Measuring US aggregate output and output gap using large datasets

Matteo Barigozzi
London School of Economics
m.barigozzi@lse.ac.uk

Matteo Luciani
Federal Reserve Board
matteo.luciani@frb.gov

July 21, 2018

Abstract

We propose new measures of US aggregate output and output gap computed by means of a Non-Stationary Dynamic Factor model estimated on a large dataset of macroeconomic indicators, combined with a non-parametric Trend-Cycle decomposition of the factors. We find that: (i) since 2010 output growth was on average 0.4% higher than measured by GDP, the difference being been concentrated in the first quarter of the year; (ii) while for several consecutive years before the financial crisis the economy operated above its potential, as of 2017:Q4 there is still slack in the economy. Both our measures are robust to data revisions.

JEL classification: C32, C38, C55, E0.
Keywords: Gross Domestic Output; Output Gap; Non-stationary Approximate Dynamic Factor Model; Trend-Cycle Decomposition.

*Disclaimer: the views expressed in this paper are those of the authors and do not necessarily reflect the views and policies of the Board of Governors or the Federal Reserve System.

1 Introduction

There are two fundamental, and intimately related, issues in monetary policy, and macroeconomics in general: measuring appropriately aggregate output; and, decomposing it into potential output and output gap. Indeed, while growth of aggregate output affects our assessment of current macroeconomic conditions as well as our decisions about the future, decomposing aggregate output into potential output and output gap is a critical task for both monetary and fiscal policy, as the former is a key input for long-term projections, and the latter can be an important tool to measure slack in the economy.

There are two main difficulties related to these tasks. First, measuring aggregate output accurately is difficult. The Bureau of Economic Analysis publishes two measures: Gross Domestic Product (GDP), which tracks expenditures on goods and services, and Gross Domestic Income (GDI), which tracks income received by those who produce output. The two measures are the same in theory, but not in practice, suggesting that each is measured with error. Second, even if an accurate measure of aggregate output were available, both potential output and the output gap are unobservable, hence a statistical or economic model must be used to conduct such decomposition.

In this paper, we tackle both issues "without pretending to have too much *a priori* economic theory" (Sargent and Sims, 1977), and thus letting the data speak as freely as possible. To this end, we first disentangle common and idiosyncratic dynamics by estimating a Non-Stationary Dynamic Factor Model on a large datasets of US macroeconomic indicators, thus getting an estimate of aggregate output. Then, we decompose aggregate output into potential output and output gap by disentangling common trends from common cycles by means of a non-parametric Trend-Cycle decomposition of the latent common factors. Our methodology builds on three features. First, macroeconomic time series are characterized by two main stylized facts: co-movements and non-stationarity (Lippi and Reichlin, 1994a). Second, by aggregating a large amount of information, that is by analyzing large datasets, it is possible to separate macroeconomic fluctuations from sectoral dynamics and measurement error (Forni et al., 2000; Bai and Ng, 2002; Stock and Watson, 2002). A Non-Stationary Dynamic Factor Model is capable to address these features in a parsimonious and realistic way. Third, the dynamics of trends and cycles might be more rich than just the standard random walk vs. causal ARMA distinction (Lippi and Reichlin, 1994a) and, in this respect, a non-parametric Trend-Cycle decomposition has the advantage that it does not require us to specify a law of motion for the trend and the cycles.

The first part of our analysis is about measuring aggregate output, to which we refer as Gross Domestic Output (GDO). We assume that long-run GDP and GDI are driven exclusively, and by the same amount, by factors common to the whole economy, while the discrepancy between GDP and GDI is only temporary and driven by idiosyncratic factors and measurement errors. Our estimate of GDO shows that since 2010, output growth was on average 0.4 percentage point higher than the estimate by the Bureau of Economic Analysis. Our measure's higher growth has been concentrated in the first quarter of the year, suggesting that weakness in the GDP's first-quarter (Q1) growth over the past several years, to which many authors refer to as residual seasonality (e.g. Rudebusch et al., 2015), is not a matter of shifting growth across quarters, rather it is a matter of missing Q1 growth that is not made up later in the year.

The second part of our analysis is about measuring the output gap. We compare our estimate with the one produced by the Congressional Budget Office (CBO), which estimates potential output as that level of output consistent with current technologies and normal utilization of capital and labor, and the output gap as the residual part of output. Although these two estimates of output gap are obtained in completely different ways, in practice they look very similar and are comparable for most of the sample considered, but from the late nineties to the end of the sample in 2017:Q4, when they differ, sometimes significantly. In particular: according to our estimate between 2002:Q1 and 2007:Q4 the output gap was on average 3.7 percentage points higher than estimated by the CBO, thus suggesting that throughout this period the economy operated above its potential, and hence that that level of output was unsustainable. Moreover, compared to the CBO model, our model suggests that growth after the financial crisis is due primarily to permanent factors. As a consequence, our output gap estimate indicates that as of 2017:Q4 there is still slack in the economy. By contrast, the two measures interpret roughly in the same way the years of the financial crisis as primarily due to the cyclical component.

The third part of our analysis investigates the real-time properties of our proposed measures. Indeed, there has been a lot of debate on the reliability of end-of-sample estimates of the output gap in real-time (e.g. Orphanides and van Norden, 2002), but the same problem matters also for estimates of GDO. To investigate this issue we build real-time data vintages back till September 2013, and we compute the end-of-sample revisions of both our GDO and output gap estimates. The results of this exercise are clear: our estimates revise less than, or comparably with, other commonly used measures, thus confirming the intuition that data revisions are by nature idiosyncratic, and therefore, by pooling a large amount of data, it is possible to get a robust estimate of the signal while parsing out the noise.

Although aggregate output is one of the most important concepts in macroeconomic theory, there has been surprisingly nearly no academic research on how to measure it appropriately. Two exceptions are represented by the Bureau of Economic Analysis and the Philadelphia Fed who, in recent years, proposed to combine GDP and GDI so to mitigate measurement error, and thus came up with a better estimate of aggregate output (see Council of Economic Advisers, 2015, and Aruoba et al., 2016, respectively). Compared to this approach based only on GDP and GDI growth data, we go one step further—and to the best of our knowledge we are the first doing so—in that our estimate of aggregate output is not only consistent with national account statistics, but also consistent with the rest of the economy—i.e., it is consistent with the labor market, financial conditions, inflation, industrial production, etc.

By contrast, since the seminal paper of Beveridge and Nelson (1981) there has been a wide interest in decomposing aggregate output into a trend and a cycle. In the last 30 years, many papers have suggested different ways to obtain a Trend-Cycle decomposition of aggregate output (e.g. Stock and Watson, 1988; Lippi and Reichlin, 1994a; Gonzalo and

Granger, 1995; Garratt et al., 2006; Creal et al., 2010), and some of them proposed to use non-stationary low-dimensional factor models to estimate the output gap (e.g. Fleischman and Roberts, 2011; Jarociński and Lenza, 2018), and secular trends (e.g. Antolin-Diaz et al., 2017). Compared to those works, which typically consider estimates of the output gap using only highly aggregated variables such as GDP, the unemployment rate, and PCE price inflation, we include several other indicators, thus capturing information coming from a wider spectrum of the economy. Finally, Aastveit and Trovik (2014) and Morley and Wong (2017) are the sole two papers that have used a high-dimensional dataset for estimating the output gap, and to do so they estimate a stationary factor model and a large Bayesian VAR, respectively. Compared to those works, we explicitly address non-stationarity of the data, thus we take into account also long-run co-movements when building our measures.

The rest of this paper is structured as follows. In Section 2 we present the Non-Stationary Dynamic Factor model and in Section 3 we describe the model setup. Then, in Section 4 we present our estimate of aggregate output. In Section 5 we move to the output gap. More precisely, in Section 5.1 we present the non-parametric Trend-Cycle decomposition that we use, and in Section 5.2 we present our estimate of the output gap and we compare it with the one produced by the CBO. In Section 6 we study the real-time properties of our suggested measures, and we also present a comparison of our output gap estimate with the most common univariate methods used in the literature. To conclude, in Section 7 we discuss our findings and the advantages and limitations of our methodology, and we propose directions for further research.

2 Non-Stationary Dynamic Factor Model

The goal of this paper is to estimate aggregate output and to decompose it into potential output and output gap. To do so, we first disentangle common and idiosyncratic dynamics by using a Non-Stationary Approximate Dynamic Factor Model (NS-DFM) estimated on a large dataset of US macroeconomic indicators. Then, we disentangle common trends from common cycles by applying a non-parametric Trend-Cycle decomposition to the estimated latent common factors. This section provides a presentation of the NS-DFM originally developed in Barigozzi et al. (2016b), which is a generalization to the case of unit-roots of the dynamic factor model proposed by Stock and Watson (2005) and Forni et al. (2009) for stationary data. Closely related models have been studied by Bai (2004) and Bai and Ng (2004). The Trend-Cycle decomposition is discussed in Section 5.1.

Dynamic factor models are based on the idea that fluctuations in the economy are due to a few macroeconomic or common shocks ($\mathbf{u}_t$), affecting the whole economy, and to sectorial and local shocks, influencing just a part of the economy (Forni et al., 2000). Therefore, each variable in a macroeconomic dataset (x_{it}) is decomposed into the sum of a macroeconomic or common component (χ_{it}), which is driven by the macroeconomic shocks, and an idiosyncratic component (ξ_{it}), which is driven by both sectorial/local shocks and by measurement error, where the common and the idiosyncratic component are assumed to be independent at all leads and lags.

Formally, let us consider a panel of n time series $\{\mathbf{x}_t = (x_{1t} \cdots x_{nt})' : t = 1, \ldots, T\}$, such that $\mathsf{E}[x_{it}] = 0$, for any i and t and $\mathbf{x}_t \sim I(1)$, that is at least one of its components has a unit root. Then,

$$x_{it} = \chi_{it} + \xi_{it}, \tag{1}$$

$$\chi_{it} = \sum_{k=0}^{s} \boldsymbol{b}_{ik} \boldsymbol{f}_{t-k}, \tag{2}$$

$$\boldsymbol{f}_t = \sum_{k=1}^{\ell} \boldsymbol{\mathcal{A}}_k \boldsymbol{f}_{t-k} + \mathbf{u}_t, \qquad \mathbf{u}_t \overset{i.i.d.}{\sim} \mathcal{N}(\mathbf{0}, \mathbf{Q}), \tag{3}$$

where $\boldsymbol{f}_t = (f_{1t} \cdots f_{qt})'$ are the common latent factors capturing co-movements across series and across time, $\boldsymbol{b}_{ik} = (b_{i1k} \cdots b_{iqk})$ are the q factor loadings at lag k, $s \geq 1$ and $\ell \geq 1$ are finite integers, and $\mathbf{Q}$ is a $q \times q$ positive definite covariance matrix.

Non-stationarity in the data is modeled by means of two main assumptions:

1. The dynamics of the factors $\boldsymbol{f}_t$ is governed by $(q - d)$ unit roots with $d > 0$. This assumption is consistent with the idea that, while there are multiple—i.e., q—forces that are capable of generating macroeconomic fluctuations, only a few of them—i.e., $(q - d)$—are capable of permanently affecting the economy. Technically, we assume that the coefficients of (3) are such that $\det(\mathbf{I}_q - \sum_{k=1}^{p} \boldsymbol{\mathcal{A}}_k z^k) = 0$ has $(q - d)$ roots in $z = 1$, while all other d roots lie outside the unit circle. This is equivalent to saying that: (i) $\boldsymbol{f}_t$ is driven by $(q - d)$ common trends, or (ii) $\boldsymbol{f}_t$ is a cointegrated vector with cointegration rank d. As a consequence, the common component has a non-stationary part driven by common trends, and a stationary (cyclical) part.

2. Some (but not all) idiosyncratic components have a unit root. Indeed, assuming that all idiosyncratic components are stationary, would imply that any $(q - d + 1)$-dimensional vector would be cointegrated, an assumption which is very restrictive and in general not supported by the data—see also the theory and empirical results presented in Barigozzi et al. (2016a,b).

To conclude the presentation of the model, note that if we take first differences of (2)-(3), then, by substituting (3) into (1)-(2), we obtain a special case of the generalized dynamic factor model for $\Delta\mathbf{x}_t$, as originally proposed by Forni et al. (2000). Our model is then identifiable by means of assumptions on the asymptotic behavior, as $n \to \infty$, of the eigenvalues of the spectral density matrix of $\Delta\mathbf{x}_t$. In particular, we require the idiosyncratic vector $\Delta\boldsymbol{\xi}_t$ to be weakly serially and cross-correlated, while the factors $\Delta\boldsymbol{f}_t$ are assumed to have a dynamically pervasive effect on $\Delta\mathbf{x}_t$. These constraints are imposed by requiring that: (i) all eigenvalues of the spectral density of $\Delta\boldsymbol{\xi}_t$ stay bounded at all frequencies as n increases; and, (ii) the q largest eigenvalues of the spectral density of the common component vector $\Delta\boldsymbol{\chi}_t$ diverge with n at all frequencies but at zero-frequency where–due to the presence of common trends–only $(q-d)$ eigenvalues diverge—see Forni et al. (2000), Forni et al. (2009), and Barigozzi et al. (2016b) for further details. These results justify from a methodological point of view our choice of using large datasets.

2.1 Estimation

The focus of the first part of this paper is on estimating the Gross Domestic Output (GDO), which in our framework is given by the common components of GDP and GDI under the constraint that these two are equal, thus $\text{GDO}_t := \chi_{\text{GDP},t} = \chi_{\text{GDI},t}$. Therefore, in order to estimate GDO, we need to estimate both the common factors and their loadings under the above constraint. To this end, we estimate the NS-DFM by Quasi-Maximum Likelihood, as described in this section.

Estimation of (1)-(3) is carried out by recurring to its so-called "static" representation—see also Stock and Watson (2005), Bai and Ng (2007), and Forni et al. (2009) in the stationary setting. The same approach can be adopted in a non-stationary setting. Specifically, let $\mathbf{F}_t = \mathbf{K}(\boldsymbol{f}_t' \cdots \boldsymbol{f}_{t-s}')'$, for some $r \times r$ invertible matrix $\mathbf{K}$, then we can re-write the NS-DFM, in vector notation, as

$$\mathbf{x}_t = \boldsymbol{\Lambda}\mathbf{F}_t + \boldsymbol{\xi}_t, \tag{4}$$

$$\mathbf{F}_t = \sum_{k=1}^{p} \mathbf{A}_k \mathbf{F}_{t-k} + \mathbf{H}\mathbf{u}_t, \qquad \mathbf{u}_t \overset{i.i.d.}{\sim} \mathcal{N}(\mathbf{0}, \mathbf{Q}), \tag{5}$$

where $\boldsymbol{\Lambda}$ is an $n \times r$ matrix of factor loadings, with rows given by $\boldsymbol{\lambda}_i' = (\boldsymbol{b}_{i0}' \cdots \boldsymbol{b}_{is}')\mathbf{K}^{-1}$, $\mathbf{H}$ is $r \times q$, and $\mathbf{F}_t$ is an r-dimensional vector of latent factors with $r = q(s+1)$. Finally, for identification of the common component, we require $\frac{\boldsymbol{\Lambda}'\boldsymbol{\Lambda}}{n} \to \mathbf{I}_r$ as $n \to \infty$, and $\mathbf{F}_t$ to have positive definite covariance.

In this paper, we follow the standard practice in factor models literature and we estimate (4)–(5) without identifying $\mathbf{K}$ explicitly. The rationale for this approach is that, being the idiosyncratic component, $\boldsymbol{\xi}_t$, the same in (1) and in (4), then the common components are also the same, that is both representations (4)-(5) and (1)-(3) are equivalent for what concerns the space spanned by the common factors. In particular, both representations impose the same constraints on the co-movement of the data, which are expressed via constraints on the behavior of the eigenvalues of the spectral densities of $\Delta\boldsymbol{\chi}_t$ and of $\Delta\boldsymbol{\xi}_t$ (see previous section). In other words, we can consider (4)–(5) as a reduced form of the NS-DFM.

We estimate the parameters of (4)-(5) by Quasi-Maximum Likelihood, implemented through the Expectation Maximization (EM) algorithm, where in the E-step the factors $\mathbf{F}_t$ are estimated with the Kalman Smoother. This approach in the stationary setting was originally proposed by Shumway and Stoffer (1982) and Watson and Engle (1983), and further developed and studied by Doz et al. (2012) and Bai and Li (2016), when considering large datasets.[1] The case of non-stationary data has been considered in fewer works—see e.g. Quah and Sargent (1993) and Seong et al. (2013). The theoretical properties of the estimators we use here are studied in Barigozzi and Luciani (2017), where consistency of the estimated loadings and factors is proved as $n, T \to \infty$.

[1] Recent applications of this approach include Reis and Watson (2010), Bańbura and Modugno (2014), Juvenal and Petrella (2015), Luciani (2015), and Coroneo et al. (2016).

We opt for Quasi-Maximum Likelihood estimation rather than, for example, Principal Component Analysis as in Bai and Ng (2004) and Barigozzi et al. (2016b), since in this way we are able to impose the restriction $\text{GDO}_t := \chi_{\text{GDP},t} = \chi_{\text{GDI},t}$, as well as other constraints. In particular, in estimating (4)-(5) we impose the following restrictions.

1. The loadings of GDP and GDI are imposed to be equal: $\boldsymbol{\lambda}_{\text{GDP}} = \boldsymbol{\lambda}_{\text{GDI}}$. In this way, we enforce equality of the common components.

2. The law of motion of the factors is specified as an unrestricted VAR of order 2, that is we do not impose cointegration in the common factors—see the discussions in Sims et al. (1990) for the case of observed data and Barigozzi et al. (2016b) for the factor model case. This choice has no consequences for the estimation of GDO, as we are just interested in taking into account the dynamics of the factors for estimating the common component. By contrast, when turning to the output gap, we show in Appendix A that the implications of cointegration of $\boldsymbol{f}_t$ for $\mathbf{F}_t$ could be exploited to refine the estimate.

3. The non-stationary idiosyncratic components are treated as additional latent states following univariate random walks, while the stationary idiosyncratic components are treated as white noise:

$$\xi_{it} = \rho_i \xi_{it-1} + e_{it}, \qquad e_{it} \overset{i.i.d.}{\sim} \mathcal{N}(0, \sigma_i^2), \quad \text{and} \quad \rho_i = \begin{cases} 1 & \text{if} \quad \xi_{it} \sim I(1), \\ 0 & \text{if} \quad \xi_{it} \sim I(0). \end{cases}$$

4. The covariance matrix of the vector of idiosyncratic innovations $\boldsymbol{e}_t = (e_{1t} \cdots e_{nt})'$ is constrained to be diagonal with entries $(\sigma_1^2 \cdots \sigma_n^2)$.

Notice that, by assuming (i) no autocorrelation of stationary idiosyncratic components and (ii) no cross-correlation among all idiosyncratic component, we are in fact estimating a mis-specified model. Both assumptions could in principle be relaxed by: (i) adding additional latent states for each stationary idiosyncratic component that is autocorrelated—see e.g. Bańbura and Modugno (2014); and (ii) estimating a full invertible idiosyncratic covariance matrix, possibly by means of regularization techniques—see e.g. Bai and Liao (2016), although in a slightly different setting. That said, increasing the dimensions of the latent states vector and/or of the parameters space might increase too much the complexity of estimation. For this reason we here rely on the arguments put forward by Doz et al. (2011, 2012), who show how the effects of mis-specification of the idiosyncratic dependence structure are, at least asymptotically as $n \to \infty$, negligible, thus justifying the high-dimensional setting from a methodological point of view.

3 Model setup

Our analysis is carried out on a large macroeconomic dataset comprising $n = 103$ quarterly series from 1960:Q1 to 2017:Q4 describing the US economy. Specifically, our dataset

Table 1: PERCENTAGE OF EXPLAINED VARIANCE

	1	2	3	4	5	6	7	8	9	10
q	33.4	46.0	53.4	59.0	63.6	67.4	70.6	73.3	75.7	77.9
r	23.1	34.0	42.4	48.0	52.0	55.5	58.4	60.7	62.9	65.0

This table reports the percentage of total variance explained by the q largest eigenvalues of the spectral density matrix of $\Delta\mathbf{x}_t$ and by the r largest eigenvalues of the covariance matrix of $\Delta\mathbf{x}_t$.

includes national account statistics, industrial production indexes, various price indexes including CPIs, PPIs, and PCE price indexes, as well as core CPI/PCE (i.e. CPI/PCE excluding food and energy), various labor market indicators including indicators from both the household survey and the establishment survey, as well as labor cost and compensation indexes, monetary aggregates, credit and loans indicators, housing market indicators, interest rates, the oil price, and the S&P500 index.

Broadly speaking, we take log-transforms of all variables in levels which are not already expressed in percentage points, then all the variables that are $I(1)$ are not transformed, while all the variables that are $I(2)$ are differenced once. Moreover, a linear trend is estimated where necessary before estimating the model. This final transformation deserves an explanation. In Section 2 we presented the model by making the simplifying assumption that $\mathsf{E}[x_{it}] = 0$, i.e., no deterministic component. This assumption is clearly unrealistic, as several macroeconomic time series exhibit a linear deterministic trend, in which case the model for an observed time series y_{it} would read $y_{it} = a_i + b_i t + \boldsymbol{\lambda}_i'\mathbf{F}_t + \xi_{it}$, where $x_{it} = \boldsymbol{\lambda}_i'\mathbf{F}_t + \xi_{it}$. In practice, we remove a_i and b_i by least square regression before estimating the factor model, and in order to choose whether or not to de-trend a variable we test for significance of the sample mean of Δy_{it}. In other words, we allow $b_i = 0$ for some i, thus ruling out the possibility of a common deterministic trend. The complete list of variables and transformations is reported in Appendix C.

Before estimating the model we need to tackle two main preliminary issues. First we need to determine the numbers of common factors q and r. Second, we need to choose which idiosyncratic components to model as random walk, and which as white noises.

In order to estimate q, we use the test by Onatski (2009) and the information criterion by Hallin and Liška (2007), both exploiting the behavior of the eigenvalues of the spectral density matrix of $\Delta\mathbf{x}_t$ averaged across all frequencies. Both methods indicate $q = 3$. Having determined q, and by virtue of the restriction $r = q(s + 1)$, we can choose r such that the share of variance explained by the factors $\mathbf{F}_t$ coincides with the share of variance explained by the q factors $\boldsymbol{f}_t$—see also D'Agostino and Giannone (2012). By looking at Table 1 we can clearly see that $r \simeq 2q$, and therefore in our benchmark specification we set $q = 3$ and $r = 6$. An alternative way to select r is to resort to one of the many available methods based on the behavior of the eigenvalues of the covariance matrix of $\Delta\mathbf{x}_t$ such as, for example, the information criterion of Bai and Ng (2002), which for our dataset gives results in line with our choice.

In order to choose which idiosyncratic components to model as random walk, for each variable we test the null-hypothesis of unit root by means of the test proposed by Bai and

Ng (2004). Then, if we fail to reject the null, we set $\rho_i = 1$, while if we reject, we set $\rho_i = 0$ (see Appendix C for details). This approach is applied to all variables in the dataset except for GDP, GDI, the unemployment rate, the Federal funds rate, and CPI, core CPI, PCE, and core PCE inflation, for which we impose a priori $\rho_i = 0$. That is, while for most of the variables in the dataset we let the data determine what is driving their long run dynamics, we impose the constraint that the long-run dynamics of GDP, GDI, unemployment rate, Federal funds rate, and CPI, core CPI, PCE, and core PCE inflation is driven exclusively by aggregate macroeconomic forces, with the idiosyncratic component accounting only for short-run dynamics.

4 Gross Domestic Output

A fundamental issue in economics is the measurement of aggregate output, GDO. Historically, GDO has been measured mainly by GDP, but GDP, which tracks all expenditures on final goods and services produced, is just an estimate of GDO. An equally acceptable estimate of the concept of GDO is represented by GDI, which tracks all income received by those who produced the output. GDP is almost always preferred to GDI, the main reason being that it is generally released one month before GDI. However, it has been shown that GDI reflects the business cycle fluctuations in true output growth better than GDP, and that GDI is better than GDP in recognizing the start of a recession (Nalewaik, 2010, 2012).

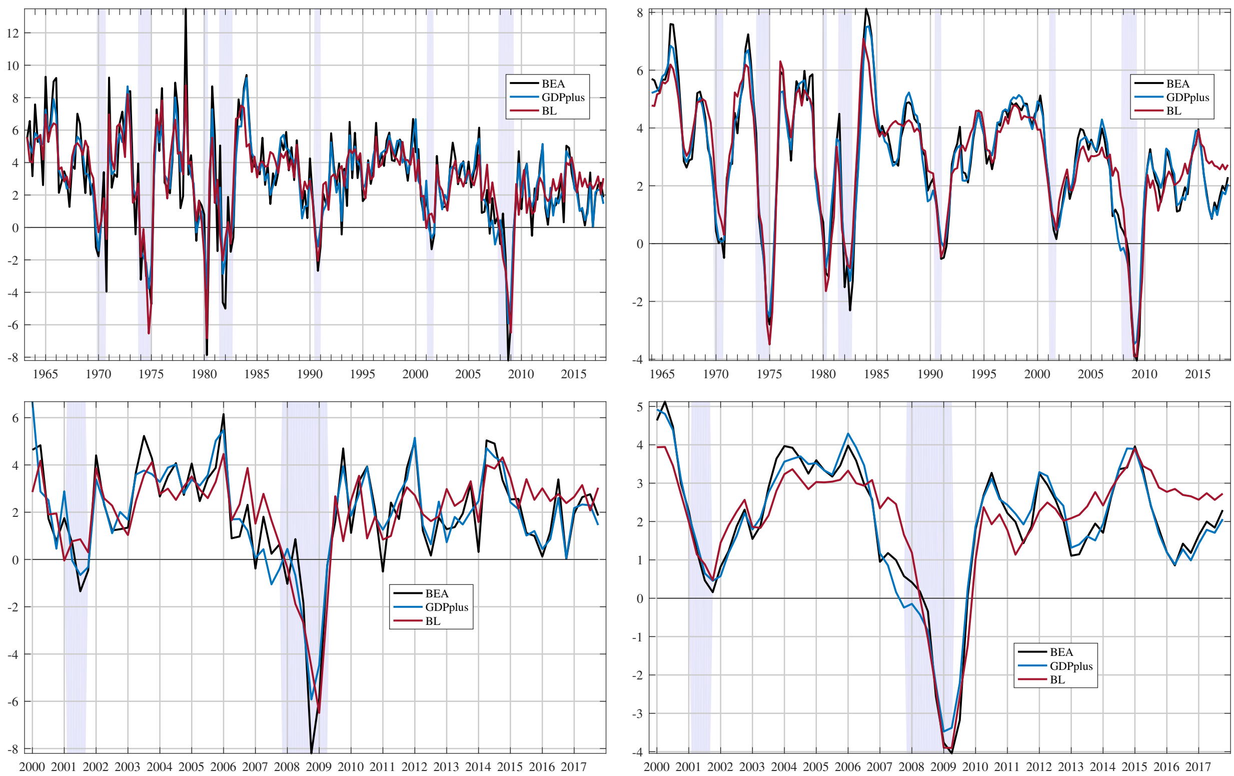
In recent years, there has been interest in combining GDP and GDI to come up with a better estimate of GDO: the rationale for doing so is that the difference between GDP and GDI is exclusively the result of measurement error—using the NIPA table definition "statistical discrepancy"—as these two statistics are in fact measuring the same thing. Since November 2013, the Philadelphia Fed releases an estimate of GDO, called "GDPplus" proposed by Aruoba et al. (2016), which is defined as the common factor driving the dynamics of GDP and GDI growth. Similarly, since July 2015, the Bureau of Economic Analysis (BEA) has released "the average of GDP and GDI", which the Council of Economic Advisers refers to as GDO (Council of Economic Advisers, 2015).

In this paper, we introduce a new measure of GDO obtained from the NS-DFM estimated on the large dataset of US macroeconomic indicators described in the previous section. In our framework, GDO is defined as that part of GDP and GDI that is driven by the aggregate macroeconomic forces, i.e., the common component. In particular, to estimate GDO we impose two restrictions: first, as already discussed in Section 2.1, GDP and GDI respond to macroeconomic factors in the same way, so that the common component of GDP is equal to the common component of GDI, $\text{GDO}_t := \chi_{\text{GDP},t} = \chi_{\text{GDI},t}$.[2] And, second, the long run dynamics of GDP and GDI are entirely driven by the common factors, so that both sectorial and local factors affecting the economy, and measurement error, have just a temporary effect, i.e., the idiosyncratic components of GDP and GDI are stationary,

[2] This restriction is indeed corroborated by the data, as even if we do not impose it, the estimated $\chi_{\text{GDP},t}$ and $\chi_{\text{GDI},t}$ are nearly identical. In numbers, the standard deviation of $(\Delta x_{\text{GDP},t} - \Delta x_{\text{GDI},t})$ is 1.93, while the standard deviation of $(\Delta\chi_{\text{GDP},t} - \Delta\chi_{\text{GDI},t})$ is reduced to 0.25.

Figure 1: Gross Domestic Output

Quarterly annualised percentage change | *4-quarter percentage change*



This figure reports different estimates of GDO. Black line: "the average of GDP and GDI" released by the BEA; blue line: "GDPplus" released by the Philadelphia Fed; red line: our estimate.

$\xi_{\text{GDP},t} \sim I(0)$ and $\xi_{\text{GDI},t} \sim I(0)$.

Compared to the approach of the BEA and the Philadelphia Fed based on GDP and GDI growth data only, there are two advantages in our approach. First, by analyzing a large amount of information, our estimate of GDO incorporates information coming from a wider spectrum of the economy, thus accounting for most sources of common variation, such as, for example, monetary policy, oil price, and technology shocks. In this way, our measure is not only consistent with national account statistics, but also consistent with the rest of the economy—i.e., it is consistent with the labor market, financial conditions, inflation, industrial production, etc. Second, by using data in levels, we are able to impose that GDP and GDI evolve around a common stochastic trend, and thus we can impose that the discrepancy between these two statistics is only temporary. In other words we are assuming that both GDP and GDI are correct estimators of GDO up to transitory dynamics.

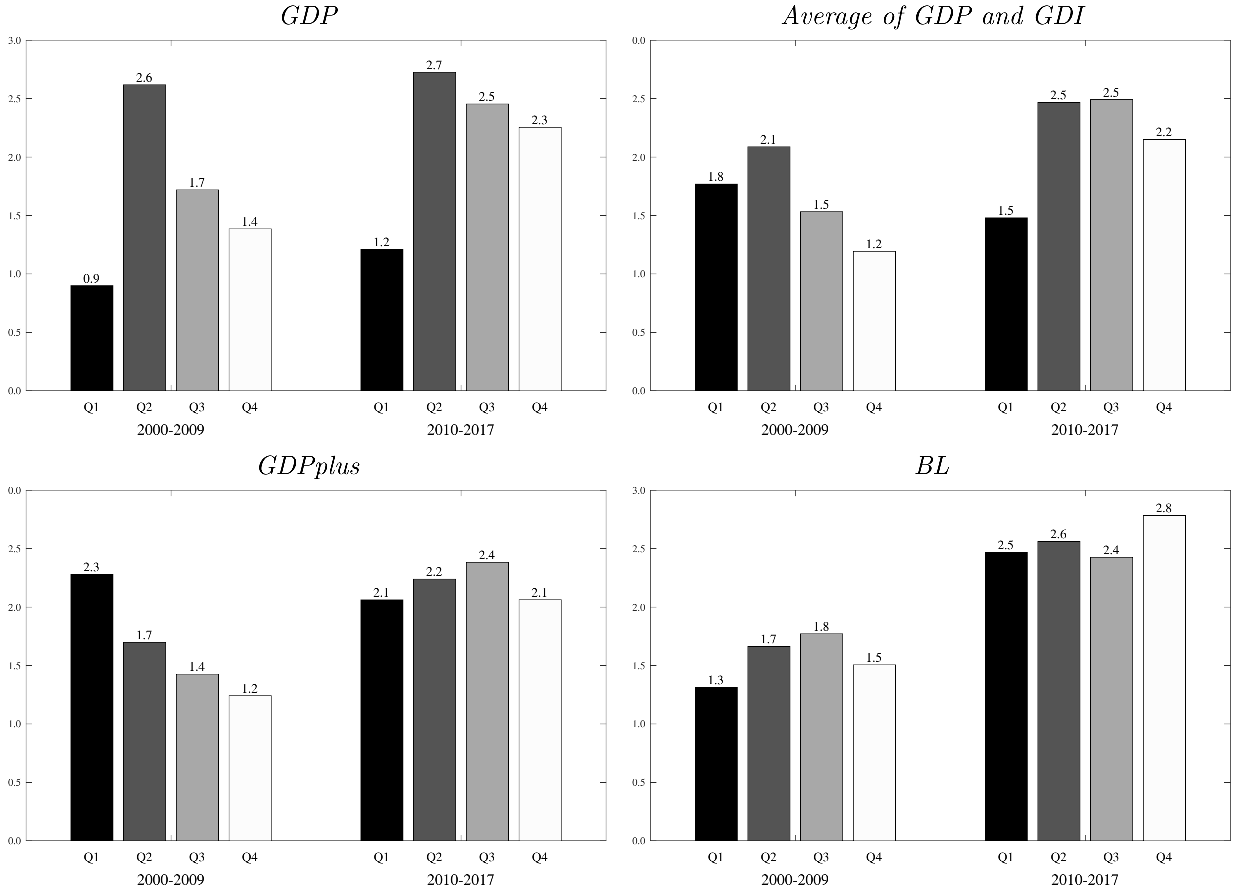
Figure 1 shows our proposed estimate of GDO (red line), together with "GDPplus" (blue line), and the "average of GDP and GDI" (black line), where the plots on the left column show quarterly annualized percentage changes, while the plots on the right column show 4-quarter percentage changes. Overall, the three measures are very similar, which is not surprising, as they are attempting to estimate the same quantity, with our estimate of GDO that appears to be smoother than the other two—this is the case since by construction we capture all low frequency movements of GDP and GDI. However, despite the similarities, two important differences emerges.

First, our measure of GDO does not show any kind of residual seasonality in the last seventeen years, where the term "residual seasonality" refers to the presence of "lingering seasonal effects even after seasonal adjustment processes have been applied to the data" (Moulton and Cowan, 2016). Indeed, Figure 2 shows average quarterly annualized GDO growth by quarter for our measure (lower-right plot), for "GDPplus" (lower-left plot), and for "the average of GDP and GDI" (upper-right plot). Mainly motivated by the fact that since 2010 GDP growth in Q1 has been on average more than 1 percentage point lower than in the other quarters (upper-left plot of Figure 2), in recent years there has been lots of discussion on whether US GDP exhibits residual seasonality or not. The profession is not in agreement on this issue, as some authors (e.g. Gilbert et al., 2015; Lengermann et al., 2017) conclude that US GDP does not exhibit residual seasonality, while others (e.g. Rudebusch et al., 2015; Lunsford, 2017) find evidence of residual seasonality—see Moulton and Cowan (2016) for a technical discussion on causes and remedies for residual seasonality in US GDP. As can be clearly seen from the lower-right plot of Figure 2, our estimate of GDO exhibits no residual seasonality whatsoever in the last seventeen years.

Second, our estimate of GDO in the recent years gives a different picture of the economy than the one given by "GDPplus" and "the average of GDP and GDI" (see the lower plots in Figure 1). According to our estimate, since 2010 quarterly annualized GDO growth was on average about 0.4 percentage point higher than estimated by the BEA and the Philadelphia Fed. Compared to the BEA estimates most of this difference comes from a higher estimate of GDO growth in the first quarter, whereas compared to the Philadelphia Fed estimate this difference spreads on all quarters (see Figure 2). Two main conclusions can be drawn from these results: first, based on the commonality in the data, the US economy grew at a faster pace than measured by national account statistics. This is true in particular for 2015 a year characterized by remarkable improvements in the labor market, improvements that are likely to be captured by our measure, while they are not necessarily captured by national account statistics.[3] Second, the low level of GDP in Q1 since 2010 is not a matter of shifting growth across quarters, rather it is a matter of missing Q1 growth that is not made up later in the year.

[3] In 2015 on average (quarterly annualized) GDP growth was 2.0%, the BEA average GDP-GDI growth was 1.8%, and GDPplus growth was 1.7%. At the same time, 2015 was characterized by remarkable improvements in the labor market, as payroll employment increased on average by about 228,000 units per months, the unemployment rate fell from 5.7% in 2016Q4 to 5.0% in 2017:Q4, and the number of long-term unemployed civilians decreased by about 729,000 units. We affirm that our measure of GDO, which in 2015 grew on average 2.8% at an annual rate, captures these improvements in the labor market, because if we exclude all the labor market variables contained in our dataset and we re-estimate the model, the average estimated (quarterly annualized) growth drops to 2.0%.

Figure 2: Residual seasonality



This figure reports average growth at an annual rate by quarter for GDP, "the average of GDP and GDI" released by the BEA , the Philadelphia Fed estimate of GDO (GDPplus), and our estimate of GDO (BL).

5 Output gap

Having estimated the space spanned by the common factors, and GDO, our goal is to estimate the output gap, which in our framework is the cyclical component of GDO. Therefore, our goal is to decompose the common factors into common trends and common cycles.

Since the seminal work of Beveridge and Nelson (1981), the issue of decomposing GDP into a trend and a cycle has been a central question in both time series econometrics and policy analysis. This is not surprising, as long-run trends are mainly influenced by supply-side factors, while short-run cycles are mainly associated with demand-side factors, and therefore different estimates of the trend and of the cycle can lead to different policy recommendations. Given the relevance of the issue, in the last 30 years, many papers have suggested different ways to obtain a Trend-Cycle decomposition of GDP. Roughly speaking, those works can be grouped under two main approaches: one based on univariate methods (e.g. Watson, 1986; Lippi and Reichlin, 1994b; Morley et al., 2003; Dungey et al., 2015), and another using multivariate techniques (e.g. Stock and Watson, 1988; Lippi and Reichlin, 1994a; Gonzalo and Granger, 1995; Garratt et al., 2006; Creal et al., 2010).

In this paper we propose a Trend-Cycle decomposition based on principal component analysis of the long run covariance matrix of the factors. Being non-parametric this decomposition has the advantage that it does not require us to specify a law of motion for the trend and the cycles, as instead, for example, in Harvey (1985) and Hasenzagl et al. (2018). Moreover, being our analysis based on a large dataset, we cannot use such structural approaches directly in the EM estimation, but we instead apply our decomposition on the common factors, thus guaranteeing that we actually recover trends and cycles that are common to the whole dataset.

5.1 Trend-Cycle decomposition via eigen-analysis

We now describe the estimation of common trends and common cycles. By comparing (5) with (3), it is possible to show that, since $\mathbf{F}_t$ is driven by the same q shocks that drive $\boldsymbol{f}_t$, then $\mathbf{F}_t$ has $(q-d)$ unit roots—see Appendix B.2. Therefore, $\mathbf{F}_t$ is driven by $(q-d)$ common trends and it admits the factor representation:

$$\mathbf{F}_t = \boldsymbol{\Phi}\mathbf{T}_t + \boldsymbol{\Gamma}_t,$$

where $\boldsymbol{\Gamma}_t$ is stationary, $\boldsymbol{\Phi}$ is $r \times (q-d)$ with full column rank and $\mathbf{T}_t$ is the vector of $(q-d)$ common trends, i.e. with all components $I(1)$—see e.g. the Theorem in Escribano and Peña (1994) for a derivation. Different choices of $\boldsymbol{\Phi}$ lead to different definitions of common trends.

Here we assume $\mathbf{T}_t$ and $\boldsymbol{\Gamma}_t$ to be uncorrelated and we estimate the elements of $\mathbf{T}_t$ as the first $(q-d)$ principal components of the estimated factors $\widehat{\mathbf{F}}_t$—see also Proposition 1 in Bai (2004) and Theorem 1 in Peña and Poncela (2006). Specifically, let

$$\widehat{\mathbf{S}} = \frac{1}{T^2}\sum_{t=1}^{T}\widehat{\mathbf{F}}_t\widehat{\mathbf{F}}_t'$$

and denote by $(\widehat{\boldsymbol{\Phi}}\ \widehat{\boldsymbol{\Phi}}_\perp)$ the $r \times r$ matrix with columns given by the normalized eigenvectors of $\widehat{\mathbf{S}}$, ordered according to the decreasing value of the corresponding eigenvalues, and such that $\widehat{\boldsymbol{\Phi}}$ is $r \times (q-d)$, $\widehat{\boldsymbol{\Phi}}_\perp$ is $r \times (r-q+d)$, $\widehat{\boldsymbol{\Phi}}'\widehat{\boldsymbol{\Phi}} = \mathbf{I}_{(q-d)}$, $\widehat{\boldsymbol{\Phi}}_\perp'\widehat{\boldsymbol{\Phi}}_\perp = \mathbf{I}_{(r-q+d)}$ and $\widehat{\boldsymbol{\Phi}}'\widehat{\boldsymbol{\Phi}}_\perp = \mathbf{0}$. Then, the estimator of common trends is given by the projection:

$$\widehat{\mathbf{T}}_t = \widehat{\boldsymbol{\Phi}}'\widehat{\mathbf{F}}_t. \tag{6}$$

While, by projecting $\widehat{\mathbf{F}}_t$ onto the columns of $\widehat{\boldsymbol{\Phi}}_\perp$ we obtain our estimator of common cycles:

$$\widehat{\mathbf{G}}_t = \widehat{\boldsymbol{\Phi}}_\perp'\widehat{\mathbf{F}}_t. \tag{7}$$

It can be shown that $\widehat{\mathbf{G}}_t$ belongs to the cointegration space of $\mathbf{F}_t$ and it is therefore stationary—see Theorem 1 in Zhang et al. (2018). Therefore, in our definition the common cycles represent deviations from long-run equilibria—see also e.g. Johansen (1991) and Kasa (1992) for similar definitions.[4]

From definitions (6) and (7), the estimated Trend-Cycle decomposition of the common component becomes:

$$\widehat{\chi}_{it} = \widehat{\boldsymbol{\lambda}}_i' \widehat{\boldsymbol{\Phi}} \widehat{\mathbf{T}}_t + \widehat{\boldsymbol{\lambda}}_i' \widehat{\boldsymbol{\Phi}}_\perp \widehat{\mathbf{G}}_t, \tag{8}$$

and the output gap is defined as the cycle component of GDO: $\widehat{\boldsymbol{\lambda}}_{\text{GDP}}' \widehat{\boldsymbol{\Phi}}_\perp \widehat{\mathbf{G}}_t = \widehat{\boldsymbol{\lambda}}_{\text{GDI}}' \widehat{\boldsymbol{\Phi}}_\perp \widehat{\mathbf{G}}_t$.

We conclude with some important remarks. First, notice that our definition (6) of common trends is more general than the usual definition entailed by the multivariate Beveridge-Nelson decomposition by Stock and Watson (1988). Indeed, $\widehat{\mathbf{T}}_t$ is in general not constrained to be a pure vector random walk, and therefore it can accommodate for more complex dynamics, which is a desirable property advocated for by many authors—see e.g. Lippi and Reichlin (1994a). Second, under definition (7) the components of $\widehat{\mathbf{G}}_t$ are common cycles in the sense of Vahid and Engle (1993). Third, it can be proved that, unless we impose further restrictions, the vector $\mathbf{F}_t$ has a rank of cointegration c, which is in general not identified but is such that $d \leq c \leq (r - q + d)$—see Proposition 3 and Appendix B in Barigozzi et al. (2016a) for details.[5] . Therefore, should c be lower than $(r - q + d)$ then some component of $\widehat{\mathbf{G}}_t$ would be redundant and the question of how to identify the non-redundant cycles would become important. This aspect is discussed in Appendix A.

5.2 Results

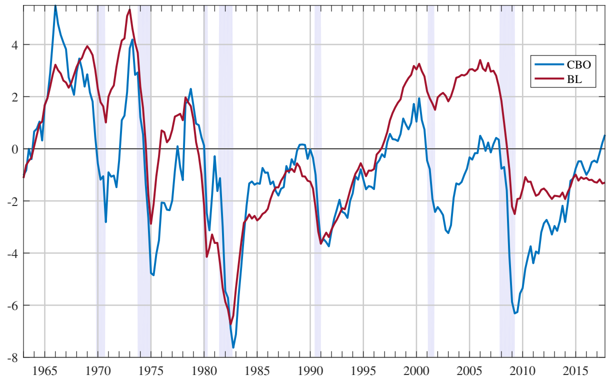
In order to apply the non-parametric Trend-Cycle decomposition just discussed, we first have to determine the number of common trends $(q - d)$. To do so, we use the criterion by Barigozzi et al. (2016b), which exploits the behavior at the zero-frequency of the eigenvalues of the spectral density matrix of $\Delta \mathbf{x}_t$. This criterion indicates the presence of $(q - d) = 1$ common trend, which is in line with many theoretical models assuming a common productivity trend as the sole driver of long-run dynamics (e.g. Del Negro et al., 2007). As a consequence, we indirectly estimate $d = 2$.

Figure 3 shows our measure of the output gap (red line), together with the one produced by the Congressional Budget Office (CBO) (blue line), where the upper-left plot shows the level of the output gap, while the upper-right plot shows the 4-quarter percentage change of the output gap. Overall, we can see that our estimate of the output gap is remarkably similar to that of the CBO. This is a result *per se*, as our estimate of the output gap is very different from that of the CBO from both a technical and an interpretation point of view. Indeed, while the CBO constructs the output gap so that its level has a specific economic meaning, our measure of the output gap is simply the deviation of output from its long-run stochastic trend—i.e., those fluctuations of aggregate output that will disappear in the

[4] Other decompositions, based on different definitions of cycles than the one used here, are, for example, in Gonzalo and Granger (1995) and Gonzalo and Ng (2001).

[5] Note that if $r = q$, then the standard results that $c = d$ holds and the cointegration rank is identified.

Figure 3: Output gap



The blue line is the level of the output gap estimated by the CBO, while the red line is our estimate.

long-run. In particular, our estimate of the output gap seems more suitable to answer the question "which part of current growth is due to temporary factors?", while the measure of the CBO is certainly more suitable as a gauge of inflation pressure.[6]

Despite the evident similarity between the two output gap estimates shown in Figure 3, since the late nineties the two measure diverge, sometimes significantly. In order to investigate the reasons of these discrepancies we split the discussion in two parts: we first discuss the period from the late nineties to the financial crisis included, and then we discuss the period after until 2017:Q4.

From the late nineties to the financial crisis

The two output gap measure started to diverge somewhat significantly at the end of 1997, and in 2000:Q2 our measure was 1.3 percentage points higher than estimated by the CBO. Starting in 2000:Q3 both measure declines, though the CBO measure declined much more—between 2000:Q3 and the trough of the dot com recession in 2001:Q4, the CBO measure declined by 4.4 percentage points, while our measure by 1.8 percentage points. Then, while according to the CBO the level of the output gap was negative between 2002:Q1 and 2005:Q4, according to our measure in that same period the output gap was positive—on average 4.1 percentage points higher than estimated by the CBO. Therefore, according to our estimate the level of the output gap right before the financial crisis in 2007:Q4

[6] The CBO estimates potential output and the output gap by using the so-called "production function approach" (Kiley, 2013) according to which potential output is that level of output consistent with current technologies and normal utilization of capital and labor, and the output gap is the deviation of output from potential output. Specifically, the CBO model is based upon a textbook Solow growth model, with a neoclassical production function. Labor and productivity trends are estimated by using a variant of the Okun's law, so that actual output is above its potential (the output gap is positive), when the unemployment rate is below the natural rate of unemployment, which is in turn defined as the non-accelerating inflation rate of unemployment (NAIRU), i.e., that level of unemployment consistent with a stable inflation—for further details see Congressional Budget Office (2001). Notice that also for the CBO the output gap is assumed to revert to zero in the long-run as it imposes in its forecast that in 10 years the output gap will be zero—see e.g. Congressional Budget Office (2004).

was 2.3%, while according to the CBO was 0.3%, and hence we estimate that the level of slack in the economy at the trough of the crisis in 2009:Q2 was -2.5%, approximately 3.8 percentage points higher than estimated by the CBO.

What does explain this divergence of the two estimates in the 2000s? We believe that this divergence is related to the way in which the two considered measures are defined. Indeed, this period was characterized by stable and low inflation—on average core PCE price inflation between 2001:Q1 and 2007:Q4 was approximately 1.9%. Accordingly, the CBO estimates that slack is positive (i.e., the output gap negative). By contrast, our measure, which is not specifically tight to inflation, but it is more broadly influenced by the co-movement in the data, estimates that a part of aggregate output was transitory. This makes sense given that the years before the financial crisis were characterized by several factors that proved indeed transitory, such as the housing boom, a historically high share of sub-prime loan origination (Haughwout and Okah, 2009), and a large amount of equity withdrawal from housing (Fuster et al., 2017). And, since our model includes a large number of variables, including housing indicators as well as financial variables, these transitory factors are captured by our model—see also Borio et al. (2017) and Morley and Wong (2017) for similar results.[7]

From the financial crisis to the end of the sample

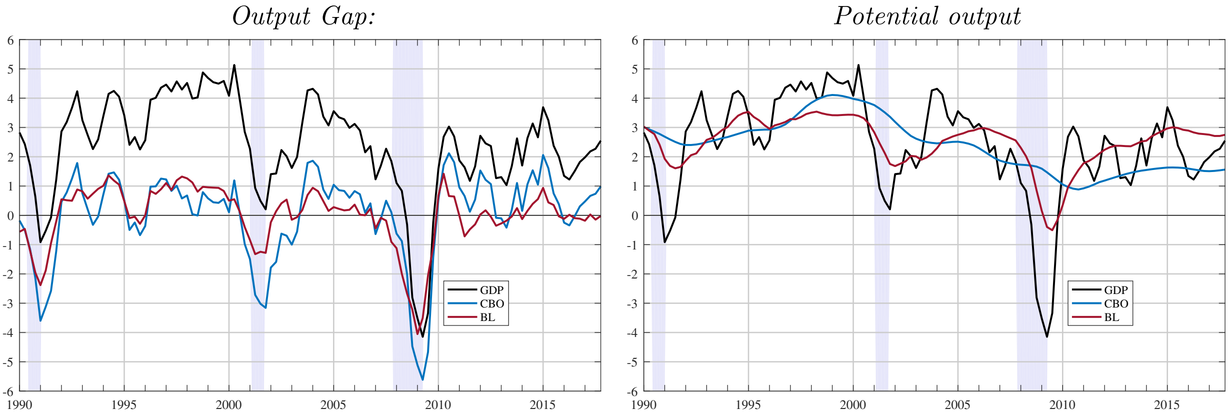
Our estimate of the output gap in 2017:Q4 is -1.3%, whilst the CBO estimate in the same quarter is 0.5%. The question then is: how is justifiable an output gap of -1.3% given that in 2017:Q4 the unemployment rate was 4.1% and GDP growth was 2.9%?

To answer this question, Figure 4 shows the decomposition of the 4-quarter percentage change of GDP (black line) into the part attributable to the output gap (left plot), and the part attributable to potential output (right plot). By looking at Figure 4, it is clear how our model attributes most of the recent output growth to the trend component, as since 2010 the role of the cyclical component is marginal. This contrasts with the CBO estimate of potential growth, according to which potential growth in the US has decreased since the late nineties. Our finding is consistent with the results in Coibion et al. (2017).

To summarize, compared to the CBO model, our model interprets the recent growth as more solid, i.e. based on permanent factors. But there is more on that. Figure 5 reports the estimate of the unemployment gap, i.e. the cyclical common component of the unemployment rate, estimated with our model (the red line), together with the estimate of the unemployment gap published by the CBO (the blue line). By comparing Figure 3 and Figure 5, we can clearly see how the output gap and the unemployment gap estimated with our model are sending contrasting signals at the end of the sample: the output gap

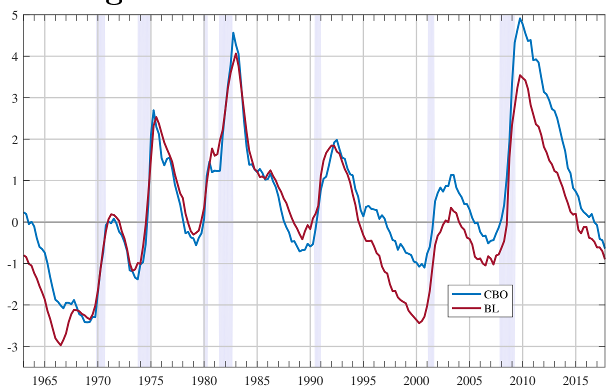
[7] Specifically, Borio et al. (2017) argue that the conceptual association between output gap and inflation is too narrow because it neglects the role of financial factors. Therefore, they build a measure of output gap based on a multivariate filter that includes also financial variables and find that the output gap before the financial crisis was in the neighborhood of 2.0%. Similarly, when using their benchmark 23-variable BVAR, Morley and Wong (2017) find that the output gap before the financial crisis was in the neighborhood of 2.0%, while their estimate increase in the 3-4% range when using a 138-variable model.

Figure 4: Decomposition of GDP growth
4-quarter percentage change



The left plot shows the 4-quarter percentage change of the output gap, while the right plot shows the 4-quarter percentage change of potential output. In each plot, the red line is our estimate, while the blue line is the CBO estimate. In both plots the black line is the 4-quarter percentage change of GDP. Note that the sum of the red lines in the two plots is our estimate of GDO, while the sum of the blue lines is GDP.

Figure 5: Unemployment gap



The blue line is the unemployment gap estimated by the CBO, while the red line is our estimate.

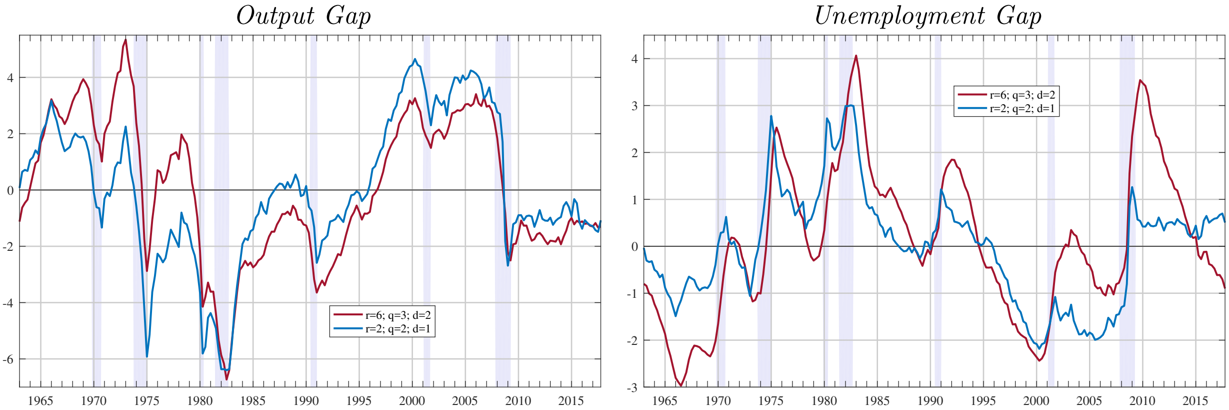
is indicating that there is still slack in the economy, whilst the unemployment gap is suggesting that the US economy is operating above potential.[8]

How is it possible that the output gap and the unemployment gap send different signals? To answer this question, we must first point out that, while in our model it can happen that the output gap and the unemployment gap send different signals, in the CBO model this is nearly impossible. Indeed, in our model, the output gap and the unemployment gap are driven by a five-dimensional cycle process $\widehat{\mathbf{G}}_t$ defined in (7), whereas in the model used by the CBO there is one single cycle, and moreover the output gap and the unemployment gap are essentially related on-to-one via the Okun's law.[9] Now, while on average our

[8] It should be noticed that in our results this is not always the case. For example, before the financial crisis the output gap and the unemployment gap estimated with our model were both indicating that the economy was quite hot.

[9] This feature of a one-to-one relationship between the output gap and the unemployment gap is also

Figure 6: OUTPUT GAP AND UNEMPLOYMENT GAP
DO WE REALLY NEED MORE THAN ONE CYCLE?



The red line in the left/right plot is the level of the output/unemployment gap estimated with our benchmark model specification, i.e., $r = 6$, $q = 3$, and $d = 2$, while the blue line is the output/unemployment gap obtained when the model is estimated by imposing a univariate cycle, i.e., $r = 2$, $q = 2$, and $d = 1$.

estimated output gap and unemployment gap are related approximately in the same way as those estimated by the CBO—the estimated Okun's low coefficient on our gap estimates is -0.5, while the one obtained on CBO estimates is -0.6—this relation holds less tightly for our estimates—the R^2 for our Okun's law is 0.6, while that for the CBO Okun's law is 0.8. Therefore, to answer the question, it is possible that the output gap and the unemployment gap send different signals (1) because they are not tightly related by construction, and (2) because the five cycles in $\widehat{\mathbf{G}}_t$ are not restricted to impact the output gap and the unemployment gap in the same way.

The above discussion prompts a new question: is it really necessary to have more than one cycle? What if there is just one trend and one cycle, i.e., $r = q = 2$ and $d = 1$, which is the configuration imposed by most of the literature? To answer this question, in Figure 6 we compare estimates of the output gap and of the unemployment gap when $r = q = 2$ and $d = 1$, implying the presence of just one cycle, with our benchmark specification, i.e., $r = 6$, $q = 3$ and $d = 2$, implying the presence of a five dimensional cycle process. The answer is clear: while for the estimation of the output gap this does not really matter, in order to have a sensible estimate of the unemployment gap we need more than one cycle and this is true especially towards the end of the sample. Hence, we conclude that data speaks against the standard practice in the literature of imposing a cycle of dimension one, and in favor of more than one cycle. In Figure D3 in Appendix D, we also show that such conclusion is robust to several other configurations of q, d, and r.

To conclude this section, in Table 2, we show the percentage of variance of the common cyclical components of some variables of interest, explained by each component $\widehat{G}_{jt}$ of the five-dimensional cycle process $\widehat{\mathbf{G}}_t$. Inspection of these results confirms both that cycles are loaded differently by different variables and that more than one cycle is necessary for fully

typical of all the unobserved component models in which there is a single cycle, as for example in Fleischman and Roberts (2011) and Jarociński and Lenza (2018).

Table 2: Variance decomposition
Percentage of total variance of the common cyclical component

	$\widehat{G}_{1t}$	$\widehat{G}_{2t}$	$\widehat{G}_{3t}$	$\widehat{G}_{4t}$	$\widehat{G}_{5t}$
GDP	85.8	8.6	5.4	0.0	0.2
Unemployment rate	84.0	11.3	0.7	0.3	3.6
PCE price inflation	0.3	8.3	69.5	21.3	0.5
Core PCE price inflation	0.6	6.5	90.9	1.6	0.4
Fed funds rate	0.0	2.8	95.4	0.0	1.7
WTI price inflation	3.7	3.6	1.3	75.4	15.9
S&P500	47.9	11.0	36.2	0.9	4.0

explaining the variance of the cyclical components of GDP and of the unemployment rate. However, it has to be stressed that, since in general $\widehat{\mathbf{G}}_t$ is not identified unless we impose some additional constraint, we are not attempting here to identify each of its components.

6 Real-time estimation of aggregate output and the output gap

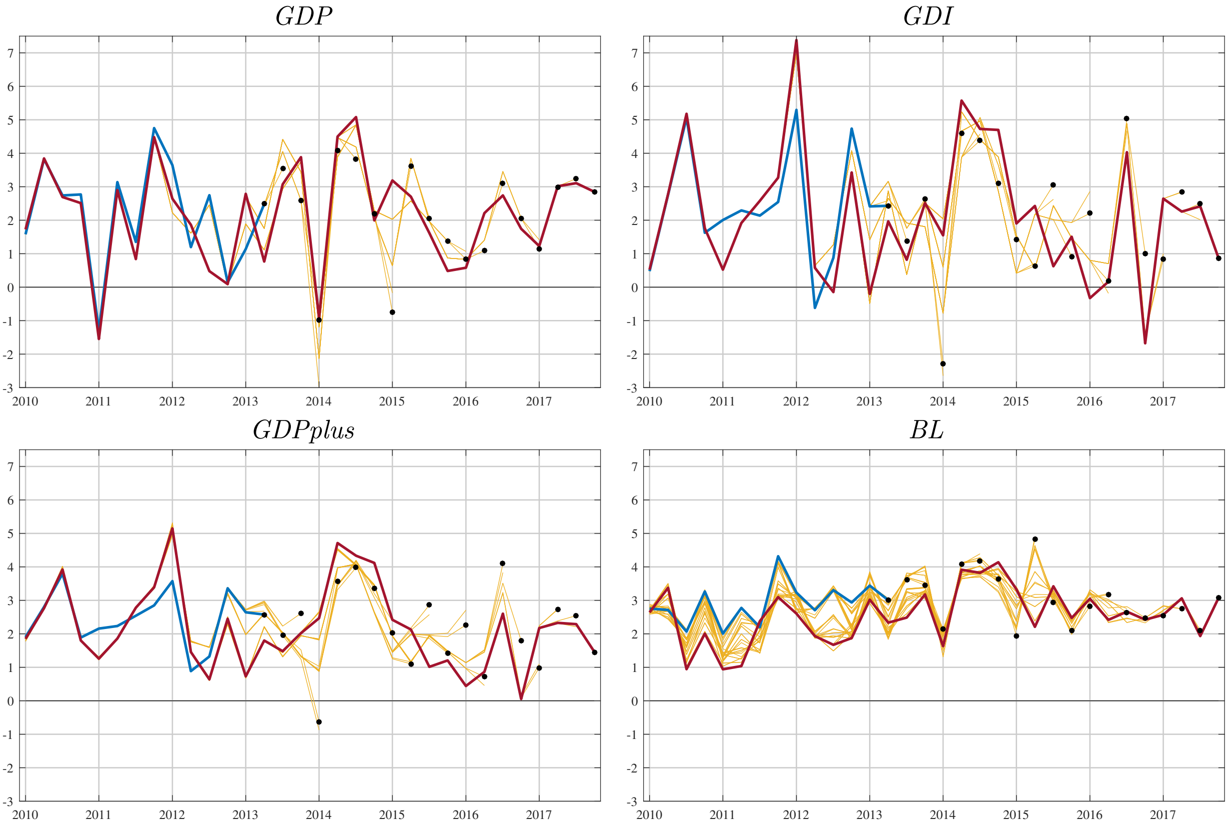
Since the seminal work of Orphanides and van Norden (2002), who show that end-of-sample revisions of GDP are of the same order of magnitude as the output gap, there has been a lot of debate on the reliability of such estimates in real-time. In this section we address this issue, and we analyze the real-time properties of both our proposed measures of aggregate output and output gap.

The analysis in this section is based on real-time data vintages of our 103 variables dataset starting in September 13 2013, that is after the 2013 NIPA comprehensive data revisions, and ending on March 30 2018, i.e., the vintage of data that we use to produce the result in the rest of the paper. We estimate aggregate output and the output gap on 7 vintages per year, where each vintage corresponds approximately to a release of GDI, for a total of 33 data vintages. Details on the construction of the real-time data vintages, and on the sources used are in Appendix C.1.

Figure 7 shows real-time estimates of the four different measures of GDO that we have already presented in Section 4. In each plot of Figure 7 the blue line is the estimate obtained with the first vintage of data (September 13 2013), the red line is the estimate obtained with the last vintage of data (March 30 2018), which corresponds to the results shown in the rest of the paper, and the yellow lines are the estimates obtained with all the remaining 31 vintages. Finally, in each plot the 19 black dots represent the estimate of GDO for quarter Q and year Y obtained with the vintage of data ending at quarter Q and year Y corresponding to the first release of GDI for quarter Q and year Y. For example, the second black dot from the left is the estimate of GDO for 2013:Q3 obtained using the data as of December 13 2013, which contain the first release of GDI for 2013:Q3.

Table 3 reports the standard set of descriptive statistics that is used to evaluate data revisions in the output gap literature (see e.g., Orphanides and van Norden, 2002; Edge and

Figure 7: Gross Domestic Output
Real-time estimation



In each plot the blue line is the estimate of GDO obtained with the vintage of data from September 13 2013, the red line is the estimate obtained with the vintage of data from March 30 2018, and the yellow lines are the estimate obtained with all the remaining 31 vintages. Finally, in each plot the 19 black dots represent the estimate of GDO for quarter Q and year Y obtained with the vintage of data ending at quarter Q and year Y corresponding to the first release of GDI for quarter Q and year Y.

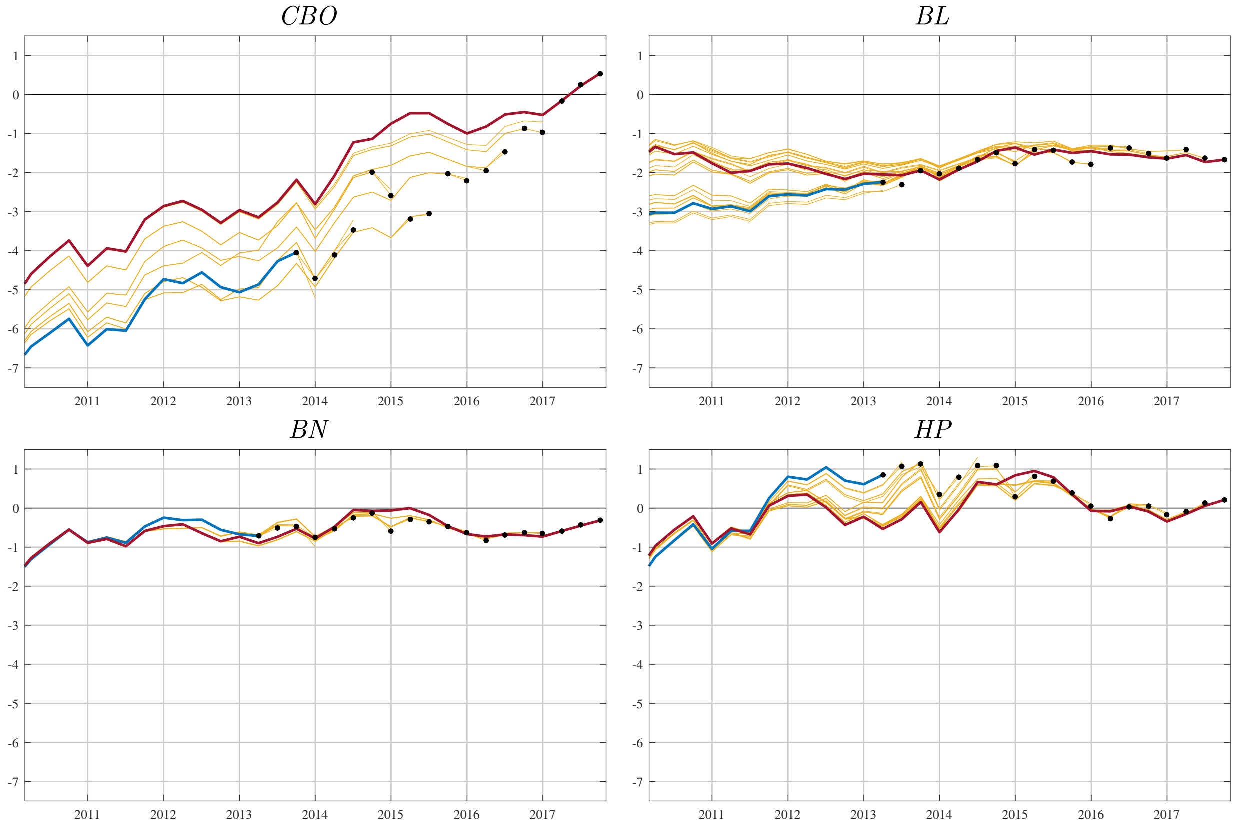
Table 3: Statistics on Gross Domestic Output revisions

	GDP	GDI	BEA	GDPplus	BL
Mean	0.13	0.05	0.09	-0.08	-0.17
Mean Absolute Value	0.73	1.15	0.84	0.93	0.53
Standard Deviation	1.18	1.61	1.06	1.24	0.81
Root Mean Squared Error	1.16	1.57	1.03	1.21	0.81

Rudd, 2016), but for GDO revisions. Here with the term "revision" we mean the difference between the real-time estimate, i.e., the black dot in Figure 7, and the "final" estimate, i.e., the value indicated by the red line in Figure 7.

By looking at Figure 7 and Table 3 three main conclusions can be drawn: first, our estimate of GDO revises less than those provided by the BEA or the Philadelphia Fed in that the mean absolute value of the revision, the standard deviation of the revision and the Root Mean Squared Error (RMSE) of the revision are smaller than those of the other

Figure 8: Output gap
Real-time estimation



In each plot the blue line is the estimate of the output gap obtained with the vintage of data from September 13 2013, the red line is the estimate obtained with the vintage of data from March 30 2018, and the yellow lines are the estimate obtained with all the remaining 31 vintages. Finally, the 19 black dots represent the estimate of the output gap for quarter Q and year Y obtained with the vintage of data ending at quarter Q and year Y corresponding to the first release of GDI for quarter Q and year Y. Note that, for CBO the blue line represents the estimate available as of April 11 2014.

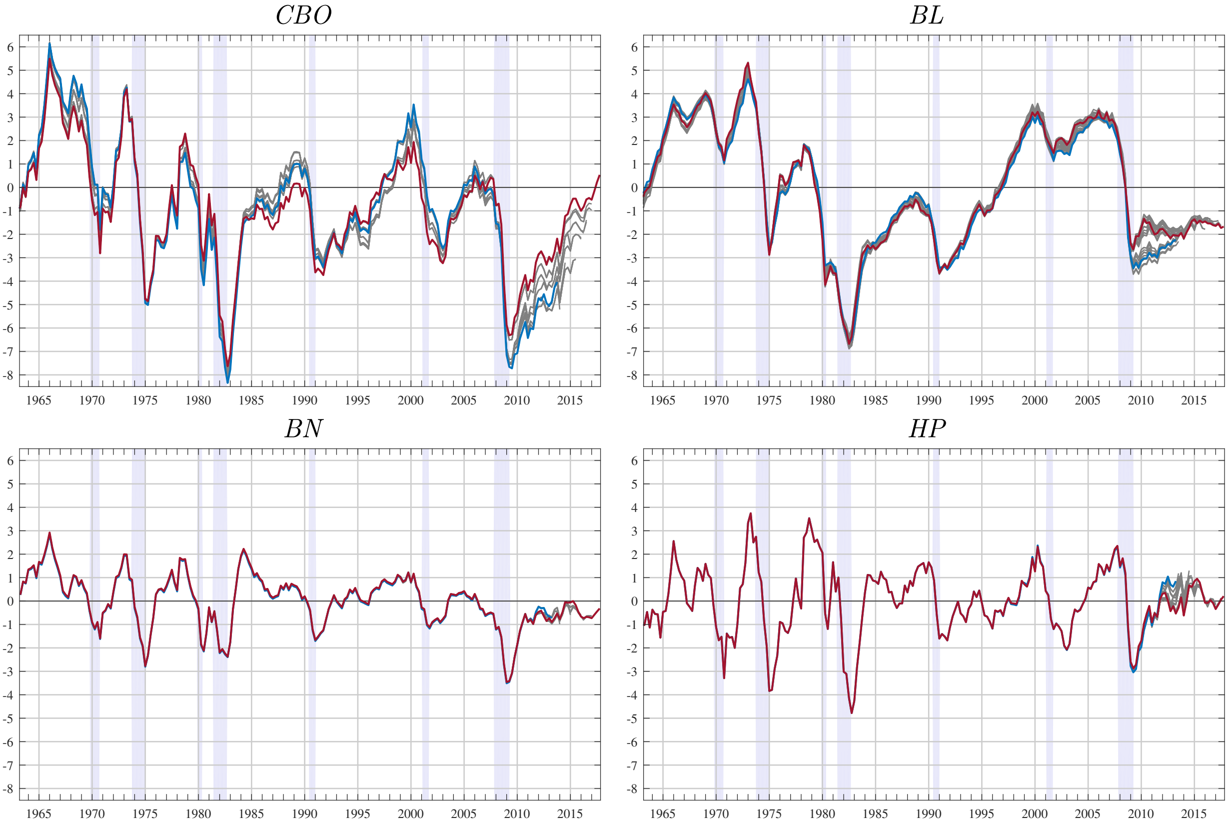
Table 4: Statistics on output gap revisions

	CBO	BN	HP	BL
Mean	1.26	0.04	-0.32	0.02
Mean Absolute Value	1.26	0.11	0.42	0.13
Standard Deviation	0.9	0.17	0.54	0.18
Root Mean Squared Error	1.53	0.17	0.62	0.17

measures. This result confirms the intuition that data revisions are by nature idiosyncratic, and that by pooling a large amount of data it is possible to partially parse them out. Second, all the main conclusions drawn from the analysis in Section 4 are robust to data revision.

Moving to the output gap, Figure 8 shows real-time estimates of the output gap, while Table 4 shows the standard descriptive statistics for the revisions. We compare the real-time performance of our estimate of the output gap with the one of the CBO, as well

Figure 9: Output Gap
Real-time estimation

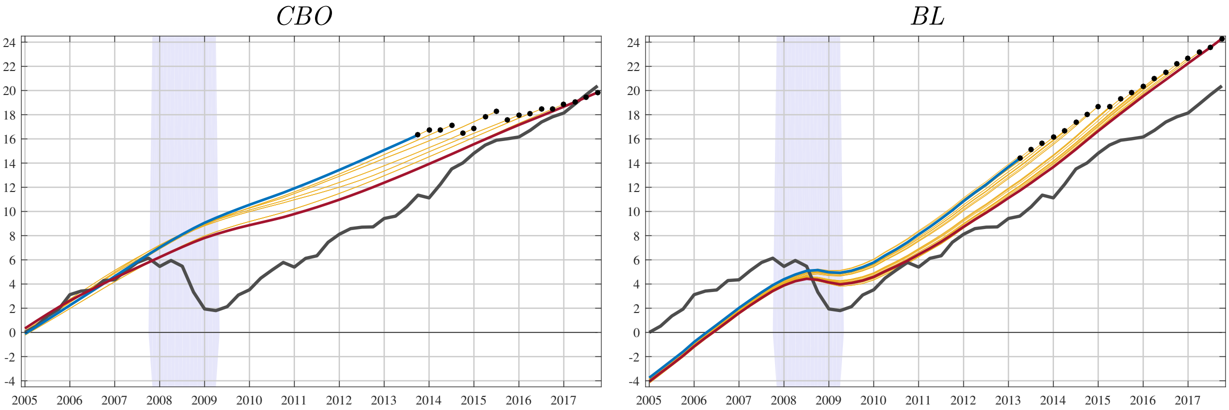


In each plot the blue line is the estimate of the output gap obtained with the vintage of data from September 13 2013, the red line is the estimate obtained with the vintage of data from March 30 2018, and the gray lines are the estimate obtained with all the remaining 31 vintages. Note that, for CBO the blue line represents the estimate available as of April 11 2014.

as with two classic univariate estimates. The first univariate estimate is obtained with the Hodrick-Prescott (HP) filter with a smoothing parameter set at 1600, the value used in the literature for quarterly data. The second univariate estimate is obtained with a Beveridge-Nelson (BN) decomposition estimated as in Kamber et al. (2018), who develop a "BN filter" with low signal-to-noise ratio that is imposed by estimating a univariate AR model with Bayesian methods.

By looking at Figure 8 and Table 4 four main conclusions can be drawn: first, our estimate of the Output Gap revises substantially less compared to both the CBO estimate and HP filter estimate, while it revises about the same as the estimate obtained with the BN decomposition. Second, all the main conclusions drawn from the analysis in Section 5 are robust to data revision (see also Figure 9). Third, by comparing the estimate obtained on the first vintage of data (the blue line) with the one obtained on the last vintage (red line), it is interesting to notice that our model and the CBO model revised in the same direction, i.e., both models interpreted incoming data as signaling that there was less slack in the economy. Last, and not related specifically with the real-time evaluation, as we can see from Figure 9 the peaks and the troughs of the output gap estimated by our

Figure 10: Potential output
Real-time estimation



In each plot the blue line is the estimate of potential output obtained with the vintage of data from September 13 2013, the red line is the estimate obtained with the vintage of data from March 30 2018, and the yellow lines are the estimate obtained with all the remaining 31 vintages. The 19 black dots represent the estimate of the output gap for quarter Q and year Y obtained with the vintage of data ending at quarter Q and year Y corresponding to the first release of GDI for quarter Q and year Y. Finally, the dark gray line is real GDP. Note that for CBO the blue line represents the estimate available as of April 11 2014. The unit of measure is log-deviation from 2005:Q1 GDP. This means that in 2010:Q1 the level of real GDP was 4% higher than in 2005:Q1, while our estimate of potential output obtained with the vintage of data from September 13 2013 was 6% higher.

measure are broadly consistent with those estimated with univariate benchmarks, however the amplitude of the fluctuations of our estimated output gap is much larger.

To conclude, by looking at Figure 8 we can see that the output gap estimate obtained with the HP filter revises a lot, often revising positive numbers to negative ones. This kind of revisions are those that led Orphanides and van Norden (2002) to conclude that output gap estimates are unreliable in real-time. By contrast, our real-time exercise shows that the output gap estimates obtained with our model are pretty robust.

Finally, Figure 10 shows real-time estimates of potential output. Notice that, while for the CBO decomposes GDP into the sum of potential output and the output gap, in our model this is not the case due to the presence of the idiosyncratic component, and hence potential output is the common trending component of GDO. From Figure 10 it can be clearly seen that both our model and the CBO model revised estimates of the level of the potential output downwards. However while we estimate that the impact of the crisis has resulted in a level shift in potential output, the CBO estimates not only a level shift, but also a decrease in potential growth (see also the right plot in Figure 4). Our result are in line with those of Coibion et al. (2017), who also estimated that potential output fell during the financial crisis, but then since 2010 has started to grow at approximately the same pace as before the crisis.[10]

[10] Coibion et al. (2017) argue that CBO estimates of potential GDP "are failing to adequately distinguish between permanent and transitory shocks". They estimate potential output by using the method proposed by Blanchard and Quah (1989) according to which potential output responds only to permanent shocks and not to transitory ones.

7 Conclusions

In this paper we measure US aggregate output, and we decompose it into potential output and output gap. To do so, we first disentangle common and idiosyncratic dynamics by using a Non-Stationary Approximate Dynamic Factor Model (NS-DFM) estimated on a large dataset of US macroeconomic indicators. Then, we disentangle common trends from common cycles by applying a non-parametric Trend-Cycle decomposition based on the eigen-analysis of the long run covariance matrix of the estimated latent common factors.

The first part of our analysis is about measuring aggregate output: our estimate shows that since 2010, output growth was on average 0.4 percentage point higher than Bureau of Economic Analysis estimates. Our measure's higher growth has been concentrated in the first quarter of the year, suggesting that weakness in the GDP's first-quarter growth over the past several years is not a matter of shifting growth across quarters, rather it is a matter of missing Q1 growth that is not made up later in the year.

The second part of our analysis is about measuring the output gap: our estimate suggests that for several consecutive years before the financial crisis the economy operated above its potential, and hence that level of output was unsustainable. Moreover, our estimate suggests that growth after the financial crisis is due primarily to permanent factors. As a consequence, our output gap estimate indicates that as of 2017:Q4 there is still slack in the economy.

The last part of our analysis investigates the real-time properties of our suggested measures and we show that our estimates revise less than, or comparably with, other commonly used measures. This confirms the intuition that data revisions are by nature idiosyncratic, and therefore, by pooling a large amount of data, it is possible to get a robust estimate of the signal while parsing out the noise.

Our analysis has been so far deliberately entirely data driven, and we have been careful in imposing the least possible amount of restrictions to let the data speak freely. This approach has undeniably some important merits, as estimation of aggregate output seems to fit naturally in our framework, and the Trend-Cycle decomposition that we obtain for our estimate of aggregate output is economically sensible and capable to capture important aspects of aggregate output dynamics. However, we believe that identifying the common cycles, as well as imposing economically meaningful constraints, is an essential step forward. Our view is that one way to proceed is to consider Bayesian estimation of the model, so that our economic and statistical knowledge of the data can be included by means of suitable priors. All this is the subject of our current research.

Acknowledgements

We thank for helpful comments the participants to the: "Central Bank Forecasting Conference", IAAE annual conference, Université du Québec à Montréal (UQAM), June 2018, "10^{th} ECB Workshop on Forecasting Techniques", European Central Bank, Frankfurt, June 2018, Barcelona GSE Summer Forum, June 2018, Bank of Italy, Rome, April 2018, ASSA

Meetings, Philadelphia, January 2018, Central Bank Forecasting Conference Federal Reserve of St. Louis, November 2017. We would like also to thank: Stephanie Aaronson, Gianni Amisano, Marco Lippi, Lucrezia Reichlin, and John Roberts. Finally, we are indebted with Luke Van Cleave for his invaluable help with the construction of real-time data vintages, and to Cassandra Duchan for excellent research assistance.

References

Aastveit, K. A. and Trovik, T. (2014). Estimating the output gap in real time: A factor model approach. *The Quarterly Review of Economics and Finance*, 54:180–193.

Antolin-Diaz, J., Drechsel, T., and Petrella, I. (2017). Tracking the slowdown in long-run GDP growth. *The Review of Economics and Statistics*, 99:343–356.

Aruoba, S. B., Diebold, F. X., Nalewaik, J., Schorfheide, F., and Song, D. (2016). Improving GDP measurement: A measurement-error perspective. *Journal of Econometrics*, 191:384–397.

Bai, J. (2004). Estimating cross-section common stochastic trends in nonstationary panel data. *Journal of Econometrics*, 122:137–183.

Bai, J. and Li, K. (2016). Maximum likelihood estimation and inference for approximate factor models of high dimension. *Review of Economics and Statistics*, 98:298–309.

Bai, J. and Liao, Y. (2016). Efficient estimation of approximate factor models via penalized maximum likelihood. *Journal of Econometrics*, 191:1–18.

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

Bai, J. and Ng, S. (2004). A PANIC attack on unit roots and cointegration. *Econometrica*, 72:1127–1177.

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

Bańbura, M. and Modugno, M. (2014). Maximum likelihood estimation of factor models on datasets with arbitrary pattern of missing data. *Journal of Applied Econometrics*, 29:133–160.

Barigozzi, M., Lippi, M., and Luciani, M. (2016a). Dynamic factor models, cointegration, and error correction mechanisms. FEDS 2016-18, Board of Governors of the Federal Reserve System.

Barigozzi, M., Lippi, M., and Luciani, M. (2016b). Non-stationary dynamic factor models for large datasets. FEDS 2016-24, Board of Governors of the Federal Reserve System.

Barigozzi, M. and Luciani, M. (2017). Common factors, trends, and cycles in large datasets. FEDS 2017-111, Board of Governors of the Federal Reserve System.

Beveridge, S. and Nelson, C. R. (1981). A new approach to decomposition of economic time series into permanent and transitory components with particular attention to measurement of the 'business cycle'. *Journal of Monetary Economics*, 7:151–174.

Blanchard, O. J. and Quah, D. (1989). The dynamic effects of aggregate demand and supply disturbance. *The American Economic Reivew*, 79:655–673.

Borio, C., Disyatat, P., and Juselius, M. (2017). Rethinking potential output: embedding information about the financial cycle. *Oxford Economic Papers*, 69:655–677.

Coibion, O., Gorodnichenko, Y., and Ulate, M. (2017). The cyclical sensitivity in estimates of potential output. Technical Report 23580, National Bureau of Economic Research.

Congressional Budget Office (2001). CBO's method for estimating potential output: An update.

Congressional Budget Office (2004). A summary of alternative methods for estimating potential GDP. CBO Background Paper.

Coroneo, L., Giannone, D., and Modugno, M. (2016). Unspanned macroeconomic factors in the yield curve. *Journal of Business and Economic Statistics*, 34:472–485.

Council of Economic Advisers (2015). A better measure of economic growth: Gross Domestic Output (GDO). CEA Issue Brief.

Creal, D., Koopman, S. J., and Zivot, E. (2010). Extracting a robust US business cycle using a time-varying multivariate model-based bandpass filter. *Journal of Applied Econometrics*, 25:695–719.

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

Del Negro, M., Schorfheide, F., Smets, F., and Wouters, R. (2007). On the fit of New Keynesian models. *Journal of Business and Economic Statistics*, 25:123–143.

Doz, C., Giannone, D., and Reichlin, L. (2011). A two-step estimator for large approximate dynamic factor models based on Kalman filtering. *Journal of Econometrics*, 164:188–205.

Doz, C., Giannone, D., and Reichlin, L. (2012). A quasi maximum likelihood approach for large approximate dynamic factor models. *The Review of Economics and Statistics*, 94(4):1014–1024.

Dungey, M., Jacobs, J. P., Tian, J., and Van Norden, S. (2015). Trend in cycle or cycle in trend? New structural identifications for unobserved-components models of US real GDP. *Macroeconomic Dynamics*, 19:776–790.

Edge, R. M. and Rudd, J. B. (2016). Real-time properties of the Federal Reserve's output gap. *The Review of Economics and Statistics*, 98:569–583.

Escribano, A. and Peña, D. (1994). Cointegration and common factors. *Journal of Time Series Analysis*, 15:577–586.

Fleischman, C. A. and Roberts, J. M. (2011). From many series, one cycle: improved estimates of the business cycle from a multivariate unobserved components model. FEDS 2011-046, Board of Governors of the Federal Reserve System.

Forni, M., Giannone, D., Lippi, M., and Reichlin, L. (2009). Opening the black box: Structural factor models versus structural VARs. *Econometric Theory*, 25:1319–1347.

Forni, M., Hallin, M., Lippi, M., and Reichlin, L. (2000). The Generalized Dynamic Factor Model: Identification and estimation. *The Review of Economics and Statistics*, 82:540–554.

Fuster, A., Geddes, E., and Haughwout, A. (2017). Houses as ATMs no longer. Federal Reserve Bank of New York Liberty Street Economics blog, February 15. http://libertystreeteconomics.newyorkfed.org/2017/02/houses-as-atms-no-longer.html.

Garratt, A., Robertson, D., and Wright, S. (2006). Permanent vs transitory components and economic fundamentals. *Journal of Applied Econometrics*, 21:521–542.

Gilbert, C., Morin, N., Paciorek, A. D., and Sahm, C. R. (2015). Residual seasonality in GDP. FEDS Notes 2015-05-14, Board of Governors of the Federal Reserve System.

Gonzalo, J. and Granger, C. (1995). Estimation of common long-memory components in cointegrated systems. *Journal of Business and Economic Statistics*, 13:27–35.

Gonzalo, J. and Ng, S. (2001). A systematic framework for analyzing the dynamic effects of permanent and transitory shocks. *Journal of Economic Dynamics and Control*, 25:1527–1546.

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

Harvey, A. C. (1985). Trends and cycles in macroeconomic time series. *Journal of Business & Economic Statistics*, 3:216–227.

Hasenzagl, T., Pellegrino, F., Reichlin, L., and Ricco, G. (2018). A model of the Fed's view on inflation. CEPR Discussion papers DP12564.

Haughwout, A. F. and Okah, E. (2009). Below the line: Estimates of negative equity among nonprime mortgage borrowers. *Economic Policy Review*, 15:31–43. Federal Reserve Bank of New York.

Jarociński, M. and Lenza, M. (2018). An inflation-predicting measure of the output gap in the euro area. *Journal of Money Credit and Banking*. forthcoming.

Johansen, S. (1991). Estimation and hypothesis testing of cointegration vectors in Gaussian vector autoregressive models. *Econometrica*, 59:1551–80.

Juvenal, L. and Petrella, I. (2015). Speculation in the oil market. *Journal of Applied Econometrics*, 30:1099–1255.

Kamber, G., Morley, J., and Wong, B. (2018). Intuitive and reliable estimates of the output gap from a Beveridge-Nelson filter. *Review of Economics and Statistics*, 100:550–566.

Kasa, K. (1992). Common stochastic trends in international stock markets. *Journal of Monetary Economics*, 29:95–124.

Kiley, M. T. (2013). Output gaps. *Journal of Macroeconomics*, 37(C):1–18.

Lengermann, P., Morin, N., Paciorek, A. D., Pinto, E., and Sahm, C. R. (2017). Another look at residual seasonality in GDP. FEDS Notes 2017-07-28, Board of Governors of the Federal Reserve System.

Lippi, M. and Reichlin, L. (1994a). Common and uncommon trends and cycles. *European Economic Review*, 38:624–635.

Lippi, M. and Reichlin, L. (1994b). Diffusion of technical change and the decomposition of output into trend and cycle. *The Review of Economic Studies*, 61:19–30.

Luciani, M. (2015). Monetary policy and the housing market: A structural factor analysis. *Journal of Applied Econometrics*, 30:199–218.

Lunsford, K. G. (2017). Lingering residual seasonality in GDP growth. Economic Commentary 2017-06, Federal Reserve Bank of Cleveland.

Morley, J., Nelson, C. R., and Zivot, E. (2003). Why are unobserved component and Beveridge-Nelson trend-cycle decompositions of GDP so different. *The Review of Economics and Statistics*, 85:235–243.

Morley, J. and Wong, B. (2017). Estimating and accounting for the output gap with large Bayesian vector autoregressions. CAMA Working Papers 2017-46, Centre for Applied Macroeconomic Analysis.

Moulton, B. R. and Cowan, B. D. (2016). Residual seasonality in GDP and GDI: Findings and next steps. *Survey of Current Business*, 96:1–6.

Nalewaik, J. J. (2010). The income- and expenditure-side measures of output growth. *Brookings Papers on Economic Activity*, 1:71–106.

Nalewaik, J. J. (2012). Estimating probabilities of recession in real time using GDP and GDI. *Journal of Money Credit and Banking*, 44:235–253.

Onatski, A. (2009). Testing hypotheses about the number of factors in large factor models. *Econometrica*, 77:1447–1479.

Orphanides, A. and van Norden, S. (2002). The unreliability of output-gap estimates in real time. *The Review of Economics and Statistics*, 84:569–583.

Peña, D. and Poncela, P. (2006). Nonstationary dynamic factor analysis. *Journal of Statistical Planning and Inference*, 136:1237–1257.

Proietti, T. (1997). Short-run dynamics in cointegrated systems. *Oxford Bulletin of Economics and Statistics*, 59:405–422.

Quah, D. and Sargent, T. J. (1993). A dynamic index model for large cross sections. In *Business cycles, indicators and forecasting*. University of Chicago Press.

Reis, R. and Watson, M. W. (2010). Relative goods' prices, pure inflation, and the Phillips correlation. *American Economic Journal Macroeconomics*, 2:128–157.

Rudebusch, G. D., Wilson, D., and Mahedy, T. (2015). The puzzle of weak first-quarter growth. Economic Letter 2015-16, Federal Reserve Bank of San Francisco.

Sargent, T. J. and Sims, C. A. (1977). Business cycle modeling without pretending to have too much a priori economic theory. In *New methods in business cycle research*. Federal Reserve Bank of Minneapolis.

Seong, B., Ahn, S. K., and Zadrozny, P. A. (2013). Estimation of vector error correction models with mixed-frequency data. *Journal of Time Series Analysis*, 34:194–205.

Shumway, R. H. and Stoffer, D. S. (1982). An approach to time series smoothing and forecasting using the EM algorithm. *Journal of Time Series Analysis*, 3:253–264.

Sims, C., Stock, J. H., and Watson, M. W. (1990). Inference in linear time series models with some unit roots. *Econometrica*, 58:113–144.

Stock, J. H. and Watson, M. W. (1988). Testing for common trends. *Journal of the American Statistical Association*, 83:1097–1107.

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

Stock, J. H. and Watson, M. W. (2005). Implications of dynamic factor models for VAR analysis. Working Paper 11467, NBER.

Vahid, F. and Engle, R. F. (1993). Common trends and common cycles. *Journal of Applied Econometrics*, 8:341–360.

Watson, M. W. (1986). Univariate detrending methods with stochastic trends. *Journal of Monetary Economics*, 18:49–75.

Watson, M. W. (1994). Vector autoregressions and cointegration. In Engle, R. and McFadden, D., editors, *Handbook of Econometrics, Volume IV*. Elsevier Science.

Watson, M. W. and Engle, R. F. (1983). Alternative algorithms for the estimation of dynamic factor, mimic and varying coefficients regression models. *Journal of Econometrics*, 23:385–400.

Zhang, R., Robinson, P., and Yao, Q. (2018). Identifying cointegration by eigenanalysis. *Journal of the American Statistical Association*. forthcoming.

Appendix A Alternative estimation of the output gap

In the Trend-Cycle decomposition of Section 5.1, the vector of common cycles $\widehat{\mathbf{G}}_t$ is of dimension $(r-q+d)$, and, as we have seen, it belongs to the cointegration space of $\widehat{\mathbf{F}}_t$ by construction. However, as mentioned in Section 5.1, the constraint $\mathbf{F}_t = \mathbf{K}(\boldsymbol{f}_t \cdots \boldsymbol{f}_{t-s})'$ has a specific implication for the cointegration rank of $\mathbf{F}_t$, which could be exploited in estimating the output gap. Indeed, it can be easily verified that $\mathbf{F}_t$ must have cointegration rank d as $\boldsymbol{f}_t$, regardless of the choice of $\mathbf{K}$—see Appendix B.3. Therefore, in principle $\widehat{\mathbf{G}}_t$ could be reduced to a process of dimension d without loosing information. This is not an easy task, since in order to isolate d cycles we would need to maximize their contribution to the dynamics of $\widehat{\mathbf{G}}_t$ at all frequencies, as opposed to the contribution of common trends which is by construction maximized at the zero frequency.

In this section, we consider two alternative ways of reducing the dimension of the common cycles $\widehat{\mathbf{G}}_t$. For simplicity of exposition, we use the dimensions considered in the empirical analysis, thus $r = 6$, $q = 3$, and $(q-d) = 1$, so that $\widehat{\mathbf{G}}_t$ has dimension $(r-q+d) = 5$, and the estimated cycles vector must have dimension $d = 2$. In other words, there are $(r-q) = 3$ redundant processes in $\widehat{\mathbf{G}}_t$.

A first simple way to reduce the dimension of $\widehat{\mathbf{G}}_t$ consists in keeping as common cycles only the first two components of $\widehat{\mathbf{G}}_t$. The rationale for this choice is based on the fact that since we extract the trend and the cycles by principal component analysis on the long-run covariance of $\widehat{\mathbf{F}}_t$, then the resulting principal components are ordered according to their contribution (in first differences) to the lowest frequencies. Low frequencies are in turn the predominant ones in real variables as GDP and the unemployment rate, and indeed, as already shown by the results of Table 2, the first two cycles explain almost all variation of the common cyclical component of those variables. In particular, the matrix $(\widehat{\boldsymbol{\Phi}}\ \widehat{\boldsymbol{\Phi}}_\perp)$ of ordered normalized eigenvectors of the long run covariance matrix $\widehat{\mathbf{S}}$ can be further partitioned according to the corresponding eigenvalues of $\widehat{\mathbf{S}}$ in decreasing order, such that $(\widehat{\boldsymbol{\Phi}}\ \widehat{\boldsymbol{\Phi}}_\perp) = (\widehat{\boldsymbol{\Phi}}\ \widehat{\boldsymbol{\Phi}}^*\widehat{\boldsymbol{\Phi}}^*_\perp)$, where $\widehat{\boldsymbol{\Phi}}$ is 6×1, $\widehat{\boldsymbol{\Phi}}^*$ is 6×2, and $\widehat{\boldsymbol{\Phi}}^*_\perp$ is 6×3 with $\widehat{\boldsymbol{\Phi}}^{*\prime}\widehat{\boldsymbol{\Phi}}^*_\perp = \mathbf{0}$. We label this method as BL-PCA and we define the two-dimensional process of common cycles as the projection

$$\widehat{\mathbf{C}}_t^{\text{BL-PCA}} = \widehat{\boldsymbol{\Phi}}^{*\prime}\widehat{\mathbf{F}}_t,$$

which it implies the Trend-Cycle decomposition of the common component:

$$\widehat{\chi}_{it} = \widehat{\boldsymbol{\lambda}}_i'\widehat{\boldsymbol{\Phi}}\widehat{\mathbf{T}}_t + \widehat{\boldsymbol{\lambda}}_i'\widehat{\boldsymbol{\Phi}}^*\widehat{\mathbf{C}}_t^{\text{BL-PCA}} + \widehat{\boldsymbol{\lambda}}_i'\widehat{\boldsymbol{\Phi}}^*_\perp\widehat{\boldsymbol{\Phi}}^{*\prime}_\perp\widehat{\mathbf{F}}_t. \tag{A1}$$

Note that the last term has covariance of rank $(r-q) = 3$ and it would not be present if we had $r = q$.

A second way to reduce the dimension of $\widehat{\mathbf{G}}_t$ consists in looking for the two-dimensional projection of $\widehat{\mathbf{G}}_t$ with maximum spectral density. Consider the VAR(2):

$$\widehat{\mathbf{G}}_t = \boldsymbol{\mathcal{A}}_1\widehat{\mathbf{G}}_{t-1} + \boldsymbol{\mathcal{A}}_2\widehat{\mathbf{G}}_{t-2} + \boldsymbol{v}_t, \tag{A2}$$

where $\boldsymbol{v}_t \overset{i.i.d.}{\sim} \mathcal{N}(\mathbf{0}_5, \boldsymbol{\Sigma}_v)$ and $\det(\mathbf{I}_5 - \boldsymbol{\mathcal{A}}_1 z - \boldsymbol{\mathcal{A}}_2 z^2) \neq 0$ for $|z| \leq 1$, since $\widehat{\mathbf{G}}_t$ is stationary.

Once we estimate (A2) we can compute the residuals $\widehat{\boldsymbol{v}}_t$ and an estimate of their covariance matrix $\widehat{\boldsymbol{\Sigma}}_v$. Denote as $\widehat{\boldsymbol{\mathcal{H}}}$ the 5×2 matrix having as columns the first two leading normalized eigenvectors of $\widehat{\boldsymbol{\Sigma}}_v$. We label this method as BL-PCA and the two-dimensional process of common cycles can then be found by projecting $\widehat{\mathbf{G}}_t$ onto the space spanned by the columns of $\widehat{\boldsymbol{\mathcal{H}}}$:

$$\widehat{\mathbf{C}}_t^{\text{BL-VAR}} = \widehat{\boldsymbol{\mathcal{H}}}'\widehat{\mathbf{G}}_t.$$

The estimated Trend-Cycle decomposition of the common component is then given by

$$\widehat{\chi}_{it} = \widehat{\boldsymbol{\lambda}}_i'\widehat{\boldsymbol{\Phi}}\widehat{\mathbf{T}}_t + \widehat{\boldsymbol{\lambda}}_i'\widehat{\boldsymbol{\Phi}}_\perp\widehat{\boldsymbol{\mathcal{H}}}\widehat{\mathbf{C}}_t^{\text{BL-VAR}} + \widehat{\boldsymbol{\lambda}}_i'\widehat{\boldsymbol{\Phi}}_\perp\widehat{\boldsymbol{\mathcal{H}}}_\perp\widehat{\boldsymbol{\mathcal{H}}}_\perp'\widehat{\mathbf{G}}_t, \tag{A3}$$

where $\widehat{\boldsymbol{\mathcal{H}}}_\perp$ is 5×3 and such that $\widehat{\boldsymbol{\mathcal{H}}}_\perp'\widehat{\boldsymbol{\mathcal{H}}} = \mathbf{0}$. Again, if we had $r = q$ the last term on the right-hand-side of (A3) would disappear.

To give an intuition of what are the implications of the two methods, we show in Figure A1 the spectral densities of the estimated common trend $\Delta\widehat{\mathbf{T}}_t$ (blue line which is the same in both plots), together with the spectral densities of the two cycle components of $\Delta\mathbf{C}_t^{\text{BL-PCA}}$ (red lines, left plot) and $\Delta\mathbf{C}_t^{\text{BL-VAR}}$ (red lines, right plot), and those of the first differences of the three residual processes (black lines). Moreover, in Table A1, we also report for some selected variables the percentage of variance of the cyclical common components explained by the two-dimensional cycle processes estimated with the two different methods.

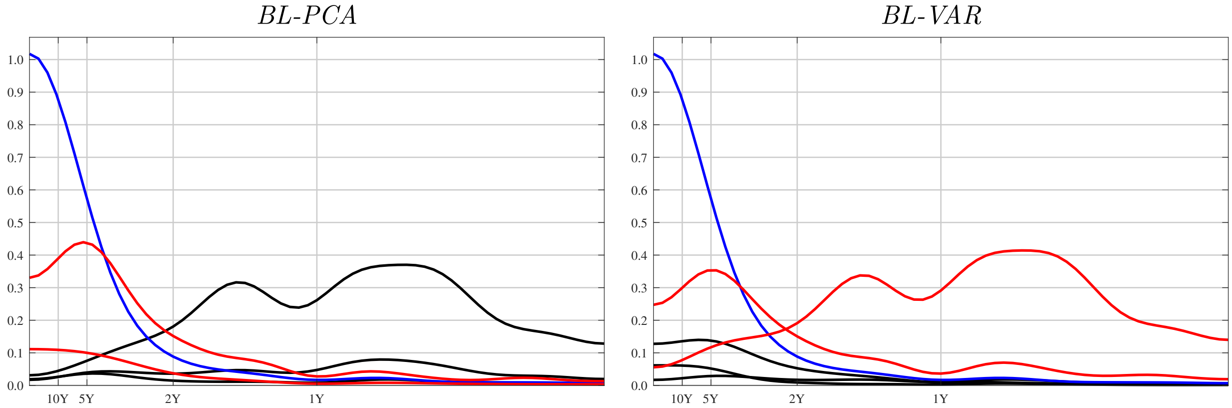
Both identification strategies have pros and cons, and a priori it is very difficult to determine which one is better. In particular, the main advantage of BL-PCA is simplicity and the fact that it requires no further estimation. On the other hand, by construction this method cuts out high frequencies (typical for example of nominal and financial variables, see the third to fifth column of Table 2), an exclusion for which there is no theoretical ground. This is clear from the left plot of Figure A1, where we see that $\Delta\mathbf{C}_t^{\text{BL-PCA}}$ does not contribute to frequencies corresponding to periods shorter than two years. By contrast, the main cons of BL-VAR identification is that it requires an additional estimation step, thus implying a potential loss of efficiency of the estimator. On the other hand, its main advantages are that all frequencies are captured, and that the three residual processes account (by construction) for a very small proportion of total variance. This can be seen by looking at the right plot in Figure A1 and at the last column of Table A1. Last, by looking at Table A2, which shows the correlation between each cycle $\widehat{G}_{jt}$ and each identified cycle $\widehat{C}_{jt}^{\text{BL-PCA}}$ or $\widehat{C}_{jt}^{\text{BL-VAR}}$ we clearly see that $\widehat{C}_{2t}^{\text{BL-VAR}}$ is nearly identical to $\widehat{G}_{1t} \equiv \widehat{C}_{1t}^{\text{BL-PCA}}$, while $\widehat{C}_{1t}^{\text{BL-VAR}}$ is a linear combination of all five $\widehat{G}_{jt}$ cycles.

To conclude, Figure A2 reports the estimates of the output gap and the unemployment gap obtained using the BL-PCA or BL-VAR method as defined by the second term on the right hand side of (A1) or (A3), respectively. Starting with the output gap, the three estimates are very similar until the financial crisis. Indeed, when the two cycles are identified using BL-VAR the output gap reach a trough about 2.2 percentage points lower than estimated with both our benchmark model, and with BL-PCA identification. This make sense as when we identify the cycles with BL-VAR we capture higher frequencies typical of the oil price (see Table A1), which in 2018 plunged from about \$130 in July to \$40 in

Table A1: Variance decomposition
Percentage of total variance of the common cyclical component

	$\widehat{\mathbf{C}}_t^{\text{BL-PCA}}$	$\widehat{\mathbf{C}}_t^{\text{BL-VAR}}$
GDP	94.4	91.9
Unemployment rate	95.4	11.0
PCE price inflation	8.7	35.3
Core PCE price inflation	7.1	15.1
Fed funds rate	2.8	8.4
WTI price inflation	7.3	51.0
S&P500	58.9	14.6

Figure A1: Spectral densities of common cycles



The plots report the spectral densities of the common trend $\Delta\widehat{\mathbf{T}}_t$ (blue line), the common cycles $\Delta\widehat{\mathbf{C}}_t^{\text{BL-PCA}}$ (red lies in the left plot) and $\Delta\widehat{\mathbf{C}}_t^{\text{BL-VAR}}$ (red lines in the right plot), and the residual cycles given by the third term on the right hand side of (A1) or (A3) (black lines). On the horizontal axis of the plots in the left column we report periods τ measured in years such that the corresponding frequencies are given by $\theta = 2\pi/(4\tau)$.

December of the same year. Moving on, by looking at the right plot in Figure A2 we can clearly see that the cycles identified with BL-VAR fail to provide a sensible estimate of the unemployment gap. In particular, it looks like these cycles are missing some low-mid frequencies that are very much relevant for the unemployment gap, whereas it is capturing high-frequencies that have nothing to do with it. This interpretation is supported by both the spectral density in the right plot of Figure A1, from which we can see that there is a bulk of fluctuations with period longer than five years that are not captured by BL-VAR, and by Table A2, from which we can see that the fifth cycle $\widehat{C}_{5t}^{\text{BL-VAR}}$, which is excluded, is very much related with $\widehat{G}_{2t}$, which in turns explains about 11% of the variance of the unemployment gap.

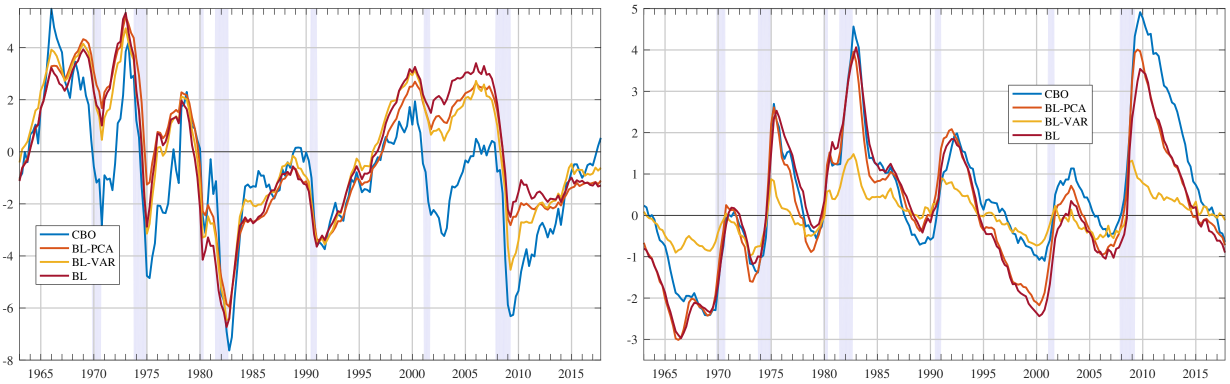
To summarize, in Section 5.2 we have shown that the five different cycles that drive our common cyclical components, are loaded in a different way by the different variables in the dataset. As we discussed in this section, in principle, this five different cycles can be reduced to two cycles, in that theoretically there are three redundant cycles. However, in practice this task poses a tradeoff. From a purely econometric point of view the right way

Table A2: Correlation between cycles

	$\widehat{G}_{1t}$	$\widehat{G}_{2t}$	$\widehat{G}_{3t}$	$\widehat{G}_{4t}$	$\widehat{G}_{5t}$
$\widehat{C}_{1t}^{\text{BL-VAR}}$	-0.54	0.32	0.11	0.76	0.15
$\widehat{C}_{2t}^{\text{BL-VAR}}$	0.99	0.08	0.09	0.00	0.09
$\widehat{C}_{3t}^{\text{BL-VAR}}$	0.10	-0.30	-0.92	0.01	0.23
$\widehat{C}_{4t}^{\text{BL-VAR}}$	-0.85	-0.37	0.24	-0.15	0.25
$\widehat{C}_{5t}^{\text{BL-VAR}}$	-0.32	0.94	-0.04	-0.10	0.06
$\widehat{C}_{1t}^{\text{BL-PCA}}$	1.00	0.00	0.00	0.00	0.00
$\widehat{C}_{2t}^{\text{BL-PCA}}$	0.00	1.00	0.00	0.00	0.00

Figure A2: Output gap and unemployment gap
Alternative Estimations

Output Gap *Unemployment Gap*



The left/right plot shows the level of the output/unemployment gap estimated by the CBO (blue line), together with the estimates obtained with the BL-PCA (orange line) and BL-VAR (yellow line) identification methods, and our estimates of Section 5.2 (red line).

to proceed would be to maximize the variance explained at all frequencies as in the BL-VAR approach. However, by doing so we capture frequencies that seem to be not relevant for real macroeconomic variables, while leaving out important slow moving components.

Appendix B Properties of $\mathbf{F}_t$

Hereafter, for simplicity we consider the case $s = 1$, thus $r = 2q$. Moreover, we assume $p = 2$ in (5) as in our empirical application, such that as shown below we have $\ell = 3$ in (3).

B.1 Reduced and structural form

The reduced form (4)-(5) can be written as

$$\mathbf{x}_t = \mathbf{\Lambda}\mathbf{F}_t + \boldsymbol{\xi}_t, \tag{B4}$$

$$\begin{pmatrix} \mathbf{F}_t \\ \mathbf{F}_{t-1} \end{pmatrix} = \begin{pmatrix} \mathbf{A}_1 & \mathbf{A}_2 \\ \mathbf{I}_r & \mathbf{0}_r \end{pmatrix} \begin{pmatrix} \mathbf{F}_{t-1} \\ \mathbf{F}_{t-2} \end{pmatrix} + \begin{pmatrix} \mathbf{H} \\ \mathbf{0}_{r\times q} \end{pmatrix} \mathbf{u}_t, \tag{B5}$$

with $\mathbf{\Lambda} = (\boldsymbol{\lambda}_1 \cdots \boldsymbol{\lambda}_n)'$ the $n \times r$ loadings matrix. Similarly consider the structural form (2)-(3), where, for convenience, in the VAR(3) we write twice the same equation:

$$\mathbf{x}_t = \mathbf{B}_0\boldsymbol{f}_t + \mathbf{B}_1\boldsymbol{f}_{t-1} + \boldsymbol{\xi}_t, \tag{B6}$$

$$\begin{pmatrix} \boldsymbol{f}_t \\ \boldsymbol{f}_{t-1} \\ \boldsymbol{f}_{t-1} \\ \boldsymbol{f}_{t-2} \end{pmatrix} = \begin{pmatrix} \boldsymbol{\mathcal{A}}_1 & \boldsymbol{\mathcal{A}}_2 & \mathbf{0}_q & \boldsymbol{\mathcal{A}}_3 \\ \mathbf{I}_q & \mathbf{0}_q & \mathbf{0}_q & \mathbf{0}_q \\ \mathbf{I}_q & \mathbf{0}_q & \mathbf{0}_q & \mathbf{0}_q \\ \mathbf{0}_q & \mathbf{I}_q & \mathbf{0}_q & \mathbf{0}_q \end{pmatrix} \begin{pmatrix} \boldsymbol{f}_{t-1} \\ \boldsymbol{f}_{t-2} \\ \boldsymbol{f}_{t-2} \\ \boldsymbol{f}_{t-3} \end{pmatrix} + \begin{pmatrix} \mathbf{I}_q \\ \mathbf{0}_q \\ \mathbf{0}_q \\ \mathbf{0}_q \end{pmatrix} \mathbf{u}_t, \tag{B7}$$

where $\mathbf{B}_0 = (\boldsymbol{b}_{01} \cdots \boldsymbol{b}_{0n})'$, $\mathbf{B}_1 = (\boldsymbol{b}_{11} \cdots \boldsymbol{b}_{1n})$ are both $n \times q$.

For the two representations to be equivalent there must exist an invertible $r \times r$ matrix $\mathbf{K}$ such that

$$\mathbf{F}_t = \mathbf{K}(\boldsymbol{f}_t'\,\boldsymbol{f}_{t-1}')', \qquad (\boldsymbol{f}_t'\,\boldsymbol{f}_{t-1}')' = \mathbf{K}^{-1}\mathbf{F}_t, \tag{B8}$$

$$\mathbf{\Lambda} = (\mathbf{B}_0\,\mathbf{B}_1)\mathbf{K}^{-1}, \qquad (\mathbf{B}_0\,\mathbf{B}_1) = \mathbf{\Lambda}\mathbf{K}. \tag{B9}$$

By comparing (B4)-(B5) with (B6)-(B7) and using (B8)-(B9), we have the parameters of the reduced form as functions of those of the structural form

$$\mathbf{A}_1 = \mathbf{K}\begin{pmatrix} \boldsymbol{\mathcal{A}}_1 & \boldsymbol{\mathcal{A}}_2 \\ \mathbf{I}_q & \mathbf{0}_q \end{pmatrix}\mathbf{K}^{-1}, \qquad \mathbf{A}_2 = \mathbf{K}\begin{pmatrix} \mathbf{0}_q & \boldsymbol{\mathcal{A}}_3 \\ \mathbf{0}_q & \mathbf{0}_q \end{pmatrix}\mathbf{K}^{-1}, \qquad \mathbf{H} = \mathbf{K}\begin{pmatrix} \mathbf{I}_q \\ \mathbf{0}_q \end{pmatrix}. \tag{B10}$$

B.2 Number of common trends

The VAR(3) polynomial in (B7) has a Smith-McMillan form (see e.g. Watson, 1994), which without loss fo generality can be written as:

$$\boldsymbol{\mathcal{A}}(L) = \boldsymbol{\mathcal{V}}(L)\begin{pmatrix} (1-L)\mathbf{I}_{(q-d)} & \mathbf{0}_{(q-d)\times d} \\ \mathbf{0}_{d\times(q-d)} & \mathbf{I}_d \end{pmatrix}, \tag{B11}$$

where $\boldsymbol{\mathcal{V}}(L)$ is a $q \times q$ polynomial matrix of order 2 and such that $\det(\boldsymbol{\mathcal{V}}(z)) \neq 0$ for $|z| \leq 1$. Then, as assumed, there are $(q-d)$ unit roots driving the dynamics of $\boldsymbol{f}_t$ and we have $\text{rk}(\boldsymbol{\mathcal{A}}(1)) = d$, which is the cointegration rank of $\boldsymbol{f}_t$.

Now consider the Wold representation for $\Delta \boldsymbol{f}_t$, which is obtained by inverting (B7):

$$\Delta \boldsymbol{f}_t = \boldsymbol{\mathcal{C}}(L)\mathbf{u}_t = (1-L)[\boldsymbol{\mathcal{A}}(L)]^{-1}\mathbf{u}_t, \tag{B12}$$

and by using (B11),

$$\boldsymbol{\mathcal{C}}(L) = (1-L)[\boldsymbol{\mathcal{A}}(L)]^{-1} = \begin{pmatrix} (1-L)\mathbf{I}_d & \mathbf{0}_{d\times(q-d)} \\ \mathbf{0}_{(q-d)\times d} & \mathbf{I}_{(q-d)} \end{pmatrix} [\boldsymbol{\mathcal{V}}(L)]^{-1}\mathbf{u}_t,$$

from which we see that $\text{rk}(\boldsymbol{\mathcal{C}}(1)) = (q-d)$, that is $(q-d)$ shocks have a permanent effect on $\boldsymbol{f}_t$. Hence, the number of shocks with non-vanishing long run effects is equal to the number of unit roots and common trends.

Since $\mathbf{F}_t \sim I(1)$, because of (B8), then it must have unit roots. From the results in Barigozzi et al. (2016a) we have that the unit roots can be at most q, hence in general we have $(q-\tau)$ unit roots for some τ such that $0 \leq \tau \leq q$. Then, the singular VAR(2) in (B5) has an MA representation which using again the Smith-McMillan form of matrix polynomials reads as

$$\Delta \mathbf{F}_t = \mathbf{C}(L)\mathbf{H}\mathbf{u}_t = \begin{pmatrix} (1-L)\mathbf{I}_\tau & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{I}_{(q-\tau)} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & (1-L)\mathbf{I}_{(r-q)} \end{pmatrix} \mathbf{U}(L)\mathbf{H}\mathbf{u}_t, \tag{B13}$$

where $\mathbf{U}(L)$ is an $r \times r$ infinite polynomial matrix with $\det(\mathbf{U}(z)) \neq 0$ for $|z| \leq 1$ and with no poles in $z = 1$. As before, the number of shocks with non-vanishing long run effects is equal to the number of unit roots and common trends. We now show that in our model $\tau = d$.

Now, using (B8) and (B10) , we have

$$\begin{aligned} \mathbf{K}^{-1}\Delta \mathbf{F}_t &= \mathbf{K}^{-1}\mathbf{C}(L)\mathbf{K}\mathbf{K}^{-1}\mathbf{H}\mathbf{u}_t \\ &= \mathbf{K}^{-1} \begin{pmatrix} (1-L)\mathbf{I}_\tau & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{I}_{(q-\tau)} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & (1-L)\mathbf{I}_{(r-q)} \end{pmatrix} \mathbf{U}(L)\mathbf{K} \begin{pmatrix} \mathbf{I}_q \\ \mathbf{0}_q \end{pmatrix} \mathbf{u}_t. \end{aligned} \tag{B14}$$

Let us partition the first q rows of $\mathbf{K}^{-1}$ as follows

$$[\mathbf{K}^{-1}]_{1:q,:} = [\boldsymbol{\mathcal{K}}_1 \ \boldsymbol{\mathcal{K}}_2 \ \boldsymbol{\mathcal{K}}_3],$$

where $\boldsymbol{\mathcal{K}}_1$ is $q \times \tau$, $\boldsymbol{\mathcal{K}}_2$ is $q \times (q-\tau)$, and $\boldsymbol{\mathcal{K}}_3$ is $q \times (r-q)$. Define $\widetilde{\boldsymbol{\mathcal{K}}}_j(L) = \boldsymbol{\mathcal{K}}_j\mathbf{U}(L)\mathbf{K}$, for $j = 1, 2, 3$ which have also square summable coefficients and therefore no unit root. Then, from (B13) and since $\mathbf{U}(L)$ has no unit root we have

$$\begin{aligned}\Delta \boldsymbol{f}_t = [\mathbf{K}^{-1}]_{1:q,:}\Delta\mathbf{F}_t &= \big[(1-L)\boldsymbol{\mathcal{K}}_1\ \boldsymbol{\mathcal{K}}_2\ (1-L)\boldsymbol{\mathcal{K}}_3\big]\mathbf{U}(L)\mathbf{K}\begin{pmatrix}\mathbf{I}_q\\ \mathbf{0}_q\end{pmatrix}\mathbf{u}_t\\ &= \big[(1-L)\widetilde{\boldsymbol{\mathcal{K}}}_1(L)\ \widetilde{\boldsymbol{\mathcal{K}}}_2(L)\ (1-L)\widetilde{\boldsymbol{\mathcal{K}}}_3(L)\big]\begin{pmatrix}\mathbf{I}_q\\ \mathbf{0}_q\end{pmatrix}\mathbf{u}_t\\ &= \big[(1-L)\widetilde{\boldsymbol{\mathcal{K}}}_1(L)\ \widetilde{\boldsymbol{\mathcal{K}}}_2(L)\big]\mathbf{u}_t = \boldsymbol{\mathcal{D}}(L)\mathbf{u}_t. \end{aligned} \tag{B15}$$

Clearly $\text{rk}(\boldsymbol{\mathcal{D}}(1)) = (q-\tau)$. However, since the innovations $\mathbf{u}_t$ in (B15) are the same as in (B12), then we must have $\boldsymbol{\mathcal{D}}(L) = \boldsymbol{\mathcal{C}}(L)$, and since $\text{rk}(\boldsymbol{\mathcal{C}}(1)) = (q-d)$, then we must have $\tau = d$. Therefore, $\mathbf{F}_t$ has $(q-d)$ unit roots as $\boldsymbol{f}_t$.

B.3 Cointegration rank

Although one might be tempted to say that as a consequence of (B13) the cointegration rank of $\mathbf{F}_t$ is d, this is general not the case. Indeed, following Barigozzi et al. (2016a) we have that the cointegration rank c is such that $d \le c \le (r-q+d)$. We now show that under the restriction in (B8) we indeed have $c = d$.

The vector $\boldsymbol{f}_t$ is cointegrated thus it admits a VECM representation which for simplicity we consider with 2 lags, since this will imply a VECM(1) and therefore a VAR(2) for the factors $\mathbf{F}_t$ as implemented in our empirical analysis. Thus,

$$\Delta \boldsymbol{f}_t = -\mathbf{a}\mathbf{b}'\boldsymbol{f}_{t-3} + \boldsymbol{\Gamma}_1\Delta\boldsymbol{f}_{t-1} + \boldsymbol{\Gamma}_2\Delta\boldsymbol{f}_{t-2} + \mathbf{u}_t, \tag{B16}$$

where $\mathbf{a}$ and $\mathbf{b}$ are $q \times d$.

First assume that the restriction $\mathbf{F}_t = (\boldsymbol{f}_t'\boldsymbol{f}_{t-1}')'$ holds. Since we model $\mathbf{F}_t$ as a VAR(2) we know that we must have the VECM(1) representation

$$\Delta\mathbf{F}_t = -\boldsymbol{\alpha}\boldsymbol{\beta}'\mathbf{F}_t + \mathbf{M}\Delta\mathbf{F}_{t-1} + \mathbf{H}\mathbf{u}_t, \tag{B17}$$

where $\boldsymbol{\alpha}$ and $\boldsymbol{\beta}$ are $r \times c$ with $c < r$ and $\mathbf{H}$ is $r \times q$. The aim of this section is to find c, which is the cointegration rank for $\mathbf{F}_t$ when (B16) holds. In order to do this we look for the expressions of $\mathbf{M}$, $\boldsymbol{\alpha}$, $\boldsymbol{\beta}$, and $\mathbf{H}$ as functions of the parameters $\mathbf{a}$, $\mathbf{b}$, $\boldsymbol{\Gamma}_1$, and $\boldsymbol{\Gamma}_2$ in (B16). Let us write $\boldsymbol{\alpha} = (\boldsymbol{\alpha}_1'\ \boldsymbol{\alpha}_2')'$ and $\boldsymbol{\beta} = (\boldsymbol{\beta}_1'\ \boldsymbol{\beta}_2')'$ where $\boldsymbol{\alpha}_1$, $\boldsymbol{\alpha}_2$, $\boldsymbol{\beta}_1$, $\boldsymbol{\beta}_2$ are all $q \times c$. We also denote as $\mathbf{M}_{ij}$ for $i,j = 1,2$ the four $q \times q$ blocks of $\mathbf{M}$ and as $\mathbf{H}_1$ and $\mathbf{H}_2$ the two $q \times q$ blocks of $\mathbf{H}$. Following Proietti (1997), we define the $(2r+c)$-dimensional vector

$$\mathbf{G}_t = \begin{pmatrix}\Delta\mathbf{F}_t\\ \Delta\mathbf{F}_{t-1}\\ \boldsymbol{\beta}'\mathbf{F}_{t-2}\end{pmatrix} = \begin{pmatrix}\Delta\boldsymbol{f}_t\\ \Delta\boldsymbol{f}_{t-1}\\ \Delta\boldsymbol{f}_{t-1}\\ \Delta\boldsymbol{f}_{t-2}\\ \boldsymbol{\beta}_1'\boldsymbol{f}_{t-2} + \boldsymbol{\beta}_2'\boldsymbol{f}_{t-3}\end{pmatrix}.$$

Then, the state-space form of (B17) is given by

$$\Delta\mathbf{F}_t = \mathbf{Z}\mathbf{G}_t, \quad \mathbf{G}_t = \mathbf{T}\mathbf{G}_{t-1} + \mathbf{Z}'\mathbf{H}\mathbf{u}_t, \tag{B18}$$

with the $r \times (2r+c)$ matrix $\mathbf{Z} = (\mathbf{I}_r\ \mathbf{0}_r\ \mathbf{0}_{r\times c})$. Then,

$$\mathbf{Z}'\mathbf{H} = \begin{pmatrix} \mathbf{I}_r \\ \mathbf{0}_r \\ \mathbf{0}_{c\times r} \end{pmatrix} \begin{pmatrix} \mathbf{H}_1 \\ \mathbf{H}_2 \end{pmatrix} = \begin{pmatrix} \mathbf{H}_1 \\ \mathbf{H}_2 \\ \mathbf{0}_q \\ \mathbf{0}_q \\ \mathbf{0}_{c\times q} \end{pmatrix}.$$

and the $(2r+c)\times(2r+c)$ matrix $\mathbf{T}$ is given by

$$\mathbf{T} = \begin{pmatrix} \mathbf{M} & -\boldsymbol{\alpha}\boldsymbol{\beta}' & -\boldsymbol{\alpha} \\ \mathbf{I}_r & \mathbf{0}_r & \mathbf{0}_{r\times c} \\ \mathbf{0}_{c\times r} & \boldsymbol{\beta}' & \mathbf{I}_c \end{pmatrix} = \left(\begin{array}{cc|cc|c} \mathbf{M}_{11} & \mathbf{M}_{12} & -\boldsymbol{\alpha}_1\boldsymbol{\beta}_1' & -\boldsymbol{\alpha}_1\boldsymbol{\beta}_2' & -\boldsymbol{\alpha}_1 \\ \mathbf{M}_{21} & \mathbf{M}_{22} & -\boldsymbol{\alpha}_2\boldsymbol{\beta}_1' & -\boldsymbol{\alpha}_2\boldsymbol{\beta}_2' & -\boldsymbol{\alpha}_2 \\ \hline \mathbf{I}_q & \mathbf{0}_q & \mathbf{0}_q & \mathbf{0}_q & \mathbf{0}_{q\times c} \\ \mathbf{0}_q & \mathbf{I}_q & \mathbf{0}_q & \mathbf{0}_q & \mathbf{0}_{q\times c} \\ \hline \mathbf{0}_{c\times q} & \mathbf{0}_{c\times q} & \boldsymbol{\beta}_1' & \boldsymbol{\beta}_2' & \mathbf{I}_c \end{array}\right).$$

Now using these definitions into (B18) we have five q-dimensional equations. The first one is

$$\Delta\boldsymbol{f}_t = \mathbf{M}_{11}\Delta\boldsymbol{f}_{t-1} + \mathbf{M}_{12}\Delta\boldsymbol{f}_{t-2} - \boldsymbol{\alpha}_1\boldsymbol{\beta}_1'\boldsymbol{f}_{t-2} - \boldsymbol{\alpha}_1\boldsymbol{\beta}_2'\boldsymbol{f}_{t-3} + \mathbf{H}_1\mathbf{u}_t,$$

which is equivalent to (B16) when

$$\mathbf{M}_{11} = \boldsymbol{\Gamma}_1, \quad \mathbf{M}_{12} = \boldsymbol{\Gamma}_2, \quad \boldsymbol{\alpha}_1 = \mathbf{a}, \quad \boldsymbol{\beta}_1 = \mathbf{0}_{q\times c}, \quad \boldsymbol{\beta}_2 = \mathbf{b}, \quad \mathbf{H}_1 = \mathbf{I}_q, \quad c = d. \tag{B19}$$

The second equation is

$$\Delta\boldsymbol{f}_{t-1} = \mathbf{M}_{21}\Delta\boldsymbol{f}_{t-1} + \mathbf{M}_{22}\Delta\boldsymbol{f}_{t-2} - \boldsymbol{\alpha}_2\boldsymbol{\beta}_1'\Delta\boldsymbol{f}_{t-2} - \boldsymbol{\alpha}_2\boldsymbol{\beta}_2'\Delta\boldsymbol{f}_{t-3} - \boldsymbol{\alpha}_2\boldsymbol{\beta}_1'\boldsymbol{f}_{t-3} - \boldsymbol{\alpha}_2\boldsymbol{\beta}_2'\boldsymbol{f}_{t-4} + \mathbf{H}_2\mathbf{u}_t,$$

from which we see that we must also have

$$\mathbf{M}_{21} = \mathbf{I}_q, \qquad \mathbf{M}_{22} = \mathbf{0}_q, \qquad \boldsymbol{\alpha}_2 = \mathbf{0}_{q\times c}, \qquad \mathbf{H}_2 = \mathbf{0}_q. \tag{B20}$$

Under (B19) and (B20) the third, fourth and fifth equation in (B18) are just identities.

By imposing the restrictions in (B19)-(B20), we have the mapping between the VECM(1) for $\mathbf{F}_t$ in (B17) and the VECM(2) for $\boldsymbol{f}_t$ in (B16)

$$\mathbf{M} = \begin{pmatrix} \boldsymbol{\Gamma}_1 & \boldsymbol{\Gamma}_2 \\ \mathbf{I}_q & \mathbf{0}_q \end{pmatrix}, \qquad \boldsymbol{\alpha} = \begin{pmatrix} \mathbf{a} \\ \mathbf{0}_{q\times d} \end{pmatrix}, \qquad \boldsymbol{\beta} = \begin{pmatrix} \mathbf{0}_{q\times d} \\ \mathbf{b} \end{pmatrix}, \qquad \mathbf{H} = \begin{pmatrix} \mathbf{I}_q \\ \mathbf{0}_q \end{pmatrix}. \tag{B21}$$

If we now consider the general reduced form case in which $\mathbf{F}_t = \mathbf{K}(\boldsymbol{f}_t'\ \boldsymbol{f}_{t-1}')'$ for a $r \times r$

invertible matrix $\mathbf{K}$, then (B21) becomes

$$\mathbf{M} = \mathbf{K} \begin{pmatrix} \boldsymbol{\Gamma}_1 & \boldsymbol{\Gamma}_2 \\ \mathbf{I}_q & \mathbf{0}_q \end{pmatrix} \mathbf{K}^{-1}, \qquad \boldsymbol{\alpha} = \mathbf{K} \begin{pmatrix} \mathbf{a} \\ \mathbf{0}_{q\times d} \end{pmatrix}, \qquad \boldsymbol{\beta} = \mathbf{K}^{-1\prime} \begin{pmatrix} \mathbf{0}_{q\times d} \\ \mathbf{b} \end{pmatrix}, \qquad \mathbf{H} = \mathbf{K} \begin{pmatrix} \mathbf{I}_q \\ \mathbf{0}_q \end{pmatrix}.$$

The cointegration rank c of $\mathbf{F}_t$ is given by $\mathsf{rk}(\boldsymbol{\alpha}\boldsymbol{\beta}') = d$.

Appendix C Data Description and Data Treatment

This Appendix present the dataset used in the analysis. All variables where downloaded from Haver on March 30^{th} 2018. This vintage of data incorporates the third estimate of Gross Domestic Product in 2017:Q4, as well as the first estimate of Gross Domestic Income in 2017:Q4. None of the variables where adjusted for outliers but variables 57, 83, 87, and 94. All variables are from the USECON database but variable 103 that is from the DAILY database. All monthly and daily series are transformed into quarterly observation by simple averages.

List of Abbreviations

Source:
BLS=U.S. Department of Labor: Bureau of Labor Statistics
BEA=U.S. Department of Commerce: Bureau of Economic Analysis
ISM = Institute for Supply Management
CB=U.S. Department of Commerce: Census Bureau
FRB=Board of Governors of the Federal Reserve System
EIA=Energy Information Administration
WSJ=Wall Street Journal
CBO=Congressional Budget Office
FRBPHIL=Federal Reserve Bank of Philadelphia

F = Frequency	T = Transformation	SA	ξ = Idiosyncratic
Q = Quarterly	0 = None	0 = no	0 = $I(0)$
M = Monthly	1 = log	1 = yes	1 = $I(1)$
D = Daily	2 = $\Delta \log$		

D = Deterministic Component	U = Units
$0 = \widehat{a}_i = \frac{1}{T}\sum_{t=1}^{T} \Delta y_{it}$, $\widehat{b}_i = 0$	1000–P = Thousands of Persons
1 = OLS Detrending	1000–U = Thousands of Units
	BoC = Billions of Chained
	\$–B = Dollars per Barrel

N	Series ID	Definition	Unit	F	S	SA	T	D	ξ
1	GDPH	Real Gross Domestic Product	BoC 2009$	Q	BEA	1	1	1	0
2	GDYH	Real Gross Domestic Income	BoC 2009$	Q	BEA	1	1	1	0
3	FSH	Real Final Sales of Domestic Product	BoC 2009$	Q	BEA	1	1	1	1
4	IH	Real Gross Private Domestic Investment	BoC 2009$	Q	BEA	1	1	1	1
5	GSH	Real State & Local*	BoC 2009$	Q	BEA	1	1	1	1
6	FRH	Real Private Residential Fixed Investment	BoC 2009$	Q	BEA	1	1	1	0
7	FNH	Real Private Nonresidential Fixed Investment	BoC 2009$	Q	BEA	1	1	1	1
8	MH	Real Imports of Goods & Services	BoC 2009$	Q	BEA	1	1	1	0
9	GH	Real Government*	BoC 2009$	Q	BEA	1	1	1	1
10	XH	Real Exports of Goods & Services	BoC 2009$	Q	BEA	1	1	1	0
14	CH	Real Personal Consumption Expenditures (PCE)	BoC 2009$	Q	BEA	1	1	1	0
11	CNH	Real PCE: Nondurable Goods	BoC 2009$	Q	BEA	1	1	1	0
12	CSH	Real PCE: Services	BoC 2009$	Q	BEA	1	1	1	0
13	CDH	Real PCE: Durable Goods	BoC 2009$	Q	BEA	1	1	1	0
15	GFDIH	Real National Defense Gross Investment	BoC 2009$	Q	BEA	1	1	1	0
16	GFNIH	Real Federal Nondefense Gross Investment	BoC 2009$	Q	BEA	1	1	1	0
17	YPDH	Real Disposable Personal Income	BoC 2009$	Q	BEA	1	1	1	0
18	JI	Gross Private Domestic Investment⋆	2009=100	Q	BEA	1	2	0	0
19	JGDP	Gross Domestic Product⋆	2009=100	Q	BEA	1	2	0	1
20	LXNFU	Unit Labor Cost†	2009=100	Q	BLS	1	1	1	1
21	LXNFR	Real Compensation Per Hour†	2009=100	Q	BLS	1	1	1	1
22	LXNFC	Compensation Per Hour†	2009=100	Q	BLS	1	1	1	1
23	LXNFH	Hours of All Persons†	2009=100	Q	BLS	1	1	1	0
24	LXNFA	Output Per Hour of All Persons†	2009=100	Q	BLS	1	1	1	0
25	LXMU	Unit Labor Cost‡	2009=100	Q	BLS	1	1	1	1
26	LXMR	Real Compensation Per Hour‡	2009=100	Q	BLS	1	1	1	1
27	LXMC	Compensation Per Hour‡	2009=100	Q	BLS	1	1	1	1
28	LXMH	Hours of All Persons‡	2009=100	Q	BLS	1	1	0	1
29	LXMA	Output Per Hour of All Persons‡	2009=100	Q	BLS	1	1	1	1
30	IP	Industrial Production (IP) Index	2012=100	M	FRB	1	1	1	1
31	IP521	IP: Business Equipment	2012=100	M	FRB	1	1	1	1
32	IP511	IP: Durable Consumer Goods	2012=100	M	FRB	1	1	1	0
33	IP531	IP: Durable Materials	2012=100	M	FRB	1	1	1	1
34	IP512	IP: Nondurable Consumer Goods	2012=100	M	FRB	1	1	1	0
35	IP532	IP: nondurable Materials	2012=100	M	FRB	1	1	1	0

* Consumption Expenditures & Gross Investment

⋆ Chain-type Price Index

† Nonfarm Business Sector

‡ Manufacturing Sector

N	Series ID	Definition	Unit	F	S	SA	T	D	ξ
36	PCU	CPI-U: All Items	82-84=100	M	BLS	1	2	0	0
37	PCUSE	CPI-U: Energy	82-84=100	M	BLS	1	2	0	0
38	PCUSLFE	CPI-U: All Items Less Food and Energy	82-84=100	M	BLS	1	2	0	0
39	PCUFO	CPI-U: Food	82-84=100	M	BLS	1	2	0	0
40	JCBM	PCE: Chain Price Index	2009=100	M	BEA	1	2	0	0
41	JCEBM	PCE: Energy Goods & Services–price index	2009=100	M	BEA	1	2	0	0
42	JCNFOM	PCE: Food & Beverages–price index*	2009=100	M	BEA	1	2	0	0
43	JCXFEBM	PCE less Food & Energy–price index	2009=100	M	BEA	1	2	0	0
44	JCSBM	PCE: Services–price index	2009=100	M	BEA	1	2	0	0
45	JCDBM	PCE: Durable Goods–price index	2009=100	M	BEA	1	2	0	0
46	JCNBM	PCE: Nondurable Goods–price index	2009=100	M	BEA	1	2	0	0
47	PC1	PPI: Intermediate Demand Processed Goods	1982=100	M	BLS	1	2	0	0
48	P05	PPI: Fuels and Related Products and Power	1982=100	M	BLS	0	2	0	0
49	SP3000	PPI: Final Demand Personal Consumption Gds**	1982=100	M	BLS	1	2	0	0
50	SP3000	PPI: Finished Goods	1982=100	M	BLS	1	2	0	0
51	PIN	PPI: Industrial Commodities	1982=100	M	BLS	0	2	0	0
52	PA	PPI: All Commodities	1982=100	M	BLS	0	2	0	0
53	FMC	Money Stock: Currency	Bil. of $	M	FRB	1	2	0	0
54	FM1	Money Stock: M1	Bil. of $	M	FRB	1	2	0	1
55	FM2	Money Stock: M2	Bil. of $	M	FRB	1	2	0	0
56	FABWC	C & I Loans in Bank Credit†	Bil. of $	M	FRB	1	1	1	1
57	FABWQ	Consumer Loans in Bank Credit†	Bil. of $	M	FRB	1	1	1	1
58	FAB	Bank Credit†	Bil. of $	M	FRB	1	1	1	1
59	FABW	Loans & Leases in Bank Credit†	Bil. of $	M	FRB	1	1	1	1
60	FABYO	Other Securities in Bank Credit†	Bil. of $	M	FRB	1	1	1	1
61	FABWR	Real Estate Loans in Bank Credit†	Bil. of $	M	FRB	1	1	1	0
62	FOT	Consumer Credit Outstanding	Bil. of $	M	FRB	1	1	1	1
63	HSTMW	Housing Starts: Midwest	1000–U	M	CB	1	1	0	0
64	HSTNE	Housing Starts: Northeast	1000–U	M	CB	1	1	0	0
65	HSTS	Housing Starts: South	1000–U	M	CB	1	1	0	0
66	HSTGW	Housing Starts: West	1000–U	M	CB	1	1	0	0
67	HPT	Building Permit⋆	1000–U	M	CB	1	1	0	0
68	FBPR	Bank Prime Loan Rate	Percent	M	FRB	0	0	0	0
69	FFED	Federal Funds [effective] Rate	Percent	M	FRB	0	0	0	0
70	FCM1	1-Year Treasury Bill Yield‡	Percent	M	FRB	0	0	0	0
71	FCM10	10-Year Treasury Note Yield‡	Percent	M	FRB	0	0	0	0

* Purchased for Off-Premises Consumption

† All Commercial Banks

⋆ New Private Housing Units Authorized by

‡ at Constant Maturity

** [Finished Consumer Gds]

N	Series ID	Definition	Unit	F	S	SA	T	D	ξ
72	LP	Civilian Participation Rate: 16 yr +	Percent	M	BLS	0	0	1	0
73	LQ	Civilian Employment/Population Ratio: 16 yr +	Percent	M	BLS	0	0	1	0
74	LE	Civilian Employment: Sixteen Years & Over	1000–P	M	BLS	0	1	1	0
75	LR	Civilian Unemployment Rate: 16 yr +	Percent	M	BLS	0	0	0	0
76	LU0	Civilians Unemployed for Less Than 5 Weeks	1000–P	M	BLS	0	1	1	0
77	LU5	Civilians Unemployed for 5-14 Weeks	1000–P	M	BLS	0	1	1	1
78	LU15	Civilians Unemployed for 15-26 Weeks	1000–P	M	BLS	0	1	1	1
79	LUT27	Civilians Unemployed for 27 Weeks and Over	1000–P	M	BLS	0	1	1	1
80	LUAD	Average [Mean] Duration of Unemployment	Weeks	M	BLS	0	1	1	1
81	LANAGRA	All Employees: Total Nonfarm	1000–P	M	BLS	0	1	1	1
82	LAPRIVA	All Employees: Total Private Industries	1000–P	M	BLS	0	1	1	0
83	LANTRMA	All Employees: Mining and Logging	1000–P	M	BLS	0	1	0	1
84	LACONSA	All Employees: Construction	1000–P	M	BLS	0	1	1	1
85	LAMANUA	All Employees: Manufacturing	1000–P	M	BLS	0	1	0	1
86	LATTULA	All Employees: Trade, Transportation & Utilities	1000–P	M	BLS	0	1	1	1
87	LAINFOA	All Employees: Information Services	1000–P	M	BLS	0	1	1	1
88	LAFIREA	All Employees: Financial Activities	1000–P	M	BLS	0	1	1	1
89	LAPBSVA	All Employees: Professional & Business Services	1000–P	M	BLS	0	1	1	1
90	LAEDUHA	All Employees: Education & Health Services	1000–P	M	BLS	0	1	1	1
91	LALEIHA	All Employees: Leisure & Hospitality	1000–P	M	BLS	0	1	1	1
92	LASRVOA	All Employees: Other Services	1000–P	M	BLS	0	1	1	1
93	LAGOVTA	All Employees: Government	1000–P	M	BLS	0	1	1	1
94	LAFGOVA	All Employees: Federal Government	1000–P	M	BLS	0	1	0	1
95	LASGOVA	All Employees: State Government	1000–P	M	BLS	0	1	1	0
96	LALGOVA	All Employees: Local Government	1000–P	M	BLS	0	1	1	0
97	PETEXA	West Texas Intermediate Spot Price FOB*	$–B	M	EIA	0	2	0	0
98	NAPMNI	ISM Mfg: New Orders Index	Index	M	ISM	1	0	0	1
99	NAPMOI	ISM Mfg: Production Index	Index	M	ISM	1	0	0	1
100	NAPMEI	ISM Mfg: Employment Index	Index	M	ISM	1	0	0	1
101	NAPMVDI	ISM Mfg: Supplier Deliveries Index	Index	M	ISM	1	0	0	0
102	NAPMII	ISM Mfg: Inventories Index	Index	M	ISM	1	0	0	0
103	SP500	Standard & Poor's 500 Stock Price Index	41-43=10	D	WSJ	0	1	1	0

* Cushing, Oklahoma

Series ID	Definition	Unit	F	Source
GDPPOTHQ	Real Potential Gross Domestic Product	BoC 2009$	Q	CBO
NAIRUQ	Natural Rate of Unemployment	percent	Q	CBO
GDPPLUS	US GDPplus	percent	Q	FRBPHIL

C.1 Real-time data vintages

The real-time data vintages used in Section 6 are constructed as follows: vintages from February 24 2017 onwards where saved by us every Friday, while vintages prior to February 24 2017 are constructed using a real-time database maintained by the Division of Research and Statistics of the Board of Governors of the Federal Reserve System. All the variables were retrieved from that database but variable 66 "Housing Starts: West", which was retrieved from the ALFRED database of the Federal Reserve Bank of St. Louis. Furthermore, some missing observations for GDI and the oil price were filled by using real-time vintages downloaded from ALFRED.

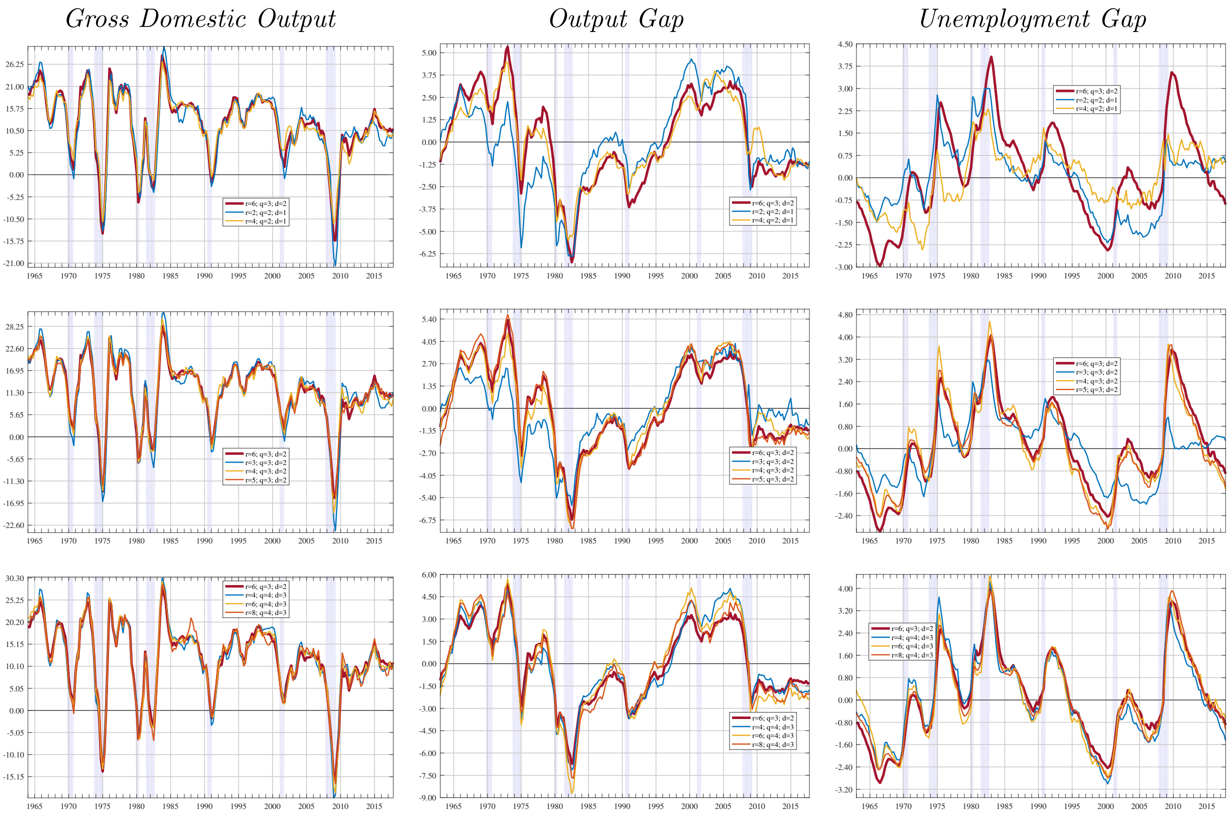
For the first four vintages we were able to retrieve only 101 variables out of 103, as we were not able to find vintages for two PPIs (variables 47 and 49).

GDO and the output gap were estimated on 7 data vintages per year corresponding to the dates after the release of GDI. The first estimate of GDI Q1 is released at the end of May, together with the second estimate of GDP Q1, while the second estimate is published in June, together with the third estimate of GDP Q1. Similarly, GDI Q2 is released in August and in September, while GDI Q3 is released in November and December. By contrast, GDI for Q4 is released only once in March together with the third estimate of GDP Q4.

The dates in which GDI is released varies from year to year, therefore we have adopted the following strategy to select the dates at which to construct the vintages: for each GDI release we select the vintage as of the first Friday after the 6^{th} of the following month. For example, for GDI Q1 which is published at the end of March, we select the vintage corresponding to the first Friday after the 6^{th} of April.

Appendix D Additional results

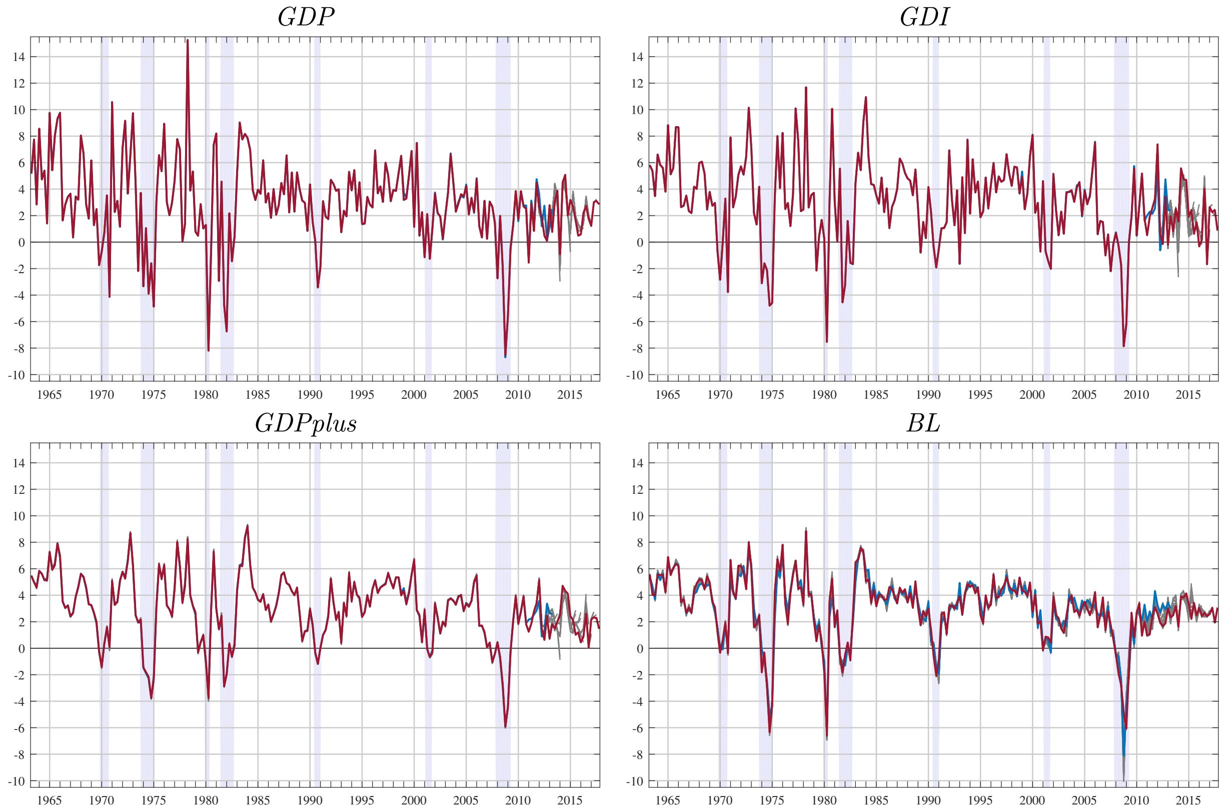
Figure D3: Robustness with respect to number of shocks/factors



In this Figure, plots on the left column show estimates of GDO, plots on the middle column show estimates of the output gap, while estimates on the right column show estimates of the unemployment gap. In each plot the red line is the estimate obtained with our benchmark model specification, i.e., $r = 6$, $q = 3$, and $d = 2$. By contrast, the blue, the yellow, and the orange lines are estimates obtained with different specifications of r, q, and d.

Figure D4: Gross Domestic Output

Real-time estimation



In each plot the blue line is the estimate of GDO obtained with the vintage of data from September 13 2013, the red line is the estimate obtained with the vintage of data from March 30 2018, and the grey lines are the estimate obtained with all the remaining 31 vintages. Note that for GDP and GDI the different lines represent not our estimate, but the data available as of. Similarly, for GDPplus the different lines represent the estimates from the Philly Fed using data available as of.