GDP Solera
The Ideal Vintage Mix

Martín Almuzara[*] Dante Amengual[†] Gabriele Fiorentini[‡]
Enrique Sentana[§]

This version: August 2023
[Link to the latest version]

Abstract

We use the information in the successive vintages of GDE and GDI to obtain an improved timely measure of US aggregate output by exploiting cointegration between the different measures taking seriously their monthly release calendar. We also combine all existing overlapping comprehensive revisions to achieve further improvements. We pay particular attention to the Great Recession and the COVID-19 pandemic, which, despite producing dramatic fluctuations, did not generate noticeable revisions in previous growth rates. Our results suggest that revised GDE estimates, unlike GDI ones, are increasingly precise and receive higher weights, but early estimates retain some influence.

Keywords: Cointegration, Comprehensive revisions, Signal extraction, US aggregate output, Vintages.
JEL Classification: E01, C32.

[*]Federal Reserve Bank of New York: martin.almuzara@ny.frb.org
[†]CEMFI: amengual@cemfi.es
[‡]Universitá di Firenze and RCEA: gabriele.fiorentini@unifi.it
[§]CEMFI and CEPR: sentana@cemfi.es

1 Introduction

Despite the recent interest in alternative measures, such as the Human Development Index or the different Gross National Happiness measures, Gross Domestic Product (GDP) remains the dominant concept to gauge the aggregate performance of an economy over a given period of time. In the United States of America, the estimates of aggregate economic activity that the Bureau of Economic Analysis (BEA) publishes as part of its National Product and Income Accounts (NIPA) are used not only by policy makers and research economists, but also by private sector agents, including households and firms, in making their production and consumption decisions, as well as their financial plans.

The BEA uses a mixture of survey, tax and other business and administrative data, as well as various indicators, which are subject to sampling errors and biases that cannot be directly assessed. As time goes by, though, the BEA acquires more and better information, and for that reason it systematically updates its measures, which results in a sequence of estimates for a given quarter known as revisions. In fact, the whole revision process is rather elaborate, and it is important to distinguish between three types: (i) successive early releases for a given quarter, usually called the "advance", "second" and "third" estimates; (ii) annual (or "final") revisions, which simultaneously update all the quarters of several previous calendar years; and (iii) occasional comprehensive revisions, which recompute the entire history of the series after a major methodological change that effectively modifies its definition. The importance of revisions should not be underestimated. For example, Orphanides (2001) convincingly argues that the use of final instead of preliminary GDP measures can lead to different monetary policy recommendations.

While in the last two decades there has been considerable progress in jointly modeling the different vintages of US GDP (see, for example, Aruoba (2008), Jacobs and van Norden (2011) and the references therein), some of these studies have ignored a second important consideration: the BEA produces not just one but two different official measures of real aggregate output and income: Gross Domestic Expenditure (GDE) and Gross Domestic Income (GDI). GDE measures activity as the sum of all final expenditures in the economy, which is reflected in the output side of the NIPAs. In turn, GDI measures activity as the sum of all income generated in production, and is therefore captured on the income side of the NIPAs (the value added approach would complete the usual trinity of GDP measurements, but the BEA does not produce real estimates at the quarterly frequency). In theory, the flows of income and expenditure should be equal, and thus, GDE and GDI should yield the same measure of economic activity. In practice, though, they differ not only due to the revisions but also because each is calculated from data from completely different sources (see Landefeld, Seskin, and Fraumeni (2008) for a review). The systematic, and at times noticeable, deviation between them – officially known as *statistical discrepancy*– was traditionally regarded by many academic economists as a curiosity in the NIPAs (see Grimm (2007) for a detailed methodological insight). However, the Great Recession led to substantially renewed interest in academic and policy circles about the possibility of obtaining more reliable economic activity figures by combining the two measures. As a consequence, various proposals for improved combinations have been discussed (see, e.g. Nalewaik (2010), Nalewaik (2011), Greenaway-McGrevy (2011), Aruoba, Diebold, Nalewaik, Schorfheide, and Song (2016) and Jacobs, Sarferaz, Sturm, and van Norden (2022)). For example, the *GDPplus* measure of Aruoba et al. (2016) is currently released on a monthly schedule by the Federal Reserve Bank of Philadelphia.

The purpose of our paper is to simultaneously tackle all these measurement issues within a single, internally coherent, signal extraction framework. Intuitively, given that GDE and GDI are based on different sources, one would expect to obtain a more accurate estimate of the underlying economic concept by making use of the static and dynamic correlation patterns in the observed series. Stone, Champernowne, and Meade (1942) is the first known reference to the signal extraction framework of our paper. Weale (1992) surveys the early literature; see also Smith, Weale, and Satchell (1998).

Despite involving a moderately large number of both latent and observed variables, our model is both flexible and parsimonious thanks to the economic and statistical discipline that we impose on the measurement errors. Although the modelling of US GDP as a unit root process rather than as a trend stationary one is now conventional (see Campbell and Mankiw (1987) and the references therein for the earlier debate), our crucial point of departure from the previous literature is that we follow Almuzara, Amengual, and Sentana (2019) and Almuzara, Fiorentini, and Sentana (2023) in imposing that (i) any two aggregate output and income measures (in logs) are cointegrated, with cointegrating vector (1,-1); and (ii) measurement errors are mean-reverting and stationary, although they may be serially correlated. Thus, we are able to focus not only in quarterly growth rates, but also assess the level of US output, which is of considerable interest in itself, particularly in regional or cross-country comparisons. Somewhat surprisingly, although the possibility of cointegration between GDP vintages was highlighted over thirty years ago by Patterson and Heravi (1991b) and Patterson and Heravi (1991a) (see also Patterson and Heravi (2004)), the subsequent literature has largely ignored this important feature of the data.

In addition, the data release calendar is at the core of our model. Specifically, we explicitly take into account that the "advance", "second" and "third" GDE estimates are published one, two and three months after the end of the quarter, respectively. Moreover, we acknowledge the fact that the timing of the quarterly releases for GDI is somewhat different, as it incorporates information from the quarterly census of employment and wages. Importantly, we also consider the annual data revisions of both series that are published in the second half of the following and subsequent years, and which typically affect the values for all the quarters of the most recent previous years. For instance, the July 2017 annual update revised all quarters for 2014, 2015 and 2016.

The final novel ingredient of our model is the combination of data from different comprehensive revisions, which take place approximately every five years based on an economic census of millions of US businesses. These revisions also incorporate changes in definitions, classifications, and statistical methodology. For example, in 2013 the BEA started counting R&D as an investment rather than as a cost, which "boosted" US GDP by over 2%. The most recent comprehensive revision we consider was published in July 2018, with a detailed analysis in a BEA paper (see Kerry, McCulla, and Wasshausen (2018)). In that report, the U.S. statistical office presented revised annual estimates for 1929-2017 and revised quarterly estimates for 1947-2017. Often, comprehensive revisions reflect either improved or totally new coverage of sectors of the economy that have become increasingly important. In addition, real GDP is usually re-based, with the reference year kept fixed during subsequent annual updates. The latest comprehensive revision was released in September 2023 while our analysis ends in the second quarter of 2023. Vintages released in July of both 2011 and 2014 are exceptions because the reference year was also revised. This resulted in a change of the GDP deflator and, in turn, a change of real GDP for the whole series since 1947.

Despite these systematic differences, the joint modeling of multiple comprehensive revisions is particularly relevant when a new one is released, which is precisely when there is very little information about the statistical properties of its successive vintages and annual revisions.

The closest paper to ours is Jacobs et al. (2022), which also use the different releases of GDE and GDI to obtain improved real-time estimates of economic activity. Nevertheless, these authors focus on growth rates and abstract from comprehensive redefinitions. One additional difference is that Jacobs et al. (2022) propose a framework to separate news from noise in the revision process along the lines of Jacobs and van Norden (2011). In Supplemental Appendix SM.F we explain how to write their news-noise model as a special case of ours. We could use the expressions we derive there to provide a decomposition of the measurement errors between "news" and "noise", a promising avenue for future research.

From the point of view of implementation, our model can be cast in linear state-space form and is therefore amenable to the use of Bayesian methods of inference for both parameters and latent variables. In particular, we develop a Gibbs sampling algorithm that tackles estimation and signal-extraction simultaneously, allowing for an efficient and conceptually simple integration of uncertainty coming from different sources. Thus, we obtain a posterior distribution for the different benchmark definitions of underlying GDP, whence we can obtain not only point estimates but also measures of dispersion that reflect the remaining uncertainty about the true value of aggregate activity. Nevertheless, given that analysts and policy makers typically focus on the evolution of the current GDP definition, we will refer to our point estimate of the most recent benchmark version as *GDPsolera* henceforth. The moniker "solera" arises because the recurrent updating of our signal extraction process is analogous to the criaderas and soleras system of sherry wine aging, whereby the final product is obtained by fractional blending inputs from different vintages over a perennial dynamic procedure that gives sherry its distinctive character (as explained by agent 007 to M in the 1971 James Bond film Diamonds are forever).

After estimating our model making the best use of all the available US data, we apply it to answer a number of empirically relevant questions. First, do comprehensive revisions modify the empirical characteristics of economic growth, such as its long-term mean or its persistence? Second, what is the contribution of the different estimates (i.e., advance, second, third, etc.) to the precision of signal extraction about economic activity? Our estimates suggest that (i) comprehensive revisions have not led to appreciable changes in the average growth rate, and that (ii) noticeable precision gains in signal extraction occur not only when the advance, second and third estimates of GDE and GDI are released but also when the annual estimates become available in subsequent years.

Finally, we provide several additional empirical exercises, including an assessment of the sensitivity of our improved estimate of economic activity to our identification assumptions and alternative specifications of the autocovariance structure of the latent variables, as well as its behavior during the COVID-19 pandemic. In this respect, we find that the real time version of *GDPsolera* provides accurate estimates of the quarters mostly affected by the pandemic, which seem to be in line with the subsequent BEA revised estimates. We also find that despite the dramatic nature of the GDP movements in 2020, our estimates of its growth rate for previous quarters are hardly affected.

The rest of the paper is organized as follows. We begin with a detailed description of the data in section 2. Section 3 introduces the model and briefly reviews our approach to estimation and filtering. Section 4 reports the empirical analysis, including the improved *GDPsolera* measure of economic activity produced by our method. Finally, we present our conclusions and directions for further research in section 5, relegating proofs and other technical details to the supplemental material.

2 Data background

Our empirical analysis uses data on the successive GDE and GDI vintages from the BEA. To get a better sense of the data, it is instructive to review the timing of the release process as it happens regularly over a typical year. Table SM.D.1 in Supplemental Appendix SM.D exemplifies the process in a recent period. Estimates for quarterly GDP are released in the following order:

(A) Advance estimate, based on data incomplete or subject to further revision by the source agency, and released at the end of the first month after the end of the quarter.

(B) Second/third estimates, which use broader and more detailed data, and are released near the end of the second and third months, respectively.

(C) Latest estimates, reflecting the results of both annual and comprehensive updates.

For GDI only second, third and latest estimates are prepared because of data availability, except for the fourth quarter of each year, for which only third and latest estimates are released, typically at the same time as the corresponding GDE figures.

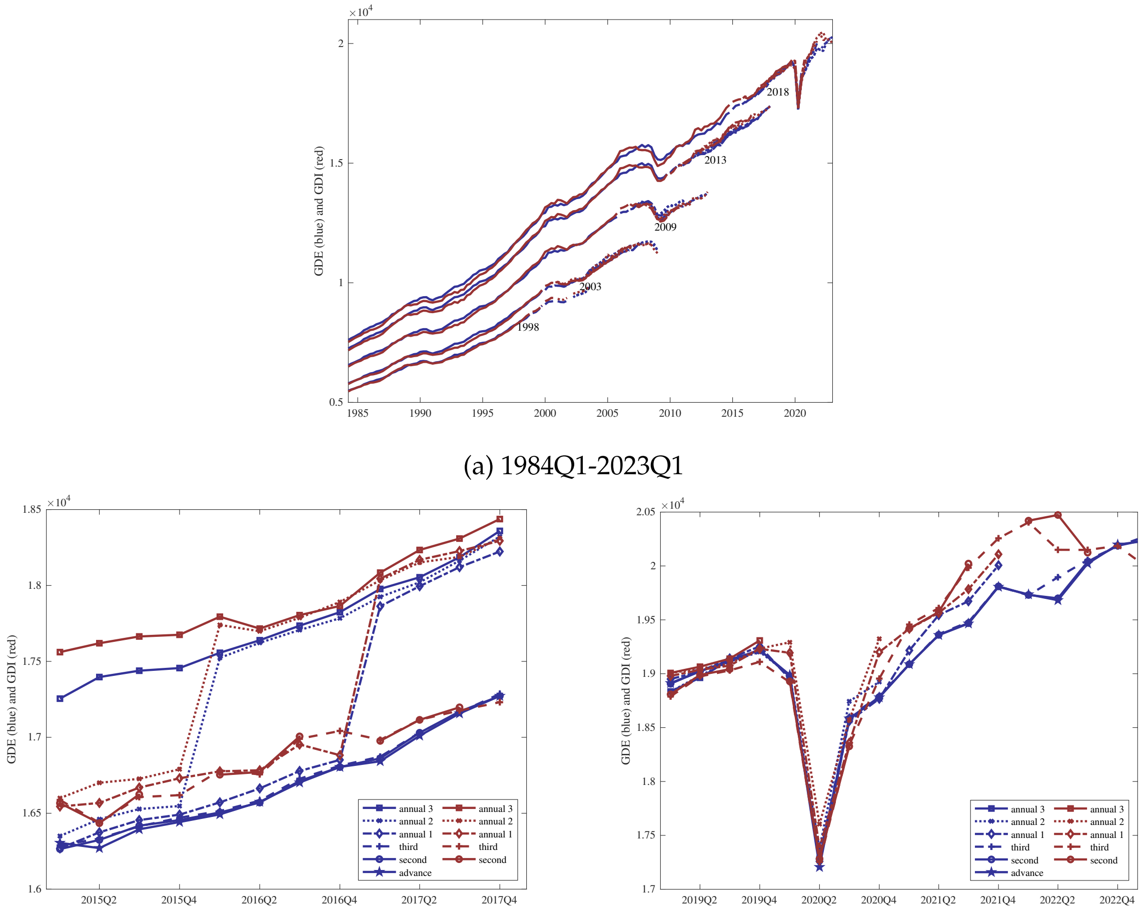
Normally, a single estimate for the latest quarter is added to the GDE/GDI series at a time, but there are two kinds of updates where multiple quarters are simultaneously updated:

(a) Annual updates, usually done in July but more recently in September, which cover at least the three most recent calendar years (e.g. the July 2017 annual update revised all quarters for 2014, 2015 and 2016). They incorporate newly available annual source data, and minor methodological changes.

(b) Comprehensive (or benchmark) updates, which are done approximately every 5 years (the actual updates took place in December 2003, July 2009, July 2013 and July 2018). They incorporate periodic data released at frequencies lower than 1 year, such as the quinquennial US Economic Census, and some major methodological changes.

In our main empirical analysis, we use the available seasonally adjusted GDE and GDI vintages over the period 1984Q1-2023Q1, including the five benchmark versions of US economic activity resulting from the comprehensive revisions in 2003, 2009, 2013 and 2018. Although the annual revision process was extended from three to five years in July 2019, we consider three annual revisions, which are the only ones available for most of our sample.

We depict the series (in levels) of different comprehensive revision releases in Figure 1.a, where we also plot data produced by early and annual revisions for the periods between two consecutive benchmark revisions. As we explained in the introduction, the vertical differences partly reflect different base years for the deflators. In turn, Figures 1.b and 1.c zoom in on two three-year subperiods to illustrate in closer detail the different measures of economic activity. The July 2018 comprehensive revision led to a thorough revision of GDE and GDI for the first subperiod (2015Q1-2017Q4), which explains the



(a) 1984Q1-2023Q1

(b) 2015Q1-2017Q4

(c) 2019Q1-2021Q4

FIGURE 1. **GDE and GDI data from the BEA (levels in B US$).** Panel (a) displays the data over the period 1984Q1 to 2023Q1. In this panel, solid lines represent data released under comprehensive revisions while dashed lines represent data produced by early and annual revisions. Panels (b) and (c) zoom into specific periods

marked differences in levels between the advance, second and third releases, and the annual ones. In contrast, no such differences appear in the second subperiod (2019Q1-2021Q4), which nevertheless shows the dramatic effects of the COVID-19 pandemic. We return to the analysis of the second period in subsection 4.1.4.

3 Model

Let x_t be the quantity of interest at time t; in our empirical analysis, US aggregate economic output (in logs) during quarter t. As most of the literature that followed Stone et al. (1942), we treat x_t as a latent variable of which only noisy measurements y_t are available. Although the components of this vector will be available after lags of different length, we will continue to use the subscript t in reference to the quarter to which they correspond (for background on output measurements, see Landefeld et al. (2008) and Nalewaik (2010, 2011)). Next, we develop the framework that will allow us to combine multiple y_t's for the purposes of obtaining an improved estimate of economic activity x_t. For the sake of clarity, we begin in subsection 3.1 with a version of our model that has no comprehensive revisions, adding them in subsection 3.2.

3.1 Modeling early and annual estimates

Let y_{it}^m be a noisy measurement of x_t, where the index i denotes type (e.g., GDE and GDI estimates) while m denotes release (e.g., early and annual estimates). This distinction is important because we will assume orthogonality of measurement errors along i but we will permit correlation over m for measurements with the same i. Orthogonality between the measurement errors of the expenditure and income estimates is not only plausible because they are based on completely different data sources, but also useful to achieve identification of the serial dependence in x_t. Still, in Supplemental Appendix SM.C we assess the robustness of our empirical results to this assumption. In contrast, correlation between the measurement errors of different releases of the same measure is to be expected, as they share revised versions of the same data sources.

The model is given by the set of measurement equations

$$y_{it}^m = x_t + v_{it}^m, \;\; m = 1, \ldots, M_i, \quad i = 1, \ldots, N,$$

where v_{it}^m is the measurement error in y_{it}^m. For each i, we collect $y_{it}^1, \ldots, y_{it}^{M_i}$ into the vector y_{it} and stack $y_{1t}, \ldots, y_{Nt}$ into y_t. Defining v_{it}, for each i, and v_t likewise, we obtain,

$$y_t = 1_{M\times 1} x_t + v_t, \tag{1}$$

where $M = \sum_{i=1}^{N} M_i$ and $1_{M\times 1}$ is an M-dimensional vector of ones.

In this context, we assume that the following conditions hold:

Assumption 1. *(a)* Δx_t *is I(0);*

(b) $v_{1t}, \dots, v_{Nt}$ *are I(0);*

(c) $\Delta x_t, v_{1t}, \dots, v_{Nt}$ *are orthogonal across blocks at all leads and lags.*

We make assumption 1(a) because y_t measures economic activity in (log) levels. We take the definition of I(0) process from the multivariate generalization of the one in Stock (1994): Consider a time series $\omega_t = \sum_{\ell=0}^{\infty} \Theta_\ell \varepsilon_{t-\ell}$, with Θ_ℓ an $n \times n$ matrix and ε_t and n-dimensional vector. Then, ω_t is I(0) if (i) ε_t is a weakly stationary vector martingale difference sequence, (ii) $\sum_{\ell=0}^{\infty} \Theta_\ell$ is nonsingular, and (iii) $\sum_{\ell=0}^{\infty} \ell \|\Theta_\ell\| < \infty$. Together with assumption 1(b), it implies that y_t is cointegrated with cointegration rank $M-1$, so that any set consisting of $M-1$ pairwise differences among the y_{it}^m is a basis for the cointegration space. Cointegration is a very plausible assumption for aggregate measurement problems originally highlighted by Patterson and Heravi (1991b) and Patterson and Heravi (1991a). In fact, assuming that the growth rates in y_t follow a strictly invertible covariance stationary process necessarily implies that the different measures of x_t would diverge in the long run, which is implausible (see Almuzara et al. (2023) for additional discussion).

On the other hand, Assumption 1(c), which allows for dynamic dependence within blocks but rules out dependence between shocks to the signal and the different measurement errors, is key for nonparametric identification, as asserted in the following proposition, whose proof can be found in Supplemental Appendix SM.A:

Proposition 1. *Under assumption 1, if* $N > 1$*, the autocovariance matrices of* $\Delta x_t, v_{1t}, \dots, v_{Nt}$ *are nonparametrically identified from the autocovariance matrices of* Δy_t*.*

Our empirical analysis has $N = 2$, as we use GDE and GDI measurements of output. $N = 1$ may be relevant for other applications. In those cases, identification can be achieved by imposing restrictions on the cross-dependence among $v_{1t}^1, \dots, v_{1t}^{M_1}$ (e.g., assuming $v_{1t}^{m_1}$ and $v_{1t}^{m_2}$ orthogonal at all leads and lags), or by sufficiently tight parametric assumptions.

3.2 Modeling comprehensive revisions

Our approach to modeling comprehensive revisions is to treat each version of the variable of interest introduced by the revision process as a different latent variable, while at the same time allowing for strong dependence among them. Thus, we naturally generalize the multiple measurements - single latent variable models in the literature (e.g., Weale (1992), Smith et al. (1998), Aruoba et al. (2016), Almuzara et al. (2019), and Almuzara et al. (2023)) to a situation in which there are multiple latent variables of interest.

Let C be the number of benchmark versions. Rather than a single variable, our extended model makes x_t a vector, $x_t = (x_{1t}, \ldots, x_{Ct})'$. Here x_{ct} represents the hypothetical value of economic output that could be measured with the definitions and methods adopted for the comprehensive revision c if the data sources and measuring tools were perfect. For example, the first three elements of x_t would treat R&D as a cost while the last two as an investment, as we explained in the introduction.

While analysts and policy makers typically focus on the latest version x_{Ct}, there are important reasons for modeling $x_{1t}, \ldots, x_{Ct}$ jointly: first, older definitions of economic activity are important from a historical perspective because, after all, those were the only ones available at the time; second, understanding the impact of comprehensive revisions on the static and dynamic characteristics of the growth rates in aggregate economic activity is particularly relevant too; finally, there is also substantial interest in quickly learning about the dynamics of the measurement errors in the most recent version, which might lead to improved inferences about x_{Ct} itself.

Measurement equation. Let δ_{it}^{m} be a $1 \times C$ array that has 1 in entry c if y_{it}^{m} measures x_{ct} and 0 otherwise. The array δ_{it}^{m} is deterministically time-varying but known, and can be easily computed by comparing the year of the comprehensive revisions and the exact release date of y_{it}^m. Our model postulates that

$$y_{it}^m = \delta_{it}^m x_t + v_{it}^m, \quad i = 1, \ldots, N, \ m = 1, \ldots, M_i.$$

Concatenating δ_{it}^m vertically to conform with y_{it} and y_t, we obtain the $M_i \times C$ array δ_{it} and the $M \times C$ array δ_t, which lead to the measurement equation

$$y_t = \delta_t x_t + v_t. \tag{2}$$

Equation (2) generalizes (1) into a deterministically time-varying measurement equation. Importantly, some entries of y_t may be missing because either they have not been released yet even though they will be released in the future according to the BEA protocol or because old methods are not applied to the computation of new estimates. Nevertheless, our estimation and filtering algorithms can perfectly accommodate the presence of such missing data.

Identification revisited. We adopt assumption 1 without change, except that Δx_t is a vector process now. Because the measurement equation is time-varying, the spectrum of y_t depends on t. However, given that the time-variation is deterministic, this entails a trivial form of non-stationarity from the point of view of identification. In our empirical analysis, moreover, there is a subvector of y_t that is stationary since there is a time-invariant block in δ_t. This allows us to establish identification through a generalization of proposition 1 applied to the time-invariant block. We state sufficient conditions for non-parametric identification in proposition 2, whose proof is also in appendix SM.A.

Proposition 2. *Suppose there are indices i_1, i_2 ($i_1 \neq i_2$) and matrices E_{i_1}, E_{i_2} such that (a) $E_{i_1} y_t$ and $E_{i_2} y_t$ are nonempty subvectors of $y_{i_1,t}$ and $y_{i_2,t}$, respectively, (b) $E_{i_1}\delta_t$ and $E_{i_2}\delta_t$ are time-invariant, and (c)* $\operatorname{rank}(E_{i_1}\delta_t) = \operatorname{rank}(E_{i_2}\delta_t) = C$. *Then, under assumption 1, the autocovariances of $\Delta x_t, v_{1t}, \ldots, v_{Nt}$ are nonparametrically identified from those of Δy_t.*

As an example, consider a model with $C = 2$ versions of economic activity. Suppose

$N = 2$ with $M_1 = M_2 = 2$ and $\delta_t = (I_2 \;\; I_2)'$ for all t. The measurement equation is

$$\begin{pmatrix} y_{1t}^1 \\ y_{1t}^2 \\ y_{2t}^1 \\ y_{2t}^2 \end{pmatrix} = \begin{pmatrix} 1 & 0 \\ 0 & 1 \\ 1 & 0 \\ 0 & 1 \end{pmatrix} \begin{pmatrix} x_{1t} \\ x_{2t} \end{pmatrix} + \begin{pmatrix} v_{1t}^1 \\ v_{1t}^2 \\ v_{2t}^1 \\ v_{2t}^2 \end{pmatrix}.$$

This setup clearly satisfies the conditions of proposition 2 with $i_1 = 1, i_2 = 2, E_{i_1} = (I_2 \;\; 0_{2\times2})$, and $E_{i_2} = (0_{2\times2} \;\; I_2)$. Consequently, the autocovariance matrices of $\Delta x_t, v_{1t}, v_{2t}$ are identified from those of Δy_t. Some intuition can be gained by first considering the subsystems

$$\begin{pmatrix} y_{1t}^c \\ y_{2t}^c \end{pmatrix} = 1_{2\times1} x_{ct} + \begin{pmatrix} v_{1t}^c \\ v_{2t}^c \end{pmatrix}, \quad c = 1, 2.$$

Proposition 1 can be applied and immediately delivers the marginal serial dependence structure of the processes $\Delta x_{1t}, \Delta x_{2t}, v_{1t}^1, v_{1t}^2, v_{2t}^1, v_{2t}^2$.

Next, we can recover the cross-autocovariances of the two signals by observing that

$$\mathrm{Cov}\Big(\Delta x_{1t}, \Delta x_{2,t-\ell}\Big) = \mathrm{Cov}\Big(\Delta y_{1t}^c, \Delta y_{2,t-\ell}^c\Big)$$

holds for $c = 1, 2$ and all ℓ. Finally, we have that for $i = 1, 2$ and all ℓ,

$$\mathrm{Cov}\Big(\Delta v_{it}^1, \Delta v_{i,t-\ell}^2\Big) = \mathrm{Cov}\Big(\Delta y_{it}^1, \Delta y_{i,t-\ell}^2\Big) - \mathrm{Cov}\Big(\Delta x_{1t}, \Delta x_{2,t-\ell}\Big).$$

In our empirical analysis we rely on $C = 5$ versions of both GDE and GDI, in addition to their early and latest estimates. This implies that, for all t, δ_t contains two distinct blocks equal to I_C each corresponding to GDE and GDI measurements, respectively, so the conditions in proposition 2 are automatically satisfied. Consequently, the joint dynamics of Δx_t are non-parametrically identified. One qualification worth making is that because past benchmark versions are discontinued, we are learning about the joint autocorrelation structure of x_t within the period in which they overlap. This amounts to a long period in our sample, spanning 1984Q1 to 2003Q2 (the time of the first comprehensive revision), yet a period that excludes the instabilities from the Great Recession or the COVID-19 pandemic.

3.3 Specification, estimation and filtering

Although the dynamics of x_t is non-parametrically identified, to implement our empirical analysis we specify a parametric model for the serial dependence of $\Delta x, v_1, \ldots, v_N$ that satisfies assumption 1 and, at the same time, is relatively simple to estimate.

Specifically, we model Δx_t as a diagonal VAR(1),

$$\text{(3)} \qquad \begin{aligned} \Delta x_t &= \mu_x + \operatorname{diag}(\rho_x)\,\Delta x_{t-1} + \mathrm{Ch}(\Sigma_x)\varepsilon_{xt}, \\ \varepsilon_{xt} &\overset{iid}{\sim} N(0_{C\times 1}, I_C), \end{aligned}$$

where $\mathrm{Ch}(M)$ denotes the lower-triangular Cholesky matrix of M. We collect the unknown parameters of the Δx_t process in $\theta_x = (\mu_x, \rho_x, \Sigma_x)$. In principle, there could be differences in the mean, persistence and variance of economic growth across versions, which will allow us to empirically test whether comprehensive revisions had any impact on the static or dynamic properties of US output.

In practice, we work with the level process x_t rather than with its first differences so that we can impose cointegration, as explained in Supplemental Appendix SM.B.3. For that reason, we model the initial condition for the level as:

$$x_1 \sim N(\mu_{x_1}, \Sigma_{x_1}),$$

independent of ε_{xt} for all t. This accommodates potential differences in levels between the elements of x_t, which adequately captures the use of deflators with a different base year, among other things. We will treat μ_{x_1} and Σ_{x_1} as known and take Σ_{x_1} to reflect a diffuse prior over x_1. A relatively easy-to-implement alternative would be to estimate μ_{x_1}.

For the measurement errors of type i, we also postulate a diagonal VAR(1) model:

$$\text{(4)} \qquad \begin{aligned} v_{it} &= \operatorname{diag}(\rho_i) v_{i,t-1} + \mathrm{Ch}(\Sigma_i)\varepsilon_{it}, \\ \varepsilon_{it} &\overset{iid}{\sim} N(0_{M_i\times 1}, I_{M_i}), \end{aligned}$$

and place the unknown parameters of this process into $\theta_i = (\rho_i, \Sigma_i)$ for i = GDE, GDI. Autocorrelated measurement errors in levels capture the persistent but stationary serial dependence observed in the statistical discrepancies. We also allow for variation in the persistence and variance across different releases.

We specify diagonal VARs and not unrestricted VARs because the pattern of overlap and missing data of different measurements calls for a parsimonious representation. In contrast, unrestricted specifications for Σ_x, Σ_{GDE} and Σ_{GDI} allow for flexible forms of cross-sectional dependence in the different innovations. A thorough assessment of model fit and diagnostics of these modelling choices can be found in Supplemental Appendix SM.C.

Estimation and filtering Our objective is to conduct inference on parameters θ and latent variables $x_1, \ldots, x_T$. We adopt a Bayesian approach because it allows us to easily integrate both estimation and filtering uncertainty when performing signal extraction in a unified, internally consistent framework.

Model (2), (3) and (4) can be represented as a linear state-space model. Using a set of Gaussian-inverse Wishart priors for model parameters, we propose a Gibbs sampling algorithm that allows us to approximate the posterior distribution of parameters and latent variables in a computationally efficient manner. The algorithm relies on standard techniques for state-space simulation smoothing (e.g., Durbin and Koopman (2002)) and multivariate linear regression with unknown covariance matrices.

A detailed discussion of our priors, state-space representation, estimation algorithm, and approach to filtering and inference can be found in Supplemental Appendix SM.B.

4 GDP solera: empirical analysis

In the next subsection, we provide a thorough empirical assessment of our model and carry out some robustness exercises. Afterwards, we use it to analyse several important empirical issues related to both the level and evolution of the US economy. Readers mostly interested in the latter subsection can safely skip the first one.

4.1 Model assessment

4.1.1 Parameter estimates and their stability across comprehensive revisions

Table SM.B.1 in Supplemental Appendix SM.B.5 summarizes the posterior distributions of the model parameters and compares those distributions to their priors. As can be seen, the data seems informative about all the parameters.

A crucial assumption of our approach is the cointegration between the different aggregate measures, as originally suggested by Patterson and Heravi (1991b), Patterson and Heravi (1991a), and Patterson and Heravi (2004). In this respect, we find that the GDE measurement errors have low persistence (point estimates of first-order autocorrelations between 0.1 and 0.2) while GDI measurement errors have higher persistence (point estimates in the range 0.4-0.6), even though they are certainly stationary. Since these are errors for the *(log) level of GDP*, the error for the *growth rate* is anti-persistent, more so for GDE than for GDI, which means that if growth estimates based on GDE and GDI overstate true GDP growth in one quarter, they will tend to understate it in the next one (and vice versa). Importantly, in Supplemental Appendix SM.C.3 we show that the parameter estimates of these measurement errors generate autocorrelation patterns that match the empirical autocorrelations of the statistical discrepancies for the second and third early estimates, and the first, second and third annual estimates.

We also find that the most recently revised estimates of GDE are the most precise ones: the standard deviations of the measurement errors are ≈ 0.5 (in percentage points of the level of GDP) for the early estimates , ≈ 0.3 for the annual estimates, and ≈ 0.2 for the comprehensive estimates. In contrast, GDI estimates are noisier than GDE, and do not improve much with revisions, with standard deviations ≈ 0.6 for almost all of the measurement errors.

In addition, our results indicate that the measurement errors of early GDE estimates have high correlation with each other (≈ 0.85), low correlations with the first annual estimates, and negative correlation with the remaining future revisions. In turn, the measurement errors of the early GDI estimates have low correlations with each other and with future revised estimates, while those of the revised estimates are highly correlated among themselves (≈ 0.8).

By working with all comprehensive revisions simultaneously, we can also use the posterior distribution of the common signal parameters to assess whether there has been any change in the static and dynamic properties of economic activity as a result of the GDP redefinitions. In this respect, a noteworthy observation is that the unconditional means of the growth rates of the five different benchmark versions of US aggregate economic activity that the BEA has produced so far are remarkably similar, even though the comprehensive revision process has certainly affected the levels of US GDP, as we saw in Figure 1. In contrast, its persistence seems to have become somewhat smaller more recently, which is perhaps not surprising in view of the unusual nature of the 2020 COVID-19 recession. We will study the potential effects of this change in sections 4.1.4 and 4.2.4 below.

4.1.2 Precision gains from using all releases for a given comprehensive revision

The root mean square error (RMSE)

$$\sqrt{V_t^\tau} = \sqrt{\mathrm{Var}\left(\mathbb{E}\left[x_{ct} \middle| \mathcal{Y}^\tau\right] - x_{ct}\right)},$$

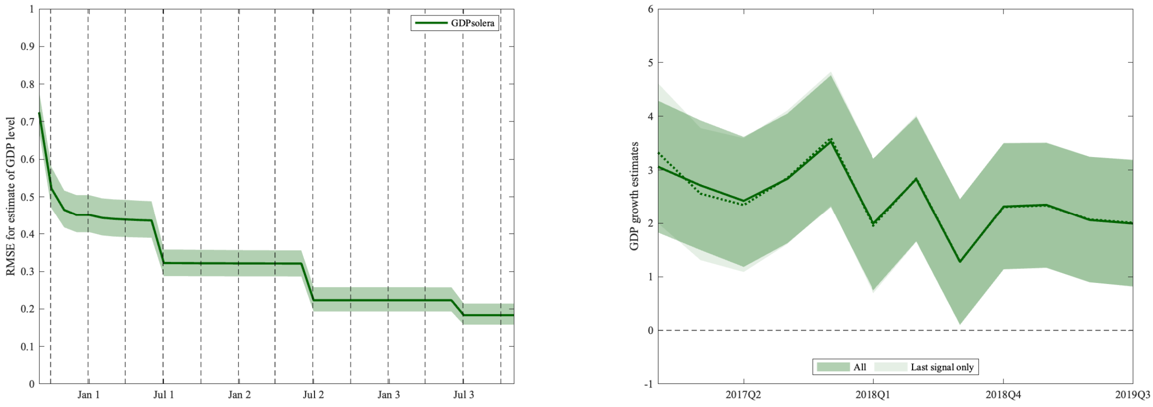
where $\mathcal{Y}^\tau$ denotes the σ-algebra generated by all measurements available until month τ, measures the precision of our signal extraction procedure. Figure 2.a reports this RMSE for $c = 5$ and a fixed t as a function of τ for a sequence of 39 months starting in October of year t, which is when the advance GDE estimate for the third quarter becomes available, under the assumption that no comprehensive revision takes place during those three years and a quarter. In computing this figure we maintain the joint posterior distribution of the model parameters fixed at its estimate in September 2018 to focus on the precision gains of the smoother as new data becomes available. Consequently, the annual revisions correspond to July 2019, 2020 and 2021.

As expected, the release of the advance GDE figure almost halves the RMSE of the prediction of the third quarter growth rate made at the end of September. Nevertheless, substantial precision gains also occur when the second and third estimates of GDE and GDI are released. Moreover, there are further gains when the annual estimates become available in July of the following three years. Still, the non-singular nature of our dynamic model, combined with the fact that the BEA does not attempt to reconcile the GDE and GDI figures, implies that there is a positive floor to the RMSE, which will not go to zero regardless of the number of subsequent annual revisions.

Exactly the same pattern arises if we repeat this exercise for the first and second quarters of year t in April and July, respectively, but not for the fourth quarter, which shows a slightly different initial pattern (not reported here) because there is no second GDI release in February.

4.1.3 Effects of combining all comprehensive revisions

To assess the effect of using data from all comprehensive revisions simultaneously, we have also estimated the single signal version of the model in Section 3 using only the data from most recent comprehensive revision. Figure 2.b reports the posterior medians of GDP growth generated by our MCMC estimation and filtering procedure and their point-wise 90% credible sets based on both datasets for the period 2017Q1 to 2019Q4, together with the actual GDE and GDI figures. In this case, we maintain the joint posterior distributions of the parameters of the models with either one or five signals fixed at their estimates in January 2022. As can be seen, the use of the five comprehensive revisions results in not only significantly tighter bands around the smoothed estimates of economic activity but also a smoother temporal evolution for those estimates.



(a) RMSE improvements

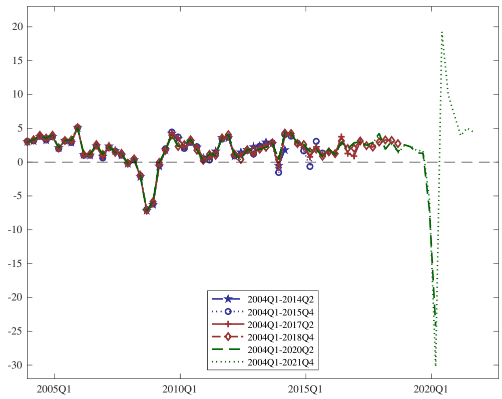
(b) Signal extraction comparison

FIGURE 2. **Accuracy of the *GDPsolera* estimate.** Panel (a) shows the root mean square error improvements from using all releases for a given comprehensive revision The solid green line is the posterior median of the RMSEs. Panel (b) compares signal extraction for Δx_{Ct} using either one or multiple comprehensive revisions. The solid line is the median of Δx_{Ct} given $y_{1:T}$ when all signals are used while the dashed line refers to its conditional median when only the most recent signal is used. Shaded areas are pointwise 90%-probability intervals.

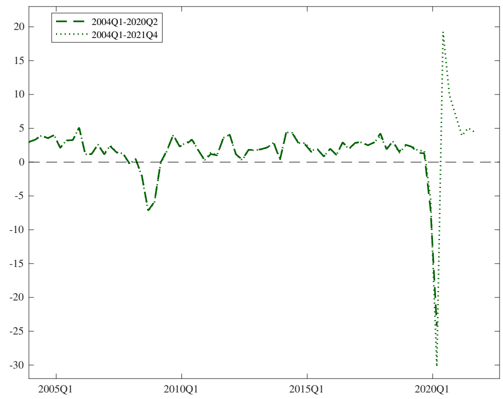
4.1.4 The stability of *GDPsolera* releases

Figure 3 reports the smoothed estimates for US GDP growth from six different solera releases, which we have recursively estimated as follows. The first series uses data until January 2012 to provide estimates up to 2011Q4. Similarly, the second series provides estimates up to 2013Q4 using data until January 2014, and so forth, until the sixth series, which represents estimates of GDP growth until 2021Q4 using the data available at the BEA website at the end of January 2022. Thus, some of the data that was missing in the earlier versions progressively becomes available in the later ones. As can be seen in panel (a), which depicts the six series starting in 2004Q1, all estimates display close paths until 2010Q1.

Still, the growth rates estimates for the last few quarters of each series are somewhat different from the corresponding estimates in the next ones, an effect that it is very likely due to the smoothing embedded in our filtering algorithm, which systematically



(a) Stability with respect to 2018 comprehensive revision and pandemic data



(b) Stability with respect to pandemic data

FIGURE 3. **Stability of *GDPsolera* releases.** The first release uses data until January of 2012 to provide estimates up to 2011Q4. The second provides estimates up to 2013Q4 using data until January of 2014; and so on until the sixth which, using data until January of 2022, delivers estimates of GDP growth up to 2021Q4. Panel (b) displays the two most recent ones.

reassesses the past after observing the future.

Additionally, the two most recent solera releases that we display in green present a different pattern from the others in the second quarters of 2011 and 2012. These differences can be explained by the fact that the data underlying those last two series incorporate modifications to the GDP definition resulting from the comprehensive revision the BEA released in July 2018.

Indeed, panel (b), which only reports the two most recent series in panel (a), shows an extremely similar pattern between them even though the most recent version of *GDPsolera* includes data from the pandemic. Therefore, the post pandemic estimates for the pre-pandemic period are remarkably stable to the inclusion of the large 2020 outliers, which affect not only the output of the simulation smoother for fixed parameter estimates but also the posterior distribution of the parameter estimates.

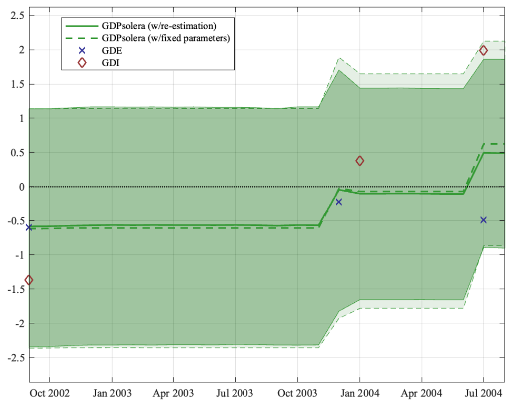
4.2 The evolution of the US economy

4.2.1 Analysis of some specific quarters

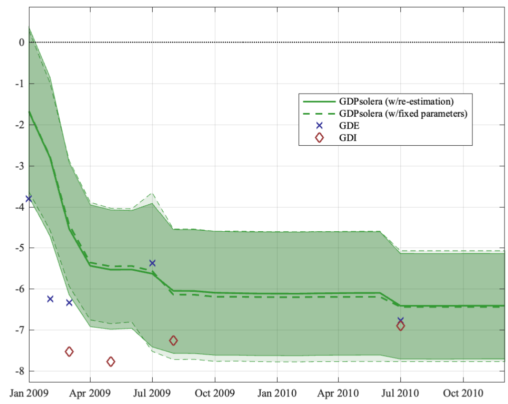
Next, we shed some light on the effect of data revisions as well as the arrival of information for subsequent periods on the estimates of US GDP growth rate at a fixed date t both through the smoother and the real time re-estimation of the model parameters in three specific quarters of interest: 2001Q1, 2008Q4 and 2019Q2.

We chose the first one because of the political controversy surrounding what at the time some Republican politicians called the "Clinton recession", in marked contrast to the NBER Business Cycle Dating Committee, which officially dated the peak of the previous ten year expansionary phase in March 2001. Although the BEA only publishes vintage data from September 2002 onwards, Figure 4a, which uses blue crosses and red diamonds to represent GDE and GDI estimates, respectively, shows that the data initially available suggested that GDP growth had already turned negative in the first quarter of 2001. However, the comprehensive revision that became available in December 2003 is more ambiguous, with GDE and GDI growth rates having different signs. If anything, the subsequent annual revision released in July 2004 increases the degree of ambiguity. Not surprisingly, when one looks at the solid and dashed lines in that figure, which represent the posterior medians with and without parameter re-estimation, and the shaded areas, which display the corresponding 90% point-wise credible bands, the only conclusion that one can draw is that the uncertainty is too large to determine the sign of the GDP growth rate in 2001Q1 unequivocally.

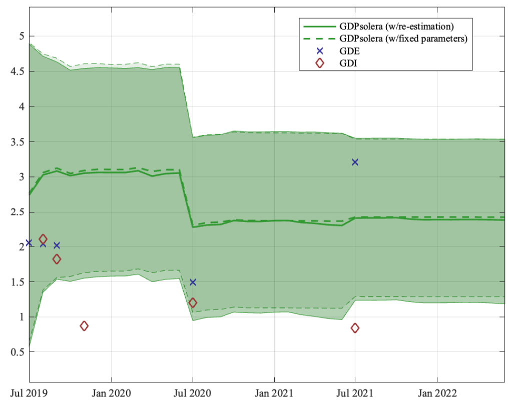
Our next example focuses on 2008Q4, the worst quarter of the Great Recession, which we analyze in Figure 4b. Although the advanced GDE estimate the BEA released initially pointed to a serious but not dramatic recession, subsequent releases justify the adjective ''Great". Nevertheless, this figure also shows the adjustment of the posterior median of our *GDPsolera* growth estimate as soon as we process the third releases of GDE and GDI, which is in line with the evidence we observed in Figure 2. In addition, Figure 4b



(a) Δx_{2001Q1}



(b) Δx_{2008Q4}



(c) Δx_{2019Q2}

FIGURE 4. **Real time filtering of Δx_t.** The dashed and solid lines are posterior medians with and without parameter re-estimation while the shaded areas represent 90%-point-wise credible bands. Data releases for GDE (blue crosses) and GDI (red diamonds) are displayed too.

also shows the effect that the comprehensive revision of July 2009 had on the precision of the estimates, and especially the annual revision of July 2010, which reduced further not only the growth rate but also the width of the credible sets.

Our third and final example focuses on 2019Q2, a relatively normal quarter despite the fact that some Federal Reserve officials had previously expressed concerns about a potential deceleration of the economy. This quarter is also interesting because it allows us to explicitly assess the effect of the pandemic data on our parameter estimates. As Figure 4c shows, the estimates of economic growth were noticeably revised downwards after the annual update that the BEA released in July 2020. However, a substantial part of this reduction was reversed following the July 2021 annual revision. Interestingly, the width of the credible sets goes down fairly slowly, which probably reflects the fact that the unprecedented GDP fluctations in 2020Q2 and 2020Q3 increased the uncertainty of the parameter estimates.

4.2.2 Weights of the *GDPsolera* estimates

Like in any linear state-space model, our smoother potentially assigns some weight to all past, present and future measurements in order to come up with the best possible estimates of the underlying variables of interest. Therefore, for each observation there is an entire matrix of weights with both a cross-sectional dimension reflecting the different measurements for that quarter and a time-series dimension capturing past and future quarters. In our context, though, those weights change from month to month as different measurements become available, so in practice, the entire matrix of weights is evolving over time, although with a clear recurrent pattern.

To provide a snapshot of this matrix, we have created a table showing the weights for different vintages of GDE and GDI. For comparison purposes, we focus on the same quarter we considered in Figure 2, with the **Month** column indicating how many months after the end of Q3 have elapsed. To compute these weights in a numerically efficient manner, we fix the parameter estimates at their posterior means and successively perturb each measure by one unit, comparing the resulting smoothed quantity to the one obtained with the actual data.

A basic insight of signal extraction is that one should give more weight to those measurements which are less noisy and less correlated with the rest. As we saw before, soon after the end of a quarter, GDE estimates are slightly less noisy than GDI but their measurement errors are more correlated with each other. These two forces roughly compensate by the time of the release of the third estimate, so the weights to GDE and

TABLE 1. Weights of *GDPsolera estimates*

	GDE weights						GDI weights				
Month	Adv	2nd	3rd	An1	An2	An3	2nd	3rd	An1	An2	An3
1	.72										
2	.20	.29					.37				
3	.11	.17	.15				.24	.22			
4	.08	.15	.12				.20	.18			
5	.09	.13	.12				.20	.18			
6	.09	.14	.12				.20	.17			
7	.09	.14	.12				.20	.17			
8	.09	.14	.12				.20	.17			
9	.09	.14	.12				.20	.17			
10	.05	.08	.07	.27			.11	.09	.12		
22	.06	.09	.08	.12	.40		.05	.04	.05	.00	
34	.07	.10	.08	.07	.23	.29	.03	.03	.04	.01	-.02

GDI are balanced. But once the first annual revision occurs, the estimates of GDE/GDI receive high/low weights, as GDI remains noisier than GDE after multiple revisions. Nevertheless, the weights are not merely concentrated on the latest available estimate, with early estimates retaining some influence.

However, *GDPsolera* depends not only on the different GDE and GDI measures for the quarter of interest, but also on the measures for previous and successive quarters through the smoother.

For that reason, we also look at the split of the weigths across:

a. GDE and GDI regardless of the quarter, and

b. the quarterly lags and leads irrespective of the measure.

Total weights on GDE exceed those on GDI, eventually by a long margin, unlike the contemporaneous weights in the previous table, which were balanced, at least during

TABLE 2. Weights of *GDPsolera* estimate on quarter-t data

Month	GDE	GDI	Lag Q	Cont Q	Lead Q
1	.84	.16	.28	.72	
2	.63	.37	.14	.86	
3	.58	.42	.11	.89	
4	.63	.37	.14	.73	.13
5	.64	.36	.14	.72	.14
6	.63	.37	.14	.72	.14
7	.63	.37	.14	.72	.14
8	.64	.36	.14	.72	.14
9	.65	.35	.14	.72	.14
10	.74	.26	.11	.78	.11
22	.92	.08	.05	.90	.05
34	.96	.04	.03	.94	.03

the first twelve months. The reason is that GDE contributes more to the dynamic term, which is in turn explained by the lower persistence of its measurement errors.

On the other hand, dynamic weights are reasonably small, symmetric between leads and lags, and become minimal when subsequent annual revisions are available.

4.2.3 Why not give all the weight to the most recent releases?

The BEA does it best to improve GDP measurements as new information sources come along. However, this does not mean that there is no information on the past releases of GDE and GDI.

We assess the informational efficiency of the most recent estimates by testing if the revisions are correlated to past growth rates computed from earlier estimates, which we justify in Supplemental Appendix SM.E.

Using growth rates of GDE and GDI we get the following correlations:

$$\mathrm{Corr}(\Delta y_{\mathrm{GDESec}}, \Delta y_{\mathrm{GDPSec}} - \Delta y_{\mathrm{GDPAdv}}) = 0.45 \ _{[0.25,0.58]}$$

$$\mathrm{Corr}(\Delta y_{\mathrm{GDEThr}}, \Delta y_{\mathrm{GDPThr}} - \Delta y_{\mathrm{GDPAdv}}) = 0.47 \ _{[0.26,0.69]}$$

$$\mathrm{Corr}(\Delta y_{\mathrm{GDEThr}}, \Delta y_{\mathrm{GDPThr}} - \Delta y_{\mathrm{GDPSec}}) = 0.48 \ _{[0.25,0.61]}$$

$$\mathrm{Corr}(\Delta y_{\mathrm{GDIThr}}, \Delta y_{\mathrm{GDIThr}} - \Delta y_{\mathrm{GDISec}}) = 0.39 \ _{[0.18,0.57]}$$

Given that the figures in brackets correspond to 95%-confidence intervals for each correlation computed via block bootstrap, we can conclude that all of them are significantly different from zero. Moreover, a joint test of the null hypothesis that these correlations are simultaneously zero rejects with level below 1%. In other words, there is strong evidence that the optimal estimate of x_t should not discard the previous vintages.

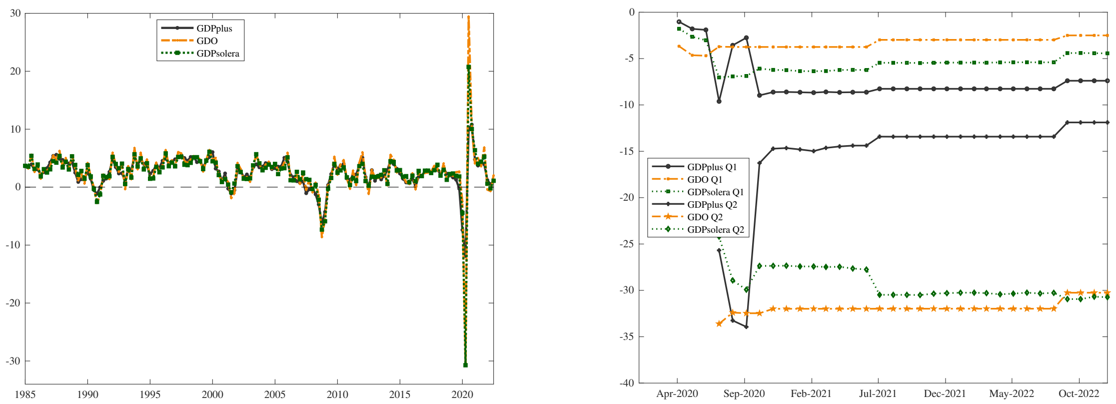
If x_t is a sum of components and each y_i is a sum of noisy measurements of those components, correlations between the latest estimate and revisions may arise if some components are missing and imputed in certain ways.

4.2.4 Comparison with *GDO* and *GDPplus*

We also compare our measure of economic activity – *GDPsolera* – with the simple arithmetic average of the expenditure and income measures reported by the BEA since 2015 as *GDO*, as well as with *GDPplus* initially proposed by Aruoba et al. (2016), and released on a monthly basis by the Federal Reserve Bank of Philadelphia since August 2013. Unfortunately, we cannot compare our measure to the *GDP*++ series of Jacobs et al. (2022) because their vintage data is unavailable.

To begin with, we look at the smoothed estimates of GDP between the first quarter of 1985 and the fourth quarter of 2021. To construct our solera measure, we use the data released by the BEA by the end of January 2022, while for *GDPplus* we use the release that uses the same dataset to level the playing field. Given that *GDPplus* is based on the mostly recently available estimates of GDE and GDI rather than on multiple vintages, it can use data from 1960Q1. Nevertheless, this should not affect too much their estimates in recent years.

We plot the estimated annualized growth rates in panel (a) of Figure 5. As can be seen, the two Kalman filter-based estimates are quite close to each other with a contemporaneous correlation of 0.86, and an average annualized growth rate of 2.61% for *GDPplus* and 2.54% for *GDPsolera* over the entire sample period. GDO, on the other hand, has contemporaneous correlation of 0.96 with both GDE and GDI, and a standard deviation of 4.2% compared to 4.4% for both GDE and GDI.



(a) Comparison of most recent historical estimates

(b) First two quarters of 2020 estimates

FIGURE 5. **Comparing *GDPplus*, GDO and *GDPsolera***. Panel (a) displays *GDPplus*, GDO and *GDPsolera* series estimated using data until January 2022. Panel (b) displays GDP estimates at the beginning of the COVID-19 outbreak revised from April 2020 to January 2022.

Nevertheless, our solera estimates are clearly more volatile than *GDPplus*, with a standard deviation that is 40% larger than. The smoothness of *GDPplus* results in relatively more conservative estimates of the large fall and rise of economic activity after the start of the COVID-19 outbreak.

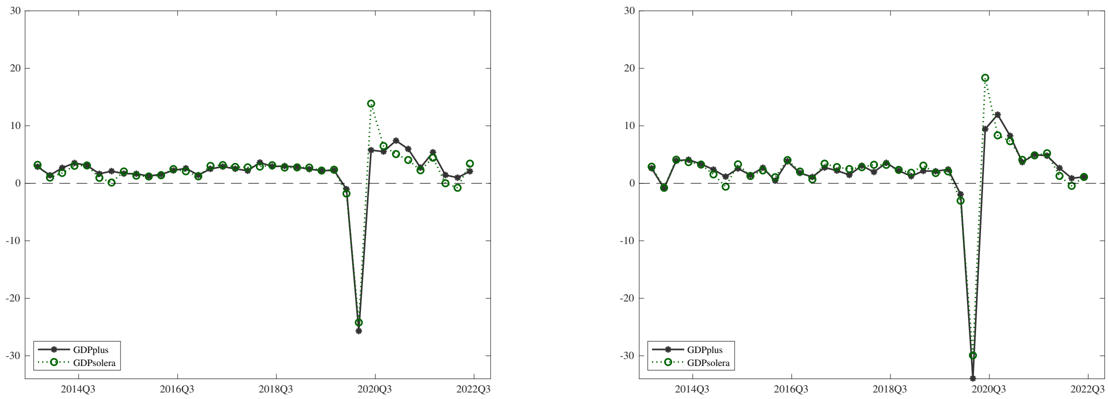
To shed further light on this, we report in panel (b) of Figure 5 the two real-time estimates of economic activity for 2020Q1 and 2020Q2 using the data available at the time. Perhaps not suprisingly, for 2020Q1 both estimators of GDP are in agreement, and remain quite stable as new information became available. In contrast, the estimators for 2020Q2 are very different and this difference increased in October 2020 when the BEA published the advance GDE estimate for 2020Q3. This is perhaps not surprising in view of the tables of weights that we have presented in section 4.2.2, as our approach is more cross-sectional than time-series, in the sense that it combines multiple measures for the same quarter, relying less in smoothing. Interestingly, the most recent figures produced by the BEA for the COVID-19 recession are closer to the *GDPsolera* series. Nevertheless, it must be acknowledged that the extremely atypical size of the pandemic shock is a challenge to linear Gaussian state-space models, which makes the comparison difficult.

In our last exercise, we compare the concurrent online estimates of GDP growth rates generated by *GDPplus* and our procedure. Specifically, we consider estimates for each quarter based on the information available one month after the end of that quarter, by which time only the "advance" GDE estimate is available, which implies that GDO could not offer any precission gains. In addition, we also look at the estimates of the same GDP growth rates obtained three months after the end of the quarter, which also make use of the "second" and "third" estimates of GDE and GDI released by the BEA, which gives empirical meaning to GDO. Panels (a) and (b) of Figure 6 displays these two set of results. Interestingly, the real time *GDPsolera* and *GDPplus* estimates appear to be more similar than the historical ones we saw in Figure 5, and in line with the "third" GDO estimates. Still, we can observe a few differences in the first two quarters of 2015 afected by the 2018 comprehensive revision, and at the end of the sample, starting after the 2020Q2 drop and the 2020Q3 rebound.

4.2.5 Post-pandemic evolution

The COVID-19 recession was the most dramatic decline in economic activity seen in at least a century but also the shortest: US GDP sank from about 19.2 trillions in 2019Q4 to about 17.4 in 2020Q2, then rebounded to 18.6 in 2020Q3.

There is less agreement among the different estimates about the strength of the recovery and expansion that followed. Panels (a) and (b) of Figure 7 display this disagreement



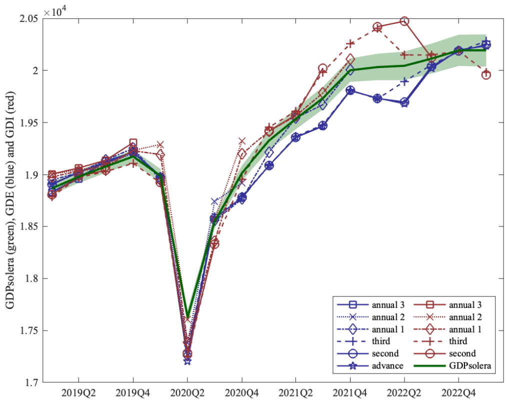
(a) GDP growth estimates one month after the end of the quarter

(b) GDP growth estimates three months after the end of the quarter

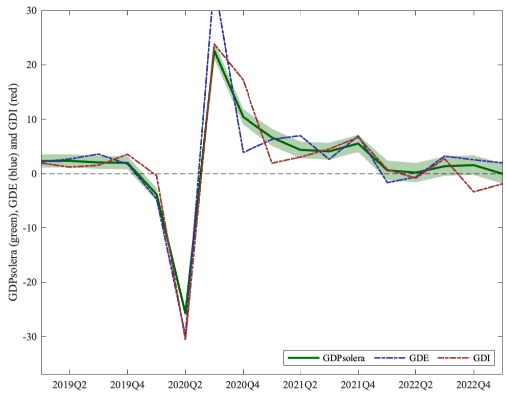
FIGURE 6. **Nowcast: *GDPplus* versus *GDPsolera*.** Panel (a) displays *GDPplus* and *GDPsolera* series estimated with information available one month after the end of the quarter when only advance of GDE is available for the most recent quarter. *GDPplus* for 2018Q4 was released in February 2019. Panel (b) displays *GDPplus* and *GDPsolera* series estimated with information available three months after the end of the quarter.

for the level and growth rate of GDP, respectively, alongside the corresponding *GDPsolera* estimates and their 90% (pointwise) credible intervals. The width of those intervals crucially depends on the availability of annually revised estimates. For 2019-20, we have second and third annual revisions, but for 2021 we only have the first one, while for 2022 we have none. Given that growth rates are multiplied by 4 to express them in annual terms, mechanically there would be more uncertainty about growth rates than about levels.

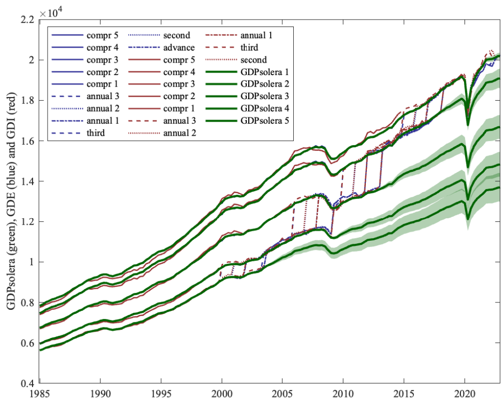
Based on the point estimates, it appears the economy surpassed its 2019Q4 level in 2021Q1 and continued growing rapidly afterwards, at rates that ranged from 4.5% to 7% during each quarter of 2021. Then, abruptly, it decelerated in 2022Q1 and was flat for the rest of 2022 around the 20 trillion mark. The suddenness of the deceleration is remarkable: growth declined from 5.5% in 2021Q4 to 0.7% in 2022Q1 and -0.1% in 2022Q2. This implies a drop of 5.6pp in the lapse of two quarters, similar to what is typically observed in recessions.



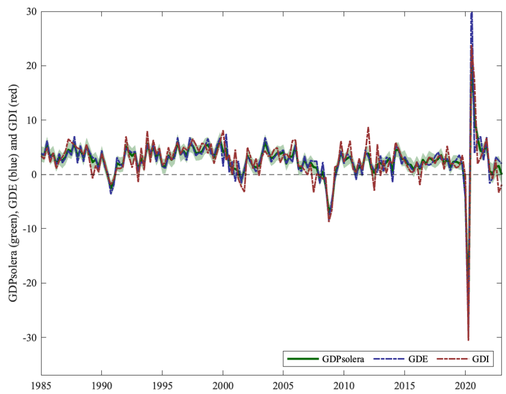
(a) Post-pandemic GDP levels



(b) Post-pandemic GDP growth rates



(c) Full-sample GDP levels



(d) Full-sample GDP growth rates

FIGURE 7. ***GDPsolera* together with GDE and GDI data from the BEA**. For the current version of economic activity, we display different GDE and GDI growth measurements alongside our *GDPsolera* estimate (solid green line) and 90% pointwise credible bands. Panels (a) and (b) display levels and growth rates, respectively. Panels (c) and (d) report analogous results over the entire sample period for comparison purposes.

If we were to take GDE estimates at face value, though, the deceleration would be even more dramatic, as they indicate growth was 7% in 2021Q4, -1.6% in 2022Q1 and -0.6% in 2022Q2. It is difficult to rule out two quarters of negative growth, but it is not the best guess according to our point estimates. In fact, the GDP estimate for 2022Q1 lies below the lower end of the 90%-CI while the GDP estimate for 2022Q3 growth (3.1%) is at the upper end of its corresponding interval. On the other hand, GDI estimates initially painted a different picture: 2022Q1 GDI growth was estimated at 2.1% in May (marked down to 0.7% in a September off-schedule revision) while 2022Q2 was 0.1% in August (then revised to -0.8%).

Our results suggests that the ups and downs in these numbers are mostly noise, and that, instead, economic activity came to a sudden stop early in 2022 and was stagnant and not very volatile since then — a scenario in between the contraction of GDE estimates and the more gradual deceleration in initial GDI estimates.

5 Conclusion

We make the best use of the information in the different vintages of GDE and GDI from the current comprehensive revision to obtain an improved timely measure of US aggregate output by imposing cointegration between the different measures and taking seriously their monthly release calendar. We also combine overlapping comprehensive revisions to achieve further improvements.

We express our model in linear state-space form, and use Bayesian methods of inference for both parameters and latent variables. Specifically, we develop a Gibbs sampling algorithm that tackles estimation and signal extraction simultaneously, allowing for an efficient and conceptually simple integration of uncertainty coming from different sources. Thus, we obtain a posterior distribution for the underlying GDP measure, whence we can obtain not only point estimates but also measures of dispersion.

The estimated parameters of our dynamic state-space model suggest that comprehensive revisions have not changed the long-run growth rate of US GDP, but they have somewhat lowered the persistence of its shocks. We also find that revised GDE estimates are increasingly precise and receive higher weights, unlike GDI ones, but early estimates retain some influence. In this respect, we clearly reject the null hypothesis of lack of predictability for the revisions in the second and third releases of GDE, as well as in the third release of GDI.

Our results also suggest that noticeable precision gains in signal extraction occur not only when the advance, second and third estimates of GDE and GDI are released but also when the annual estimates become available in July of the subsequent years. We also observe that the use of the five comprehensive revisions not only results in significantly tighter bands around the smoothed estimates of economic activity, but also a smoother temporal evolution for those estimates.

In addition, we pay particular attention to certain recent episodes, including the Great Recession, and the COVID-19 pandemic, which, despite producing dramatic fluctuations, does not generate noticeable revisions in previous growth rates.

Although the objective of our analysis is not the creation of a real time activity index (see e.g. Lewis, Mertens, Stock, and Trivedi (Forthcoming) and the references therein), combining our approach with either high frequency data or additional quarterly variables constitutes a promising avenue for further research. Assessing the effect of incorporating the seasonally unadjusted GDE and GDI data that the BEA has released since 2018 to our empirical results would also provide a valuable addition.

Similarly, the potential forecasting improvements of the model we propose in this paper for the early releases of GDE and GDI would be worth investigating, as they would provide an external validity check on our modelling approach. In this respect, another potential extension would allow for a more flexible autocorrelation structure, as well as conditional heteroskedasticity and non-normal shocks, although the latter would require replacing the analytical Kalman filter by a numerical non-linear one.

Finally, it would be interesting to apply our approach to the different components of GDE and GDI, as well as other macroeconomic series subject to revisions, like the Non-farm Payroll Employment figures or the Chained Consumer Price Index for All Urban Consumers released by the US Bureau of Labor Statistics.

Acknowledgments

We would like to thank Boragan Aruoba, Gabriel Pérez Quirós, António Rua, Dongho Song and audiences at the RCEA Webinar Series, the 3rd Italian Workshop of Econometrics and Empirical Economics (Rimini), Banco de Portugal, the XII Time Series Econometrics Workshop (Zaragoza), Universitat de les Illes Balears, the 2022 North American (Miami) and European (Milan) Summer Meetings of the Econometric Society, American University in Cairo, Bank of Spain, XI Encuentro de la SEU (Montevideo), Universidad de Alicante, the 2023 Climate Finance and the Hydrogen Economy Conference (ICADE) and the 2023 Barcelona Workshop in Financial Econometrics (ESADE) for useful comments and suggestions. We are also thankful to Joël Marbet, Utso Pal Mustafi and Clara Arroyo for excellent research assistance. Three anonymous referees provided very useful feedback. Of course, the usual caveat applies. The second and fourth authors gratefully acknowledge financial support from the Spanish Ministry of Science and Innovation through grant PID2021-128963NB-I00, while the third one is grateful to MIUR through the PRIN project "High-dimensional time series for structural macroeconomic analysis in times of pandemic". The views expressed in this paper are those of the authors and do not necessarily reflect the position of the Federal Reserve Bank of New York or the Federal Reserve System.

References

Almuzara, M., D. Amengual, and E. Sentana (2019): "Normality tests for latent variables," *Quantitative Economics*, 10, 981–1017.

Almuzara, M., G. Fiorentini, and E. Sentana (2023): "Aggregate output measurements: a common trend approach," in *Essays in Honor of Joon Y. Park: Econometric Methodology in Empirical Applications, Advances in Econometrics*, ed. by Y. Chang, S. Lee, and J. Miller, Emerald, vol. 45B.

Aruoba, S. B. (2008): "Data revisions are not well behaved," *Journal of Money, Credit and Banking*, 40, 319–340.

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

Campbell, J. and G. Mankiw (1987): "Are output fluctuations transitory?" *Quarterly Journal of Economics*, 102, 857–880.

Durbin, J. and S. J. Koopman (2002): "A simple and efficient simulation smoother for state space time series analysis," *Biometrika*, 89, 603–615.

Greenaway-McGrevy, R. (2011): "Is GDP or GDI a better measure of output? A statistical approach," Working Paper 2011–08, Bureau of Economic Analysis.

Grimm, B. T. (2007): "The statistical discrepancy," Working Paper 2007-01, Bureau of Economic Analysis.

Jacobs, J. P. A. M., S. Sarferaz, J. Sturm, and S. van Norden (2022): "Can GDP measurement be further improved? Data revision and reconciliation," *Journal of Business and Economic Statistics*, 40, 423–431.

Jacobs, J. P. A. M. and S. van Norden (2011): "Modeling data revisions: Measurement error and dynamics of "true" values," *Journal of Econometrics*, 161, 101–109.

Kerry, P., S. H. McCulla, and D. B. Wasshausen (2018): "Improved estimates of the national income and product accounts: Results of the 2018 comprehensive update," *Survey of Current Business*, 98.

Landefeld, J. S., E. P. Seskin, and B. M. Fraumeni (2008): "Taking the pulse of the economy: Measuring GDP," *Journal of Economic Perspectives*, 22, 193–216.

Lewis, D. J., K. Mertens, J. H. Stock, and M. Trivedi (Forthcoming): "Measuring real activity using a weekly economic index," *Journal of Applied Econometrics*.

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

——— (2011): "The income- and expenditure-side measures of output growth – An update through 2011Q2," *Brookings Papers on Economic Activity*, 2, 385–402.

Orphanides, A. (2001): "Monetary policy rules based on real-time data," *American Economic Review*, 94, 964–984.

Patterson, K. D. and S. M. Heravi (1991a): "Are different vintages of data on the components of GDP co-integrated?: Some evidence for the United Kingdom," *Economics Letters*, 35, 409–413.

——— (1991b): "Data Revisions and the Expenditure Components of GDP," *The Economic Journal*, 101, 887–901.

——— (2004): "Revisions to Official Data on U.S. GNP: A Multivariate Assessment of Different Vintages," *Journal of Official Statistics*, 20, 573.

Smith, R. J., M. R. Weale, and S. E. Satchell (1998): "Measurement error with accounting constraints: Point and interval estimation for latent data with an application to U.K. gross domestic product," *Review of Economic Studies*, 65, 109–134.

Stock, J. H. (1994): "Unit roots, structural breaks and trends," in *Handbook of Econometrics*, ed. by D. McFadden and R. Engle, Elsevier, vol. 4, chap. 46.

Stone, R., D. G. Champernowne, and J. E. Meade (1942): "The precision of national income estimates," *Review of Economic Studies*, 9, 111–125.

Weale, M. (1992): "Estimation of data measured with error and subject to linear restrictions," *Journal of Applied Econometrics*, 7, 167–174.