — | 9.84 | 0.48 | [0.17; 0.79] | 83.96 |
| *A*3 | 0.95 | 68.31*** | 1.13 | 28.97 | 0.42 | [0.00; 0.83] | 69.64 |
| *Panel E: Stock markets* | | | | | | | |
| *E*3 | 9.15*** | — | 0.02 | 20.39 | 0.42 | [0.12; 0.72] | 70.28 |
| *E*6 | 9.95 | 559.59*** | 63.33 | 164.25 | 0.48 | [0.09; 0.87] | 63.98 |
| *A*2 | 0.59 | 29.26*** | — | 6.81 | 0.40 | [0.14; 0.66] | 87.05 |
| *A*3 | 0.72 | 96.35*** | 4.14 | 23.26 | 0.41 | [0.13; 0.69] | 76.19 |

*Notes*: The asymptotic critical values (95%) for the SMS($\rho_0$) test are: $n=2$: 3.84; $n=3$: 7.81; $n=6$: 25.00. The test statistics SPPS(i) and SPPS(ii) are calculated using equations (31) and (32) in Harding and Pagan (2006, p. 70). SPPS(ii) is only performed conditional upon rejection of SPPS(i). The SMS($\rho_0$) hypothesis is tested using equation (4) in this paper. The 95% bootstrapped critical value $CV_B$ (95%) is determined using the non-parametric block bootstrap procedure. If SMS($\rho_0$) is not rejected according to $CV_B$ (95%), we report the common synchronization index $\hat{\rho}$ and its confidence interval $[\rho_-; \rho_+]$. Asymptotic rejections of SPPS at the 1% significance level are indicated by ***. Whereas the SMS($\rho_0$) hypothesis is tested at the 5% significance level. PC refers to the percental contribution of the first principal component to the overall variation in the data. SPPS, strong perfect positive synchronization; SMS, strong multivariate synchronization; ECRI, Economic Cycle Research Institute.

one tests in a second stage whether the binary cyclical correlations in equation (1) are all equal to unity (SPPS(ii)). Thus, Table 2 only reports test statistics for SPPS(ii) in case SPPS(i) is not rejected.[26] According to Table 2, either the necessary condition for SPPS(i) or the sufficient condition SPPS(ii) is rejected at the 1% significance

[26]CV are not reported but are available upon request. The null hypothesis SPPS (ii) is strongly rejected with $p$-values well below 1%.

level in a majority of cases. This should not surprise because the requirement that macroeconomic cycles are perfectly synchronized is rather restrictive.

Next, we test the 'weaker' null hypothesis of SMS of order $\rho_0$, i.e. SMS ($\rho_0$). The GMM test outcomes and accompanying CV are also reported in Table 2.[27] The CV for the GMM test [equation (4)] is determined using the earlier described block bootstrap. Reported closed intervals $[\rho_-, \rho_+]$ contain all values of $\rho_0$ that lead to non-rejection of the null hypothesis of SMS($\rho_0$). The common synchronization parameter estimate ($\hat{\rho}_0$) minimizes the GMM test in equation (4) over a grid for $\rho = \rho_0$ ranging from $-0.99$ to 0.99 and provided the null hypothesis of SMS($\rho_0$) is not rejected, i.e. $W(\hat{\rho}_0) \leq \text{CV}(95\%)$.

In order to better grasp the relation between the bivariate and multivariate outcomes in Tables 1 and 2, consider, e.g. the index of multivariate sychronization $\hat{\rho} = 0.54$ for the industrial production series of the European 'core' countries $E3$. The corresponding bivariate correlations for this country trio (Table 1) are of the same order of magnitude. Thus, it should not surprise that the null hypothesis $\rho$[Italy, France] $= \rho$[Italy, Germany] $= \rho$[France, Germany] cannot be rejected: all three values fall within the confidence interval [0.29, 0.80] of the common synchronization index. Turning to the outcomes of the SMS($\rho_0$) test, the null hypothesis of a common synchronization index cannot be rejected in a majority of cases despite the dispersion of bivariate cycle correlations. Also, the polar case of SMNS, $\rho = 0$, falls in the non-rejection intervals $[\rho_-, \rho_+]$ for only 6 out of 19 cases. This provides additional justification to allow for the 'intermediate' case of 'imperfect' multivariate synchronization. The SMS($\rho_0$) outcomes also confirm that using asymptotic vs. bootstrapped CV may lead to different conclusions. For example, upon testing for IPI synchronization in $E6$ and $A3$, the presence of a common synchronization index is rejected by the asymptotic GMM test but not by its bootstrapped counterpart. This can be explained by the size distortion in the multivariate version of the asymptotic GMM test whereas the bootstrap mitigates the distortion problem for medium-sized sets of countries.

Upon comparing the results across country sets and time series, the ECRI and IPI outcomes are very different indeed. The multivariate synchronization index is lowest for the ECRI data and we cannot reject the null hypothesis of SMNS. The wideness of ECRI confidence intervals as compared with the IPI intervals may again be due to the relative lack of cyclical peaks and troughs in the ECRI data.[28] The differences in point estimates and confidence intervals are in line with our bivariate outcomes previously discussed.

As cross-Atlantic synchronization differences are concerned, Table 2 shows that industrial production and employment evolve in a more synchronous way for the North American country block $A3$ as compared with the European country sets $E3$

[27] Because it only makes sense to test for multivariate synchronization when $n > 2$, the columns for the $W$-test and the CV are left empty in the $A2$ row.

[28] The average business cycle length in ECRI data is 34.15, 44.4 and 42.9 months for $E3$, $A2$ and $A3$, respectively; whereas the average cycle lengths are approximately half for IPI data. Otherwise stated, industrial production contains much more cyclical information.

TABLE 3

*Multivariate synchronization: estimation/test results (1982–2007)*

| | *SPPS* | | | *SMS($\rho_0$)* | | | | |
|---|---|---|---|---|---|---|---|---|
| *Countries* | *SPPS(i)* | *SPPS(ii)* | $W(\hat{\rho}_0)$ | $CV_B$ | $\hat{\rho}_0$ | $[\hat{\rho}_-;\hat{\rho}_+]$ | *PC (%)* | *MC (P-value)* |
| *Panel A: ECRI* | | | | | | | | |
| *E*3 | 4.21 | 20.84*** | 3.77 | 122.83 | 0.56 | [−0.08; 0.99] | — | — |
| *A*2 | 0.53 | 1.34*** | — | 7.55 | 0.6 | [0.17; 0.99] | — | — |
| *A*3 | 1.75 | 7.04*** | NA | NA | NA | NA | — | — |
| *Panel B: Industrial production* | | | | | | | | |
| *E*3 | 2.46 | 44.58*** | 2.61 | 40.18 | 0.50 | [0.07; 0.92] | 73.37 | |
| *E*6 | 7.12 | 376.32*** | 74.74 | 538.26 | 0.59 | [0.09; 0.98] | 52.51 | |
| *A*2 | 6.53*** | — | — | 5.98 | 0.63 | [0.14; 0.99] | 88.07 | |
| *A*3 | 8.46*** | — | NA | NA | NA | NA | 72.22 | |

*Notes*: The asymptotic critical values (95%) for the SMS($\rho_0$) test are: $n=2$: 3.84; $n=3$: 7.81; $n=6$: 25.00. The test statistics SPPS(i) and SPPS(ii) are calculated using equations (31) and (32) in Harding and Pagan (2006, p. 70). SPPS(ii) is only performed conditional upon rejection of SPPS(i). The SMS($\rho_0$) hypothesis is tested using equation (4) in this paper. The 95% bootstrapped critical value $CV_B$ (95%) is determined using the non-parametric block bootstrap procedure. If SMS($\rho_0$) is not rejected according to $CV_B$ (95%), we report the common synchronization index $\hat{\rho}$ and its confidence interval $[\rho_-;\rho_+]$. Asymptotic rejections of SPPS at the 1% significance level are indicated by ***. Whereas the SMS($\rho_0$) hypothesis is tested at the 5% significance level. PC refers to the percentual contribution of the first principal component to the overall variation in the data. NA indicates that the SMS($\rho_0$) test cannot be performed due to the non-invertibility of $V$ in equation (4). SPPS, strong perfect positive synchronization; SMS, strong multivariate synchronization; ECRI, Economic Cycle Research Institute; MC, multiple correlation.

and $E6$. Moreover, the North American block with the UK ($A3$) exhibits higher business cycle and employment synchronization than without the UK, which seems to suggest that the UK is part of a North American cycle indeed.

On the contrary, the unemployment series of the broader group of European countries $E6$ seem to be more synchronous than the Anglosaxon unemployment series.[29] Finally, stock market synchronization is of comparable magnitude for the different country sets. This seems to confirm that (western) international financial markets have become strongly integrated over the last 25 years. Admittedly, these claims have to be taken with a grain of salt because the reported confidence intervals are relatively wide due to the size-corrected bootstrap, especially for the larger groups of countries.

As an additional robustness check, we also performed a multiple correlations (MC) test and a principal component (PC) analysis towards identifying comovements in macroeconomic time series using the raw data (see Anderson, 1984).[30] MC are found to be highly significant in all cases. The PC analysis determines the percentage

[29]The lower European employment synchronization may simply reflect that the demographic evolution in European countries is less homogenous than in North America. As for the differences in business cycle and unemployment synchronization, a full-fledged structural analysis of the underlying driving forces would be needed in order to say something meaningful about the observed cross-continental differences.

[30]We thank an anonymous referee for this valuable suggestion.

variation in the series PC that can be attributed to the first (corresponding to the largest eigenvalue). PC is found to be high in all cases. Moreover, high (low) values of synchronization seem to coincide with a high (low) value of PC. For example, the lowest synchronization index and the lowest value of PC are both observed for the employment series of the $E6$ country set.

Finally, it is often argued that the business cycle characteristics may be time varying.[31] Applying a structural change test to the Harding–Pagan framework as in Candelon, Piplack and Straetmans (2008) is non-desirable in this context because of the insufficient length of our time series and the lack of power of the stability test applied in that paper (for more details on the small sample behaviour of stability tests within the Harding–Pagan GMM framework, see Candelon *et al.*, 2008). As a second best approach, we limit ourselves to replicating Table 2 for the business cycle proxies ECRI and IPI and over the subsample (1982–2007), which amounts to $T = 104$ observations.

Table 3 provides us with broadly similar results as in Table 2, suggesting no significant differences between the full sample and the subsample synchronization indices. Subsample point estimates are slightly higher but full sample and subsample estimates still fall well in each other's confidence intervals. The shorter sample confidence intervals are wider by construction.

## V. Conclusion

In this paper, we proposed a GMM framework in the spirit of Harding and Pagan (2006) in order to measure the degree of multivariate synchronization between macroeconomic time series. Accurately measuring the degree of business cycle synchronization is of potential importance to policy-makers because of the theoretical link between business cycle synchronization and the sustainability (survival probability) of currency unions in the longer run. Moreover, policy-makers and regulators typically like to know the magnitude of business cycle synchronization and whether synchronization changed over time because of the potentially destabilizing effects of recession spillovers from one country to another (cf. the recent US subprime crisis and the resulting debate on the extent to which a possible US recession will impact the rest of the world).

Prior to calculating a measure of cyclical synchronization we classified macroeconomic time series into 'boom' and 'bust' periods using the BBQ dating algorithm (except for ECRI data that directly offers business cycle information). We subsequently applied the Harding and Pagan (2006) GMM framework but allowed for intermediate values of multivariate synchronization [SMS of order $\rho_0$ or SMS($\rho_0$)]. More specifically, we allowed for a value of the common synchronization index between $-1$ and 1 whereas Harding and Pagan only tested against the polar null hypotheses 'SPPS' or 'SMNS'. However, it is rather unlikely that cycles in real or

[31] See in particular Del Negro and Otrok (2004), Canova *et al.* (2007) and Giannone and Reichlin (2006).

financial variables are either perfectly synchronized or completely independent across countries. Moreover, our approach also produces an estimate of the multivariate synchronization index $\rho_0$ $(-1<\rho_0<1)$.

Before putting the test to work in an empirical application, we performed a Monte Carlo experiment for a variety of representative DGP in order to evaluate the (small sample) size and power properties of the procedure. We found that the asymptotic version of the SMS($\rho_0$) test potentially suffers from over-rejection. We illustrated that the size distortion becomes particulary severe when the number of time series (countries) considered is larger relative to the length of the time series. As a remedy to this problem, we proposed a block bootstrap procedure to determine the small sample CV of the GMM test. The bootstrap is purely non-parametric in that we do not make any parametric assumptions about the DGP of the business cycle. Moreover, the bootstrap is performed in blocks so as to preserve the persistence of the cycles and their co-cyclicality in the bootstrapped samples.

The bootstrap was shown to reduce the size distortion to a satisfactory level and to produce accurate CV provided the time series length $T$ is sufficiently long relative to the number of countries $n$. In the small sample power study of the bootstrap-based test, we made use of the small sample CV so as to obtain size-adjusted power values. We found that the test for imperfect multivariate synchronization of order $\rho_0$ (SMS($\rho_0$)) already exhibits satisfactory small sample power against alternatives for $\rho_0$.

We applied the business cycle synchronization framework to nine developed economies (seven European and two North American) and for a variety of macroeconomic time series. However, applying the bootstrap-based test to a full set of nine countries would not work as our simulation experiments have convincingly shown (the size distortion would not be sufficiently mitigated by the bootstrap algorithm). Instead, we applied the synchronization measures and tests to economically meaningful subsets of countries. Inspired by previous literature, we separately identified business cycle synchronization for European and North American countries. First, we were able to reject the Harding–Pagan test towards SPPS in a majority of cases. Next, we tested the weaker null hypothesis of SMS of order $\rho_0$ (SMS($\rho_0$)). Using bootstrapped CV, we found that a common multivariate synchronization index is justified for most of the considered country subsets. More specifically, we found that North American business cycle synchronization seems to dominate European business cycle synchronization (upon including the UK in the set of North American countries). Surprisingly, we found the opposite result for the employment figures. As for stock markets, their synchronization estimates seem to be of comparable magnitude on both sides of the Atlantic. This may be interpreted as evidence that international financial markets are well integrated. Also, European synchronization measures seem relatively robust to adding or deleting countries. Finally, we also investigated the robustness of our results over time, i.e. are the synchronization results time-varying? The first best thing would have been to implement stability testing procedures but these lack power for the sample sizes under consideration. We therefore limited ourselves to a comparison of subsample results that were found to lie close to their full sample counterparts.

*Final Manuscript Received: April 2009*

## References

Anderson, T. W. (1984). *An Introduction to Multivariate Statistical Analysis*, John Wiley and Sons, New York.

Andrews, D. W. K. and Buchinsky, M. (2000). 'A three-step method for choosing the number of bootstrap repetitions', *Econometrica*, Vol. 68, pp. 23–51.

Artis, M. (2003). 'Reflections on the optimal currency area (OCA) criteria in the light of EMU', *International Journal of Finance and Economics*, Vol. 8, pp. 297–307.

Artis, M., Krolzig, H. M. and Toro, J. (2004). 'The European business cycle', *Oxford Economic Papers*, Vol. 56, pp. 1–44.

Breitung, J. and Candelon, B. (2001). 'Is there a common business cycle? New insight from a frequency domain analysis', *DIW Vierteljahrsheft*, Vol. 3, pp. 331–338.

Bry, G. and Boschan, G. (1971). *Cyclical Analysis of Time Series: Selected Procedures and Computer Programs*, NBER, New York.

Candelon, B., Piplack, J. and Straetmans, S. (2008). 'On measuring synchronization of bulls and bears: the case of East Asia', *Journal of Banking and Finance*, Vol. 32, pp. 1022–1035.

Canova, F., Ciccarelli, M. and Ortega, E. (2007). 'Similarities and convergence in G-7 cycles', *Journal of Monetary Economics*, Vol. 54, pp. 850–878.

Christiano, L. and den Haan, W. (1996). 'Small sample properties of GMM for business cycle analysis', *Journal of Business and Economic Statistics*, Vol. 14, pp. 309–327.

Croux, C., Forni, M. and Reichlin, L. (2001). 'A measure of comovement for economic variables: theory and empirics', *Review of Economics and Statistics*, Vol. 83, pp. 232–241.

Del Negro, M. and Otrok, C. (2004). *Dynamic Factor Models with Time-Varying Parameters*, Manuscript, Federal Reserve Bank of Atlanta.

Giannone, D. and Reichlin, L. (2006). *Trends and Cycles in the Euro Area: How Much Heterogeneity and Should we Worry about It?*, European Central Bank Working Paper Series No. 595.

Gourieroux, C., Holly, A. and Monfort, A. (1982). 'Likelihood ratio test, Wald test, and Kuhn–Tucker test in linear models with inequality constraints on the regression parameters', *Econometrica*, Vol. 50, pp. 63–80.

de Haan, J., Inklaar, R. and Jong-a-Pina, R. (2005). *Will Business Cycles in the Euro Area Converge? A Critical Survey of Empirical Research*, CCSO Working Paper No. 08.

Hall, A. R. and Sen, A. (1999). 'Structural stability testing in models estimated by generalized method of moments', *Journal of Business and Economic Statistics*, Vol. 17, pp. 335–348.

Hamilton, J. D. (1989). 'A new approach to the economic analysis of nonstationary time series and the business cycle', *Econometrica*, Vol. 57, pp. 357–384.

Harding, D. and Pagan, A. R. (2002). 'Dissecting the cycle: a methodological investigation', *Journal of Monetary Economics*, Vol. 49, pp. 365–381.

Harding, D. and Pagan, A. R. (2006). 'Synchronization of cycles', *Journal of Econometrics*, Vol. 132, pp. 59–79.

Horowitz, J. (2001). 'The bootstrap in econometrics', in Heckman J. J. and Leamer E. E. (eds), *Handbook of Econometrics*, Elsevier Science B.V., pp. 3159–3228.

Hughes-Hallet, A. and Richter, C. (2004). *A Time-Frequency Analysis of the Coherence of US Business Cycle and the European Business Cycle*, CEPR Working Paper No. 4751.

Koenker, R. and Machado, J. A. F. (1999). 'GMM inference when the number of moment conditions is large', *Journal of Econometrics*, Vol. 93, pp. 327–344.

Koopman, S. J. and Azevedo, J. V. (2003). *Measuring Synchronization and Convergence of Business Cycles*, Tinbergen Institute Discussion Paper No. 052.

Newey, W. K. and West, K. D. (1987). 'A simple, positive semi-definite heteroscedasticity and autocorrelation consistent covariance matrix', *Econometrica*, Vol. 55, pp. 703–708.

Watson, M. W. (1994). 'Business cycle durations and postwar stabilization of the US economy', *American Economic Review*, Vol. 84, pp. 24–46.