JDEMETRA+ Nowcasting[1]

Macroeconomic Monitoring and Visualizing News

NATIONAL BANK OF BELGIUM
Research & Development

Philippe Charles
David De Antonio Liedo[2]
Mats Maggi
Jean Palate

Abstract. This article presents the first open source IT solution for nowcasting and reading news with dynamic factor models. As illustrated in our workhorse example, the software allows us to extend the limits of currently established practices. The *nowcasting* model proposed for the US economy is, to the best of our knowledge, the first one that accounts for the joint behavior of quantities and prices. The model also provides a join interpretation of the forecast revisions for multiple horizons in terms of the unexpected part of both new data releases and *revisions* to past data, which become available in real time. For instance, a worse than expected inflation release will have an impact on the forecasting updates for GDP, but the sign of that impact will depend on the remaining news too. The reason is that we can have both positive supply and negative demand disturbances underlying the bad surprise in inflation data.

Keywords: JDEMETRA+, GDP and inflation interactions, state-space, business cycles, timeliness, data revisions, real-time, forecasting

JEL: C87, E31, E32, E37

---

[1] Licensed under the EUPL (http://ec.europa.eu/idabc/eupl).
The last updated version of the JD+ software, which has been designed for the analysis of seasonal data, can be downloaded here:
http://www.cros-portal.eu/content/jdemetra.
The *nowcasting* tool is distributed as a plug-in and resides in the web site of the National Bank of Belgium: http://www.nbb.be/app/dqrd/jdemetra/jdplugins-1.5.3.zip

[2] Corresponding author. Email: david.deantonioliedo@nbb.be. 1000 Boulevard de Berlaimont 14, 1000 Bruxelles, Belgium. This work has benefited from discussions with Raf Wouters, Geert Langenus, Marta Banbura, and seminar participants at the NBB and the CFE 2014 in Pisa. All errors are our own responsibility.

1 Introduction

The meteorological term nowcasting has become increasingly popular in economics over the last years. Unlike nowcasting users in meteorology, who base their decisions on the current weather along with forecasts for a period of zero to six hours ahead, institutions responsible for economic policy need to make important decisions without directly observing the current state of the economy.

The first papers that formalize the real-time forecasting process are due to Evans (2005) and Giannone et al. (2008). They use the term nowcasts to refer to predictions of the most recent past, the present, and the nearest future. In this paper, we describe a set of open source modules for nowcasting and analyzing news with dynamic factor models, as described by Banbura and Modugno (2010). These modules are integrated in the JDEMETRA+ framework developed at the National Bank of Belgium.

Our nowcasting library[3] aims to simplify the practice of forecasting in real-time by helping analysts and researchers to communicate their forecasts revisions in real time. Moreover, users are able to save the models and data vintages and share them with other analysts. Our implementation is based on two pillars. First, users can easily specify and estimate a broad range of dynamic factor models, which are internally casted in state-space form and estimated via maximum likelihood. For this purpose, we combine the EM algorithm proposed by Banbura and Modugno (2010) with numerical optimization methods. The second pillar of our implementation is a user friendly graphical interface that represents in a transparent manner how the model based expectations for all variables change as a consequence of news embedded in different data releases. This feature simplifies the use and interpretation of the model in real-time forecasting applications regardless its complexity.

Understanding the role of news is a crucial point for analysts and policy institutions that produce forecast and are requested to explain modifications in their assessment on the basis of new information that becomes available. The nowcasting literature has aimed since its origins to understand the impact of data releases in forecasts for economic growth. Since the work by Giannone et al. (2008), much of the empirical research has drawn conclusions on the usefulness of monthly surveys at nowcasting economic growth without being able to quantify their precise role. In the framework of a small model with only one factor, Camacho and Pérez-Quirós (2010) suggested that the importance of a given data release could be measured by how much it contributes in the estimation of the driving factor, following the contributions analysis

[3] JDEMETRA+ is Open Source time series software written in Java. It is mostly used in statistical agencies for the analysis of seasonality (X12, TRAMO-SEATS, or structural models). This article focuses on the new Nowcasting library, which currently supports the use dynamic factor models.

of Harvey and Koopman (2003). However, the formal analysis of news in the nowcasting context was brought by Banbura and Modugno (2010) on the basis of concepts originated in control engineering. Their approach to handle this problem has been incorporated in the JDEMETRA+ module for news along with an interactive graphical interface that allows the user to assess the contribution of all data releases (including data revisions) at updating the expected evolution path for all variables in the system. To the best of our knowledge, we are the first ones proposing a feasible method to disentangle the impact of data revisions from the whole set of news.

Finally, the empirical application describes a nowcasting model for real activity growth and inflation in the US. Although the idea has been proposed by Aruoba and Diebold (2010), this paper contains the first implementation of a nowcasting application providing a joint *interpretation* of the fluctuations in prices and quantities[4]. This example serves well to illustrate that the JDEMETRA+ nowcasting platform aims to extend the limits of the currently established practices and promote an exchange of ideas. Users can easily introduce their own expertise in the form of models, which can be saved along with the data vintages that are available in real time. Through this platform, researchers will also be able to easily share their forecasting knowledge and explore alternative datasets or model specifications.

Overall, this implies forecasters with different backgrounds, independently of their technical expertise, can avoid dealing with complex algorithms and spend more of their valuable time trying to better understand the large amount of information at their disposal. This new technology can help to revisit classic applications such as studying the role of financial markets data at forecasting inflation, e.g. see Stock and Watson (2003) for a survey. Our choice to publish this software as an Open Source solution also establishes the basis for cooperation with external participants, which could lead to significant improvements in the library or the implementation of alternative methods that are common in empirical macroeconomics and finance, such as Bayesian VARs with mixed frequency data (e.g. Schorfheide and Song, 2014), models with Markov-Switching parameters or stochastic volatility. Another topic that would be particularly relevant to statistical agencies would be the introduction of techniques such as temporal disaggregation, as in Frale et al. (2011), who propose to convert quarterly national accounts into monthly frequency.

[4] The model by Banbura et al. (2011) specifies a block of inflation variables linked to an "inflation specific" factor which is assumed to be independent from real quantities. Thus, their framework does not take into account the interactions between prices and quantities.

2 Basic Model Specification

Consider the following representation of the multivariate process $\{ y_t \}$ governing the dynamics of a given set of *N* observables over time, e.g. months, $t = 1, \ldots, T$:

$$y_t = Z\alpha_t + \xi_t \qquad with \;\; \xi_t \sim N(0, R) \qquad [1]$$

$$\alpha_t = T_1\alpha_{t-1} + \cdots + T_p\alpha_{t-p} + u_t \quad with \;\; u_t \sim N(0, Q) \qquad [2]$$

The first equation represents the observables as a function of the vector of k unobserved factors α_t plus a vector of idiosyncratic measurement errors ξ_t. The second equation defines the dynamics of the factors, which follow a covariance stationary VAR process of order *p*. Thus, the dynamic interactions between all the factors are given by the matrices $T_1, \ldots, T_p$ and the underlying shocks u_t, which are uncorrelated with the measurement errors. Interestingly, the idiosyncratic nature of the measurement errors, whose covariance matrix R is diagonal, implies that all the co-movements in the data can be accounted for by fluctuations in the latent factors.

More parsimonious parameterizations are possible, as it will be clarified in the empirical application. For example one can assume that a given factor does not load on specific variables, i.e. the matrix of loadings will have zeroes in the columns corresponding to that factor. Alternative parameterizations are also possible. Consider for instance the possibility that the number of stochastic terms in u_t is smaller than the size k of the state vector, suggesting a reduced number of shocks spreading throughout the economy. In this case, we would need to parameterize Q and T in terms of the actual number of shocks, *q*.

Concept	Size	Definition
$\mathbf{y_t}$	$\mathbf{N \times 1}$	Observed data
$\boldsymbol{\alpha_t}$	$\mathbf{k \times 1}$	Underlying factors
$\boldsymbol{\xi_t}$	$\mathbf{N \times 1}$	Measurement Error
$\mathbf{u_t}$	$\mathbf{k \times 1}$	Shocks to the factors
Z	$\mathbf{N \times k}$	Loadings
$\mathbf{T_1, \ldots, T_p}$	$\mathbf{k \times k}$	VAR coefficients in the motion equation for the factors
R	$\mathbf{N \times N}$	Covariance of the measurement error
Q	$\mathbf{k \times k}$	Covariance of the shocks to the factors

3 Estimation, Forecasting and Analysis of News

The data vector y_t may contain missing observations simply because macroeconomic indicators are not necessarily released at the same time. In the case that the model is specified at the monthly frequency, the presence of quarterly data also generates additional missing observations. In this context, which will be carefully described below, a variable such as GDP is of course only available for each quarter and not for each month. This variable, which aims to measure the flow of economic activity over a quarter, is observed every three months, i.e. y_t^{GDP} is missing for $t \neq 3,6,9,12, \ldots$ The standard Kalman filter can handle this complication and evaluate the likelihood via prediction-error decomposition. Estimating the parameter values that maximize the likelihood is conceptually simple, but complex models may require a heavy use of the optimization capabilities implemented in JDEMETRA+. Once the model has been specified and estimated, the so-called Kalman smoother is used to calculate projections, as highlighted for example by Durbin and Koopman (2001). The projections for any variable k at any point in time t conditional on the information set available at any date v will be represented throughout this document as $E[y_{k,t} \mid \mathcal{F}_v]$

The practice of forecasting in real time requires incorporating changes in the information set that involve news. The Kalman filter can also be used in this context to formalize the impact of news in the forecast revisions. Thus, updates in forecasts can be expressed as a linear combination of the news identified by the model (see equations [1] and [2]):

$$E[y_{k,t} \mid \mathcal{F}_{\text{updated}}] - E[y_{k,t} \mid \mathcal{F}_{\text{old}}] = \sum_{j=1}^{J} w_j^{k,t} \left(y_{i_j,t_j} - E\left[y_{i_j,t_j} \middle| \mathcal{F}_{\text{old}} \right] \right) \qquad [\ 3]$$

where the weights $w_j^{k,t}$ associated to each one of the J news is specific to each variable k and time period t. The news itself is defined by the difference between the released indicator and its expected value conditional on the previous information set $\mathcal{F}_{\text{old}}$. Banbura and Modugno (2012) explain in detail how to compute these weights. As opposed to the solution proposed by Harvey and Koopman (2003) to understand the impact of the variables in the forecasts[5], the JDEMETRA+ implementation focuses on the concept of news, which also fits within the state-space modeling framework.

[5] Harvey and Koopman (2002)'s insight has been followed in nowcasting applications by Banbura and Rünstler (2011) and Camacho and Pérez-Quirós (2010).

4 Advanced Specification Options

Concept	Representation of the link to the factors	JDEMETRA+ code
monthly growth	$\boldsymbol{\alpha_t}$	M
quarterly growth	$\alpha_t^Q \cong (\frac{1}{3}\alpha_t + \frac{2}{3}\alpha_{t-1} + \alpha_{t-2} + \frac{2}{3}\alpha_{t-3} + \frac{1}{3}\alpha_{t-4})$	Q
year-on-year growth	$\alpha_t^Y = \alpha_t + \alpha_{t-1} + \alpha_{t-2} + \alpha_{t-3} + \alpha_{t-4} + \alpha_{t-5} + \alpha_{t-6} + \alpha_{t-7} + \alpha_{t-8} + \alpha_{t-9} + \alpha_{t-10} + \alpha_{t-11}$	YoY

4.1 Measurement Equation Type "M"

All variables measuring the monthly growth rate of the economy will be directly linked to the factors α_t , exactly as specified in the measurement equation [1]:

$$y_t = Z\alpha_t + \xi_t$$

The next two subsections explain how to introduce variables expressed in terms of percentage changes both with respect to the previous quarter and with respect to the same month of the previous year.

4.2 Measurement Equation Type "Q"

This option will be very familiar to practitioners that exploit monthly indicators of economic activity as regression variables to obtain a nowcast for GDP, which is typically expressed in terms of growth rates over a whole quarter. Thus, the measurement equation [2], which relates all monthly indicators with a latent factor representing the underlying monthly growth rate of the economy, cannot be used. The measurement equation for variables measuring quarterly growth rates will be defined as follows:

$$y_t^{GDP} = Z_{GDP}\alpha_t^Q + \xi_t^{GDP},$$

where α_t^Q represents the underlying quarterly growth rate of the economy. A linear approximation can be used under the assumption that one third of the economic flow of added value registered over the quarter is given by the geometric average over the three months. Thus, the approximation

$$\alpha_t^Q \cong (\frac{1}{3}\alpha_t + \frac{2}{3}\alpha_{t-1} + \alpha_{t-2} + \frac{2}{3}\alpha_{t-3} + \frac{1}{3}\alpha_{t-4})$$

allows us to rewrite the measurement equation above as a linear combination of the underlying monthly growth rates:

$$y_t^{GDP} = Z_{GDP} \frac{1}{3} (\alpha_t + 2\alpha_{t-1} + 3\alpha_{t-2} + 2\alpha_{t-3} + \alpha_{t-4}) + \xi_t^{GDP}$$

Albeit may seem to be a complex representation even in the simplest case where α_t contains a single factor, the only parameters to be estimated in the measurement equation for GDP are the loading Z_{GDP} and the variance of the measurement error component ξ_t^{GDP}. The fact that y_t^{GDP} is available only every three months, i.e. missing for $t \neq 3,6,9,12, \dots$ does not pose any technical difficulty thanks to the availability of alternative indicators.

Further Remarks

In contrast to our state-space approach where the full system is modeled simultaneously, the commonly used bridge equations[6] for nowcasting involves a simple regression of the quarterly growth rate of GDP, y_t^{GDP}, on aggregated values of the higher frequency indicators, i.e. α_t^Q would be replaced by the quarterly growth rate of a predictor variable. The drawback is that some of that higher frequency data may not be available for the whole quarter, which requires the use of auxiliary forecasting models to fill-in the gaps. This clearly adds an extra layer of complexity to the methodology. The solution of using bridge equations to calculate direct forecasts without the need to use auxiliary models for the indicators would be also problematic because the resulting forecasting equation will change depending on the information set available. This is an undesirable feature because this implies that updates in the forecasts are driven by both new data releases and changes in the model. Thus, an analysis of news such as the one I will present in Section 8 is not possible in the context of bridge equations. Finally, inside the broad class of bridge equation models for nowcasting, the so-called MIDAS approach introduced by Ghysels, Santa-Clara and Valkanov (2004) and its multiple parameterizations[7] provide a parsimonious solution to link our target low frequency variable with indicators released at a higher frequency. The MIDAS aggregation for monthly data is also consistent with the aggregation scheme presented above to link a factor representing economic growth over the quarter with monthly factors, although the original MIDAS formulation involves explanatory variables and not unobserved factors[8]. The key advantage of MIDAS is that the aggregation can be implemented in a much more flexible way. The drawback of this univariate approach is that the large-scale analysis of news illustrated in this article would not be valid anymore.

[6] See for example Ingentino and Trehan (1996), Baffigi, Golinelli and Parigi (2004) or Diron (2008).

[7] See the survey by Andreou, Ghysels and Kourtellos (2010) for an overview.

[8] See Marcellino and Schumacher (2010) for an example where the MIDAS regressions are augmented with factors extracted from large cross-section.

4.3 Measurement Equation Type "YoY"

Some variables, such as the results of many surveys that represent how business or consumer sentiments have changed with respect to one year ago, are likely to have a strong correlation with year-on-year growth rates. In our framework, this has the implication that they should be linked to the cumulative sum of the factors for twelve months, i.e. year-on-year growth rate of the factor, as in the nowcasting model developed by Camacho and Pérez-Quirós (2010). The European Commission (2006) explicitly states that the guiding principle for the selection of questions in the different surveys is the aim to achieve as high as possible coincident correlation of the confidence indicators with year-on-year growth of the reference series. De Greef and Van Nieuwenhuyze (2009) also emphasize the coincident correlation of the NBB Business Survey with year-on-year GDP growth rates. The measurement equation for variables measuring year-on-year growth rates will be defined as follows:

$$y_t^{Survey} = Z_{Survey}\alpha_t^{YoY} + \xi_t^{Survey}$$

where

$$\alpha_t^{YoY} = \alpha_t + \alpha_{t-1} + \alpha_{t-2} + \alpha_{t-3} + \alpha_{t-4} + \alpha_{t-5} + \alpha_{t-6} + \alpha_{t-7} + \alpha_{t-8} + \alpha_{t-9} + \alpha_{t-10} + \alpha_{t-11}$$

As in the previous case, measurement equations of this type increase the complexity of the model without the need to estimate more parameters. The only parameter that needs to be estimated is the factor loading Z_{Survey} and the variance of the measurement error.

Further Remarks

Variables concerning expectations about the future are useful because they may contain relevant information about the current unobserved state, i.e. factors, of the economy. However, rather than linking it to the *current* factors, one could consider the possibility to link it directly to the *expected* state. This option is not currently available in JDEMETRA+, but it could be a feasible extension.

Consider a measure of inflation expectations for the next twelve months, $\pi_{t|t+12}^{Survey}$.

This variable should be linked to the latent inflation factor as follows:

$$\pi_{t+12|t}^{Survey} = Z_{Survey}\, E[\alpha_{t+12}^{YoY} | \alpha_t] + \xi_{t+12}^{Survey} \quad [4]$$

where α_{t+12}^{YoY} is defined as the year-on-year growth rate of the factor:

$$\alpha_{t+12}^{YoY} = \alpha_{t+12} + \alpha_{t+11} + \alpha_{t+10} + \alpha_{t+9} + \alpha_{t+8} + \alpha_{t+7} + \alpha_{t+6} + \alpha_{t+5} + \alpha_{t+4} + \alpha_{t+3} + \alpha_{t+2} + \alpha_{t+1}$$

Without loss of generality, assuming that the transition equation in the state-space model [2] is a first order VAR we obtain a very simple expression for the conditional expectation required to specify measurement equation [4]:

$$E[\alpha_{t+h} | \alpha_t] = T^h \alpha_t \; => \; E[\alpha_{t+12}^{YoY} | \alpha_t] = (T + T^2 + \cdots + T^{12})\, \alpha_t$$

This implies the measurement equation should be specified as follows:

$$\pi_{t+12|t}^{Survey} = Z_{Survey} (T + T^2 + \cdots + T^{12}) \; \alpha_t + \xi_{t+12}^{Survey} \qquad [5]$$

Modeling expectations correctly helps to better estimate T, which determines the transition dynamics of the whole economic system. Alternatively, it would also be possible to estimate the reduced form expression

$$\pi_{t+12|t}^{Survey} = Z_{Survey}^{*} \; \alpha_t + \xi_{t+12}^{Survey},$$

where the state α_t loads on Z_{Survey}^{*} and not on Z_{Survey}, defined above. If you consider for simplicity that α_t and $E[\alpha_{t+12}^{Y} | \alpha_t]$ where observed variables, both loadings would be linked, i.e. $Z_{Survey}^{*} = Z_{Survey} (T + T^2 + \cdots + T^{12})$, and the fit for $\pi_{t+12|t}^{Survey}$ would be exactly the same in both equations [4] and [5]. However, the motivation for introduction data on expectations is the suspicion that they add fundamental information about the state of the economy that would be otherwise missed by the model.. Because data on expectations is relevant for policy, we would propose to specify the measurement equation as in [5]. Ideally, expectations for any horizon *h* could be incorporated in the model's information set. This would be possible with an interface that allows the user to select the measurement type "YoY" together with the option link the variable with expectations for a given forecast horizon, *h*, which has been set equal to 12 in this case.

Table 1: Data Selection

Real Activity	Inflation
Real Gross Domestic Product	Consumer Price Index for All Urban Consumers: All Items
U.S. Department of Commerce: Bureau of Economic Analysis	U.S. Department of Labor: Bureau of Labor Statistics
Quarterly; start: 1947Q1	Monthly; start: January 1947
"Advance" and "Prelimimary" releases available 30 and 60 days, respectively, after the end of the quarter. "Final" release available with a delay of 85 days, but subsequent revisions may be large.	*Released about 16 days after the month ends*
Real personal income excluding current transfer receipts	Producer Price Index: Finished Goods
U.S. Department of Commerce: Bureau of Economic Analysis	U.S. Department of Labor: Bureau of Labor Statistics
Monthly; start: January 1959	Monthly; start: April 1947
Released around 30 days after the end of the month	*Released about 14 days after the month ends*
All Employees: Total nonfarm	Gross Domestic Product: Implicit Price Deflator
U.S. Department of Labor: Bureau of Labor Statistics	U.S. Department of Commerce: Bureau of Economic Analysis
Monthly; start: January 1939	Quarterly; start: 1947Q1
Released on the first Friday after the month ends	*"Advance" and "Prelimimary" releases available 30 and 60 days, respectively, after the end of the quarter. "Final" release available with a delay of 85 days.*
Industrial Production Index	Nonfarm Business Sector: Compensation Per Hour
Board of Governors of the Federal Reserve System	U.S. Department of Labor: Bureau of Labor Statistics
Monthly; start: January 1919	Quarterly; start: 1947Q1
Available around 16 days after the end of the month	*First release 30 days after the end of the quarter; Second release 60 days after the end of the quarter (subsequent revisions may be significant)*
Real Manufacturing and Trade Industries Sales	Crude Oil Prices: West Texas Intermediate (WTI) - Cushing, Oklahoma
Federal Reserve Bank of St. Louis	U.S. Department of Energy: Energy Information Administration
Monthly; start: January 1967	Monthly; start: January 1986
Available about 60 days after the month ends	*Available daily*
Initial Claims	S&P GSCI Non-Energy Spot - PRICE INDEX
U.S. Department of Labor: Employment and Training Administration Monthly; start: January 1967	Monthly; start: December 1969
Released four days after the week ends	*Available daily*

5 A Joint Model for Nowcasting US inflation and GDP

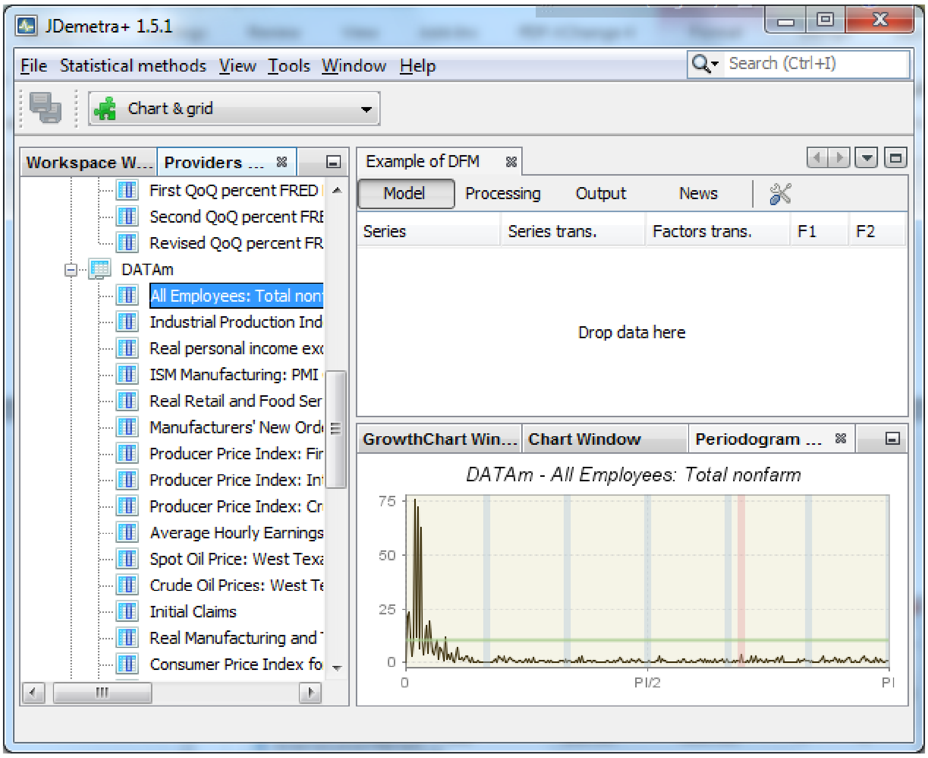
Aruoba and Diebold (henceforth A&D) (2010) have proposed to extract monthly factors from two data sets containing macroeconomic prices and quantities, respectively, and discuss the interactions between inflation and real activity. Instead of extracting those factors separately from the two different datasets, as they propose, we will show how JDEMETRA+ can be used to define the interaction between real activity and inflation within a state-space model with two factors. This approach is more suitable to formalize the discussion of whether supply or demand shocks are behind a given recession and quantify the historical contribution of supply shocks at different stages of the business cycle. Table 1 presents the selection of variables made by A&D, which contains some of the most monitored real activity and inflation indicators at the monthly frequency.

5.1 Model Specification to Extract the Latent Real Activity Growth Rate

A&D use the indicators in the left panel of Table 1 to extract a smooth index of real economic activity. Those indicators are also used by the so-called ADS Index published by the Federal Reserve Bank of Philadelphia, which uses a simplified implementation of the model proposed by Aruoba, Diebold and Scotti (2009). The model used by A&D to extract that signal from the monthly data differs from the state-space representation originally proposed in JDemetra+ not only because their measurement equation contains lags of the observed variables, but also because their approach requires the idiosyncratic errors of quarterly variables to be explicitly disaggregated in terms of the monthly measurement errors, even if they are unobserved (see Mariano and Murasawa, 2003). Nevertheless, our approach is not fundamentally different, since JDEMETRA+ also aims to represent GDP growth as a weighted average of latent factors, i.e. code “Q”, as described in section 4.3. For the rest of the variables, we use the code “M”, which implies that the growth rate of each one of the indicators loads contemporaneously on the factor representing the latent growth rate of the economy.

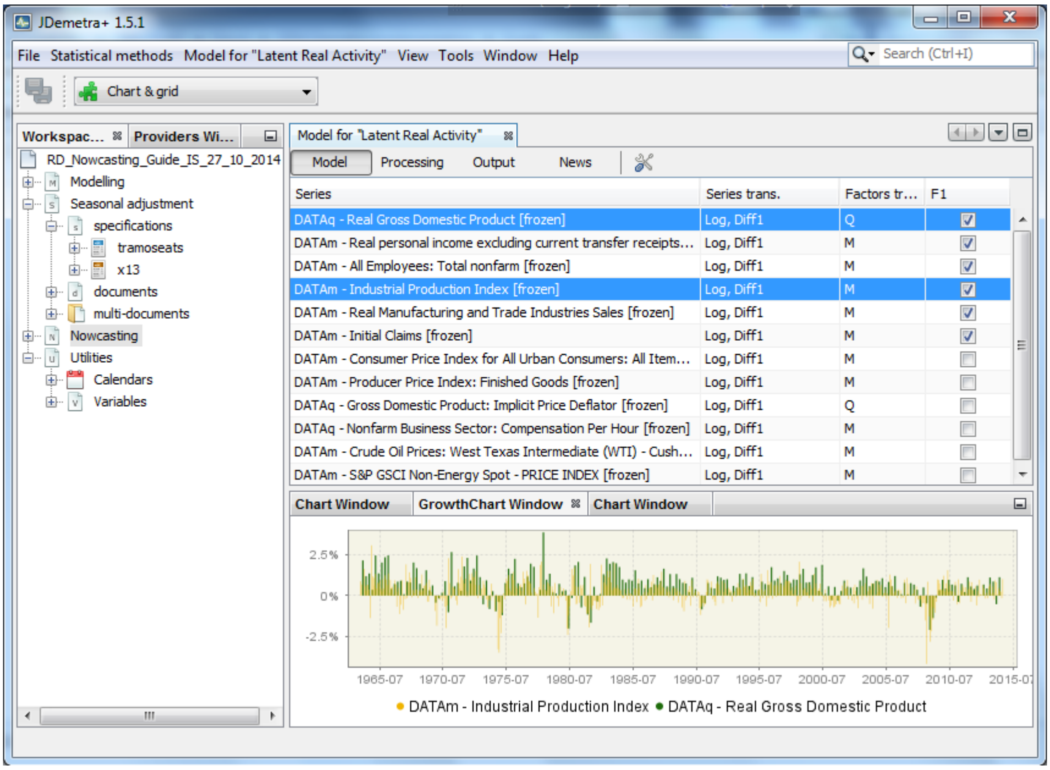
The user simply needs drag and drop the series, as shown in Figure 1. Then, one should simply tick those to be incorporated in the model, and transform them so that they are consistent with the way they are linked to the factors, as explained in Section 4, i.e. “Q”, “M” or “YoY”. Figures 2 (a, b and c) below describes three simple specifications of our measurement equation. In the case of Figure 2(a), we have specified only one factor, and all series transformed in log differences “Log, Diff1”. The series can also be seasonally adjusted by simply adding the option “sa” before or after the “Log” instruction or converted in year-on-year growth rates by using the option “Log, DiffY”. Figure 3 describes the specification of the transition equation. By clicking on the icon highlighted in the screenshot below, users will choose the number of factors here (i.e. “Equations count=1”) and the number of lags in the VAR process that defines the transition equation. As in A&D, we choose three lags (i.e. “Lags count=3”).

Figure 1: Drag and Drop the Data into the Model Space



The “Providers tab” allows you to access the source data so that you can drag and drop it into the model space. Before selecting the series to be incorporated in the model, one can perform some basic analysis. For example, by using the “GrowthChart” one can compare the monthly or yearly growth rate of different time series even if they do not have the same frequency. The “Periodogram” window can also be used to have an overview of the cyclical properties of the data and check that the series have been correctly adjusted for seasonality and calendar effects (frequencies corresponding to the blue and red shadow, respectively).

Figure 2 (a): Model Specification to extract “Latent Real Activity”

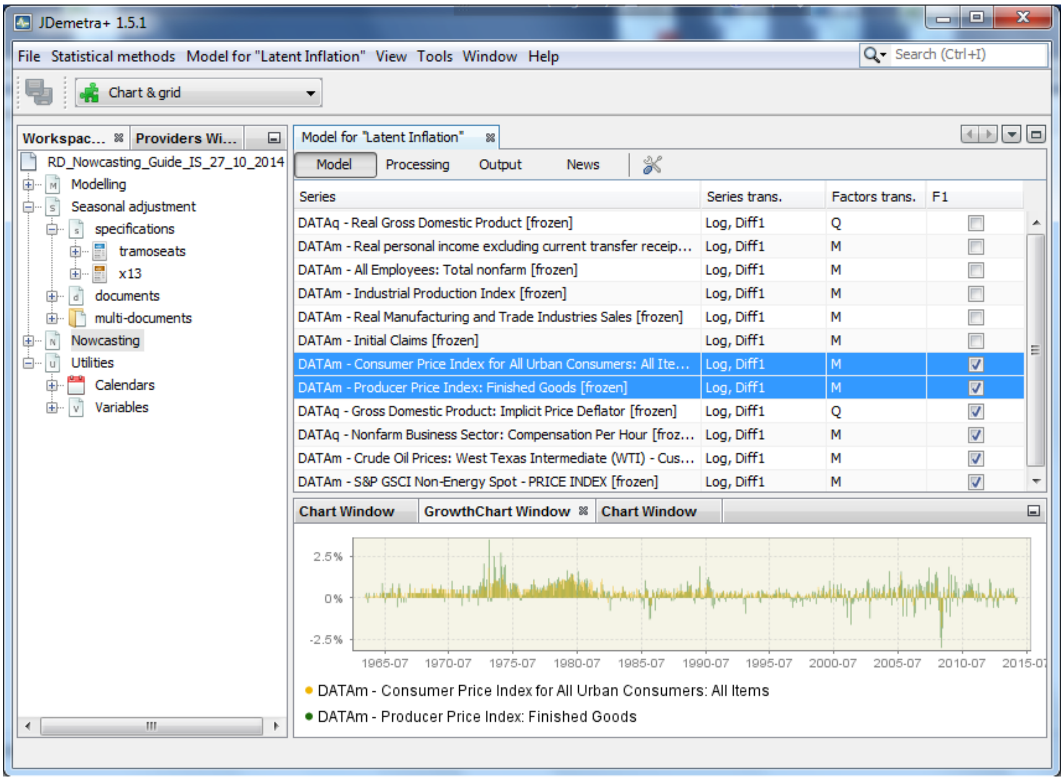


Two steps are required to specify the measurement equation. The first step to extract the “latent real activity” factor is to select all measures of real growth that are associated to it. Next, all series can be transformed into growth rates (Log, Diff1), so that the factor actually represents the latent growth rate for each month. Thus, we need to make sure that each series is correctly linked to either the monthly growth rate of the factors (code “M”) or the quarterly rate (code “Q”), depending on whether it refers to a month or to the whole quarter. The code “YoY” for the factors can be used to represent year-on-year growth rates of the series (Log, DiffY), although that was not necessary in this example. The series can also be seasonally adjusted by simply adding the option “sa” before or after the “Log” code.

5.2 Model Specification to Extract the Latent Inflation Rate

Following A&D, we use the indicators in the right panel of Table 1 to extract a smooth index of inflation. Figure 2(b) shows that all concepts related to inflation have been ticked. All the details regarding the specification of the measurement and transition equations have been described in the previous subsection, 5.1.

Figure 2 (b): Model Specification to extract "Latent Inflation"



Two steps are required to specify the measurement equation. The first step to extract the "latent inflation" factor is to select all measures of inflation that are associated to it. Next, all indexes can be transformed into growth rates (Log, Diff1), so that the factor actually represents the latent inflation rate for each month. Thus, we need to make sure that each series is correctly linked to either the monthly growth rate of the factors (code "M") or the quarterly rate (code "Q"), depending on whether it refers to a month or to the whole quarter. The code "YoY" for the factors can be used to represent year-on-year growth rates of the series (Log, DiffY), although that was not necessary in this example. The series can also be seasonally adjusted by simply adding the option "sa" before or after the "Log" code.

———

5.3 Simultaneous Extraction of the Activity and Inflation Latent Factors

We will assume that variables measuring real activity, such as real GDP growth, will load on two factors (select "Equations count=2"; Figure 3). The first factor specified in Figure 2(c), "F1", can be interpreted as a deflator (β_t^Q) while the second factor, which is represented by "F2", will be related to nominal activity (α_t^Q)

$$y_t^{GDP} = Z_{GDP}\alpha_t^Q - \Lambda_{GDP}\beta_t^Q + \xi_t^{GDP}, \qquad [5]$$

where the sign of the so-called factor loadings Z_{GDP} and Λ_{GDP} is left unrestricted. Monthly variables such as industrial production, or employment, will be treated in the same fashion:

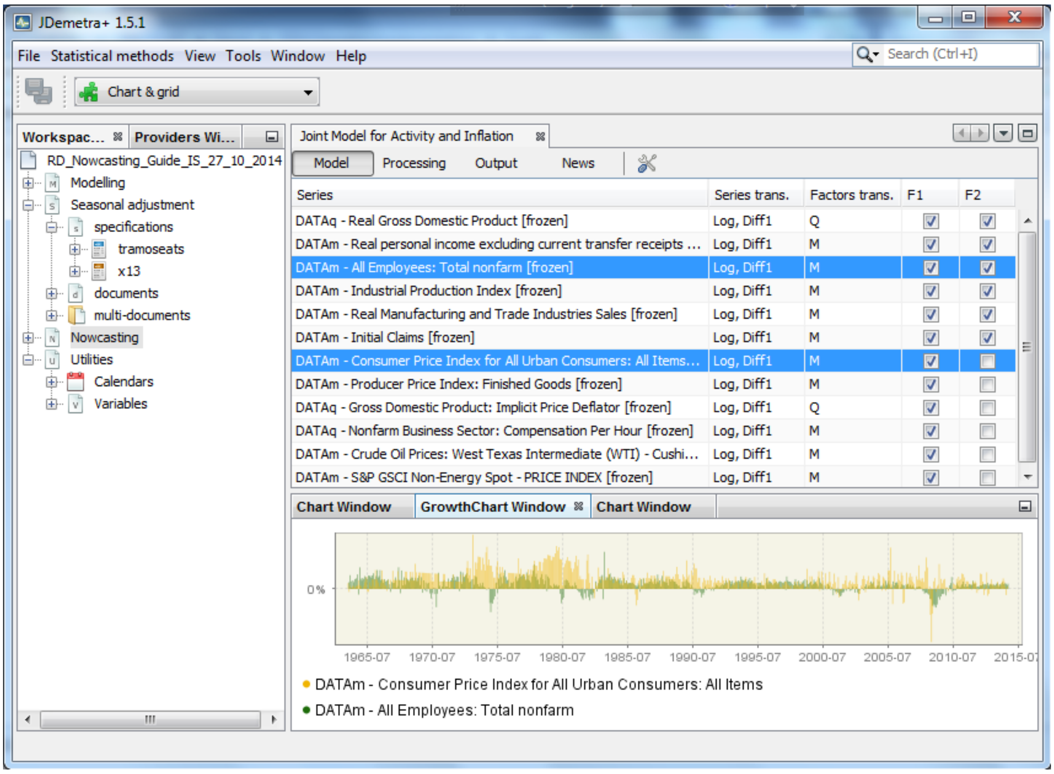
$$y_t = Z\,\alpha_t - \Lambda\,\beta_t + \xi_t \qquad [6]$$

Conversely, all measures of inflation will load exclusively on the second factor:

$$\pi_t = \Lambda_\pi \beta_t + \xi_t^\pi \qquad [7]$$

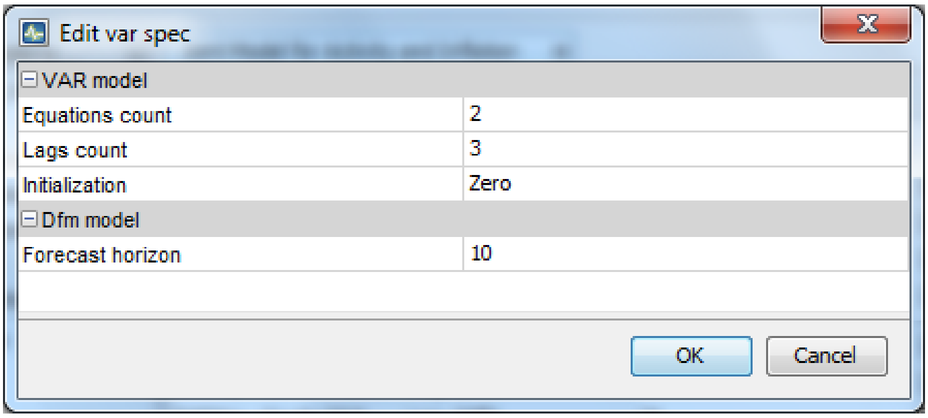
Since both latent factors, α_t and β_t follow an unrestricted VAR and their innovations may be correlated with each other, as specified in equation [2], our restrictions in the factor loadings do not represent a serious constrain. The assumption that β_t, and not α_t, has a contemporaneous effect in inflation simply implies that changes in prices driven by changes in α_t can already be accounted for by the latent inflation factor β_t. The question now is whether we can identify the structural shocks underlying the fluctuations in both factors.

Figure 2 (c): Model Specification to extract "Activity and Inflation Factors"



Two steps are required to specify the measurement equation. The first step to extract the latent "activity growth" and "inflation" factors is to select all measures that are have a associated to them. Next, all series can be transformed into growth rates (Log, Diff1), so that the factor actually represents the latent growth rate for each month. Thus, we need to make sure that each series is correctly linked to either the monthly growth rate of the factors (code "M") or the quarterly rate (code "Q"), depending on whether it refers to a month or to the whole quarter.

Figure 3: Specifying the transition equation



Specifying the transition equation, at the current development stage, only allows for an unrestricted VAR with a certain number of lags.

5.4 Structural Analysis (SVAR)

A structural interpretation requires an orthogonalization of the covariance matrix of the factor innovations Q in the transition equation [2] that is compatible with our definition of supply and demand disturbances. JDEMETRA+ performs this decomposition using a simple Choleski scheme that is consistent with the triangular structure we have imposed on the factor loadings of the measurement equation:

$$\begin{pmatrix} \pi_t \\ y_t \end{pmatrix} = \begin{pmatrix} \Lambda_\pi & 0 \\ \Lambda & Z \end{pmatrix} \begin{pmatrix} \beta_t \\ \alpha_t \end{pmatrix} + \begin{pmatrix} \xi_t^\pi \\ \xi_t \end{pmatrix}$$

We will show that defining demand shocks as the underlying forces that push nominal activity without having a contemporaneous effect on inflation, the resulting supply shocks turn out to generate a negative correlation between output and inflation, which is consistent with common wisdom. Note first that the transition equation for the factors,

$$\begin{pmatrix} \beta_t \\ \alpha_t \end{pmatrix} = \begin{pmatrix} T_{11} & T_{12} \\ T_{21} & T_{22} \end{pmatrix} \begin{pmatrix} \beta_{t-1} \\ \alpha_{t-1} \end{pmatrix} + \begin{pmatrix} u_{\beta,t} \\ u_{\alpha,t} \end{pmatrix},$$

can be written in terms of the “structural” shocks $u^*_{\beta,t}$ and $u^*_{\alpha,t}$, which are now independent and have unitary variance. This alternative VAR representation is obtained by pre-multiplying all terms of the transition equation by the inverse of the Cholesky factor C:

$$\begin{pmatrix} \beta_t^* \\ \alpha_t^* \end{pmatrix} = \begin{pmatrix} C_{11} & 0 \\ C_{21} & C_{22} \end{pmatrix}^{-1} \begin{pmatrix} \beta_{t-1} \\ \alpha_{t-1} \end{pmatrix}$$

$$\begin{pmatrix} u^*_{\beta,t} \\ u^*_{\alpha,t} \end{pmatrix} = \begin{pmatrix} C_{11} & 0 \\ C_{21} & C_{22} \end{pmatrix}^{-1} \begin{pmatrix} u_{\beta,t} \\ u_{\alpha,t} \end{pmatrix}$$

Thus, we obtain:

$$\begin{pmatrix}\beta_t^*\\ \alpha_t^*\end{pmatrix} = \begin{pmatrix}C_{11} & 0\\ C_{21} & C_{22}\end{pmatrix}^{-1}\begin{pmatrix}T_{11} & T_{12}\\ T_{21} & T_{22}\end{pmatrix}\begin{pmatrix}C_{11} & 0\\ C_{21} & C_{22}\end{pmatrix}\begin{pmatrix}\beta_{t-1}^*\\ \alpha_{t-1}^*\end{pmatrix} + \begin{pmatrix}u_{\beta,t}^*\\ u_{\alpha,t}^*\end{pmatrix}.$$

Next, the measurement equation is also written in terms of the transformed factors. It turns out that the Cholesky factorization of the covariance of the reduced form innovations $u_{\beta,t}$ and $u_{\alpha,t}$, i.e. $Q = CC'$, is the only way in which the loadings restrictions can be satisfied:

$$\begin{pmatrix}\pi_t\\ y_t\end{pmatrix} = \begin{pmatrix}\Lambda_\pi & 0\\ \Lambda & Z\end{pmatrix}\begin{pmatrix}C_{11} & 0\\ C_{21} & C_{22}\end{pmatrix}\begin{pmatrix}\beta_t^*\\ \alpha_t^*\end{pmatrix} + \begin{pmatrix}\xi_t^\pi\\ \xi_t\end{pmatrix}$$

By looking at both blocks separately, we can easily show that the structural shock underlying α_t^*, i.e. we will call $u_{\alpha,t}^*$ a demand shock, does not have a contemporaneous effect in inflation variables. Conversely, the shock underlying β_t^*, i.e. we will denote $u_{\beta,t}^*$ as a supply shock, does have an impact in both prices π_t and quantities y_t:

$$\begin{aligned}\pi_t &= & \Lambda_\pi C_{11}\beta_t^* &+ \xi_t^\pi\\ y_t &= Z\,C_{22}\;\alpha_t^* + & (\Lambda C_{11} + ZC_{21})\,\beta_t^* &+ \xi_t\end{aligned}$$

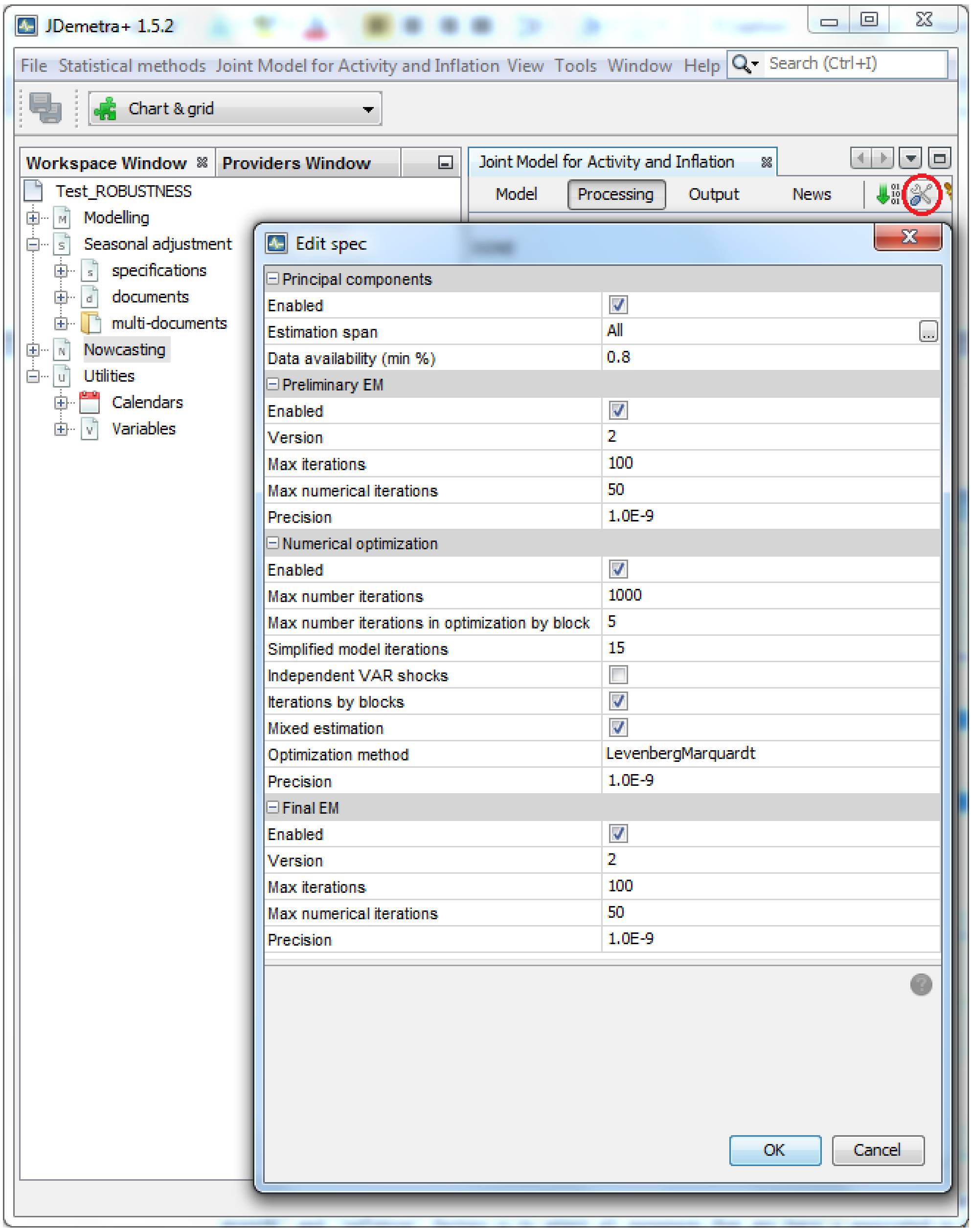
We insist that our naïve identification assumption does not impose any restriction in the sign of the impulse response functions. We will come back to this issue in more detail in the next section where the estimation results will be discussed.

JDEMETRA+ has the potential to incorporate alternative identification schemes, but the current implementation will automatically run the Cholesky decomposition. The structural interpretation will therefore remain valid only when the loadings conform to the kind of triangular structure described above.

5.5 Estimation Process

Once the measurement and the transition equations [1]-[2] of the model have been specified, we need to estimate the factor loadings Z , and the VAR parameters T_1, T_2 and T_3 along with Q, the covariance matrix of the innovations, and the diagonal covariance of the idiosyncratic measurement errors R. The Kalman filter algorithms implemented in JDEMETRA+ can handle the particularities of the nowcasting problem and evaluate the likelihood via prediction-error decomposition, for a given point in the parameter space. The objective of the estimation process is to combine several optimization methods that allow us to find the vector of parameters that maximizes the likelihood.

Figure 4: Estimation options



The method of principal components can be used to obtain an estimator of the factors or simply to have a starting value to initialize either the EM algorithm or the numerical optimization. The EM algorithm itself can also be used either in isolation, as in Banbura and Modugno (2010), or as a starting value for the numerical optimization. This is in our view the most sensible approach, since the EM algirhtm can be very slow in the neibourhood of the maximum likelihood estimator. The numerical optimization can be decomposed in two steps by using the option “Iterations by blocks”, emulating the logic of the EM algorithm. Both the Broyden–Fletcher–Goldfarb–Shanno (BFGS) and Levenberg-Marquardt algorithms are at the user’s disposal. The use of a final EM algorithm is unnecessary.

As shown in Figure 4, we can enable the option to estimate the model parameters by using the principal components of our panel as estimates of the factors, as in Giannone, Reichlin and Small (2008) or Stock and Watson (2002). In order to achieve a higher degree of efficiency, one can use those estimates as starting value for a more elaborate optimization procedure. In this case, we may want to enable the EM algorithm, as in Banbura and Modugno (2010) or use directly a numerical optimization procedure, as in Camacho and Pérez-Quirós (2010). This option has turned out to be much faster in all the examples we have tested thanks to the multithreading ability of our software, which is able to reduce the execution time by exploiting multiple processors in parallel. However, big models (i.e. in our context this means hundreds of variables, and multiple factors) are quickly estimated mixing both methods, i.e. using the EM algorithm in a first step and subsequently proceed with numerical optimization.

Regarding the numerical optimization procedure, there are two features that can be useful for complex models. First, the user can enable the option to estimate a simplified model (only one lag in the VAR) to obtain starting values and then proceed with numerical optimization for the estimation of the original model. Second, by ticking the mixed estimation box, the numerical optimization can be decomposed in two steps ("iterations by block"), emulating the logic of the EM algorithm. Both the Broyden–Fletcher–Goldfarb–Shanno (BFGS) and Levenberg-Marquardt algorithms are at the user's disposal. The use of a final EM algorithm is unnecessary. This option was introduced at an earlier development stage before the numerical procedure was fully operational.

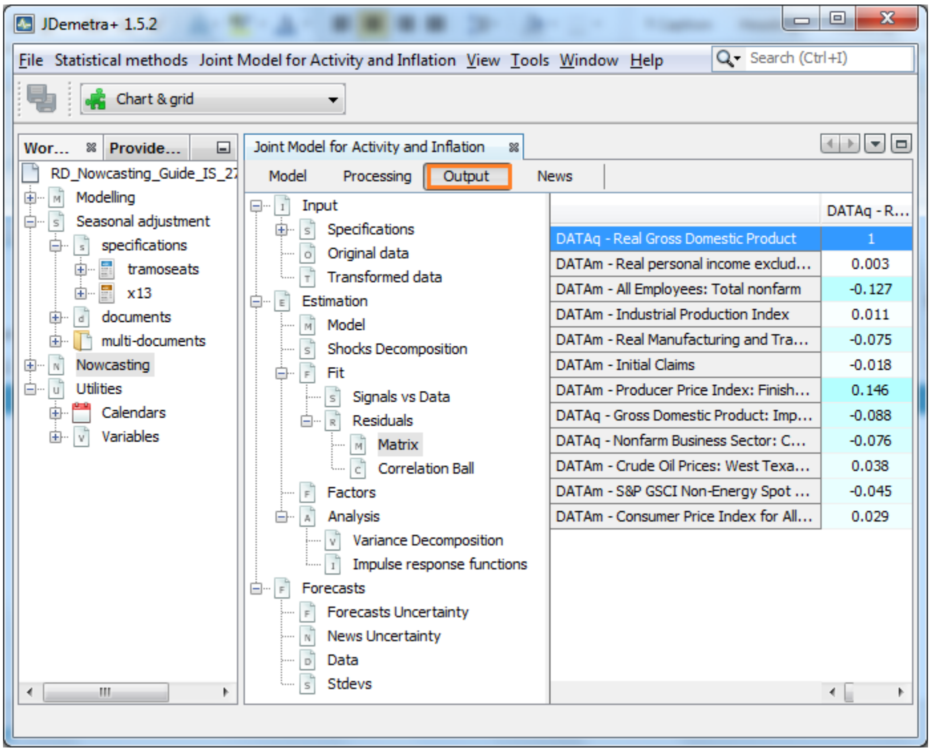
6 Estimation Results

Once the estimation process is complete, all the results can be displayed by clicking on the output tab (Figure 5). The estimation results (see Estimation/Model) show that the variables related to the real growth rate of economic activity, highlighted in grey, load on the latent inflation factor β_t with the opposite sign to the inflation variables themselves. This result suggests that one can identify innovations to β_t that can generate a negative correlation between prices and quantities, i.e. supply shocks.

Estimation of the factors

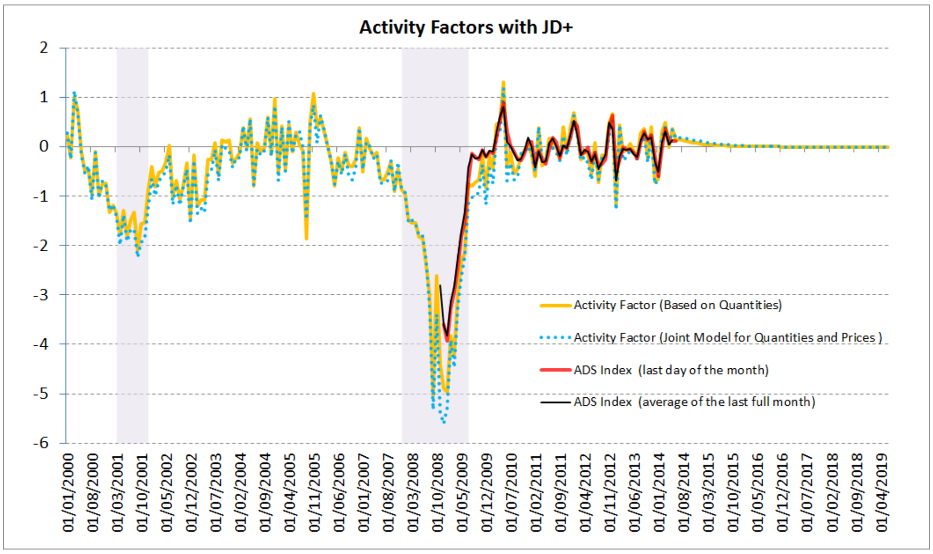
Using only the real activity monthly indicators suggested by A&D, the expected value of our factor conditional on the whole information set resembles the ADS economic activity index (see Figure 6), which have been taken from the website of the Philadelphia Fed. The JDEMETRA+ estimation of the activity factor turns out to be very similar independently of whether inflation variables have been incorporated in the model or not. Regarding the inflation factor itself, a visual examination suggests that supply shocks have not played a major role over the last recession (see upper panel of Figure 7). This contrasts with the inflationary dynamics registered by the indicator during the recessions dated in the mid-seventies and early eighties, which informally suggests the prevalence of supply factors.

Figure 5: The Output



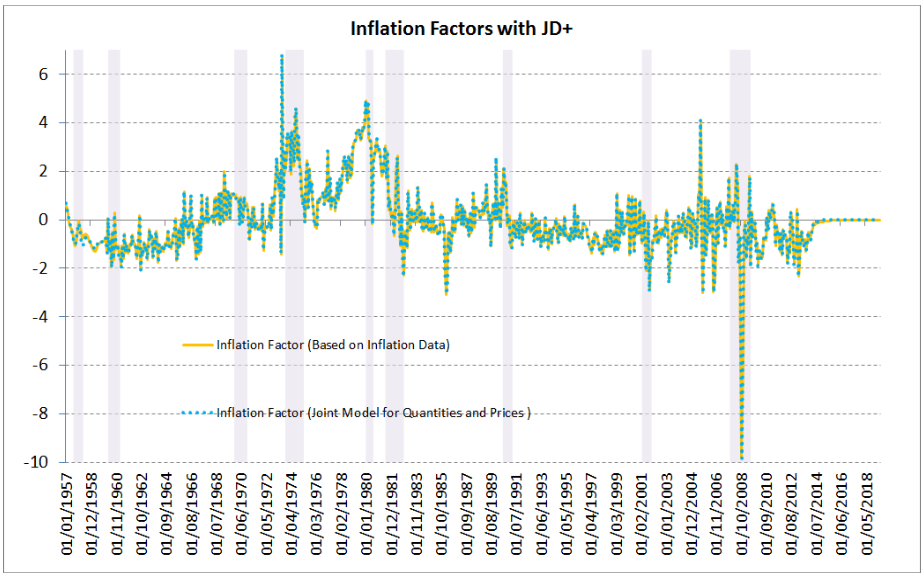
The output is divided in three blocks. First, we keep all information regarding the original data and estimation options inside “Input”. Second, all results related to the estimation are stored inside “Estimation”. This includes a display of the estimated model, plots of a historical shock decomposition, plots of all the series without the noise component, estimates of the factors, and analysis based on a decomposition of the shocks. Finally, all results related to “Forecasts” are included under a separate title.

Figure 6: Comparison with the ADS Index



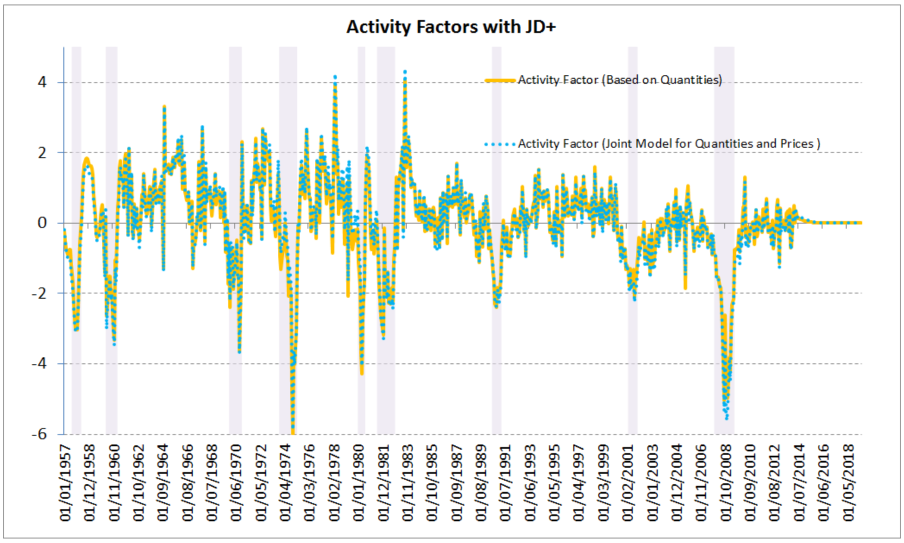
Note: The data vintage corresponds to the 30th of June, 2014. From July onwards, the factor is forecasted. The factors with confidence intervals can be visualized in the "output" tab by clicking on "Factors".

Figure 7: Inflation factor over the business cycle



Note: The data vintage corresponds to the 30th of June, 2014. From July onwards, the factor is forecasted. The factors with confidence intervals can be visualized in the "output" tab by clicking on "Factors".

Figure 8: Activity factor over the business cycle



Note: The data vintage corresponds to the 30th of June, 2014. From July onwards, the factor is forecasted. The factors with confidence intervals can be visualized in the "output" tab by clicking on "Factors".

———

Estimated factor loadings

Figure 9: Factor Loadings

Estimates of the factor loadings for all series	Sample mean	Stdev	Normalized Factors F1 (β_t)	Normalized Factors F2 (α_t)	Idiosyncratic Variance
Real Gross Domestic Product	0.008	0.010	0.02	0.07	0.39
Real personal income excluding current transfer receipts	0.002	0.006	0.15	0.30	0.80
All Employees: Total nonfarm	0.001	0.002	0.00	0.61	0.25
Industrial Production Index	0.002	0.008	0.10	0.54	0.44
Real Manufacturing and Trade Industries Sales	0.002	0.010	0.15	0.40	0.68
Initial Claims	0.001	0.049	-0.13	-0.31	0.80
Producer Price Index: Finished Goods	0.003	0.006	-0.53	-	0.46
Gross Domestic Product: Implicit Price Deflator	0.008	0.006	-0.08	-	0.25
Nonfarm Business Sector: Compensation Per Hour	0.012	0.009	-0.37	-	0.74
Crude Oil Prices: West Texas Intermediate (WTI) - Cushing, Oklahoma	0.004	0.085	-0.40	-	0.76
S&P GSCI Non-Energy Spot - PRICE INDEX	0.003	0.045	-0.02	-	1.00
Consumer Price Index for All Urban Consumers: All Items	0.003	0.003	-0.68	-	0.10

The estimation results shows that the variables related to the real growth rate of economic activity, highlighted in grey, load on the *latent inflation* factor β_t with the opposite sign to the inflation variables themselves. This result suggest that one can identify innovations to β_t that can generate a negative correlation between prices and quantities, i.e. supply shocks.

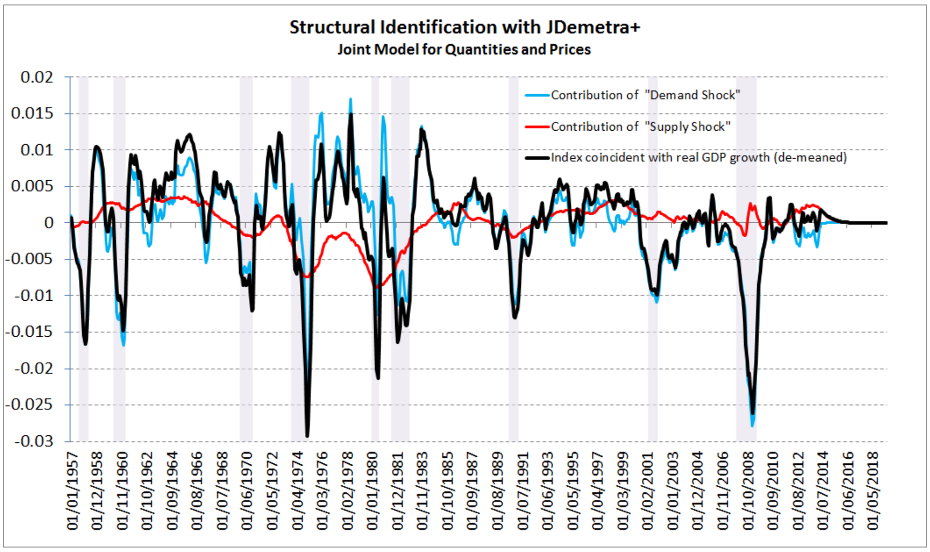
———

Structural Identification

As discussed above, a formal quantification of the role of both supply and demand determinants is possible within JDEMETRA+ if we are willing to explicitly account for the interaction between both factors. Such model, which considers both factors as a structural VAR with three lags (to be consistent with the original A&D formulation), is estimated here with maximum likelihood. The model is so simple that it does not matter whether you use the EM algorithm or a more sophisticated optimization procedure. The resulting factors remain, as discussed above, unchanged with respect to the ones obtained with the two separate panels. However, this joint model allows for a structural interpretation of the VAR innovations.

The structural identification of supply and demand shocks is conducted here using the empirical framework described in Section 5.4. Although, we recognize it is a very naïve formulation, it does not impose that supply shocks generate a negative correlation between inflation and real activity.

Figure 10: Historical Decomposition of Real GDP (de-meaned)



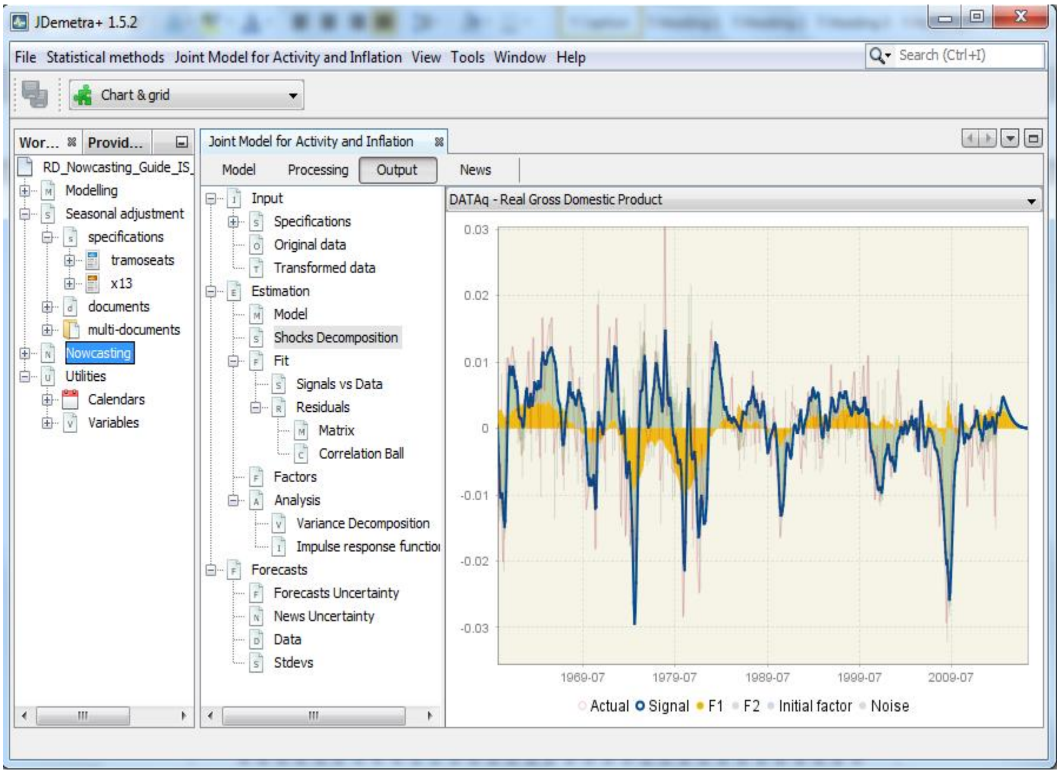
Note: The data vintage corresponds to the 30th of June, 2014. From July onwards, the factors are forecasted. The data plotted in this graph corresponds to the "Historical Shock Decomposition" graph provided by JDEMETRA+ (see Figure 11).

We are simply assuming that a demand shock does not have a contemporaneous effect in inflation variables, possibly because prices are sticky, while supply shocks can freely impact both prices π_t and quantities y_t. Ignoring the presence quarterly variables for the sake of simplicity, the structural model can therefore be written as follows:

$$\begin{pmatrix} \pi_t \\ y_t \end{pmatrix} = \begin{pmatrix} \Lambda_\pi^* & 0 \\ \Lambda^* & Z^* \end{pmatrix} \begin{pmatrix} T_{11}^* & T_{12}^* \\ T_{21}^* & T_{22}^* \end{pmatrix}^t \begin{pmatrix} \beta_0^* \\ \alpha_0^* \end{pmatrix} + \begin{pmatrix} \Lambda_\pi^* & 0 \\ \Lambda^* & Z^* \end{pmatrix} \sum_{j=0}^{t-1} \begin{pmatrix} T_{11}^* & T_{12}^* \\ T_{21}^* & T_{22}^* \end{pmatrix}^j \begin{pmatrix} u_{\beta,t}^* \\ u_{\alpha,t}^* \end{pmatrix} + \begin{pmatrix} \xi_t^\pi \\ \xi_t \end{pmatrix}$$

Thus, we can obtain a historical decomposition of out activity and inflation series in terms of the structural shocks $u_{\beta,t}^*$ (supply) and $u_{\alpha,t}^*$ (demand). The contribution of the initial state of the factors in stationary models is significant only at the beginning of the sample, i.e. the transition matrix to the power of "t" converges to zero.

Figure 11: JDemetra+ Historical Shocks Decomposition Graph



Note: The data vintage corresponds to the 27th of October, 2014. From November onwards, the factors are forecasted. This interactive graph corresponds to the actual shock decomposition provided by JDEMETERA+. All the elements can be highlighed by cliking on them and removed, for the sake of simplicity. As opposed to the simplified version plotted in Figure 10, where only the signal and its contributions are plotted, what we show here is the actual data (demeaned). Thus, the decomposition involves now the noise shocks in addition to the contributions of orthogonalized factor innovations (the structural shocks). The contribution of the initial conditions is negligible in this case, but it could play a role in models where the factors have a unit root.

Figures 10 and 11 shows the resulting decomposition of the signal underlying GDP growth rate in terms of the structural shocks, providing a quantitative support to the A&D's claim that the recessions in the mid-seventies and early eighties were largely driven by supply shocks. In turn, the contribution of supply shocks during the last recession is rather limited.

In-sample fit

So far, we have formalized A&D discussion on the source of business cycle fluctuations with special emphasis on the factors and their behavior during recessions. The JDEMETRA+ interface offers many other features that allow users to validate the model in-sample and to perform out of sample forecasts in real time. By clicking on Estimation/Fit/"Signals vs Data", we display a graph of all series together with their underlying signal. Figures 12 and 13 show the graph for GDP and employment growth.

Figure 12: JDEMETRA+ Analysis of Dynamic Factor Models (GDP)

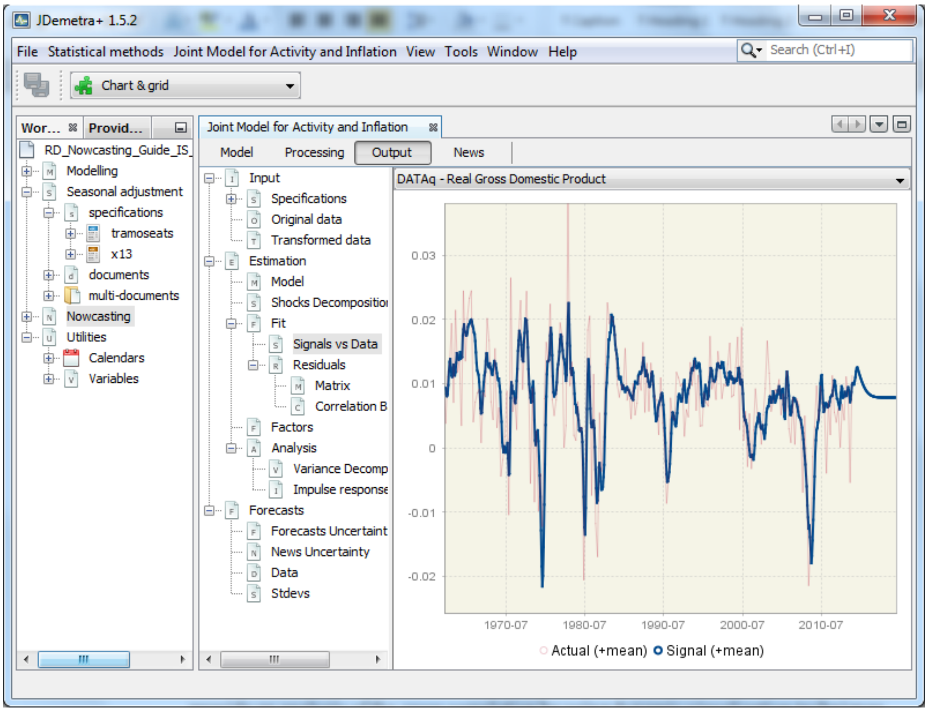
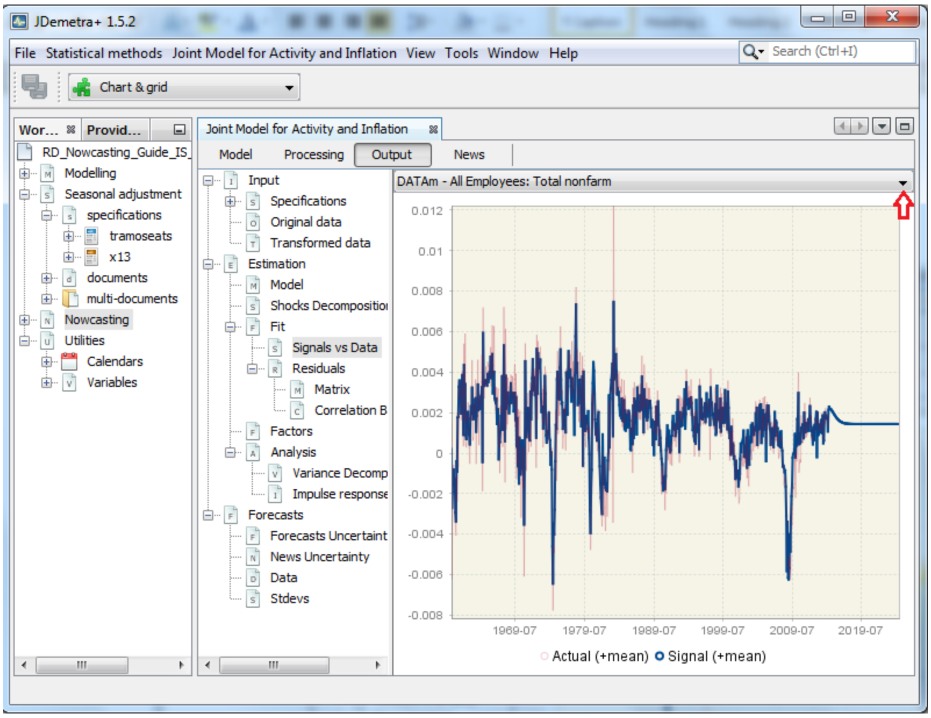


Figure 13: JDEMETRA+ Analysis of Dynamic Factor Models (Employment)



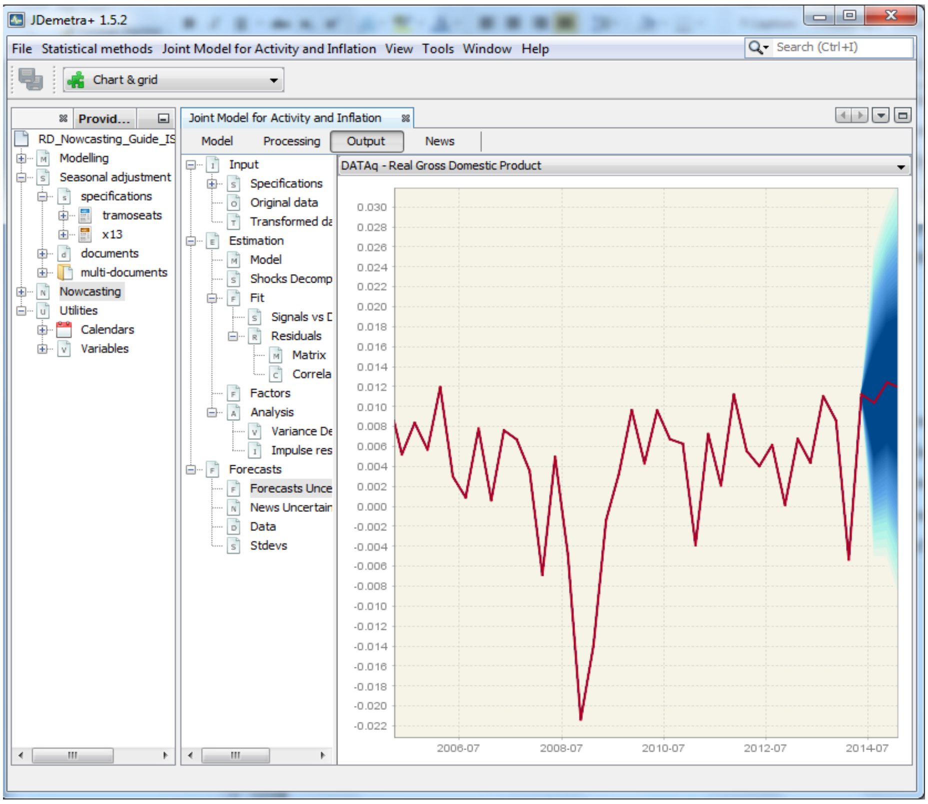
Note: The data vintage corresponds to the 27th of October, 2014. From November onwards, the data are forecasted. One can visualize the graph corresponding to any of the series that appear in the model.

The difference between both series, which is the so-called idiosyncratic measurement error, is also analyzed in detail. By clicking on Estimation/Residuals we obtain a table with the autocorrelation of the measurement errors, which will appear in red when they are considered to be statistically significant. The sections Estimation/Residuals/Matrix and …/Correlation provide an analysis of the cross correlation by using dynamic visualization techniques that can help to identify hidden patterns in the residuals. When the cross-correlation is pervasive, the model can be considered to be misspecified.

7 Forecasting

The nowcasting library has been optimized for its use in real time situations. However, the evaluation of out-of-sample forecasts is possible only after a sufficiently large forecasting record has been archived. The forecasts displayed in Figure 14 correspond to 2014Q3 and beyond.

Figure 14: Real GDP growth forecast obtained on 27th of October, 2014



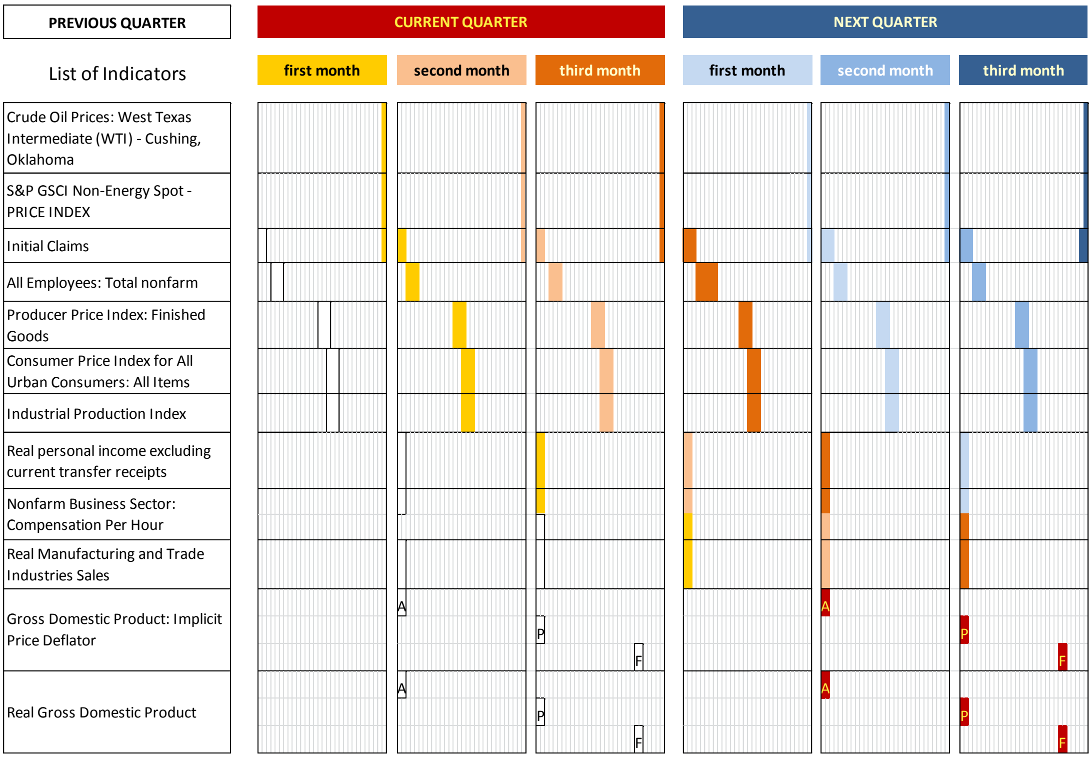
Note: The data vintage corresponds to the 27th of October, 2014. The forecasting interval represents the aggregate effect of news uncertainty (future data releases) and measurement errors.

Currently, we are developing an independent evaluation library that takes into account the calendar of macro-economic releases (see Figure 15) in order to provide realistic simulations of forecast errors in multivariate and univariate time series models. This will allow us to use pseudo out-of-sample forecasts can be used for evaluation purposes.

8 A Credible Narrative to Account for Forecasting Revisions

Many of the results mentioned so far are based on the model estimated with the vintage of data available on the 27th of October. Today, on the 12th of November, we have some new data available. We will show now the impact of that news on the forecasts for GDP at several forecast horizons. Figure 15 shows an approximate calendar of data releases. This will play a crucial role in the evaluation of forecasting accuracy.

Figure 15: Calendar of Data Releases



In this calendar, based on the information available in Table 1, we represent the approximate publication delay of all the indicators incorporated in the model. Employment, for example, is available soon after the end of the month. In order to represent data revisions in the national accounts, we make the distinction between the first release (i.e. Advanced), the second release (i.e. Preliminary) and the third release (i.e. Final). In the model, however, we simplify things by incorporating Preliminary and Final releases alone. This is a relevant modification with respect to the ADS approach.

For the moment, note that in order to represent data revisions in the national accounts, we make the distinction between the first release (i.e. Advanced), the second release (i.e. Preliminary) and the third release (i.e. Final). In the model, however, we will simplify things by incorporating Preliminary and Final releases alone. This is a new ingredient with respect to the ADS approach, where no distinction is made.

8.1 A story of news, data revisions, and inflation-output interactions

The following table summarizes the information that is going to be described in this section. The first example analyses how the forecasts obtained on October 27 are revised with the information set available on November 12. Although we have data revisions, we ignore them to focus on the concept of news. The second example does consider data revisions, but they turn out to be insignificant and the emphasis is placed on the interactions between inflation and output. In the final example, data revisions happen to play an important role:

Examples	Weights	Impacts
1: Simplified context $\mathcal{F}_{\text{old}}$: October 27 $\mathcal{F}_{\text{new}}$: November 12	Figure 17	Figure 18
2: Inflation and output interactions $\mathcal{F}_{\text{old}}$: September 11 $\mathcal{F}_{\text{new}}$: October 27	Figures 19, 20	Figure 21
3: The role of data revisions $\mathcal{F}_{\text{old}}$: November 12 $\mathcal{F}_{\text{new}}$: November 25	Figure 22, 23, 24	Figure 25

First of all, it is worth mentioning that in the three cases we follow the same procedure to update the forecast. Such an update is decomposed in terms of news and revisions to past data following four logical steps:

[1] **Archiving**. A given model based on the information set $\mathcal{F}_{\text{old}}$ can be archived, say on October 27. This implies both model and data are frozen and we are not able to perform any modification. The forecasts and all functions of this model are stored and can be retrieved at any time.

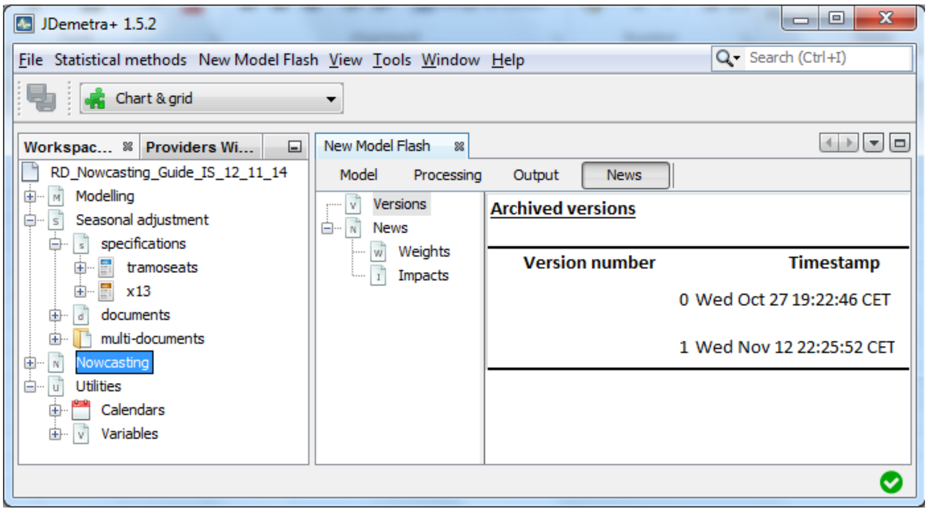
[2] **Refreshing**. On November 12, we update our database incorporating all the news and data revisions that have arrived since the last time we refreshed. *J*Demetra+ will look for all variable incorporated in the model and update their values for the whole time span.

[3] **Processing**. By clicking on the green arrow in the processing tab, we will run the Kalman filter and smoother to re-estimate the factors. Note that this process is executed using the last available version of the model, i.e. the same specification and parameters. Because we have refreshed the data in the previous step, the Kalman filter will update the forecasts.

[4] **News**. Our updated forecasts are meaningless without an economic interpretation. Thus, we are going to re-calculate those updated forecasts by expressing them as a function of the news and data revisions that have entered our information set. Computational details are provided in the appendix.

By clicking on the News tab (see Figure 16) and choosing *Versions* we can see the list of datasets that have been archived. In our case, the last archive took place in October 27 at 19:22CET, while the current update has occurred on November 12 at 22:25CET. Thus, our main goal is to compare the last archived (version 0) with the new forecast obtained with our updated data (version 1).

Figure 16: Versions

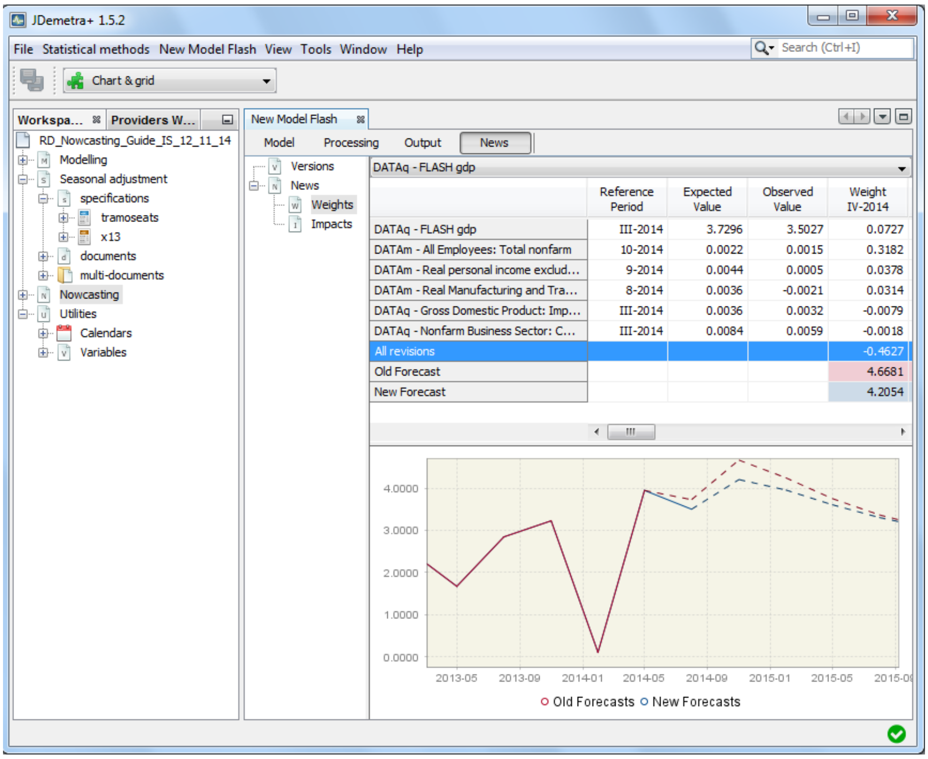


The same model (and estimation) archived for the first time on the 27th of October, is now used to analyse the news. The updating sequence consists of two steps: First, the data needs to be refresed (click on “New Model Flash” above and Refresh). Once the data has been refreshed, click on the tab “Processing” and run the model (by default, all estimation options have been unchecked to avoid re-estimating the parameters). Finally, by clicking on the tab “News”, the analysis of news is executed, as described in Section 3, for all variables and forecast horizons.

Weights and Impacts in a Simplified Context

Figure 17 can be found in the tab “News”, by clicking on News/Weights. Here, all the new data releases are compared with the forecasts of the model. From all the news (i.e. forecast errors), the largest weight corresponds in this case to the employment release, which was worse than expected by the model. Remember from Equation 3 that the sum of all news times their respective weights determines the size of the revision for GDP. Although we only show the analysis for the fourth quarter, all the subsequent quarters can be found by moving the bar towards the right. By looking at the graph below, where the updated forecasting path is compared with the old one, we can already anticipate that most of the forecast revision remains within the fourth quarter.

Figure 17: Updating Advanced GDP Projections (November 12)

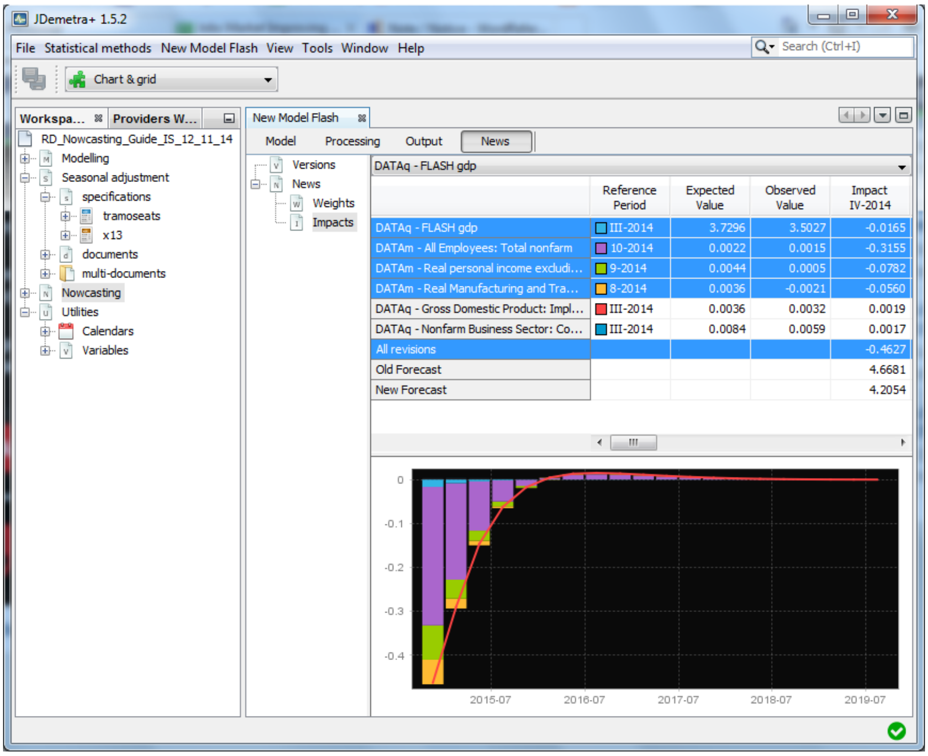


In the tab "News", click on News/Weights to compare all the new data releases with the figures expected by the model. From all the news, the largest weight (see Equation 3) corresponds in this case to the employment release, which was worst than expected by the model. The sum of all news times their respective weights equals -0.45, which corresponds to a modest downward revision of the nowcast for the Advance release in the forth quarter of the year. The graph below compares the updated forecasting path with the old one.

———

The current version of the software computes the impacts of all news, including data revisions, for all variables and forecast horizons. This computation took a couple of seconds in this simple example, but it could take a few of minutes when there are revisions to past data in all variables. Figure 18 contains the same information as Figure 17, but it does the math for you. The forecasting revision for all series is decomposed in terms of the impact of each piece of news. The plot of the forecasting revision along with the contribution of all the news turns out particularly informative. From all the news, the largest impact corresponds in this case to the employment, followed by real personal income, real manufacturing and the advance GDP release. The sum of those impacts equals -0.46 for the fourth quarter of 2014. The forecast for 2015Q1 is revised downwards by -0.3, but revisions to subsequent quarters become very small.

Figure 18: News Impacts at Updating the Advanced GDP Forecast (November 12)



In the tab "News", click on News/Impacts to decompose the forecast revisions in terms of the news. We have the information for all variables, but let's focus on GDP. From all the news, the largest impact corresponds in this case to the employment, followed by real personal income, real manufacturing and the advance GDP release. The sum of those impacts equals -0.45 for the forth quarter of 2014. The forecast for 2015Q1 is revised downards by -0.3, but subsequent revisions become very small.

———

Inflation-Output Interactions

The JDEMETRA+ news algorithm can also be used for the analysis of scenarios. Let's assume that we want to incorporate in our forecast the knowledge that CPI inflation will be zero for the rest of the year, which would be clearly below the figures expected by the model. If such an assumption aims to reflect an improvement in productivity, then our forecasts for real GDP growth will improve. If on the other hand our inflation scenario simply aims to represent a sudden deterioration of demand, growth expectations will have to be revised downwards. Fortunately, the Kalman filter provides the most likely response of the economy to the whole set of news incorporated in the system. Those forecast do not require any structural interpretation of the shocks such as the one proposed above.

Rather than building a scenario, let's consider an example that has actually occurred in real time. Figures 19 and 20 compare the old forecasts obtained on September 11 with the new predictions for output and inflation obtained after refreshing the data on October 27. We can see that real GDP growth expectations for 2014Q3 and 2014Q4 have been revised upwards by 0.5 and 2 percentage points respectively. The GDP deflator projections for those periods have been revised in the opposite direction, suggesting the presence of supply shocks. Figure 21 sheds light in the factors that account for that upward revision in GDP in spite of the deflationary pressure, which was mostly driven by oil prices. Not surprisingly, the consumer price indexes for August and September, which were lower than expected by the model, had a positive contribution in the GDP forecast revision for 2014Q3, 2014Q4 and beyond (yellow and green colors). From all the news, the largest impact corresponds in this case to the industrial production (blue) and unemployment claims (red).

The role of data revisions

In our examples, we have ignored the frequent situation in which new data releases for a series i contain important modification in its history. This implies that we will have many pieces of news for the series i, each one referring to a period $\mathrm{t_j}$ in the history that has been subject to revision: $\mathrm{y_{i_j,t_j}} - \mathrm{E}\left[\mathrm{y_{i_j,t_j}} \middle| \mathcal{F}_{\mathrm{old}}\right]$

Analyzing the impact of all those revisions at updating a given series is challenging. However, our approach helps the user to quickly identify the revisions that have a major impact. We propose to incorporate the aggregate impact of all revisions, in a new row, which can be expanded. Only in the case such impact is worth being investigated, the user will open a new window with the actual decomposition arranged by variable, in chronological order.

Let's consider a concrete example where the information set corresponding to November 12 ($\mathcal{F}_{\mathrm{old}}$) is updated on November 25 ($\mathcal{F}_{\mathrm{new}}$). Figures 22-24 shows the updated forecasts for the Advanced GDP, Deflator and Industrial production conditional on the new information set. The first two figures (22-23) also reveal that the real growth and inflation releases for Q3 were in line with the model forecasts. Thus, the forecast revisions for Q4 and beyond must be due to other news. In turn, figure 24 shows that the Industrial Production release for October has been clearly worse than expected by the model. One can also observe that the official statistic for industrial production in the month of September has suffered a downward revision too. The role played by all news and data revisions at updating a given variable can be summarized by clicking on *Impacts.* Figure 25 shows that the downward revision in GDP forecasts for Q4 and beyond is mostly driven by the bad news contained in the releases of employment and industrial production for October, and unemployment claims for November. Expanding

All Revisions allows us to analyse the role played by statistical data revisions. In this case, the largest impact corresponds to industrial production in September and the unemployment claims in October. The impact of the revisions at updating Q4 and beyond is much smaller than the impact of news. We can also observe that data revisions refering to a more distant past, i.e. very common in seasonal adjusted data, tend to have a negligible impact.

9 Conclusions

This paper presents an innovative expert system that is suitable for real-time forecasting and nowcasting applications, with a particular emphasis in decomposing the forecast revisions in terms of the unexpected component of new data releases. More than a translation of the models described in leading nowcasting applications such as Banbura and Modugno (2010) or Camacho and Pérez-Quirós (2010), the library described here proposes a re-factoring of those methods exploiting existing routines of the JDEMETRA+ environment, originally developed for the analysis of seasonal data. In particular, this library makes an extensive use of both the state-space modeling framework and dynamic graphical analysis tools that have been developed for multiple purposes.

The *nowcasting* model proposed for the US economy as an illustration is, to the best of our knowledge, the first one that accounts for the joint behavior of quantities and prices. It complements the business cycle analysis provided at the Philadelphia Fed by allowing for inflation-output interactions. Users can easily introduce their own expertise in the form of alternative methods within the class of dynamic factor models, contributing to extend the limits of the currently established practices in the nowcasting literature. All model specifications can be saved along with the data vintages that are available in real time. Thus, we hope our tool will catalyze the dissemination of research on nowcasting and real-time data analysis and provide practitioners with the means to improve the state-of-the-art. From the methodological point of view, it is also possible to implement the analysis of news and data revisions using alternative models, such as the VAR with mixed frequencies. Recent applications, such as the work by Schorfheide and Song (2014), for example, document the extent to which information improves the forecasts in real time. However, they do not provide an analytical decomposition of the forecasting revisions in terms of news.

Figure 19: Updating Advanced GDP Projections (October 27)

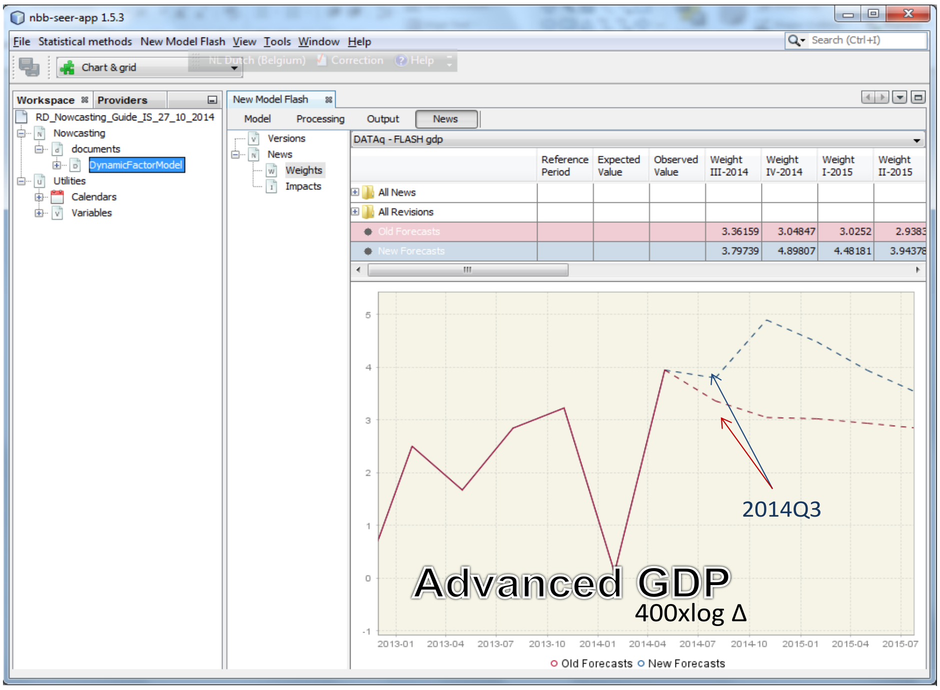
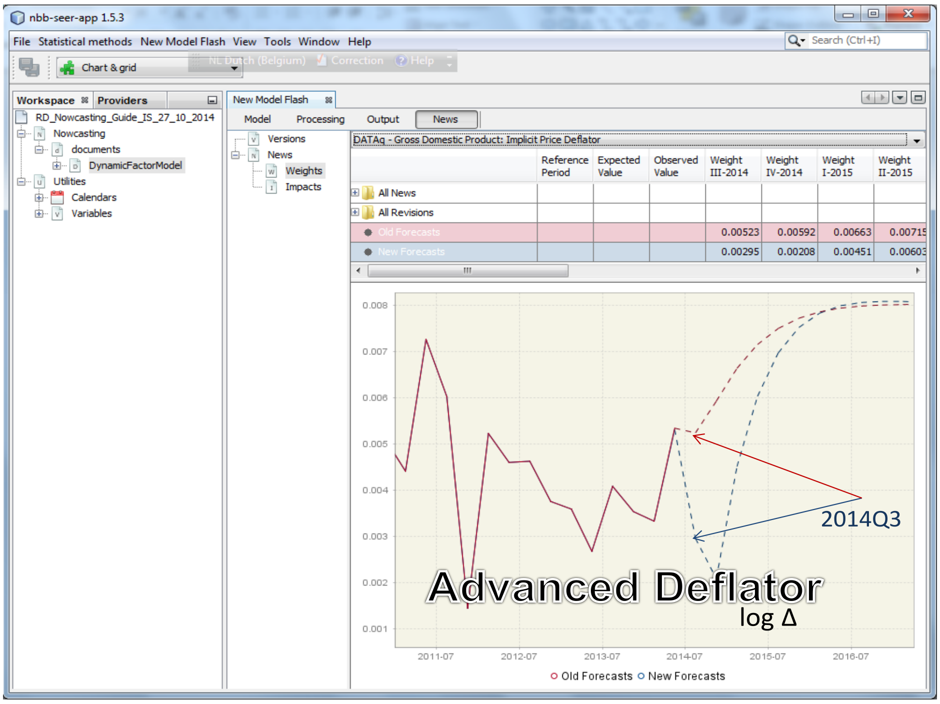
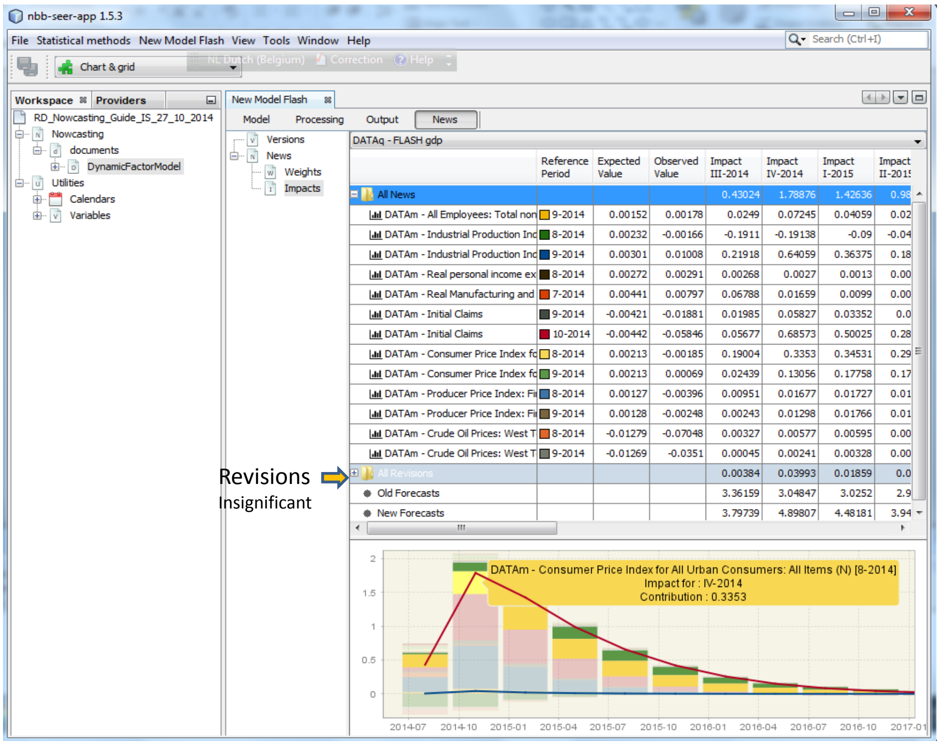


Figure 20: Updating Advanced GDP Deflator Projections (October 27)



In the tab "News", click on News/Weights. By expanding the folder *All News* we can compare all the new data releases with the figures expected by the model, and see their weight (Equation 3) at updating inflation expectations. The same holds for the data revisions. By expanding the folder *All Revisions* we can compare the revised data with the previous version. The graphs in both figures illustrate the downward and upward revisions of the forecasting path for inflation and output, respectively.

Figure 21: News Impacts at Updating Advanced GDP projections (October 27)



In the tab "News", click on News/Impacts to decompose the forecast revisions in terms of the news and revisions. We have the information for all variables, but let's focus on GDP. From all the news, the largest impact corresponds in this case to the industrial production (blue), unemployment claims (red) and the consumer price index (highlighed in yellow), which turned out to lower than predicted. Expanding the *All Revisions* help us to analyse the role played by statistical data revisions. However, their aggregate effect, summarized in the blue line, is very small.

———

Figure 22: Updating Advance Release of GDP Projections (November 25)

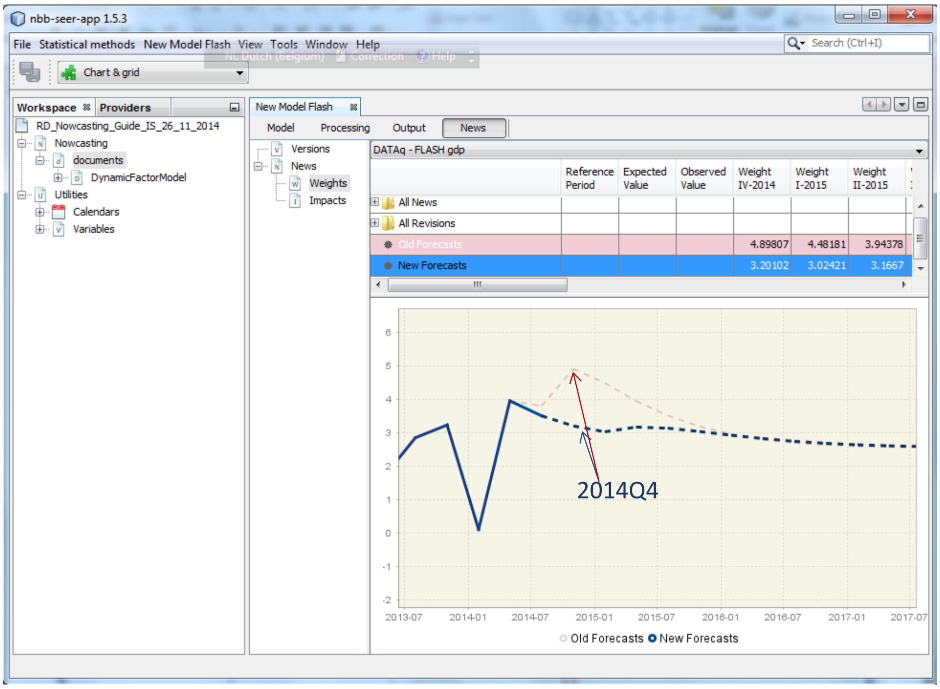
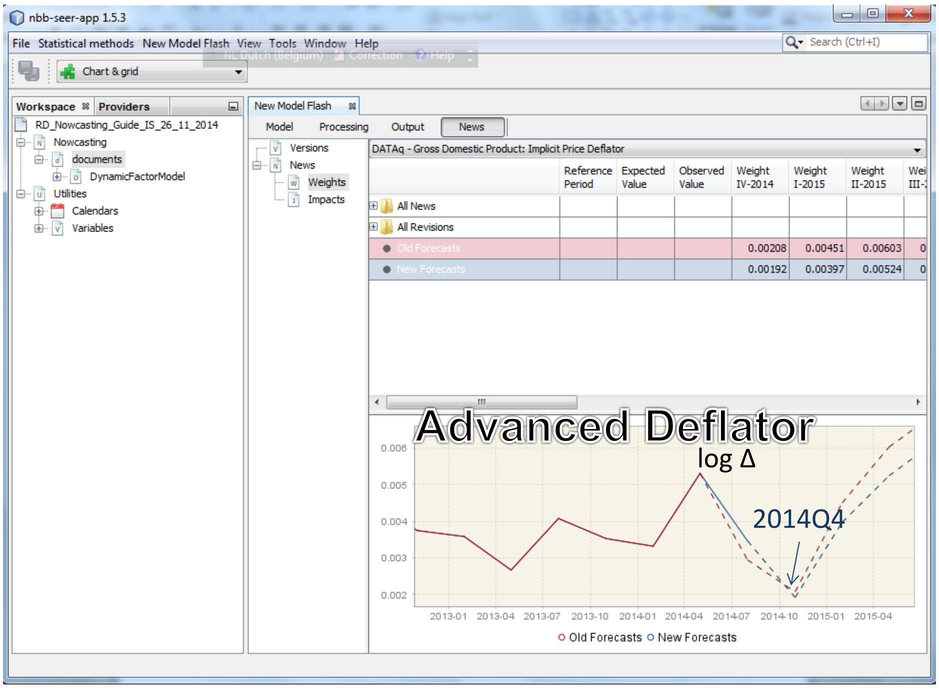
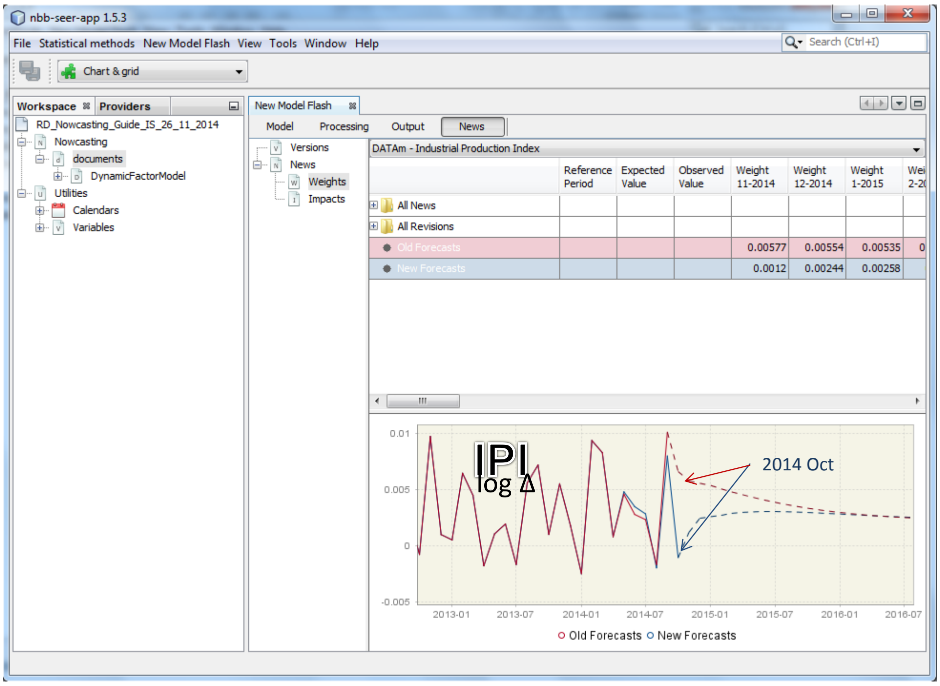


Figure 23: Updating Advance Release of GDP Deflator Projections (November 25)



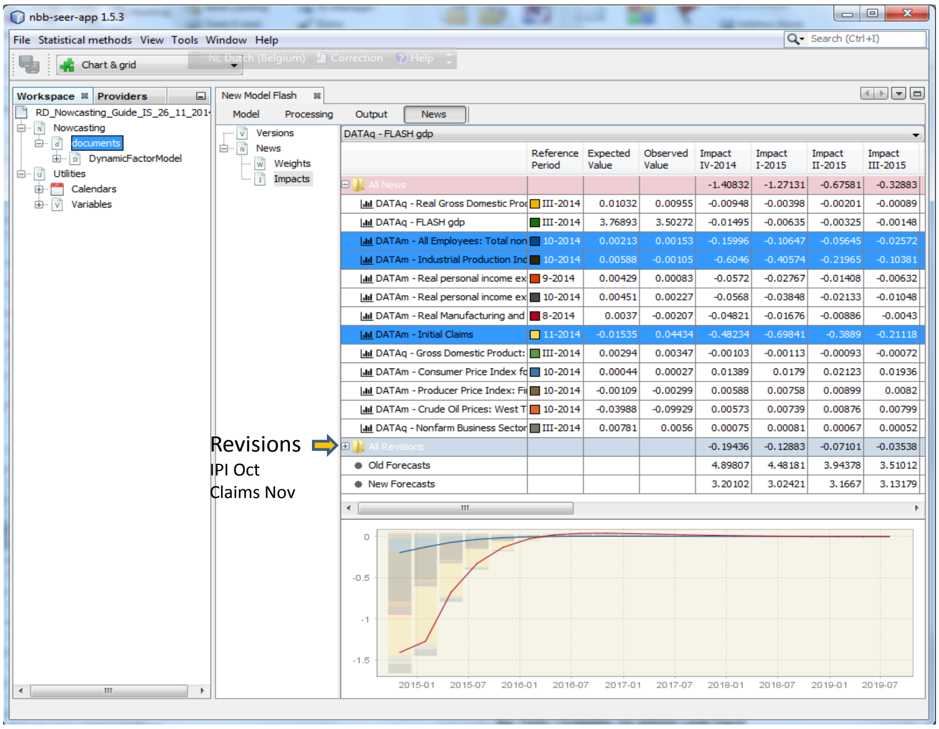
In the tab "News", click on News/Weights. By expanding the folder All News we can compare all the new data releases with the figures expected by the model, and see their weight (Equation 3) at updating inflation expectations. The same holds for the data revisions. By expanding the folder All Revisions we can compare the revised data with the previous version. The graph in Figure 21 illustrates the downward revision of GDP for Q4, while inflation expectations in Figure 22 do not change by much.

Figure 24: Updating Industrial Production Projections (November 25)



In the tab “News”, click on News/Weights. By expanding the folder *All News* we can compare all the new data releases with the figures expected by the model, and see their weight (Equation 3) at updating inflation expectations. The same holds for the data revisions. By expanding the folder *All Revisions* we can compare the revised data with the previous version.

Figure 25: News Impacts at Updating Advance GDP Projections (November 25)



In the tab "News", click on News/Impacts to decompose the forecast revisions in terms of the news and revisions. We have the information for all variables, but let's focus on GDP. From all the news, the largest impact corresponds in this case to employment, industrial production, unemployment claims, which are all highlighted in blue in the spreadsheet. Expanding the *All Revisions* help us to analyse the role played by statistical data revisions. However, their aggregate effect, summarized in the blue line is smaller than that corresponding to news, in blue.

---

10 Download the JDEMETRA+ software in any platform

- The last updated version of the software can be downloaded here
  http://www.cros-portal.eu/content/jdemetra

- Download (and unzip) the plug-ins:
  http://www.nbb.be/app/dqrd/jdemetra/jdplugins-1.5.3.zip

- Run the main application: ./bin/nbdemetra[64.exe]
  You may get an error message saying that you do not have the last version of Java. In such case, just download it here:
  http://www.oracle.com/technetwork/java/javase/downloads/jdk8-downloads-2133151.html

- Install the plug-ins:
  Main menu: « Tools Plugins »
  Select tab « Downloaded »
  Click « Add plugins… »

  ✓ Select NbDemetra-Core2 and NbDemetra-Dfm
  (from plugins repository)
  ✓ Follow the instructions.

11 References

[1] Andreou, E., E. Ghysels and A. Kourtellos (2010). "Regression models with mixed sampling frequencies" *Journal of Econometrics*, 158, pages 246-261.

[2] Aruoba S.B. & F. X. Diebold (2010). "Real-Time Macroeconomic Monitoring: Real Activity, Inflation, and Interactions" *American Economic Review,* American Economic Association, vol. 100(2), pages 20-24, May.

[3] Aruoba, S.B., F.X. Diebold and C. Scotti (2009), "Real-Time Measurement of Business Conditions" *Journal of Business and Economic Statistics* 27, pages. 417-27.

[4] Baffigi, A. & G. Roberto and G. Parigi (2004). "Bridge models to forecast the euro area GDP," *International Journal of Forecasting*, vol. 20, pages 447-460.

[5] Banbura, M., and G. Runstler (2011). "A look into the factor model black box: Publication lags and the role of hard and soft data in forecasting GDP". *International Journal of Forecasting*, 27, 333-346.

[6] Banbura M., and M. Modugno (2012). "Maximum likelihood estimation of factor models on data sets with arbitrary pattern of missing data". *Journal of Applied Econometrics*, 29, 133-160.

[7] Banbura M., D. Giannone and L. Reichlin (2010). "Nowcasting and the Real-Time Data Flow". *Handbook of Economic Forecasting*, in G. Elliot and A. Timmermann, 2.

[8] Banbura M., D. Giannone and L. Reichlin (2011). "Nowcasting". *Oxford Handbook on Economic Forecasting, of Economic Forecasting*, in Michael P. Clements and David F. Hendry, pages 193-224.

[9] Camacho M., G. Pérez-Quirós (2010). "Introducing the Euro-Sting". *Journal of Applied Econometrics*, 25, 663-694.

[10] Croushore, D. and T. Stark (2002). "Forecasting with a Real-Time Data Set for Macroeconomists". *Journal of Macroeconomics*, 24, 507-531.

[11] Croushore, D. and T. Stark (2001). "A Real-Time Data Set for Macroeconomists". *Journal of Econometrics*, 105, 111-130.

[12] Diebold, F.X. and Scotti, C. (2009), "Real-Time Measurement of Business Conditions," *Journal of Business and Economic Statistics* 27:4 (October 2009), pp. 417-27.

[13] Diron, M. (2008). "Short-term forecasts of euro area real GDP growth: an assessment of real-time performance based on vintage data," *Journal of Forecasting*, 27, pages 371-390.

[14] Doz C., D. Giannone and L. Reichlin (2012). "A Quasi Maximum Likelihood Approach for Large Approximate Dynamic Factor Models". *Review of Economics and Statistics*, 94, 1014-1024.

[15] Durbin, J and S.J. Koopman (2001). "*Time Series Analysis by State Space Methods* ". Oxford University Press.

[16] Evans M. D. D. (2005). “Where are we now? Real-Time Estimates of the Macroeconomy". *International Journal of Central Banking*, 1.

[17] Frale, C., M. Marcellino, G. L. Mazzi, and T. Proietti (2011): “EUROMIND: a monthly indicator of the euro area economic conditions,” *Journal Of The Royal Statistical Society* Series A, 174, 439–470.

[18] Ghysels, E., P. Santa-Clara, and R. Valkanov (2002). “The MIDAS touch: Mixed data sampling regression models”, Working paper, UNC and UCLA.

[19] Giannone, D., L. Reichlin and D. Small (2008). “Nowcasting: The Real-Time Informational Content of Macroeconomic Data Releases". *Journal of Monetary Economics*, 55, 665-676.

[20] Harvey, A.C. and S.J. Koopman (2003). “Computing observation weights for signal extraction and filtering”, *Journal of Economic Dynamics & Control*, 27, 1317-1333.

[21] Robert Ingenito & Bharat Trehan, 1996. "Using monthly data to predict quarterly output," *Economic Review*, Federal Reserve Bank of San Francisco, pages 3-11.

[22] Jacobs, Jan. P.A.M. and S. Van Norden (2011). “Modeling data revisions: Measurement error and dynamics of true values". *Journal of Econometrics*,

[23] 161, 101-109.

[24] Kalman, R.E (1960). “A new approach to linear fitering and prediction problems". *Journal of Basic Engineering*, 82, 35-45.

[25] Luciani, M. (2014). “Forecasting with Approximate Dynamic Factor Models:the role of non-pervasive shocks". *International Journal of Forecasting*, 30, p. 20-29

[26] Mariano R.S. and Y. Murasawa (2003). “A new coincident index of business cycles based on monthly and quarterly series". *Journal of Applied Econometrics*, 18, 427-443.

[27] Schorfheide F, and D. Song (2014). “Real-Time Forecasting with a Mixed-Frequency VAR". *Journal of Business & Economic Statistics*, forthcoming

[28] Stock J.H., and M.W. Watson (2002). “Forecasting using principal components from a large number of predictors". *Journal of the American Statistical Association*, 97, 147-162.

[29] Stock J, and M. Watson (2003). “Forecasting Output and Inflation: The Role of Asset Prices”. *Journal of Economic Literature*, 41, pages 788-829.

12 APPENDIX A: News Decomposition

Consider the following representation of two given vintages of data:

- $\mathcal{F}_v$ contains all time series of Table 1 as available at a given date v.
- $\mathcal{F}_{v+1}$ contains the same time series as available at a later date $v+1$.

Using the same notation as Banbura and Modugno (2010), the news content of the second vintage is defined by the J_{v+1} –sized vector of forecast errors I_{v+1}. Let's remove the subindex and use $J \equiv J_{v+1}$ to denote the number of news:

$$I_{v+1} = \begin{bmatrix} y_{i_1,t_1} - E\left[y_{i_1,t_1} \middle| \mathcal{F}_v\right] \\ \cdots \\ y_{i_J,t_J} - E\left[y_{i_J,t_J} \middle| \mathcal{F}_v\right] \end{bmatrix} ,$$

This vector represents is the part of the release $\mathcal{F}_{v+1}$ that is orthogonal to the information already present in $\mathcal{F}_v$. This notation can be easily understood by examining the example presented in Figure 26. Here, we have a total of $J = 5$ innovations. Two of them correspond to two consecutive months, t_1 and t_2 that become available for the same variable $i_1 = i_2 = 1$. The third innovation corresponds to a release for variable $i_3 = 2$ and refers to t_3. Note that both t_1 and t_3 correspond in this example to two innovations for the different variables but relative to the same month, which is March. Finally, the forth and fifth innovations corresponds to the last variable, N, which is *revised* for the months of March and April. Thus, data revisions have and index i corresponding to the variable they refer to and a subindex j that refers to the point in time. As a result, revisions can be represented with the same notation as news resulting from additional data releases:

$$\text{revision } i_j \quad : \quad y_{i_j,t_j} \quad - E\left[y_{i_j,t_j} \middle| \mathcal{F}_v\right], \qquad i_j \quad = N, \text{ and } t_j \quad = \text{March}$$

$$\text{revision } i_{j+1} : \quad y_{i_{j+1},t_{j+1}} - E\left[y_{i_{j+1},t_{j+1}} \middle| \mathcal{F}_v\right], \quad i_{j+1} = N, \text{ and } t_{j+1} = \text{April}$$

The forecast revision for a given variable k, $E[y_{k,t_k}|\mathcal{F}_{v+1}] - E[y_{k,t_k}|\mathcal{F}_v]$, is given by its projection on the news information set I_{v+1}:

$$E[y_{k,t_k}|I_{v+1}] = E[y_{k,t_k} \; I'_{v+1}] \; E[I_{v+1} \; I'_{v+1}]^{-1} I_{v+1} \qquad [9]$$

This linear projection determines the impact of the news. Thus, the revision can be expressed, more explicitly, as a weighted average of the different pieces of news:

$$E[y_{k,t_k}|I_{v+1}] = \sum_{j=1}^{J_{v+1}} w_j \left(y_{i_j,t_1} - E\left[y_{i_j,t_j}\middle|\mathcal{F}_v\right]\right) \qquad [10]$$

The expectations shown in expression [9], which are required to compute the weights in [10], are a function of the estimated state-space model parameters:

$$E[y_{k,t_k}\ I'_{v+1}] = \begin{bmatrix} \Lambda_k\, E\left[\left(f_{t_k} - E(f_{t_k}|\mathcal{F}_v)\right)\left(f_{t_1} - E(f_{t_1}|\mathcal{F}_v)\right)'\right]\Lambda'_{i_1} \\ \Lambda_k\, E\left[\left(f_{t_k} - E(f_{t_k}|\mathcal{F}_v)\right)\left(f_{t_2} - E(f_{t_2}|\mathcal{F}_v)\right)'\right]\Lambda'_{i_2} \\ \vdots \\ \Lambda_k\, E\left[\left(f_{t_k} - E(f_{t_k}|\mathcal{F}_v)\right)\left(f_{t_j} - E(f_{t_j}|\mathcal{F}_v)\right)'\right]\Lambda'_{i_j} \\ \vdots \\ \Lambda_k\, E\left[\left(f_{t_k} - E(f_{t_k}|\mathcal{F}_v)\right)\left(f_{t_J} - E(f_{t_J}|\mathcal{F}_v)\right)'\right]\Lambda'_{i_J} \end{bmatrix}' \qquad [11]$$

Figure 26: Two data vintages

$\mathcal{F}_v$

i =	1	2	3	...	N
january	old data	old data	old data	...	old data
february	old data	old data	old data	...	old data
march				...	old data
april				...	old data

$\mathcal{F}_{v+1}$

i =	1	2	3	...	N
january	old data	old data	old data	...	old data
february	old data	old data	old data	...	old data
march	new data t_1	new data t_3		...	revision t_4
april	new data t_2			...	revision t_5

In this stylized representation of two consecutive information sets, we have also represent revisions in old data. Macroeconomic data revisions can change both recent and historical values of a time series, which implies that a large number of innovations needs to be incorporated in I_{v+1}

The element j, l of matrix $\mathrm{E}[\mathrm{I}_{\mathrm{v}+1}\,\mathrm{I}'_{\mathrm{v}+1}]$ represents the covariance of the two innovations indexed by j and l:

$$E\left[I_{v+1}\,I^{'}_{v+1}\right]_{\{j,l\}} = \Lambda_{i_j}\,E\left[\left(f_{t_j} - E(f_{t_j}|\mathcal{F}_v)\right)\left(f_{t_l} - E(f_{t_l}|\mathcal{F}_v)\right)^{'}\right]\Lambda^{'}_{i_l} + E\left[\,\xi_{i_j t_j}\xi_{i_l t_l}\right] \qquad [12]$$

Here, we use the same notation as in Banbura and Modugno (2010), who provide details on the derivations. Note that the assumption that measurement errors are idiosyncratic implies:

$$\mathrm{E}\left[\xi_{\mathrm{i_j t_j}}\xi_{\mathrm{i_l t_l}}\right] = \begin{cases}\mathrm{R}_{\{\mathrm{j,l}\}} & \text{if } \mathrm{j} = \mathrm{l}\\ 0 & \text{if } \mathrm{j} \neq \mathrm{l}\end{cases}$$

The expression $E\left[\left(f_{t_j} - E(f_{t_j}|\mathcal{F}_v)\right)\left(f_{t_l} - E(f_{t_l}|\mathcal{F}_v)\right)'\right]$ implies that we need to compute the conditional covariance of the factors in the case that t_j and t_l refer to very distant periods of time. The state-space representation of the model is automatically enlarged in order to make sure that all those covariance terms are obtained directly by executing the Kalman smoother algorithm, i.e. the transition equation will include $t_{max} - t_{min}$, where t_{max} and t_{min} represent the most recent and oldest time period, respectively, among the set $\{t_1, \ldots, t_J,\ t_k\}$. The time index t_k represents the month for which the reaction to news of our target variable $\mathrm{y}_{\mathrm{t_k}}$ is being analyzed