GDP solera: the ideal vintage mix

Martín Almuzara
New York Fed

Dante Amengual
CEMFI

Gabriele Fiorentini
Università di Firenze

Enrique Sentana
CEMFI

BdE - CEMFI Workshop
November 22nd, 2022

Introduction

- Despite the recent interest in alternative 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 U.S., the estimates of aggregate economic activity are published by the Bureau of Economic Analysis (BEA).
- 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.
- The practical relevance of GDP revisions should not be underestimated (see, e.g., Orphanides (2001) AER).

Introduction

- In the last two decades there has been considerable progress in jointly modelling the different vintages of data (see, e.g., Aruoba (2008) *JMCB*, Jacobs and VanNorden (2011) *JoE* and the references therein).
- Some of these studies have ignored a second important consideration: the BEA produces not just one but two different measures of aggregate output and income: Gross Domestic Expenditure (**GDE**) and Gross Domestic Income (**GDI**).
- 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, and various proposals for improved combinations have been discussed: Nalewaik (2010, 2011) *BPEA*, Greenaway-McGrevy (2011) *BEA WP*, Aruoba et al. (2016) *JoE*, and Jacobs et al. (2022) *JBES*.

Introduction

- For example, the *GDPplus* measure of Aruoba et al. (2016) is currently released on a monthly basis by the Philadelphia Fed.
- The purpose of our paper is to simultaneously tackle all these measurement issues within a single, internally coherent, signal extraction framework.
- Our crucial points of departure from the previous literature are:
    1. We impose that any two aggregate expenditure and income measures (in logs) are cointegrated, with cointegrating vector (1,-1).
    2. The measurement errors are mean-reverting and stationary, although they may be serially correlated.
    3. Our empirical analysis uses data on **all** the GDE and GDI vintages from the BEA, including those that were discontinued after comprehensive revisions, taking their release calendar very seriously.
- Thus, we are able to focus not only on quarterly growth rates, but also assess the level of US output, which is important for cross-country comparisons.

Data background: Sequential releases

Estimates for quarterly GDP are released in the following order:

A. **Advance estimate**, based on incomplete data or subject to further revision by the source agency, and released near 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, and
C. **Latest (or final) estimates**, which reflect 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.

Data background: Joint releases

(a) **Annual updates**, usually in July, which cover all quarters of the most recent calendar years (e.g. in July 2017, the BEA revise all quarters for 2014, 2015 and 2016).

They incorporate newly available annual data, and minor methodological changes, but the deflator's base year is usually fixed.

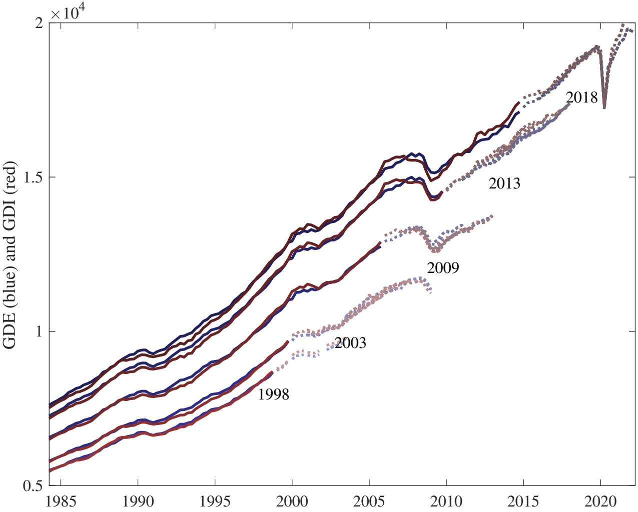
(b) **Comprehensive (or benchmark) updates**, approximately every 5 years (December 2003, July 2009, July 2013 and July 2018).

They incorporate data from a quinquennial business census and major methodological changes (e.g. in 2013, R&D started counting as an investment rather than as a cost, boosting US GDP by over 2%), and improved coverage of increasingly important sectors. Real GDP is also re-based.

Example of release schedule

Release Month	Estimate	GDE		GDI	
		New	Updated	New	Updated
Jan 2017	Advance	2016Q4			
Feb 2017	Second		2016Q4		
Mar 2017	Third		2016Q4	2016Q4	
Apr 2017	Advance	2017Q1			
May 2017	Second		2017Q1	2017Q1	
Jun 2017	Third		2017Q1		2017Q1
Jul 2017	Advance	2017Q2	2014Q1-2016Q4		2014Q1-2016Q4
Aug 2017	Second		2017Q2	2017Q2	
Sep 2017	Third		2017Q2		2017Q2
Oct 2017	Advance	2017Q3			
Nov 2017	Second		2017Q3	2017Q3	
Dec 2017	Third		2017Q3		2017Q3
Jan 2018	Advance	2017Q4			
Feb 2018	Second		2017Q4		
Mar 2018	Third		2017Q4	2017Q4	
Apr 2018	Advance	2018Q1			
May 2018	Second		2018Q1	2018Q1	
Jun 2018	Third		2018Q1		2018Q1
Jul 2018	Advance	2018Q2	1947Q1-2017Q4	2018Q2	1947Q1-2017Q4

Data releases



GDE and GDI data from the BEA. Solid lines represent data released under comprehensive revisions while dashed lines represent data produced by early and annual revisions.

The model without comprehensive revisions

- For clarity, we begin with a version of the model that has no comprehensive revisions, only the different releases and the annual revisions.
- x_t is the quantity of interest, in our case US economic output (in logs) during quarter t.
- We treat x_t as an unobservable variable of which only noisy measurements y_t are available.
- The aim is to combine multiple y_t's for the purposes of obtaining an improved estimate of economic activity.

The model without comprehensive revisions

- Define y_{it}^m as a noisy measurement of x_t, where
  - $i = 1, \ldots, N$ denotes **type** (e.g., GDE and GDI), and
  - $m = 1, \ldots, M_i$ denotes **release** (e.g., early and annual estimates).
- The model is given by the set of M measurement equations

$$y_{it}^m = x_t + v_{it}^m,$$

where v_{it}^m is the measurement error, and $M = \sum_{i=1}^N M_i$.

- In our case:
  - $N = 2$ (GDE and GDI)
  - $M_1 = 6$ (Advance, Second, Third, Latest1, Latest2 and Latest3)
  - $M_2 = 5$ (Second, Third, Latest1, Latest2 and Latest3)
- In compact vector notation:

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

The model without comprehensive revisions

Assumption

(a) Δx_t *is* $I(0)$
(b) $v_{1t}, ..., v_{Nt}$ *are* $I(0)$*;*
(c) $\Delta x_t, v_{1t}, ..., v_{Nt}$ *are mutually orthogonal at all leads and lags.*

Assumptions *(a)* and *(b)* imply that y_t is cointegrated with cointegration rank $M - 1$, while Assumption *(c)* is needed for (nonparametric) identification:

Proposition

Under the assumption above, if $N > 1$*, the autocovariances of* $\Delta x_t, v_{1t}, \ldots, v_{Nt}$ *are nonparametrically identified from the autocovariances of* Δy_t*.*

The model with comprehensive revisions: The ideal mix

- Our approach to modelling 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.
- Let C be the number of versions and $x_t = (x_{1t}, \ldots, x_{Ct})'$, where x_{ct} represents the hypothetical value of underlying economic output defined by revision c.
- For example, x_{ct} treats R&D as a cost before 2013 and as an investment afterwards.
- The dimension of x_t increases with time as new comprehensive revisions become available.
- In our empirical application we have $C = 5$ (until 2023 presumably).

The model with comprehensive revisions: The ideal mix

- Let δ_{it}^m be a $1 \times C$ row vector that has 1 in entry c if y_{it}^m measures x_{ct} and 0 otherwise.
- Our model postulates that:

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

- Stacking δ_{it}^m 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

$$y_t = \delta_t x_t + v_t.$$

- The loading matrix δ_t 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.

The model: Parametric specification

- To complete the model we need to specify the process for the latent vector x_t and the measurement errors.
- We adopt a diagonal VAR specification for Δx_t with a factor structure in the error term:

$$\Delta x_t = \mu_x + \text{diag}(\rho_x)\left(\Delta x_{t-1} - \mu_x\right) + \lambda_x \eta_{xt} + \text{diag}(\sigma_x)\varepsilon_{xt},$$
$$\eta_{xt} \overset{iid}{\sim} N(0,1) \text{ independent of } \varepsilon_{xt} \overset{iid}{\sim} N(0_{C\times 1}, I_C).$$

  and collect the unknown parameters of this process in $\theta_x = (\mu_x, \rho_x, \lambda_x, \sigma_x)$.
- 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 have any impact on the static or dynamic properties of US output.

Model structure: Parametric specification

- The initial condition for the level is modeled as $x_1 \sim N(\mu_1, \Sigma_1)$ independent of $\eta_{xt}, \varepsilon_{xt}$ for all t, which accommodates potential differences in levels between versions, partly due to the use of deflators with different base years.
- For the measurement errors of type i we postulate a parsimoniuous diagonal VAR(1) model with a factor structure in the error too:

$$v_{it} = \text{diag}(\rho_i) v_{i,t-1} + \lambda_i \eta_{it} + \text{diag}(\sigma_i) \varepsilon_{it},$$
$$\eta_{it} \overset{iid}{\sim} N(0,1) \text{ independent of } \varepsilon_{it} \overset{iid}{\sim} N(0_{M_i \times 1}, I_{M_i}).$$

  and place its unknown parameters into $\theta_i = (\rho_i, \lambda_i, \sigma_i)$.
- Autocorrelated measurement errors in levels capture the persistent but stationary serial dependence of the statistical discrepancies.
- We also allow for heterogeneity in the autocorrelations and volatilities of the different releases.

Bayesian estimation

State-space representation

For a given $\theta = (\theta_x, \theta_1, \ldots, \theta_N)$, we can easily cast the model in state-space form as:

$$y_t = H_t X_t,$$
$$X_t = C(\theta) + F(\theta) X_{t-1} + G(\theta) U_t,$$
$$U_t \overset{iid}{\sim} N(0_{(C+M+N+1)\times 1}, I_{C+M+N+1}).$$

- Our objective is to conduct inference on parameters θ and latent variables $x_1, \ldots, x_T$, so a Bayesian approach is very convenient.
- The model lends itself to stable and efficient algorithms, exploiting a Gibbs sampler for estimation and the Durbin and Koopman (2002) simulation smoother for signal extraction.

Estimation

- Given that our model is for **quarterly** data but the new figures are provided by the BEA on a monthly frequency, it can be re-estimated **every month**.
- We define *GDPsolera* as the estimate of the most recent version $\hat{x}_{ct}$ for $c = C$.
- For example, if we estimate the model now $c = 5$, but if we want to know what the results would have been with data until, say, 2014, we would use $c = 4$ because the 2018 comprehensive revision data was not yet available then.

Advantages of the ideal mix vs the most recent version

- Analysts and policy makers typically focus on the latest version, x_{Ct}, and use only the most recent vintage of data arising from the latest comprehensive revision.
- Are there any compelling reasons for using data from vintages that have been discontinued?
- In addition to the interest in older definitions of economic activity from a historical perspective, learning from older vintages about the dynamics of the measurement errors might lead to improved inferences about x_{Ct} even if one only uses the most recent version.
- Moreover, the dependence between the different versions should lead to more precise filtered and smoothed estimates.

Parameter stability

- An important question that we can address by working with all five comprehensive revisions simultaneously is whether there has been any change in the average growth rate of US GDP or its persistence.
- It seems 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 ▸ plot
- In contrast, its persistence seems to have become somewhat smaller more recently. ▸ plot
- We also find that the common shock to the different signals is more important than their idiosyncratic shocks in explaining the variance of their innovations, as one would expect from the strong cross-sectional dependence between the different comprehensive revisions that we saw before.

Precision gains from using all releases

- We compute the root mean square error (RMSE) of the GDP growth rates 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 and a quarter years. plot
- As expected, the release of the advance GDE figure has a dramatic effect relative to the RMSE of the prediction of the third quarter growth rate made at the end of September.
- Nevertheless, noticeable precision gains occur when the second and third estimates of GDE and GDI are released.
- Moreoever, there are further precision gains when the annual estimates become available in July of the subsequent years.
- Exactly the same pattern arises if we repeat this exercise for the first and second quarters in April and July, respectively, but not for the fourth quarter, which shows a slightly different initial pattern because there is no second GDI release in February.

Effects of combining all comprehensive revisions

- We have also estimated the single signal version of our model using only the most recent comprehensive revision data.
- Specifically, we have computed 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. plot
- 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.

Solera releases

- We compare the smoothed estimates for US GDP growth from six different solera releases recursively estimated every two years starting in January 2012 to provide estimates up to 2011Q4 and ending in January 2022 for 2021Q4.
- All estimates display close paths until 2010Q1. ▸ plot
- Still, the growth rates estimates for the last few quarters of each series are somewhat different from the corresponding estimates in the next ones because of the smoothing embedded in our filtering algorithm, which systematically reassesses the past after observing the future.
- The two most recent solera releases also present a different pattern in the second quarters of 2011 and 2012 because they incorporate modifications of the GDP definition resulting from the comprehensive revision the BEA released in July of 2018.
- The post-pandemic estimates for the pre-pandemic period are remarkably stable to the inclusion of the 2020 outliers. ▸ plot

Empirical results: Specific episodes

- We can also perform recursive **real time estimation** to study the impact of new data on *GDPsolera* estimates at a fixed date t.
- We do so for three specific dates: 2001Q1, 2008Q4 and 2019Q2.
- According to the NBER, March 2001 marked the peak of a ten year expansion. However, Republican politicians at the time claimed that it had finished earlier so in 2001 there was a “Clinton recession”. ▸ plot
- The Great Recession largely owes its adjective to the depth of the fall in 2008Q4, and yet, it did not seem so great to begin with. ▸ plot
- At the end of 2018, some Fed officials were concerned about a deceleration in growth after almost another decade of sustained growth, but the expansion only came to an end in 2020Q2 for a completely different reason. ▸ plot

Comparison to *GDPplus*

- We finally compare our measure of economic activity with the *GDPplus* released on a monthly basis by the Philadelphia Fed since August 2013.
- First, we look at the smoothed estimates of GDP between the first quarter of 1985 and the fourth quarter of 2021 using in both cases all the data released by the BEA by the end of January 2022. plot
- The two series 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.
- Nevertheless, our estimates have a 40% larger standard deviation.
- The smoothness of *GDPplus* results in relatively conservative estimates of the large fall and rise of economic activity after the start of the COVID 19 outbreak.

Comparison to *GDPplus*

- Finally, 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. ▸ plot
- 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 "third" estimates of GDE and GDI released by the BEA. ▸ plot
- These real time *GDP solera* and *GDPplus* estimates appear to be more similar than the historical ones we saw before.
- Still, we can observe a few differences in the first two quarters of 2015 affected by the 2018 comprehensive revision, and at the end of the sample, starting after the 2020Q2 drop.

Directions for further research

- Combining our approach with high frequency data.
- Assessing the effect of incorporating the seasonally unadjusted GDE and GDI data that the BEA has released since 2018 to our empirical results.
- Investigating the potential forecasting improvements of the model we propose in this paper for the early releases of GDE and GDI, as they would provide an external validity check on our modelling approach.
- Allowing 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.
- Considering 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.

Why "Solera"?

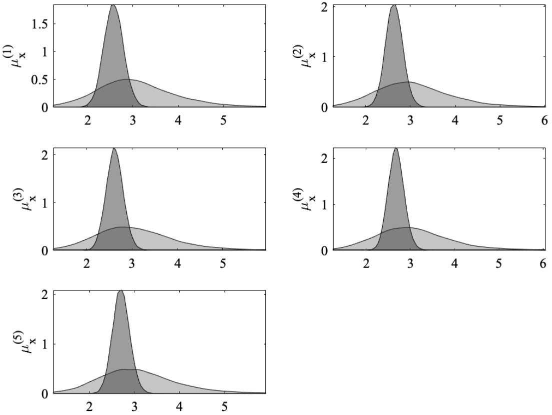
- The recurrent updating of our signal-extraction process is somewhat analogous to the *criaderas* and *soleras* system of sherry wine aging, whereby the final product is obtained by fractional blending inputs from different vintages over a perennial dynamic procedure that gives sherry its distinctive character. ▸ pic

- The *solera* system was explained by agent 007 to M in the 1971 James Bond film *Diamonds are forever*.

▸ back

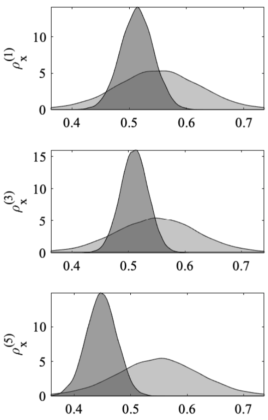
Top barrels are filled with the new vintages and wine is taken from the bottom ones

back



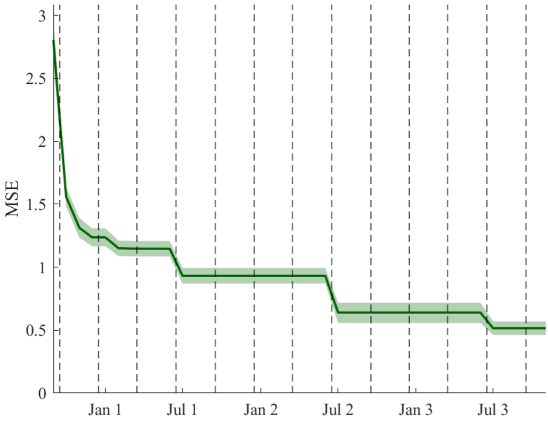
Unconditional growth rates for the five signals

back



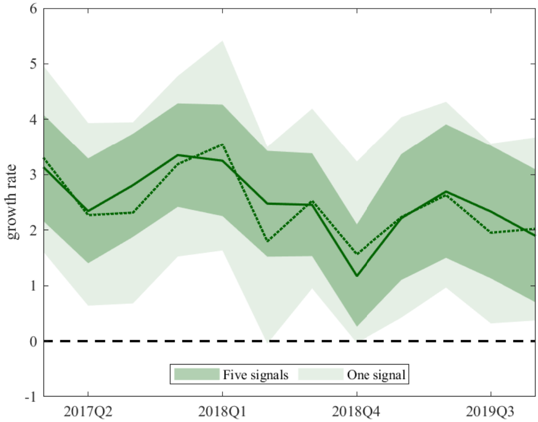
Autocorrelation coefficients for the five signals

back



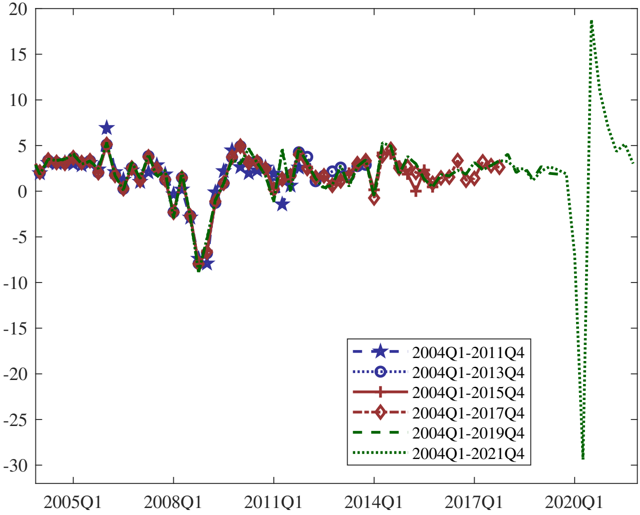
RMSE reductions from using all vintages

back



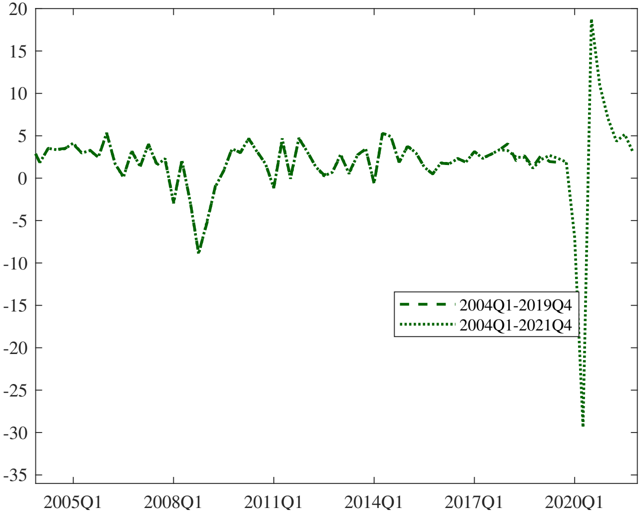
The solid line is the median of Δx_{Ct} given $y_{1:T}$ when all signals are used while the dashed line refers to its conditional median when only the most recent signal is used. Shaded areas represent t-wise 90%-probability intervals.

back



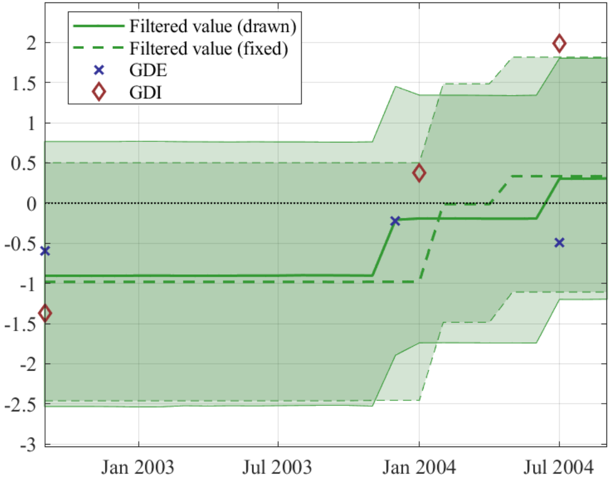
Comparison of some GDP solera historical releases

back



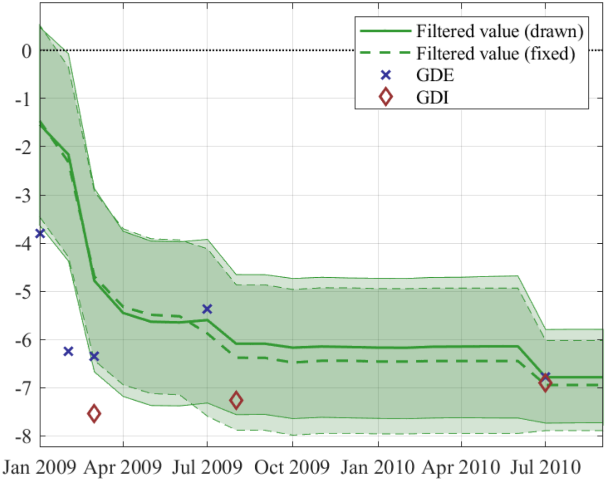
Comparison of recent GDP solera historical releases

back



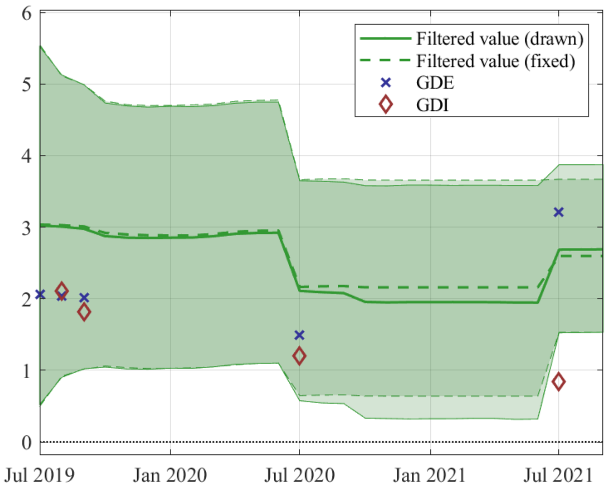
Real time filtering of $\Delta x_{2001\text{Q}1}$. Solid lines are posterior medians and the shaded areas 90%-pointwise credible bands.

▸ back



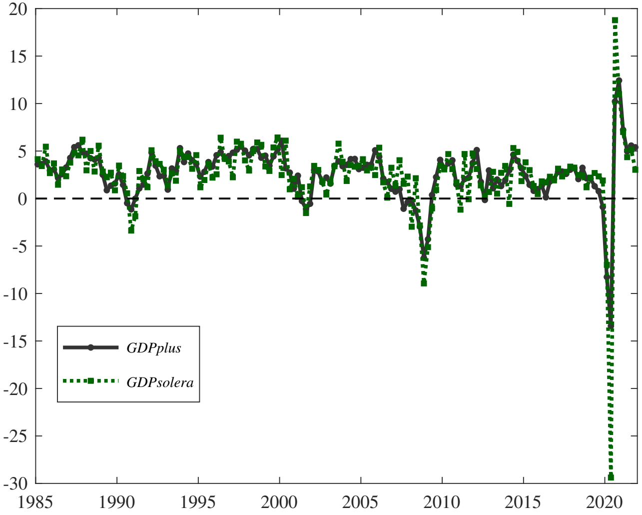
Real time filtering of Δx_{2008Q4}. Solid lines are posterior medians and the shaded areas 90%-pointwise credible bands.

▸ back



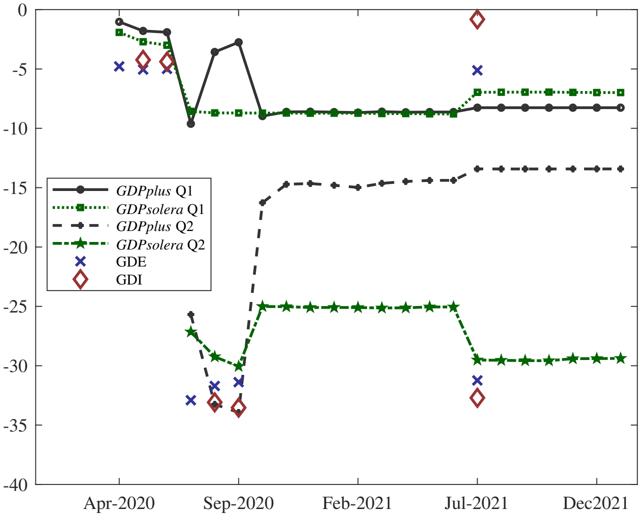
Real time filtering of Δx_{2019Q2}. Solid lines are posterior medians and the shaded areas 90%-pointwise credible bands.

▸ back



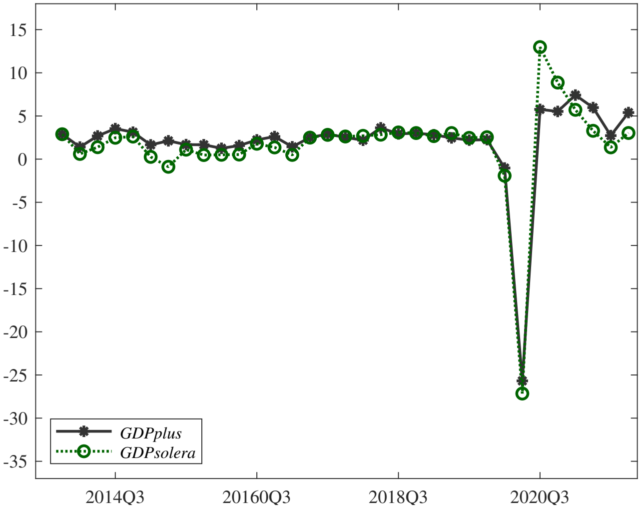
Comparison of most recent historical estimates

back



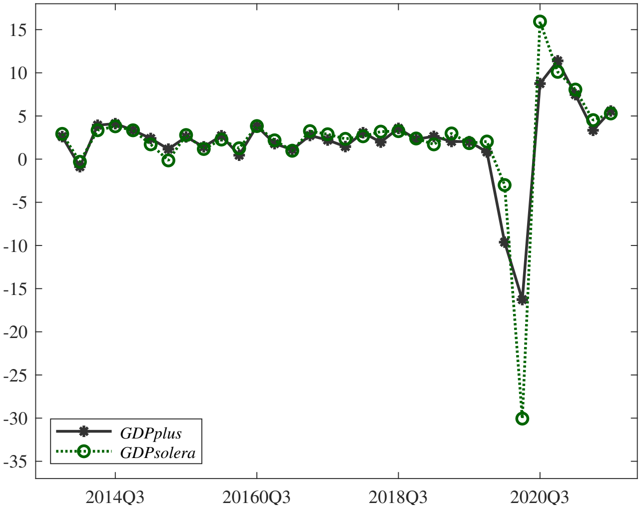
First two quarters of 2020 estimates

▸ back



GDP growth estimates one month after the end of the quarter

back



GDP growth estimates three months after the end of the quarter