How EuroCOIN is constructed

The target

The basic idea driving the construction of our coincident indicator is that the GDP is a good summary measure of economic activity. However, the GDP is affected by errors and noise disturbing the true underlying signal. Such disturbances are

- measurement errors
- local and sectoral shocks
- high-frequency, low-persistence movements

Our procedure is designed to clean the GDP from such disturbances.

The model

We model each series in the data set as the sum of two mutually independent components: the *idiosyncratic component* and the *common component*. The former is affected by shocks which are specific to that particular variable and have little or no effect on the other variables. The latter is affected by a small number of shocks which are the same for all the variables in the panel. Despite being driven by the same *common shocks*, the common components may differ to each other because of heterogeneous reactions to the shocks: responses may be small or large, positive or negative, immediate or delayed.

Formally, we represent the j-th series, x_j, as the sum

$$x_j = \chi_j + \xi_j,$$

where χ_j denotes the common component and ξ_j the idiosyncratic component. Moreover, we have $\chi_j = f_j(u_1, u_2, \ldots, u_q)$, where $u_1, u_2, \ldots, u_q$ are the q common shocks. The response functions f_j are linear in the present and lagged values of the shocks. For details and references see Appendix 1 on the GDFM.

The common component, in turn, can be represented as the sum of two independent terms: a cyclical term χ_j^C and a short-run term χ_j^S:

$$\chi_j = \chi_j^C + \chi_j^S;$$

χ_j^C is made up by smooth waves with long and medium-run periodicity, whereas χ_j^S is jagged and characterized by high-frequency volatility. This representation stems naturally from the so called "spectral representation" of a stationary series (see Appendix 2).

The EuroCOIN

Let the first variable in our panel be the European real GDP. The EuroCOIN is χ_1^C, i.e. the cyclical, common component of the European GDP.

Why smoothing the GDP by eliminating the short-run part of the common component? Monthly data are typically affected by large seasonal and higher frequency sources of variation. Both economic agents and policy makers are not particularly interested in such high-fequency changes because of their transitory nature. Washing out temporary oscillations is necessary to unveil the true underlying long-lasting tendency of the economic activity.

Why cleaning the GDP from the idiosyncratic component? We have two reasons for this: eliminating measurement errors and produce a better signal for policy makers. Firstly, national GDPs are not obtained by means of direct observation, but are at least in part the result of estimation procedures and therefore are affected by estimation errors. Moreover, data on GDP are provided quarterly by the statistical institutes; monthly figures can only be obtaind by interpolating original data, which entails additional errors. Finally, the European GDP stems from aggregation of data provided by heterogeneous sources, not all equally reliable and perfectly comparable. Summing up, the European GDP is affected by large measurement and estimation errors. Such errors are mainly idiosyncratic, since they are poorly correlated across different variables and independent from the common shocks. Secondly, the idiosyncratic component should capture both variable-specific shocks, such as shocks affecting, say, the output of a particular industrial sector, and local-specific shocks, such as for instance a natural disaster, having possibly large but geographically concentrated effects. Distinguishing between such shocks and common shocks, affecting all sectors and areas, can be useful for policy makers, who have to decide whether to carry out local and sectoral measures or common, Europe-wide interventions.

The estimation Procedure

The common and the idiosyncratic components, as well as the cyclical and the short-run components, are all unobservable variables. Hence, having defined the theoretical EuroCOIN, we have to estimate it. The estimation procedure is rather technical, but the basic idea is quite simple. The idiosyncratic components are (almost) independent of each other, and we have a huge number of variables in the cross-section. Hence if we take the cross-sectional average, positive and negative values will almost compensate by a Large-Numbers effect, and the result will be very small. By contrast, the common components, being highly correlated, will survive aggregation. Hence, by taking the average of the observable variables, we can wash out the idiosyncratic component and obtain (a combination of) the common factors. As an example, consider the simple case $x_j = u + \xi_j$. The average is

$$\bar{x} = \frac{1}{n}(x_1 + \cdots + x_n) = u + \frac{1}{n}(\xi_1 + \cdots + \xi_n),$$

where n denotes the number of variables in the cross-section. Assuming for simplicity that both u and each one of the ξ_j's have unit variance, so that for each variable the idiosyncratic component is as large as the common component, the variance of the average will be

$$1 + \frac{1}{n^2}n = 1 + \frac{1}{n}.$$

Hence with 1000 variables in the panel only 0.1 per cent of total variance is idiosyncratic and $\bar{x}$ is almost identical to u.

Taking the simple average is not the only way to eliminate the idiosyncratic components. Other aggregates such as weighted averages or other linear combinations can do the same job. This is useful because when we have q common factors we need q different averages to get them. On the other hand, an efficiency problem arise: we have to choose these averages in such a way to minimize the residual idiosyncratic components. Our estimation procedure is designed to solve this problem.

The full estimation procedure is in three steps:

- Estimating the covariances (Appendix 3) of the unobservable components
- Estimating the static factors (Appendix 4)
- Estimating the cyclical common components (Appendix 5)

Finally, data become available with different delays. Had we to wait until the last updating arrives, we would be able to compute the indicator only with a delay of four or five months. Luckily, we have a procedure to handle this end-of-sample unbalance (see Appendix 6 for details) which enable us to get provisional estimates by exploiting, for each series, the most updated information. Once the missing data become available, the final estimate is computed. This is why the indicator is subject to revision for a few months.

Appendix 1: The GDFM

The theoretical model underlying the method is the Generalized Dynamic Factor Model proposed in Forni, Hallin, Lippi and Reichlin (2000) and Forni and Lippi (2001), and further developed in Forni, Hallin, Lippi and Reichlin (2001a) and Forni, Hallin, Lippi and Reichlin (2001b). The model is related to Stock and Watson (1999) and can be seen both as a large-panel generalization of the dynamic factor model introduced by Sargent and Sims (1977), Geweke (1977) and Geweke and Singleton (1982), and as a dynamic generalization of the approximated factor model proposed by Chamberlain (1983) and Chamberlain and Rothschild (1983). Here we shall limit ourselves to a short presentation of the model, with a special emphasis placed on the issues related to the indicator; for further details we refer to the papers mentioned above.

We assume that our j-th time series, suitably transformed, is a realizations from a zero mean, wide-sense stationary process x_{jt}. Since the model is specifically designed to handle a large cross-section of time series, each process in the panel is thought of as an element from an infinite sequence, so that $j = 1, \ldots, \infty$. Moreover, all of the x's are co-stationary, i.e. stationarity holds for the n-dimensional vector process $\mathbf{x}_{nt} = (x_{1t}, \ldots, x_{nt})'$, for any n.

As in the traditional dynamic factor model, each variable is represented as the sum of two mutually orthogonal unobservable components: the 'common component', χ_{jt}, and the 'idiosyncratic component', ξ_{jt}. The common component is driven by a small number, say q, of 'factors' or shocks, u_{ht}, $h = 1, \ldots, q$, which are identical for all of the cross-sectional units. By contrast, the 'idiosyncratic components' are driven by specific shocks.

More formally, we assume

$$x_{jt} = \chi_{jt} + \xi_{jt} = \mathbf{b}_j(L)\mathbf{u}_t + \xi_{jt} = \sum_{h=1}^{q} b_{jh}(L)u_{jt} + \xi_{jt} \tag{1}$$

where χ_{jt} is the common component, $\mathbf{u}_t = (u_{1t}, \ldots, u_{qt})'$ is the vector of the common shocks, i.e. a (covariance stationary) q-vector process with non-singular spectral density matrix, and the idiosyncratic component ξ_{jt} is orthogonal to $\mathbf{u}_{t-k}$ for any k and j.

Here we specialize somewhat the GDFM by assuming that the impulse-response function $b_{jh}(L)$, $h = 1, \ldots, q$, is a s-order polynomial in the lag operator, i.e. $b_{jh}(L)u_{ht} = b_{jh0}u_{ht} + b_{jh1}u_{ht-1} + \cdots + b_{jhs}u_{ht-s}$. Notice that the coefficients $b_{jh1}, \ldots, b_{jhs}$ can differ across both j and h. Hence the model is quite flexible, in that the reaction of each variable to a given common shock may be small or large, negative or positive, immediate or delayed. Moreover, a variable can react with a given impulse-response profile, say, to shock 1 and with a completely different profile to shock 2. This can accomodate a very wide range of different behaviors of the common components χ_{jt}, $j = 1, \ldots, \infty$. In particular, with reference to the delay with which the shocks are loaded, some of them will be 'leading' with respect to the European GDP, some will be 'coincident' and some will be 'lagging'. Estimating the model enables us to see whether there are indicators (or countries) which anticipate the changes of the European GDP and to unveil the lead-lag relations within the variables in the system.

The following example can be useful to get some intuition about this. Assuming just a single shock u_t, the four dynamic loadings 1, L, -1 and $-L$ would characterize pro-cyclical and leading, pro-cyclical and lagging, counter-cyclical and leading, counter-cyclical and lagging variables. Notice that the leading variables are completely unpredictable given information at time t. The lagging variables, which are unpredictable by means of univariate modeling, can be predicted perfectly by using the leading ones. In practical situations, common shocks are more than one and dynamic responses are not so simple, so that we shall need specific criteria in order to classify variables as counter-cyclical or leading. Correspondingly, the relation between forecasting ability and "leadership" is less obvious. Nevertheless, the example provides a good intuition of the reasons why the model can perform well in forecasting.

In the traditional dynamic factor model, the idiosyncratic components are mutually orthogonal at all leads and lags. The GDFM is a generalization of the traditional model in that this orthogonality assumption is relaxed. From this point of view, the GDFM is similar to the approximate factor model of Chamberlain (1983) and Chamberlain and Rothschild (1983), the main difference being that the GDFM is dynamic.

When removing orthogonality we are left without any theoretical distinction between the common and the idiosyncratic components. In place of orthogonality, we therefore need something else. In the present context we need precisely the assumptions listed in Forni, Hallin, Lippi and Reichlin, 2001b, namely that the largest eigenvalue of the variance-covariance matrix of the vector $\xi_{nt} = (\xi_{1t}, \ldots, \xi_{nt})'$ is bounded as $n \to \infty$, whereas the $q(s+1)$ largest eigenvalues of $\chi_{nt} = (\chi_{1t}, \ldots, \chi_{nt})'$ are not. For further details we refer to that paper. Here let us simply say that bounded eigenvalues for the ξ's do not rule out cross-sectional correlation, but put limits on it. On the other hand, unbounded eigenvalues for the χ's ensure, so to speak, a minimum amount of correlation between the common components. Under these assumptions, the idiosyncratic components are such that the simple cross-sectional average vanishes in variance as $n \to \infty$, just as if they were pairwise orthogonal, while this property does not hold for the common components.

Appendix 2: The Spectral Representation

Any stationary variable can be represented as the integral of waves of different frequency, each having a random amplitude; this is the so called "spectral representation". In particular, the common component of variable j at time t, χ_{jt}, can be written as

$$\chi_{jt} = \int_{-\pi}^{\pi} e^{i\theta t} dZ_j(\theta), \tag{2}$$

where $dZ_j(\theta)$ is an "orthogonal increment process" such that $\text{cov}(dZ_j(\theta), dZ_h(\lambda)) = 0$ for $\lambda \neq \theta$ (see e.g. Brockwell and Davis, Chap. 4). By taking the integral over a specific frequency band, we can isolate the waves of a particular periodicity. In particular, we disentangle a cyclical, medium- and long-run, component, and a non-cyclical, short-run, component, by aggregating respectively waves of frequency smaller than, or larger than, a given critical frequency θ^*. Denoting with C the interval $[-\theta^*, \theta^*]$ and by S the set $[-\pi, \theta^*] \cup [\theta^*, \pi]$, we have

$$\chi_{jt} = \int_C e^{i\theta t} dZ_j(\theta) + \int_S e^{i\theta t} dZ_j(\theta) = \chi_{jt}^C + \chi_{jt}^S. \tag{3}$$

The cross-spectral density of χ_{jt}^C and χ_{ht}^S is zero everywhere, so that they are orthogonal at any lead and lag. Denoting with $S_{jh}(\theta)$ the spectral-density function of χ_{jt} and χ_{ht}, the cross-spectral density of χ_{jt}^C and χ_{ht}^C is equal to $S_{jh}(\theta)$ in C and zero elsewhere.

An equivalent representation of the components χ_{jt}^C and χ_{jt}^S is obtained by applying to χ_{jt} the pass-band filter corresponding to the selected frequency interval (see, e.g. Sargent, 1987, or Baxter and King, 1999), i.e.

$$\chi_{jt} = \chi_{jt}^C + \chi_{jt}^S = d^C(L)\chi_{jt} + d^S(L)\chi_{jt}, \tag{4}$$

where $d^S(L) = 1 - d^C(L)$ and $d^C(L)$ is the two-sided, symmetric, infinite-order, square-summable filter whose k-th coefficient is

$$d_k^C = \begin{cases} \dfrac{1}{\pi k} \sin(k \cdot \theta^*) & \text{for } k \neq 0 \\[2ex] \dfrac{\theta^*}{\pi} & \text{for } k = 0 \end{cases}$$

Appendix 3: Estimating the covariances

In the first step of our procedure, we estimate the spectral-density matrix of the common components. The theoretical basis of such procedure can be found in Forni, Hallin, Lippi and Reichlin (2000). Consistency results for the entries of this matrix as both n and T go to infinity can easily be obtained from the results in that paper. For results on consistency rates see Forni, Hallin, Lippi and Reichlin (2001a). From the estimated spectral-density matrices we can obtain all the auto-covariances and cross-covariances at all leads and lags by applying the inverse Fourier transform. Moreover, we can easily get the covariances for the cyclical and the non-cyclical components χ_{jt}^C and χ_{jt}^S simply by applying such transformation to the selected frequency band of the estimated spectra and cross-spectra.

We start by estimating the spectral-density matrix of the observable $\mathbf{x}_t = \begin{pmatrix} x_{1t} & \cdots & x_{nt} \end{pmatrix}'$. Let us denote the population matrix by $\mathbf{\Sigma}(\theta)$ and its estimate by $\hat{\mathbf{\Sigma}}(\theta)$. The estimation is accomplished by using a Bartlett lag-window of size $M = 18$, i.e. by computing the sample auto-covariance matrices $\hat{\mathbf{\Gamma}}(k)$, multiplying them by the weights $w_k = 1 - \frac{|k|}{M+1}$ and applying the discrete Fourier transform:

$$\hat{\mathbf{\Sigma}}_x(\theta) = \frac{1}{2\pi} \sum_{k=-M}^{M} w_k \cdot \hat{\mathbf{\Gamma}}(k) \cdot e^{-i\theta k}.$$

The spectra were evaluated at 101 equally spaced frequencies in the interval $[-\pi, \pi]$, i.e. at the frequencies $\theta_h = \frac{2\pi h}{100}$, $h = -50, \ldots, 50$.
Then we obtain the dynamic principal component decomposition (see Brillinger, 1981). For each frequency of the grid, we compute the eigenvalues and eigenvectors of $\hat{\mathbf{\Sigma}}(\theta)$. By ordering the eigenvalues in descending order for each frequency and collecting values corresponding to different frequencies, the eigenvalue and eigenvector functions $\lambda_j(\theta)$ and $\mathbf{U}_j(\theta)$, $j = 1, \ldots, n$, are obtained. The function $\lambda_j(\theta)$ can be interpreted as the (sample) spectral density of the j-th principal component series and, in analogy with the standard static principal component analysis, the ratio

$$p_j = \int_{-\pi}^{\pi} \lambda_j(\theta) d\theta / \sum_{j=1}^{n} \int_{-\pi}^{\pi} \lambda_j(\theta) d\theta$$

represents the contribution of the j-th principal component series to the total variance in the system. Letting $\mathbf{\Lambda}_q(\theta)$ be the diagonal matrix having on the diagonal $\lambda_1(\theta), \ldots, \lambda_q(\theta)$ and $\mathbf{U}(\theta)$ be the $(n \times q)$ matrix $\begin{pmatrix} \mathbf{U}_1(\theta) & \cdots & \mathbf{U}_q(\theta) \end{pmatrix}$, our estimate of the spectral density matrix of the vector of the common components $\boldsymbol{\chi}_t = \begin{pmatrix} \chi_{1t} & \cdots & \chi_{nt} \end{pmatrix}'$ is given by

$$\hat{\mathbf{\Sigma}}_\chi(\theta) = \mathbf{U}(\theta)\mathbf{\Lambda}(\theta)\tilde{\mathbf{U}}(\theta) \tag{5}$$

where tilde denotes conjugation and transposition. Given the correct choice of q, consistency results for the entries of this matrix as both n and T go to infinity can easily be obtained from Forni, Hallin, Lippi and Reichlin (2000). Results on consistency rates can be found in Forni, Hallin, Lippi and Reichlin (2001a).

Following Forni, Hallin, Lippi and Reichlin (2000), we identify the number of common factors q by requiring a minimum amount of explained variance: for our dataset we selected $q = 4$.

An estimate of the spectral density matrix of the vector of the idiosyncratic components $\boldsymbol{\xi}_t = \begin{pmatrix} \xi_{1t} & \cdots & \xi_{nt} \end{pmatrix}'$ is then obtained as the difference $\hat{\boldsymbol{\Sigma}}_{\xi}(\theta) = \hat{\boldsymbol{\Sigma}}(\theta) - \hat{\boldsymbol{\Sigma}}_{\chi}(\theta)$.

Starting from the estimated spectral-density matrix we obtain estimates of the covariance matrices of $\boldsymbol{\chi}_t$ at different leads and lags by using the inverse discrete Fourier transform, i.e.

$$\hat{\boldsymbol{\Gamma}}_{\chi}(k) = \frac{2\pi}{101} \sum_{h=-50}^{50} \hat{\boldsymbol{\Sigma}}_{\chi}(\theta_h) e^{i\theta_h k}.$$

Moreover, we compute estimates of the covariance matrices of the medium- and long-run component $\boldsymbol{\chi}_t^C = (\chi_{1t}^C, \ldots, \chi_{nt}^C)'$ by applying the inverse transform to the frequency band of interest, i.e. $[-\theta^*,\ \theta^*]$, where $\theta^* = 2\pi/24$, so that all periodicities shorter that two years are cut off. Precisely, letting $\boldsymbol{\Gamma}_{\chi^C}(k) = \mathrm{E}\left(\boldsymbol{\chi}_t^C \boldsymbol{\chi}_{t-k}^{C\ \prime}\right)$, the corresponding estimate will be

$$\hat{\boldsymbol{\Gamma}}_{\chi^L}(k) = \frac{2\pi}{2H+1} \sum_{h=-H}^{H} \hat{\boldsymbol{\Sigma}}_{\chi}(\theta_h) e^{i\theta_h k},$$

where H is defined by the conditions $\theta_H \leq 2\pi/24$ and $\theta_{H+1} > 2\pi/24$.

Appendix 4: Estimating the static factors

In the second step, we compute an estimate of the static factors, following Forni, Hallin, Lippi and Reichlin (2001b). With the term "static factors" we mean the $q(s+1)$ variables appearing contemporaneously in representation (factor), including the lagged u_t's, so that, say, u_{1t} and u_{1t-1} are different static factors. To be precise, the static factors are not identified in the model unless we introduce additional assumptions, so that we shall in fact estimate a vector of linear combinations of such factors, say $\mathbf{v}_t$, spanning the same information space. Such estimates, say $\hat{\mathbf{v}}_t$, are obtained as the generalized principal components of the x's, a construction which involves the (contemporaneous) variance-covariance matrices of the common and the idiosyncratic components estimated in the first step. Such generalized principal components have an important "efficiency" property: they are the contemporaneous linear combinations of the x's with the smaller idiosyncratic-common variance ratio. As shown in the paper quoted above, they can consistently approximate any point in the common-factor space, including the common components χ_{jt}'s, as $n, T \to \infty$ at a suitable rate. Similarly, we can get forecasts of the common components (and the factors themselves) by projecting χ_{jt+k} (or the k-th lead a factor) on $\hat{\mathbf{v}}_t$. This forecast approximates consistently the theoretical projection.

Starting from the covariances estimated in the first step, we estimate the static factors as linear combinations of contemporaneous values of the observable variables x_{jt}, $j = 1, \ldots, n$.

Indeed, as observed above, the static factors appearing in representation (1), i.e. u_{ht-k}, $h = 1, \ldots, q$, $k = 1, \ldots, s$, are not identified without imposing additional assumptions and therefore cannot be estimated. This however is not a problem, since we need only a set of $r = q(s+1)$ variables forming a basis for the linear space spanned by the u_{ht}'s and their lags. We can then obtain $\hat{\chi}_{jt}$ by projecting χ_{jt} on the factors and $\hat{\chi}_{jt}^C$ by projecting χ_{jt}^C on leads and the lags of the factors.

Our procedure consists in taking the first r generalized principal components of $\hat{\mathbf{\Gamma}}_\chi(0)$ with respect to the diagonal matrix having on the diagonal the variances of the idiosyncratic components ξ_{jt}, $j - 1, \ldots, n$, denoted by $\hat{\mathbf{\Gamma}}_\xi(0)$. Precisely, we compute the generalized eigenvalues μ_j, i.e. the n complex numbers solving $\det(\mathbf{\Gamma}_\chi^T(0) - z\hat{\mathbf{\Gamma}}_\xi(0)) = 0$, along with the corresponding generalized eigenvectors $\mathbf{V}_j$, $j = 1, \ldots, n$, i.e. the vectors satisfying

$$\mathbf{V}_j \hat{\mathbf{\Gamma}}_\chi(0) = \mu_j \mathbf{V}_j \hat{\mathbf{\Gamma}}_\xi(0),$$

and the normalizing condition

$$\mathbf{V}_j \hat{\mathbf{\Gamma}}_\xi(0) \mathbf{V}_i' = \begin{cases} 0 & \text{for} \quad j \neq i, \\ 1 & \text{for} \quad j = i. \end{cases}$$

Then we order the eigenvalues in descending order and take the eigenvectors corresponding to the largest r eigenvalues. Our estimated static factors are the generalized principal components

$$v_{jt} = \mathbf{V}_j' \mathbf{x}_t, \quad j = 1, \ldots, r.$$

The motivation for this strategy is that, if $\hat{\mathbf{\Gamma}}_\xi(0)$ is the variance-covariance matrix of the idiosyncratic components (i.e. the ξ_{jt}'s are mutually orthogonal), the generalized principal components are the linear combinations of the x_{jt}'s having the smallest idiosyncratic-common variance ratio (for a proof see Forni, Hallin, Lippi and Reichlin, 2001b). We diagonalize the idiosyncratic variance-covariance matrix since, as shown in the paper cited above, this gives better results under simulation when n is large with respect to T as is the case here.

By using the generalized principal components and the covariances estimated in the first step we can estimate and forecast $\boldsymbol{\chi}_t$. Precisely, setting $\mathbf{V} = (\mathbf{V}_1 \ \cdots \ \mathbf{V}_r)$ and $\mathbf{v}_t = (v_{1t} \cdots v_{rt})' = \mathbf{V}'\mathbf{x}_t$, our estimate of $\boldsymbol{\chi}_{t+h}$, $h = 0, \ldots, s$, given the information available at time t, is

$$\begin{aligned} \hat{\boldsymbol{\chi}}_{t+h} &= \hat{\mathbf{\Gamma}}_\chi(h) \mathbf{V} \left(\mathbf{V}' \hat{\mathbf{\Gamma}}(0) \mathbf{V} \right)^{-1} \mathbf{v}_t & (6) \\ &= \hat{\mathbf{\Gamma}}_\chi(h) \mathbf{V} \left(\mathbf{V}' \hat{\mathbf{\Gamma}}(0) \mathbf{V} \right)^{-1} \mathbf{V}' \mathbf{x}_t. & (7) \end{aligned}$$

In Forni, Hallin, Lippi and Reichlin (2001b) it is shown that, as both n and T go to ∞ in at a suitable rate, $\hat{\boldsymbol{\chi}}_t$ converges in probability, entry by entry, to $\boldsymbol{\chi}_t$, and $\hat{\boldsymbol{\chi}}_{t+h}$ converges to the theoretical projection of $\boldsymbol{\chi}_{t+h}$ on contemporaneous and past values of $u_{1t}, \ldots, u_{qt}$.

Appendix 5: Estimating the cyclical part of the common components

In the third and final step we use contemporaneous, past and future values of the static factors to obtain our estimate of χ_{1t}^C the cyclical component of the GDP. Precisely, we project χ_{1t}^C on $\mathbf{v}_{t-m}, \dots \mathbf{v}_{t+m}$. We do not estimate OLS, but use the projection coefficients derived by the covariance matrices of the cyclical components estimated in the first step. The lag-window size m should increase with the sample size T, but at a slower rate. Consistency of such estimator is ensured, for appropriate relative rates of m, T and n, by the fact that (a) the projection of χ_{1t}^C on the first m leads and lags of χ_{1t} is consistent beacause of consistency of $\hat{\chi}_{1t}$ and the estimated covariances involved; (b) χ_{1t} is a linear combination of the factors in $\mathbf{v}_t$, so that projecting on the factors cannot be worse than projecting on the common component itself.

Notice that here we have something like a multivariate version of the procedure by Christiano and Fitzgerald (2001) to approximate the band-pass filter. Exploiting the superior information embedded in the cross-sectional dimension enables us to obtain a very good smoothing by using a very small window ($m = 1$). This has the important consequence that we obtain a timely and reliable end-of-sample estimation, and we are not forced to revise our estimates for a long time (say 12 months or more) after the first release, as with the univariate procedure. To get an intuition of the reason why we get good results with a narrow window, consider the extreme case $m = 0$. Clearly with univariate prediction we cannot get any smoothing at all. By contrast, the static factors will include in general both contemporaneous and past values of the common shocks and can therefore produce smooth linear combinations.

Set $\mathbf{V}_t = (\mathbf{v}'_{t+m} \cdots \mathbf{v}'_t \cdots \mathbf{v}'_{t-m})'$ and

$$\mathbf{W} = \left(\underbrace{\begin{matrix} \mathbf{V} & \mathbf{0}_{n\times r} & \cdots & \mathbf{0}_{n\times r} \\ \mathbf{0}_{n\times r} & \mathbf{V} & \cdots & \mathbf{0}_{n\times r} \\ \vdots & \vdots & \ddots & \vdots \\ \mathbf{0}_{n\times r} & \mathbf{0}_{n\times r} & \cdots & \mathbf{V} \end{matrix}}_{2m+1 \text{ blocks}} \right).$$

Moreover, set $\mathbf{X}_t = (\mathbf{x}'_{t+m} \cdots \mathbf{x}'_t \cdots \mathbf{x}'_{t-m})'$, so that $\mathbf{V}_t = \mathbf{W}'\mathbf{X}_t$. The sample variance-covariance matrix of $\mathbf{X}_t$ is

$$\mathbf{M} = \begin{pmatrix} \hat{\mathbf{\Gamma}}(0) & \hat{\mathbf{\Gamma}}(1) & \cdots & \hat{\mathbf{\Gamma}}(2m) \\ \hat{\mathbf{\Gamma}}'(1) & \hat{\mathbf{\Gamma}}_0 & \cdots & \hat{\mathbf{\Gamma}}(2m-1) \\ \vdots & \vdots & \ddots & \vdots \\ \hat{\mathbf{\Gamma}}'(2m) & \hat{\mathbf{\Gamma}}'(2m-1) & \cdots & \hat{\mathbf{\Gamma}}(0) \end{pmatrix},$$

while $\mathrm{E}\left(\chi_t^C \mathbf{X}'_t\right)$ can be estimated by

$$\mathbf{R} = \begin{pmatrix} \hat{\mathbf{\Gamma}}'_{\chi^C}(m) & \cdots & \hat{\mathbf{\Gamma}}'_{\chi^C}(0) & \cdots & \hat{\mathbf{\Gamma}}_{\chi^C}(m) \end{pmatrix}.$$

Our estimate of the common cyclical components is then

$$\hat{\chi}_t^C = \mathbf{RW}\left(\mathbf{W}'\mathbf{MW}\right)^{-1}\mathbf{W}'\mathbf{X}_t. \tag{8}$$

At the end of the sample, i.e. from $T-m$ onward, we have the problem that $\mathbf{x}_{T+h}$, $h>0$, is not available. Our estimate is then obtained by substituting our forecast of the common components $\hat{\boldsymbol{\chi}}_{T+h}$, in place of $\mathbf{x}_{T+h}$ and applying the formula 8.

Appendix 6: Treatment of the end-of-sample unbalance

Our procedure to handle this problem is the following. Let T be the last date for which the whole dataset is available. Until T we estimate the static factors as explained above, i.e. by taking the generalized principal components of the vector $\mathbf{x}_{nT}$. From T onward, we use the generalized principal components of a modified n-dimensional vector $\mathbf{y}_{nT}$ which includes, for each process in the data set, only the last observed variable, in such a way as to exploit for each process the most recent information. Clearly computation will involve the estimated covariance matrices of the common and the idiosincratic component of $\mathbf{y}_{nT}$ in place of those of $\mathbf{x}_{nT}$. Having obtained an estimate of these factors, call them $\mathbf{w}_T$, we estimate χ_{1T+k}^C, $k>0$, by projecting it on $\mathbf{w}_{T-m},\ldots,\mathbf{w}_T$. Let us assume that T is the last date for which we have observations for all of the variables in the data set and that there are some variables for which we have observations until dates $T+1,\ldots,T+w$. Without loss of generality we can then reorder the variables in such a way that

$$\mathbf{x}_t = \left(\begin{array}{cccc} \mathbf{x}_t^{1\prime} & \mathbf{x}_t^{2\prime} & \cdots & \mathbf{x}_t^{w\prime} \end{array}\right),$$

where $\mathbf{x}_{jt}$, $j=1,...,w$, is such that the last available observation is dated $T+j-1$. Correspondingly, the sample covariance matrices $\hat{\mathbf{\Gamma}}(k)$ are partitioned as follows

$$\hat{\mathbf{\Gamma}}(k) = \left(\begin{array}{cccc} \hat{\mathbf{\Gamma}}^{11}(k) & \hat{\mathbf{\Gamma}}^{12}(k) & \cdots & \hat{\mathbf{\Gamma}}^{1w}(k) \\ \hat{\mathbf{\Gamma}}^{21}(k) & \hat{\mathbf{\Gamma}}^{22}(k) & \cdots & \hat{\mathbf{\Gamma}}^{2w}(k) \\ \vdots & \vdots & \ddots & \vdots \\ \hat{\mathbf{\Gamma}}^{w1}(k) & \hat{\mathbf{\Gamma}}^{w2}(k) & \cdots & \hat{\mathbf{\Gamma}}^{ww}(k) \end{array}\right).$$

A similar partition holds for $\hat{\mathbf{\Gamma}}_\chi(k)$. Our idea is simply to shift the variables in such a way as to retain, for each one of them, only the most updated observation, and compute the generalized principal components for the re-aligned vector. In such a way we are able to get information on the factors u_{hT+j}, $h=1,\ldots,q$, $j=1,\ldots,w$, and to exploit it in prediction.

Precisely, we set

$$\mathbf{x}_t^* = \left(\begin{array}{cccc} \mathbf{x}_t^{1\prime} & \mathbf{x}_{t+1}^{2\prime} & \cdots & \mathbf{x}_{t+w-1}^{w\prime} \end{array}\right).$$

Notice that the sample covariance matrices of $\mathbf{x}_t^*$ are then

$$\hat{\mathbf{\Gamma}}^*(k) = \begin{pmatrix} \hat{\mathbf{\Gamma}}^{11}(k) & \hat{\mathbf{\Gamma}}^{12}(k-1) & \cdots & \hat{\mathbf{\Gamma}}^{1w}(k-w+1) \\ \hat{\mathbf{\Gamma}}^{21}(k+1) & \hat{\mathbf{\Gamma}}^{22}(k) & \cdots & \hat{\mathbf{\Gamma}}^{2w}(k-w+2) \\ \vdots & \vdots & \ddots & \vdots \\ \hat{\mathbf{\Gamma}}^{w1}(k+w-1) & \hat{\mathbf{\Gamma}}^{w2}(k+w-2) & \cdots & \hat{\mathbf{\Gamma}}^{ww}(k) \end{pmatrix}$$

and the matrices $\hat{\mathbf{\Gamma}}_\chi^*(k)$ are defined in the same way. Then we compute the matrix $\mathbf{V}^*$ of the generalized eigenvectors of $\hat{\mathbf{\Gamma}}_\chi^*(k)$ with respect to $\hat{\mathbf{\Gamma}}_\xi(k)$ (the latter matrix is diagonal and thefore is the same for $\mathbf{x}_t$ and $\mathbf{x}_t^*$) and obtain forecasts of $\boldsymbol{\chi}_{T+h}^*$ as in equation (7):

$$\hat{\boldsymbol{\chi}}_{T+h}^* = \hat{\mathbf{\Gamma}}_\chi^*(h)\mathbf{V}^* \left(\mathbf{V}^{*\prime}\hat{\mathbf{\Gamma}}_0^*\mathbf{V}^*\right)^{-1} \mathbf{V}^{*\prime}\mathbf{x}_T^*.$$

Finally we use the forecasts in $\hat{\boldsymbol{\chi}}_{T+h}^*$, $h = 1, \ldots$ to replace missing data and to get the forecasts of $\boldsymbol{\chi}_{T+h}$, $h > w$, which are needed to apply (8).