Modeling Data Revisions

Por:
Juan Manuel Julio

Núm. 641
2011

Borradores de ECONOMÍA

Bogotá - Colombia - Bogotá - Colombia - Bogotá - Colombia - Bogotá - Colombia - Bogotá - Colombia - Bogotá - Colombia - Bogotá - Colombia

Modeling Data Revisions*

Juan Manuel Julio†

Abstract

A dynamic linear model for data revisions and delays is proposed. This model extends Jacobs & Van Norden's [13] in two ways. First, the "true" data series is observable up to a fixed period of time M. And second, preliminary figures might be biased estimates of the true series. Otherwise, the model follows Jacobs & Van Norden's [13] so their gains are extended through the new assumptions. These assumptions represent the data release process more realistically under particular circumstances, and improve the overall identification of the model.

An application to the year to year growth of the Colombian quarterly GDP reveals that preliminary growth reports under-estimate the true growth, and that measurement errors are predictable from the information available at the data release. The models implemented in this note help this purpose.

*The author indebts Norberto Rodríguez from Banco de la República for his valuable suggestions to a previous version of this paper and the help of Daniel Quintero from the Financial Accounts section of Banco de la República for providing the data set under analysis. However, any errors as well as the conclusions and opinions contained in this paper are the sole responsibility of its author and do not compromise BANCO DE LA REPUBLICA, its Board of Governors or Universidad Nacional de Colombia.
JEL: C22, C53, C82.
Keywords: Data Revisions, Now-casting, Real Time Economic Analysis.

†jjulioro@banrep.gov.co. Researcher, Banco de la República and Associate Professor, Department of Statistics, Universidad Nacional de Colombia. Bogotá D. C., Colombia

Modelando las Revisiones de Datos*

Juan Manuel Julio†

Resumen

Se propone un modelo lineal dinámico para la demora y revisión de datos. Este modelo extiende el de Jacobs & Van Norden [13] en dos direcciones. Primero, la serie de datos definitivos se observa hasta un periodo fijo de tiempo M. Y segundo, los datos preliminares pueden ser estimadores sesgados de los definitivos. Aparte de esto el modelo sigue al de Jacobs & Van Norden [13] con lo cual sus ganancias se extienden a través de los nuevos supuestos. Estos supuestos representan de manera realista el proceso de publicación de la información bajo circunstancias particulares, y mejora la identificación global del modelo.

Una applicación al crecimiento anual del PIB trimestral Colombiano muestra que los reportes preliminares del crecimiento subestiman el crecimiento definitivo, y que los errores de medición se pueden pronosticar a partir de la información disponible en cada fecha de publicación de datos. Los modelos que se implementan en este trabajo sirven para este propósito.

*El autor agradece los comentarios y sugerencias de Norberto Rodríguez del Banco de la República a una versión anterior de este trabajo y la ayuda de Daniel Quintero de la sección de Cuentas Financieras del Banco de la República por la ayuda al proveer la base de datos bajo análisis. Sin embargo, cualquier error que contenga este artículo, asi como sus conclusiones y recomendaciones son responsabilidad exclusiva de su autor y no comprometen al BANCO DE LA REPUBLICA, su Junta Directiva o la Universidad Nacional de Colombia.
JEL: C22, C53, C82.
Palabras Clave: Revisión de Datos, Now-casting, Analisis Económico en Tiempo Real.

†jjulioro@banrep.gov.co. Investigador, Banco de la República y Profesor Asociado, Departamento de Estadística, Universidad Nacional de Colombia. Bogotá D. C., Colombia

1 Introduction

The revision and delay of macroeconomic data releases have an important effect on the design and analysis of monetary and fiscal policies. Monetary policy, for instance, depends critically on the assessment of the current state of the economy and its short to medium term outlook, which summarizes in a set of indicators within which the GDP, the output gap and the inflation rate play a key role. However, the current view of the economy is blurred by the revision and delay of current and near past GDP figures, and these revisions and delays, in turn, increase the uncertainty of output gap and inflation forecasts. As a result, GDP revisions and delays distort the short to medium term outlook of the economy as well. Therefore, GDP revisions and delays increase the uncertainty over the current state of the economy and its short to medium term outlook. See Harrison et al [9].

Consequently, a policymaker that is aware of the uncertainty over the current and short to medium term outlook of the economy may elicit passive or over-smoothed policies, while a policymaker that ignores these issues, thus taking preliminary GDP figures as "true", may draw economy destabilizing policies. For this reason, models to reduce the effect of data revisions and delays on macroeconomic figures are required.

There are two polar views on the information content of ex-post revision errors $\widetilde{Y}_t - Y_t^{t+k}$, the differences between the true figures and preliminary releases. Revision errors may contain "news" or "noise". If revision errors are pure news, preliminary data releases are the optimal now-casts of the true figures, and revision errors are not forecastable from the information available at the data release. Conversely, if revision errors are pure noise, preliminary data releases are not the optimal now-casts of true figures, and revision errors can be forecasted from the information available at the data release. See Mankiw & Shapiro [14] and Arouba [2], for instance.

Furthermore, revision errors may contain "spill-over effects". Spill-overs relate to correlations between measurement errors of neighboring vintages and improve the forecasts of revision errors.

Jacobs and Van Norden [13] proposed a linear dynamic model to include, in a more realistic and parsimonious way than previous work, the dynamics of news, noise and spill-over effects in measurement errors. These authors assume that the true values are not observable but belong to a class of dynamic models like the ARIMA or the structural models families, and implicitly assume that measurement errors have zero mean.

According to these authors this model provides a framework for the "proper formulation and conduct of monetary and fiscal policy". In fact, three of the most important activities in policy design and analysis can be performed with this model: (i) data description, (ii) optimal forecast and inference, and (iii) cycle-trend decomposition, all of them in an environment of data revisions and delays.

While the assumption of non observability of the true values suits situations in which every historic figure might be revised in the future, it also conveys important modeling and interpretation issues. Three major consequences derive from this assumption. First, the dynamics of the true figures is not identified from the observable data. Second, the mean measurement error is not identified, either, and is therefore set to zero, which is is at odds with the stylized features of ex-post measurement errors. And third, the interpretation of the output gap, for instance, becomes involved. Under this assumption the output gap becomes the *unobserved* cyclical component of an *unobserved* series that follows an *unobserved* dynamics.

However, several statistical bureaus, one of which is Colombia's DANE, reset the starting date of future GDP releases with each methodology change. In this case the data release process is depicted in Figure 1 where it can be observed that the last vintage prior to the new starting date contains reports that might be regarded as true. Therefore, at every period of time t, when policy decisions are made, there is a fixed period of time $M(t)$ before which the true data is observed. See Jacobs & Van Norden [13].

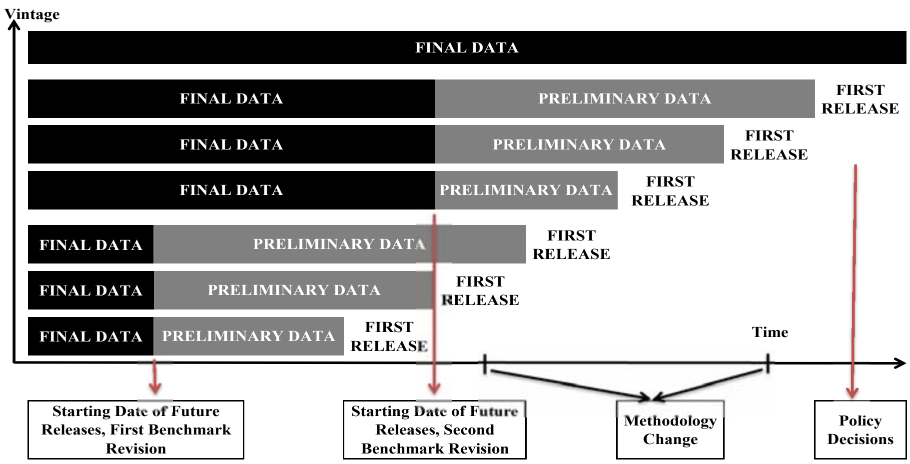

Figure 1: DANE's GDP Data Release Process.

A dynamic linear model for data revisions and delays is proposed in this paper. This model extends Jacobs & Van Norden's [13] in two ways. First, the "true" data series is observable up to a fixed period of time M. And second, preliminary figures might be biased estimates of the true series. Otherwise, the model follows Jacobs & Van Norden's [13] so their gains are extended through the new assumptions. These assumptions represent the data release process more realistically under particular circumstances, and improve the overall identification of the model.

An application to the year to year growth of the quarterly Colombian

GDP reveals features of the Colombian GDP release process that have an important effect on the use of these figures for policy purposes. First, preliminary growth figures under-estimate the true growth. And second, measurement errors contain noise and are thus predictable from the information available at the data release.

More precisely, the downward bias of the five more recent releases are 0.96%, 0.73%, 0.73%, 0.67% and 0.77%, and strong evidence in favor of the presence of noise was found. Moreover, the first data release has a statistically significant downward bias which lies in the 0.57% to 1.14% interval, on average. The models estimated in this paper provide optimal now-casts and forecasts of the true Colombian GDP growth.

Similar downward biases were found in Franses [7], Table 1 and Garratt & Vahey [8].

2 Literature Review

For a given series whose "true" values are denoted as $\widetilde{Y}_t$, statistical bureaus release a set of historical figures $\{Y_1^t, Y_2^t, \ldots, Y_{t-2}^t, Y_{t-1}^t\}$ at every period of time t. This set of preliminary and (possibly) true figures is known as the t^{th} data *vintage*. In this case a delay of one period of time to obtain the preliminary figure for the current period is assumed, and the data release schedule is represented by the following data release matrix

$$
\begin{array}{cccc}
Y_1^2 & Y_1^{t-k+1} & \ldots & Y_1^t \\
\ddots & \vdots & & \vdots \\
 & Y_{t-k}^{t-k+1} & \ldots & Y_{t-k}^t \\
 & & \ddots & \vdots \\
 & & & Y_{t-1}^t
\end{array}
$$

where each column corresponds to a data vintage.

2.1 State Space Forms for Data Revisions

Several types of models have been proposed to explain the dynamics of measurement errors. These models have conveniently been written in terms their time invariant State Space Forms, SSFs,

$$\begin{aligned}\boldsymbol{Y}_t &= \boldsymbol{d} + \boldsymbol{Z}\boldsymbol{\alpha}_t + \boldsymbol{\varepsilon}_t \\ \boldsymbol{\alpha}_{t+1} &= \boldsymbol{c} + \boldsymbol{T}\boldsymbol{\alpha}_t + \boldsymbol{R}\boldsymbol{\eta}_{t+1}\end{aligned} \tag{2.1}$$

In earlier models the observation vector of the SSF contained the $l > 1$ most recent releases in the last vintage, $\boldsymbol{Y}^t = \left[Y^t_{t-1}, Y^t_{t-2}, \ldots, Y^t_{t-l}\right]'$, and was assumed that true values are observable after $l-1$ periods of time of the first release, $\widetilde{Y}_t = Y^{t+l}_t$. See Howrey [12], Trivellato & Rettore [19], Bordignon & Trivellato [3], Patterson [17], Mariano & Tanizaki [15], Busetti [4], Harvey [10], Jacobs & Van Norden [13] and Harvey et. al. [11] for instance.

However, Jacobs & Van Norden [13] found that models based on this observation vector lack parsimony, do not permit "a clean distinction" of the properties of measurement errors, and the assumption that the true value is observable after $l-1$ periods of time of the first release, $\widetilde{Y}_t = Y^{t+l}_t$, is at odds with the stylized facts of measurement errors.

Therefore, these authors propose a linear dynamic model to include, in a more realistic and parsimonious way, the dynamics of news, noise and spill-over effects in measurement errors. In their model, the observation vector contains the releases for time t of the l most recent vintages of data, $\boldsymbol{Y}_t = \left[Y^{t+1}_t, Y^{t+2}_t, \ldots, Y^{t+l}_t\right]'$. In addition, these authors drop the assumption that the true values are observable after $l-1$ periods, $\widetilde{Y}_t = Y^{t+j}_t\ \forall j \geq l$, and assume, instead, that the true values are not observable but belong to a class of dynamic models like the ARIMA or the structural models families. The ARIMA and structural models families include a conveniently extensive variety of dynamic models for the "true" process. Finally, these authors implicitly assume that measurement errors have zero mean.

2.2 News and Noise in Revision Errors

It has been widely acknowledged that revision errors, the differences between the true ex-post figures and preliminary releases for time t, $U_t^{t+j} = \widetilde{Y}_t - Y_t^{t+j}$ for $j = 1, 2, 3, \ldots$, are not "well behaved". This observation leads to the classification of the information content of measurement errors as news or noise. See Mankiw & Shapiro [14], Arouba [2], Siklos [18] and Franses [7] for instance.

Revision errors are well behaved if they satisfy the properties of rational forecast errors and are thus regarded as "news". In this case, measurement errors do not correlate with the releases of previous vintages, $cov(U_t^{t+j}, Y_t^{t+i}) = 0$ for $i \leq j$, and, therefore, revision errors are not predictable from the information available at the time of the release. Under this circumstances, preliminary releases are the optimal now-casts of the true figures. See Mankiw & Shapiro [14] and Arouba [2] for instance.

Conversely, if revision errors lack the properties of rational forecast errors, preliminary releases are not the optimal now-casts of the true figures and revision errors are said to contain "noise". In this case $cov(U_t^{t+j}, Y_t^{t+i}) \neq 0$ which may be accomplished by setting $cov(U_t^{t+j}, U_t^{t+i}) = 0$ for all $i \neq j$.

Statistical test for the hypothesis of noise and news were developed by De Jong [5] and Mincer & Zarnowitz [16]. These tests are based on linear regressions of ex-post measurement errors on the true values and preliminary releases respectively. Although both regressions include an intercept, they are not "collective exhaustive" as both nulls may be rejected when the intercept is non zero. See Jacobs & Van Norden [13] and Arouba [2].

2.3 Spill-over Effects

Spill-over effects arise, for instance, when the revision of one figure in the vintage implies the revision of the report in neighboring vintages. Therefore, spill-over effects help forecast revision errors. See Jacobs & Van Norden [13] for instance.

3 The Statistical Model

The model is described in terms of its *time varying* SSF

$$\boldsymbol{Y}_t = \boldsymbol{d}_t + \boldsymbol{Z}_t \boldsymbol{\alpha}_t + \boldsymbol{\varepsilon}_t \tag{3.1}$$

$$\boldsymbol{\alpha}_{t+1} = \boldsymbol{T}\boldsymbol{\alpha}_t + \boldsymbol{R}\boldsymbol{\eta}_{t+1} \tag{3.2}$$

where 3.1 and 3.2 are the time varying observation equation and the time invariant state equation respectively. Standard normality and independence assumptions are imposed on the vectors of observation and state innovations, $\boldsymbol{\varepsilon}_t$ and $\boldsymbol{\eta}_\tau$, and on the initial state vector $\boldsymbol{\alpha}_0$ as well. These vectors have variance covariance matrices $\boldsymbol{H}_t$, $\boldsymbol{Q}$ and $\boldsymbol{P}_0$, respectively. See Harvey [10], Anderson & Moore [1] and Durbin & Koopman [6] for instance.

We assume that the true values are observed up to a fixed period of time $1 < M = M(T) < T$, where T is the effective sample size, and it is also assumed that measurement errors may not have zero mean under noise, $d_j = \mathrm{E}[\widetilde{Y}_t^\dagger - Y_t^{t+j}] \neq 0$.

To introduce these assumptions into Jacobs & Van Norden's model let $\widetilde{Y}_t$ be the *observed* true value of the series at time t, for $t = 1, 2, \ldots, M$, and let $\widetilde{Y}_t^\dagger$ denote the true underlying value at t, $\forall t$. Therefore,

$$\widetilde{Y}_t = \widetilde{Y}_t^\dagger \text{ whenever } 1 \leq t \leq M$$

and otherwise $\widetilde{Y}_t$ is not observed.

Let us also denote $\boldsymbol{Y}_{1,t} = \left[Y_t^{t+1}, Y_t^{t+2}, \ldots, Y_t^{t+l}\right]'$ the vector containing the reports for time t of the l more recent vintages of data. Therefore, the observation vector in 3.1 is defined as

$$\boldsymbol{Y}_t = \begin{cases} \begin{bmatrix} \widetilde{Y}_t \\ \boldsymbol{Y}_{1,t} \end{bmatrix} & \text{if } 1 \leq t \leq M \\ \boldsymbol{Y}_{1,t} & \text{if } M < t \leq T \end{cases}$$

whose size is $N_t = l + I(t)_{\{t \leq M\}}$, where $I(t)_{\{t \leq M\}}$ is the indicator function of $t \in \{t \leq M\}$.

From 3.1 it can be observed that $\boldsymbol{d}_t$, $\boldsymbol{Z}_t$ and $\boldsymbol{\varepsilon}_t$ have also N_t rows, and the covariance matrix of the observation innovations, $\boldsymbol{H}_t$, has size N_t. However, apart from their size, $\boldsymbol{d}_t$, $\boldsymbol{Z}_t$ and $\boldsymbol{H}_t$ are time invariant as we will see in the following.

Therefore, model 3.1-3.2 differs from a time invariant SSF as the size of the observation vector, N_t, is time varying. This difference, however, does not hinder the application of the Kalman filter and the prediction error decomposition. A careful tracking of matrix and vector sizes suffices for these algorithms to work in this case. See Harvey [10].

Following Jacobs & Van Norden [13], the state vector has four components,

$$\boldsymbol{\alpha}_t = \left[\widetilde{Y}_t^{\dagger}, \boldsymbol{\phi}_t', \boldsymbol{\nu}_t', \boldsymbol{\zeta}_t'\right]' \tag{3.3}$$

with sizes 1, b, l and l respectively, where the unobserved component $\boldsymbol{\phi}_t$ determines the dynamics of $\widetilde{Y}_t^{\dagger}$, and $\boldsymbol{\nu}_t$ and $\boldsymbol{\zeta}_t$ are the unobserved news and noise components respectively.

Letting $d_j = \mathrm{E}[\widetilde{Y}_t^{\dagger} - Y_t^{t+j}]$ denote the mean measurement error of the report for time t of the $t + j^{th}$ vintage, and $\boldsymbol{d}_1 = [d_1, d_2, \ldots, d_l]'$ the vector containing the mean measurement errors related to the last l vintages, for time t, by setting

$$\boldsymbol{d}_t = \begin{cases} \begin{bmatrix} 0 \\ \boldsymbol{d}_1 \end{bmatrix} & \text{if } t \leq M \\ \boldsymbol{d}_1 & \text{if } t > M \end{cases}, \tag{3.4}$$

$$\boldsymbol{Z}_t = \begin{cases} \begin{bmatrix} 1 & \mathbf{0}_{1\times b} & \mathbf{0}_{1\times l} & \mathbf{0}_{1\times l} \\ \mathbf{1}_{l\times 1} & \mathbf{0}_{l\times b} & \boldsymbol{I}_l & \boldsymbol{I}_l \end{bmatrix} & \text{if } t \leq M \\ \begin{bmatrix} \mathbf{1}_{l\times 1} & \mathbf{0}_{l\times b} & \boldsymbol{I}_l & \boldsymbol{I}_l \end{bmatrix} & \text{if } t > M \end{cases} \tag{3.5}$$

and

$$\boldsymbol{H}_t = \begin{cases} \mathbf{0}_{(l+1)\times(l+1)} & \text{if } t \leq M \\ \mathbf{0}_{l\times l} & \text{if } t > M \end{cases} \tag{3.6}$$

the observation equation 3.1 becomes

$$\begin{aligned} \widetilde{Y}_t &= \widetilde{Y}_t^{\dagger} \quad \text{if} \quad 1 \leq t \leq M \\ \boldsymbol{Y}_{1,t} &= \boldsymbol{d}_1 + \widetilde{Y}_t^{\dagger} \mathbf{1}_{l\times 1} + \boldsymbol{\nu}_t + \boldsymbol{\zeta}_t \end{aligned} \tag{3.7}$$

where the first equation states that the true figures are observed up to time M, and the second becomes "Release=Bias+Truth+News+Noise" for all t.

The state equation is determined by

$$\boldsymbol{T} = \begin{bmatrix} T_{11} & \boldsymbol{T}_{12} & \mathbf{0} & \mathbf{0} \\ \boldsymbol{T}_{21} & \boldsymbol{T}_{22} & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \boldsymbol{T}_{33} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \mathbf{0} & \boldsymbol{T}_{44} \end{bmatrix} \tag{3.8}$$

where the blocks of $\boldsymbol{T}$ have row sizes $1, b, l, l$ and column sizes $1, b, l, l$, respectively, and

$$\boldsymbol{R} = \begin{bmatrix} \boldsymbol{R}_1 & \boldsymbol{R}_3 & \mathbf{0} \\ \boldsymbol{R}_2 & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & -\boldsymbol{U}_1 \times \text{diag}(\boldsymbol{R}_3) & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \boldsymbol{R}_4 \end{bmatrix} \tag{3.9}$$

whose blocks have row sizes $1, b, l, l$ and column sizes $r - 2l, l, l$, respectively, $\boldsymbol{U}_1$ is an upper triangular matrix full of ones, $\boldsymbol{R}_3 = [\sigma_{\nu 1}, \sigma_{\nu 2}, \ldots, \sigma_{\nu l}]$, and $\boldsymbol{R}_4$ is an $l \times l$ time invariant matrix to be specified below.

Conformably, the vector of state innovations is partitioned as $\boldsymbol{\eta}_t = \left[\boldsymbol{\eta}'_{et}, \boldsymbol{\eta}'_{\nu t}, \boldsymbol{\eta}'_{\zeta t}\right]'$ where $\boldsymbol{\eta}_{et}$ are the innovations to the underlying true values, and $\boldsymbol{\eta}_{\nu t}$ and $\boldsymbol{\eta}_{\zeta t}$ are the innovations to the unobserved news and noise components respectively. In this case the variance covariance matrix of the state innovation vector is $\boldsymbol{Q} = \boldsymbol{I}_r$.

Therefore, the state equation 3.2 summarizes in

$$\begin{aligned}
\widetilde{Y}^{\dagger}_{t+1} &= T_{11}\widetilde{Y}^{\dagger}_t + \boldsymbol{T}_{12}\boldsymbol{\phi}_t + \boldsymbol{R}_1\boldsymbol{\eta}_{et} + \boldsymbol{R}_3\boldsymbol{\eta}_{\nu t} \\
\boldsymbol{\phi}_{t+1} &= \boldsymbol{T}_{21}\widetilde{Y}^{\dagger}_t + \boldsymbol{T}_{22}\boldsymbol{\phi}_t + \boldsymbol{R}_2\boldsymbol{\eta}_{et} \\
\boldsymbol{\nu}_{t+1} &= \boldsymbol{T}_{33}\boldsymbol{\nu}_t - \boldsymbol{U}_1 \times \operatorname{diag}(\boldsymbol{R}_3)\boldsymbol{\eta}_{\nu t} \\
\boldsymbol{\zeta}_{t+1} &= \boldsymbol{T}_{44}\boldsymbol{\zeta}_t + \boldsymbol{R}_4\boldsymbol{\eta}_{\zeta t} \qquad (3.10)
\end{aligned}$$

where

- The first and second equations determine the dynamics of the true underlying values of the series.
- News correlate with the true underlying series.
- Noise does not correlate with the true underlying series.
- News and noise are mutually independent and behave like VAR(1) models with identifying restrictions determined by $-\boldsymbol{U}_1 \times \operatorname{diag}(\boldsymbol{R}_3)$ and $\boldsymbol{R}_4$ respectively.

In order to understand the dynamics of news, noise and spill-over effects and their relationship with the observed data, it is advisable to study them independently. See Jacobs & Van Norden [13].

3.1 Pure News

In this case, $\boldsymbol{\zeta}_t$ and $\boldsymbol{T}_{33}$ are dropped from the model and the relevant equations become

$$\begin{aligned}
\boldsymbol{Y}_{1,t} &= \boldsymbol{d}_1 + \widetilde{Y}_t^{\dagger}\mathbf{1}_{l\times 1} + \boldsymbol{\nu}_t \\
\widetilde{Y}_{t+1}^{\dagger} &= T_{11}\widetilde{Y}_t^{\dagger} + \boldsymbol{T}_{12}\boldsymbol{\phi}_t + \boldsymbol{R}_1\boldsymbol{\eta}_{et} + \boldsymbol{R}_3\boldsymbol{\eta}_{\nu t} \\
\boldsymbol{\nu}_{t+1} &= -\boldsymbol{U}_1 \times \mathrm{diag}(\boldsymbol{R}_3)\boldsymbol{\eta}_{\nu t}
\end{aligned}$$

where the measurement errors of the l most recent consecutive vintages are the elements of $\boldsymbol{U}_t = -\boldsymbol{d}_1 - \boldsymbol{\nu}_t$, and $\mathrm{E}[\boldsymbol{U}_t] = -\boldsymbol{d}_1$.

Since

$$-\boldsymbol{U}_1 \times \mathrm{diag}(\boldsymbol{R}_3) = -\begin{bmatrix} \sigma_{\nu 1} & \sigma_{\nu 2} & \dots & \sigma_{\nu l} \\ 0 & \sigma_{\nu 2} & \ddots & \vdots \\ \vdots & \ddots & \ddots & \sigma_{\nu l} \\ 0 & \dots & 0 & \sigma_{\nu l} \end{bmatrix}$$

$\mathrm{cov}(U_t^{t+j}, Y_t^{t+i}) = 0$ for $i \leq j$. Therefore, measurement errors do not correlate with the releases of previous vintages and are thus not predictable from the information available at the time of the release.

3.2 Pure Noise

Under pure noise the measurement errors of consecutive vintages are not correlated, $\mathrm{Cov}(U_t^{t+j}, U_t^{t+j+1}) = 0 \ \forall t$ and $\forall j$.

In this case $\boldsymbol{\nu}_t$ and $\boldsymbol{R}_4$ are dropped from the state vector, and the relevant equations of the model become

$$\begin{aligned}
\boldsymbol{Y}_{1,t} &= \boldsymbol{d}_1 + \widetilde{Y}_t^{\dagger}\mathbf{1}_{l\times 1} + \boldsymbol{\zeta}_t \\
\widetilde{Y}_{t+1}^{\dagger} &= T_{11}\widetilde{Y}_t^{\dagger} + \boldsymbol{T}_{12}\boldsymbol{\phi}_t + \boldsymbol{R}_1\boldsymbol{\eta}_{et} \\
\boldsymbol{\zeta}_{t+1} &= \boldsymbol{R}_4\boldsymbol{\eta}_{\zeta t}
\end{aligned}$$

thus the vector containing the measurement errors of contiguous vintages becomes $\boldsymbol{U}_t = -\boldsymbol{d}_1 - \boldsymbol{\zeta}$. The condition for measurement errors to be noise is $\boldsymbol{R}_4 = \text{diag}(\sigma_{\zeta 1}, \sigma_{\zeta 2}, \ldots, \sigma_{\zeta l})$.

By definition, measurement errors are not correlated with the true values, which implies that measurement errors correlate with the available vintage. Therefore, revision errors, $Y_t^{t+j+1} - Y_t^{t+j}$ are forecastable.

If preliminary figures become more precise over time, the condition $\sigma_{\zeta l} \leq \sigma_{\zeta, l-1} \leq, \ldots, \leq \sigma_{\zeta 2} \leq \sigma_{\zeta 1}$ might also be imposed.

3.3 Spill-overs

Spill-over effects can be parameterized by specifying the matrices $\boldsymbol{T}_{33}$ or $\boldsymbol{T}_{44}$ of equation 3.8 for news and noise, respectively.

For instance, in the case of noise and spill-overs, simple correlation can be specified as $\boldsymbol{T}_{44} = \rho_\zeta \boldsymbol{I}_l$. In the case of higher order correlation, additional copies, $\boldsymbol{\zeta}_{t-k}$ are added to the state vector, and the corresponding matrices are specified correspondingly.

3.4 ARIMA Model Specification

The state equation 3.10 allows a variety of specifications for the dynamics of the underlying true values $\widetilde{Y}_t^\dagger$ through the appropriate parametrization of $\boldsymbol{\phi}_t$, T_{11}, $\boldsymbol{T}_{12}$, $\boldsymbol{T}_{21}$, $\boldsymbol{T}_{22}$, $\boldsymbol{R}_1$ and $\boldsymbol{R}_2$. These parameterizations include the ARIMA and the structural models families.

For instance, if $\widetilde{Y}_t^\dagger$ is assumed to be an $ARMA(1,4)$ model,

$$(1 - \phi_1 B)\widetilde{Y}_t^\dagger = (1 + \theta_1 B + \theta_2 B^2 + \theta_3 B^3 + \theta_4 B^4)e_t$$

these vectors and matrices become $T_{11} = [\phi_1]$, $\boldsymbol{T}_{12} = [\theta_1, \theta_2, \theta_3, \theta_4]$, $\boldsymbol{T}_{21} =$

$\mathbf{0}_{4\times1}$, $\boldsymbol{\phi}_t = [e_t, e_{t-1}, e_{t-2}, e_{t-3}]^T$, $\boldsymbol{R}_1 = [\sigma_e]$,$\boldsymbol{\eta}_{et} = [\eta_{et}]$,

$$\boldsymbol{T}_{22} = \begin{bmatrix} 0 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \end{bmatrix}$$

and $\boldsymbol{R}_2 = [\sigma_e, 0, 0, 0]^T$

Therefore, under news we have

$$\begin{aligned} \widetilde{Y}^{\dagger}_{t+1} &= \phi_1 \widetilde{Y}^{\dagger}_t + \sum_{i=0}^{3} \theta_{i+1} e_{t-i} + \sigma_e \eta_{et} + \boldsymbol{R}_3 \boldsymbol{\eta}_{\nu t} \\ \begin{bmatrix} e_{t+1} \\ e_t \\ e_{t-1} \\ e_{t-2} \end{bmatrix} &= \begin{bmatrix} 0 \\ 0 \\ 0 \\ 0 \end{bmatrix} \widetilde{Y}^{\dagger}_t + \begin{bmatrix} 0 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \end{bmatrix} \begin{bmatrix} e_t \\ e_{t-1} \\ e_{t-2} \\ e_{t-3} \end{bmatrix} + \begin{bmatrix} \sigma_e \\ 0 \\ 0 \\ 0 \end{bmatrix} \eta_{et} \end{aligned}$$

For the specification of other members of the ARIMA family and the structural models family see Jacobs & Van Norden [13].

4 Results

4.1 Data

The data set analyzed in this paper contains Colombian growth vintages from 2002Q2 to 2010Q1 released by DANE, the Colombian statistics bureau. These DGP releases exhibit a delay of one quarter, thus the 2002Q2 vintage, for instance, contains GDP growth reports from 1995Q1 to 2002Q1. The data set comprises two different methodologies. The first, called "base-1994" methodology, contains vintages from 2002Q2 to 2008Q1, whose reports start at 1995Q1, while the second, named "base-2000" methodology, contains vintages from 2008Q2 to 2009Q4, whose reports start at 2001Q1.

GDP growth releases are considered true after 5 years of the first release. This choice arises from the decomposition of measurement errors as between and within methodologies.

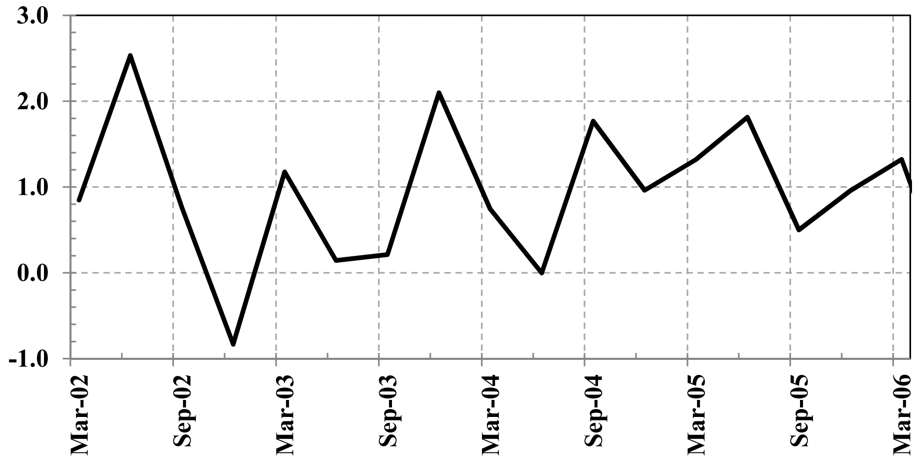

Figure 2: Final Revision Error, $\widetilde{Y}_t - Y_t^{t+1}$ in Colombian growth Releases.

4.2 News and Noise in Colombia's Growth

Methodology changes have an important effect on final revision errors. The extent of final revision error, the difference between the true growth and the first growth release $\widetilde{Y}_t - Y_t^{t+1}$, is depicted in Figure 2. Final revision errors tend to be big and positive, on average about 1%, which shows that the first release of GDP data tends to under-estimate the true growth. The highest final revision error, for instance, happens for the GDP report of 2002Q2, which was published for the first time in the 2002Q3 vintage. The final revision error for this quarter is a remarkable 2.53%, which corresponds to an initial report of 2.21% and a true one of 4.74%.

Furthermore, consecutive revisions tend to be small as Figure 3 shows. The first five releases of GDP growth intertwine closely together and thus consecutive revision errors, $Y_t^{t+j+1} - Y_t^{t+j}$, tend to be small and may also have zero mean. This contrasts sharply with the final revision error. The true growth seldom crosses the lines of the first five preliminary releases.

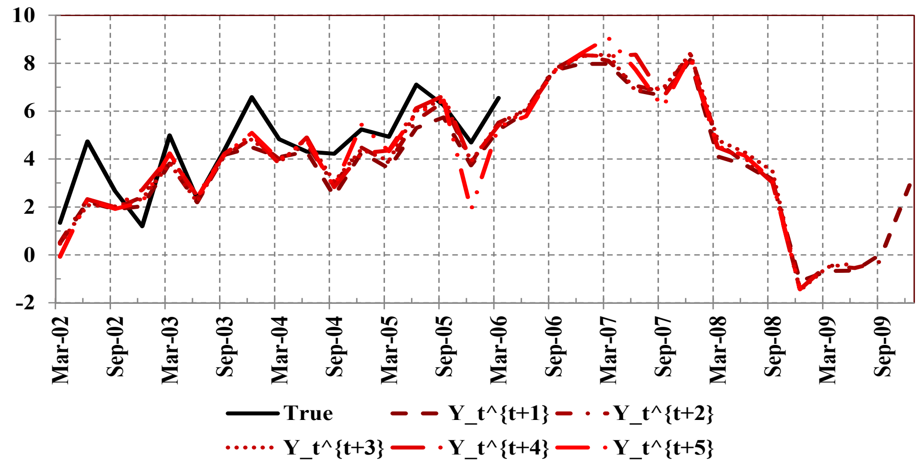

Figure 3: "True" Growth and the First Five Corresponding Releases

Evidence on the importance of news and noise in measurement errors may be found in Figure 4. This Figure displays the correlations between consecutive measurement errors $Y_t^{t+j} - Y_t^{t+j-1}$, on one hand, with the true figures $\widetilde{Y}_t$ and their current release Y_t^{t+j} on the other, for $j = 1, 2, 3, ..., 10$.

The dynamics of revision errors in Colombian growth data is complex, and a mixed model, news+noise, might be appropriate. The correlations of Figure 4 tend to be high, starting at 0.4 with a minimum of -0.4. The fact that both correlations tend to be different from zero for most of the revisions indicates the rejection of both hypothesis, pure news and pure noise. However, the fourth and eighth revisions exhibit a zero correlation with the true figures while the correlation with the current release is different from zero. This result suggests a pure noise model for these revisions. However, the ninth and tenth revisions display correlations close to zero suggesting that none of the two models, pure noise or pure news, is rejected.

Finally, Figure 5 shows evidence that suggests the presence of slight spillover effects. The Figure contains the auto-correlation of the first revision

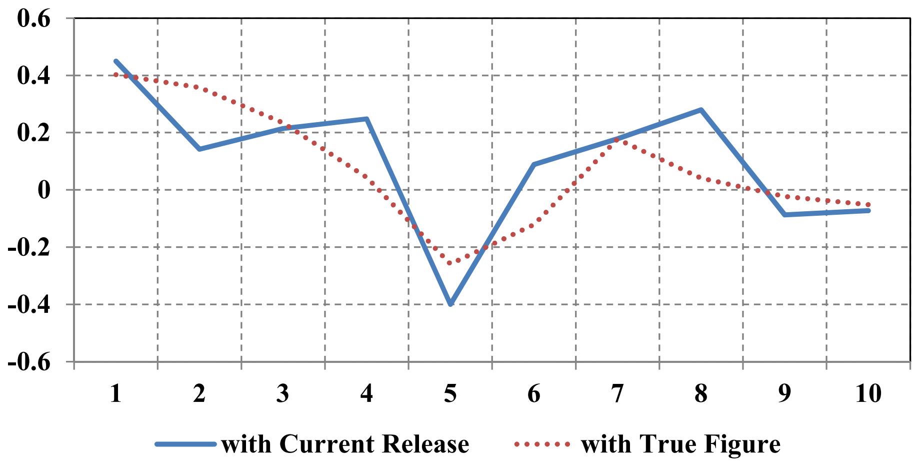

Figure 4: Correlation of Consecutive Measurement Errors

error, $Y_t^{t+2} - Y_t^{t+1}$, the second $Y_t^{t+3} - Y_t^{t+2}$, the third and fourth revision errors, all to the seventh lag. These autocorrelations tend to be small, but are enough to consider the presence of spill-overs.

Summarizing, Colombian growth data shows evidence in favor of a mixed model, noise+news, final mean measurement errors different from zero, and some evidence in favor of spill-overs.

4.3 Estimation Results

Six models were estimated for the revision of the year to year growth of the Colombian quarterly GDP. The first two contain news, the third and fourth contain noise and the last two contain both news and noise. Members of each pair differ from each other because of the inclusion of spill-overs. The observation vector comprises the releases for time t of the first five vintages of data, Y_t^{t+j} for $j = 1, 2, 3, 4, 5$ and the true figure $\widetilde{Y}_t$ until M =2006Q1. After this date the observation vector contains the releases of the five more recent vintages of data only.

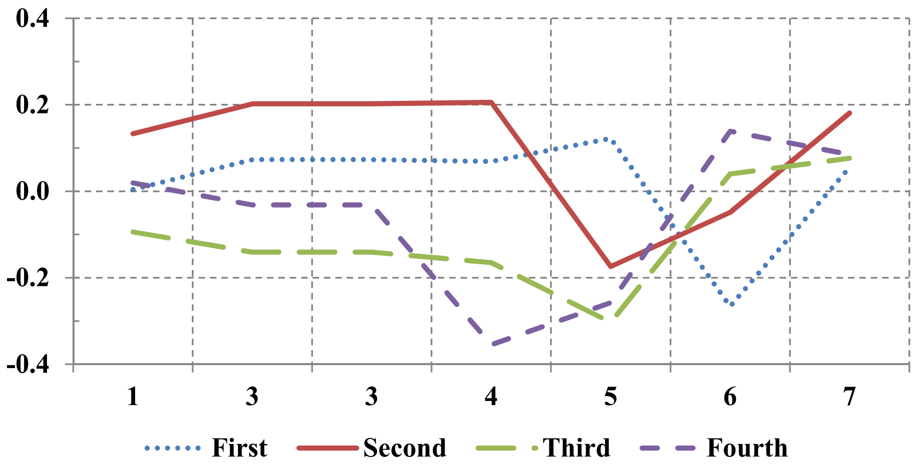

Figure 5: Autocorrelation of Measurement Errors

From the identification of the true series the model for the true underlying growth is specified as an $ARMA(1,1)$. The standard deviations in the $\boldsymbol{R}$ matrix are re-parameterized as $\sigma_e = e^{\theta_\varepsilon}$, $\sigma_{\nu,j} = e^{\theta_{\nu,j}}$ and $\sigma_{\zeta,j} = e^{\theta_{\zeta,j}}$ for $j = 1, 2, ..., l = 5$ in order to avoid restricted maximization procedures.

Parameter estimation was carried out by maximum likelihood methods based on the prediction error decomposition. The maximization of the likelihood function was performed by the Newton-Raphson method which provides a numerical approximation to the Hessian matrix from which the standard deviations of parameter estimators were derived. Moreover, the log-likelihood, AIC and BIC information criteria were calculated in order to compare the models.

Convergence to the maximum likelihood estimates of the parameters was achieved after a few steps, 6 or 7, for the first two pairs of models regardless of the starting point. In the largest models, news + noise, convergence was slower and, depending on the starting point, sometimes reached saddle points. After convenient starting values were chosen a maximum was reached in 24 iterations. Comparison of the likelihood function to those obtained over a grid of plausible parameter values suggest that a global maximum was reached.

Deterministic effects were subtracted prior to estimation so that all series have zero mean. This was performed in two steps. The long run mean of the true growth was subtracted from all series. And then, the mean difference between preliminary and true series, $Y_t^{t+j} - \widetilde{Y}_t$ was subtracted from preliminary figures. Table 1 displays the estimated mean of the true growth and the mean bias of the first five preliminary figures. Biases tend to be high, close to 1.0% and the long run mean of the true growth is 3.57%.

Positive mean bias in preliminary growth figures were found by Franses [7], Table 1.

Parameter	**Estimate**
$E[\widetilde{Y}_t]$	3.5700
$E[\widetilde{Y}_t - Y_t^{t+1}]$	0.9601
$E[\widetilde{Y}_t - Y_t^{t+2}]$	0.7332
$E[\widetilde{Y}_t - Y_t^{t+3}]$	0.7290
$E[\widetilde{Y}_t - Y_t^{t+4}]$	0.6685
$E[\widetilde{Y}_t - Y_t^{t+5}]$	0.7742

Table 1: Mean of the True Growth and Mean Bias of the First Five Preliminary Releases

4.3.1 News and Noise Models

The estimation results are contained in Tables 1 to 4. The second and third columns of Tables 2-4 contain the estimated parameter and its corresponding standard deviations for models without spill-over effects, and the fourth and fifth columns display the estimated parameter and their corresponding standard deviations for models with spill-over effects. The following results arise from these tables.

- The first five releases of the GDP growth under estimate the true growth. The mean biases $E[\widetilde{Y}_t - Y_t^{t+j}]$ are not only positive but also big in size, 0.96%, 0.73%, 0.73%, 0.66%, and 0.77% for $j = 1, 2, ..., 5$ respectively. This result shows that the measurement errors are slowly corrected during the first five releases of data and important corrections arise in the long run.

- The $AR(1)$ estimated parameters $\widehat{\phi}_1$ are between 0.82 and 0.86 which reveals a high persistence of growth innovations. The $MA(1)$ parameters $\widehat{\theta}_1$, however, are not statistically significant. The t statistics for these parameters lie in the -1.23 to -0.86 interval.

- Spill-overs have no significant effect on the dynamics of measurement errors. The t statistics for these parameters lie in the -1.27 to 0.48 interval. This result also follows from the comparison of "AIC" and "BIC" within each pair of models.

Table 2 contains the estimation results for the news models. From subsection 3.1 news innovations enter in the true underlying process peeling off information as preliminary figures become more precise. Therefore, the true underlying process has an innovation standard deviation smaller than under noise. The estimated standard deviation under news is $1.51 = e^{\theta_\varepsilon}$ while the estimated standard deviation under noise is $1.81 = e^{\theta_\varepsilon}$ which is, in turn, close to the standard deviation under news + noise. See tables 3 and 4.

There is strong evidence in favor of the presence of news. The hypothesis that news innovations are not significant is equivalent to the null of zero news innovation variance which is rejected in table 2.

	Pure News		**News + Spill-overs**	
Parameter	**Estimate**	**Std-Err**	**Estimate**	**Std-Err**
ϕ_1	0.8237	0.1051	0.8278	0.1042
θ_1	-0.1865	0.2150	-0.1906	0.2144
ρ_ν			0.0411	0.0862
θ_ε	0.4127	0.1540	0.4082	0.1539
$\theta_{\nu,1}$	-1.6253	0.1313	-1.6237	0.1314
$\theta_{\nu,2}$	-1.7431	0.1313	-1.7495	0.1318
$\theta_{\nu,3}$	-1.0057	0.1312	-1.0014	0.1316
$\theta_{\nu,4}$	-0.6943	0.1311	-0.6946	0.1312
$\theta_{\nu,5}$	0.0053	0.1788	0.0111	0.1797
log-likelihood	200.6627		200.7768	
AIC	-385.3254		-383.5535	
BIC	-347.4486		-340.9422	

Table 2: Maximum Likelihood Estimation of News Models

Table 3 contains the estimation results for the noise models. There is strong evidence in favor of the presence of noise in measurement errors. The hypothesis of no significant noise effects is equivalent to the null of zero noise innovation standard deviation which is clearly rejected from table 3. Moreover, the estimated standard deviations of the noise innovations $e^{\theta_{\zeta,j}}$ are smaller than the corresponding standard deviations of the news innovations. This result might suggest that news innovations are more important than noise innovations in the explanation of the dynamics of measurement errors.

Table 4 contains the estimation results for the news + noise models. Because of the presence of news the true underlying process innovation has an estimated standard deviation of $1.48 = e^{\theta_\varepsilon}$, close to those in Table 2.

There is strong evidence in favor of the presence of both, news and noise, in measurement errors. The null of no significant news and noise effects is clearly rejected from table 4. However, news innovations might

	Pure Noise		Noise + Spill-overs	
Parameter	**Estimate**	**Std-Err**	**Estimate**	**Std-Err**
ϕ_1	0.8672	0.1002	0.8670	0.1003
θ_1	-0.2334	0.1892	-0.2328	0.1893
ρ_ζ			-0.1093	0.0854
θ_ε	0.5883	0.1335	0.5885	0.1337
$\theta_{\zeta,1}$	-0.4128	0.1337	-0.4235	0.1339
$\theta_{\zeta,2}$	-0.4071	0.1333	-0.4147	0.1334
$\theta_{\zeta,3}$	-0.3589	0.1337	-0.3692	0.1338
$\theta_{\zeta,4}$	-0.2885	0.1338	-0.2883	0.1339
$\theta_{\zeta,5}$	-0.2214	0.1333	-0.2258	0.1334
log-likelihood	85.2595		86.0720	
AIC	-154.5189		-154.1440	
BIC	-116.6422		-111.5326	

Table 3: Maximum Likelihood Estimation of Noise Models

be relatively more important than noise innovations in the explanation of the dynamics of measurement errors. The estimated standard deviation of news innovations is more than 200 times higher than the estimated standard deviation of noise innovations for the first and second data releases. For the remaining three releases the standard deviations are similar. Therefore news innovations dominate during the first two releases but after the third release noise innovations become important.

An over all comparison of the models suggests that models in which news are present are preferred. The highest log-likelihood and smaller AIC arise in the model that includes news and noise but no spill-over effects. However, the BIC information criteria minimizes for the pure noise model without spill-overs. Since noise innovations become important after the third release, noise plays an important role in the determination of the dynamics of measurement errors. These results suggest that a model that includes both news and noise is appropriate to now-cast and forecast the true Colombian

	News + Noise		News + Noise + Spill-overs	
Parameter	**Estimate**	**Std-Err**	**Estimate**	**Std-Err**
ϕ_1	0.8496	0.0997	0.8538	0.0985
θ_1	-0.1997	0.2045	-0.2031	0.2037
ρ_ζ			0.0653	0.1345
θ_ε	0.4051	0.1494	0.3982	0.1496
$\theta_{\nu,1}$	-1.6254	0.1313	-1.6218	0.1318
$\theta_{\nu,2}$	-3.4461	4.9454	-3.0263	2.0690
$\theta_{\nu,3}$	-1.5077	0.3311	-1.5162	0.3299
$\theta_{\nu,4}$	-6.7610	333.5675	-6.4329	28.6912
$\theta_{\nu,5}$	-0.1600	0.1846	-0.1483	0.1867
$\theta_{\zeta,1}$	-7.7569	168.4589	-7.4627	16.7954
$\theta_{\zeta,2}$	-8.9184	227.1944	-8.8269	85.2251
$\theta_{\zeta,3}$	-1.7598	0.2143	-1.7842	0.2206
$\theta_{\zeta,4}$	-1.4323	0.2808	-1.4371	0.2787
$\theta_{\zeta,5}$	-0.7941	0.1500	-0.7980	0.1501
log-likelihood	208.0583		208.1758	
AIC	-390.1166		-388.3516	
BIC	-328.5669		-322.0673	

Table 4: Maximum Likelihood Estimation of News + Noise Models

growth.

4.3.2 The Final Model

In this sub-section a last feature is included to obtain the final version of the model. This feature relates to the fact that partial information is observed during the last $l-1=4$ quarters of the sample. At the last period of the sample, $t=T$, only Y_T^{T+1} is observed, at $t=T-1$ only Y_{T-1}^T and Y_{T-1}^{T+1} are available, at $t=T-2$ three preliminary releases, Y_{T-2}^{T-1}, Y_{T-2}^T and Y_{T-2}^{T+1}, are available, and at $t=T-3$ four preliminary releases, Y_{T-3}^{T-2}, Y_{T-3}^{T-1}, Y_{T-3}^T and Y_{T-3}^{T+1}, are available. For $t=M+1,\ldots,T-1$ the whole vector $\boldsymbol{Y}_{1t}$ is observed, and prior to that date $\widetilde{Y}_t$ is also observed.

In order to include the last four observations of the sample, the observation vector becomes

$$\mathbf{Y}_t = \begin{cases} \begin{bmatrix} \widetilde{Y}_t \\ \mathbf{Y}_{1,t} \end{bmatrix} & \text{if } 1 \le t \le M \\ \mathbf{Y}_{1,t} & \text{if } M < t \le T-4 \\ \begin{bmatrix} Y_t^{t+1} \\ \vdots \\ Y_t^{t+4} \end{bmatrix} & \text{if } t = T-3 \\ \begin{bmatrix} Y_t^{t+1} \\ Y_t^{t+2} \\ Y_t^{t+3} \end{bmatrix} & \text{if } t = T-2 \\ \begin{bmatrix} Y_t^{t+1} \\ Y_t^{t+2} \end{bmatrix} & \text{if } t = T-1 \\ \begin{bmatrix} Y_t^{t+1} \end{bmatrix} & \text{if } t = T \end{cases}$$

The estimation results for this model are shown in Table 5. Convergence to a maximum is achieved in just 7 iterations starting at the parameter estimates of table 4. The results of table 5 summarize as follows.

	News + Noise	
Parameter	**Estimate**	**Std-Err**
ϕ_1	0.7628	0.0934
θ_ε	0.4339	0.1379
$\theta_{\nu,1}$	-1.5358	0.1270
$\theta_{\nu,2}$	-3.2823	4.2699
$\theta_{\nu,3}$	-1.5178	0.3426
$\theta_{\nu,4}$	-7.0173	570.8885
$\theta_{\nu,5}$	-0.1682	0.1825
$\theta_{\zeta,1}$	-8.1420	116.1066
$\theta_{\zeta,2}$	-9.2174	479.9131
$\theta_{\zeta,3}$	-1.7629	0.2403
$\theta_{\zeta,4}$	-1.4256	0.2820
$\theta_{\zeta,5}$	-0.7754	0.1524
log-likelihood	209.2037	
AIC	-394.4074	
BIC	-335.2298	

Table 5: Maximum Likelihood Estimation of News + Noise Final Model

The last four observations are very informative with respect to the persistence of the true growth. The $AR(1)$ parameter estimate in Table 4 is 0.85, which reduces to 0.76 in Table 5 after introducing the last four observations. This might be due to the fact that the last four observations contain information of a rapid recovery from the world financial breakdown. The persistence estimate of Table 5 agrees with the estimation results of an $ARMA$ model for the true growth not shown in this paper.

In addition, the last four observations increase the innovation variance of the true underlying process. The standard deviation estimate of the innovations in Table 4 is $e^{0.4}$ while the innovation standard deviation in Table 5 is $e^{0.43}$. However, the rest of the parameters ar of similar value and significance.

Finally, the AIC and BIC of Table 5 are smaller than those of the second column in Table 4. Although it may suggest that the latest model is better, these results are not totally comparable as the later miss the last four observations.

Figure 6 displays the preliminary and definitive data along with the corresponding now-casts of the Colombian growth. Table 6 contains the preliminary data and their corresponding now-casts for the period of time when no definitive data is available. This Table contains the difference between the now-cast and the first data release, Now-cast-Y_t^{t+1}, in the last column.

Now-casts are always above the first data release. The last column of Table 6 shows that the bias of the first release ranges from 0.38% at 2008Q4 to 1.51% at 2007Q2. On average, preliminary figures are 0.87% below the true underlying growth. See Table 1 also.

The last column of Table 6 reveals rich dynamics resulting from noise

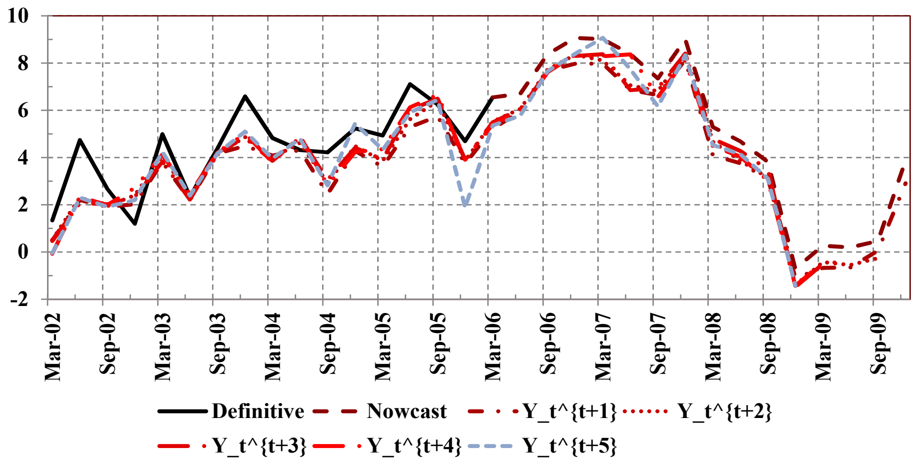

Figure 6: Observed Data and Now-casts of Year to Year growth of the Colombian Quarterly GDP

innovations.

Figure 7 shows the optimal now-cast of the GDP, its confidence interval and the first data release, Y_t^{t+1}, for each period of time t. The standard deviation of the now-cast is small, on average 0.29%, which provides a 95% confidence interval 1.14% wide. The first data release falls into the confidence interval just two times, 2008Q4 and 2009Q3. The rest of the time the first data release falls below the confidence interval. This result shows that the first data release has a statistically significant downward bias between 0.57% and 1.14%, on average.

5 Final Remarks

In several situations statistical bureaus reset the start of subsequent data releases with each benchmark methodology change. Under this circumstances the last vintage previous to the reset date contains figures that might be taken as true. In this case the data release process resembles Figure 1.

Date	**Now-cast**	$\boldsymbol{Y_t^{t+1}}$	$\boldsymbol{Y_t^{t+2}}$	$\boldsymbol{Y_t^{t+3}}$	$\boldsymbol{Y_t^{t+4}}$	$\boldsymbol{Y_t^{t+5}}$	**Now-cast-**$\boldsymbol{Y_t^{t+1}}$
2006Q2	6.70	5.96	6.06	6.06	5.96	5.79	0.74
2006Q3	8.38	7.68	7.72	7.62	7.62	7.66	0.70
2006Q4	9.07	7.97	8.36	8.30	8.35	8.40	1.10
2007Q1	9.01	7.98	8.11	8.37	8.29	9.07	1.03
2007Q2	8.38	6.87	7.06	6.85	8.36	7.74	1.51
2007Q3	7.35	6.65	6.81	6.94	6.56	6.16	0.70
2007Q4	9.00	8.14	8.41	8.40	8.14	8.34	0.86
2008Q1	5.30	4.10	4.67	4.79	4.49	4.54	1.19
2008Q2	4.69	3.76	3.86	4.28	4.04	4.15	0.92
2008Q3	3.82	3.10	3.15	3.50	3.03	3.11	0.71
2008Q4	-0.71	-1.10	-1.40	-1.48	-1.43	-1.42	0.39
2009Q1	0.27	-0.67	-0.42	-0.50	-0.43		0.94
2009Q2	0.20	-0.65	-0.54	-0.35			0.85
2009Q3	0.47	0.06	-0.26				0.41
2009Q4	3.87	2.91					0.96

Table 6: Now-casts of the Year to Year Growth of the Quarterly Colombian GDP

Moreover, it has also been observed that that preliminary releases are biased estimates of the true growth. A dynamic linear model that fits this behavior was proposed in this paper.

The model presented in this paper extends Jacobs & Van Norden's [13] in two ways. First, the "true" data series is observable up to a fixed period of time M. And second, preliminary figures might be biased estimates of the true series. Otherwise, the model follows Jacobs & Van Norden's [13] so their gains are extended through the new assumptions. These assumptions represent the data release process more realistically under particular circumstances, and improve the overall identification of the model.

By assuming that the true series is observed up to a fixed time M the overall identification of the model improves. This results from the availability of true data to identify the dynamics of the true underlying process, and

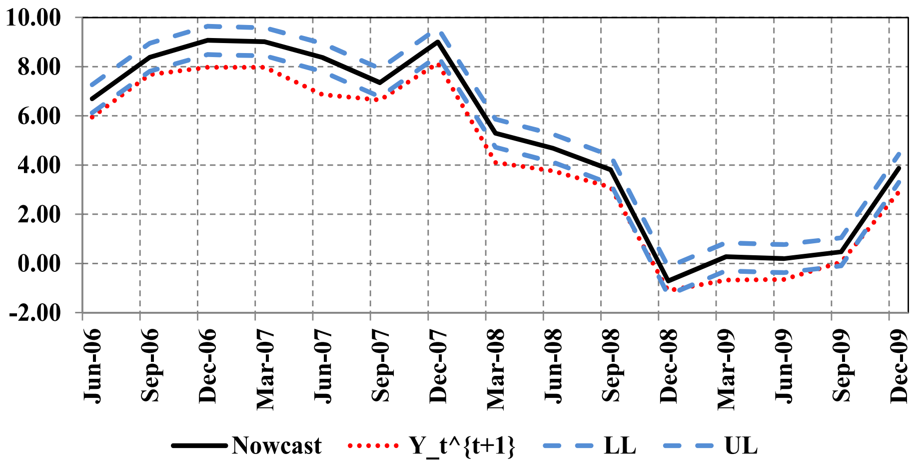

Figure 7: Now-casts of Year to Year Growth of the Colombian Quarterly GDP, Confidence Interval and the First Data Release

because mean measurement errors are also identified.

An application to the year to year growth of the quarterly Colombian GDP reveals features of the Colombian GDP release process that have an important effect on the use of these figures for policy purposes. First, preliminary growth figures under-estimate the true ones. And second, measurement errors contain noise. More precisely, the downward bias of the five more recent releases are 0.96%, 0.73%, 0.73%, 0.67% and 0.77% respectively. Moreover, the first data release has a statistically significant downward bias between 0.57% and 1.14%, on average. Therefore measurement errors are predictable from the information available at the data release.

Similar downward biases were found in Franses [7], Table 1 and Garratt & Vahey [8].

The models estimated in this paper serve the following purposes; (i) describe the dynamics of the Colombian growth, (ii) optimal inference and forecasting of the true growth, and (iii) trend-cycle decomposition of the

GDP, all in a setting of data revisions and delays.

References

[1] B. Anderson and J. Moore. *Optimal Filtering.* Prentice Hall, New Jersey, first edition, 1979.

[2] S. Arouba. Data revisions are not well behaved. *Journal of Money, Credit and Banking*, 40(2/3):319–340, 2008.

[3] S. Bordignon and U. Trivellato. The optimal use of provisional data in forecasting with dynamic models. *Journal of Business & Economic Statistics*, 7:275–286, 1989.

[4] F. Busetti. Preliminary data and econometric forecasting: an application with the bank of italy quarterly model. *Journal of Forecasting*, 25:1–23, 2006.

[5] P. De Jong. Rational economic data revisions. *Journal of Business and Economic Statistics*, 5:539–548, 1987.

[6] J. Durbin and S. Koopman. *Time Series Analysis by State Space Methods.* Oxford University Press, Oxford, first edition, 2001.

[7] P. Franses. Why is GDP tipically revised upwards? *Statistica Neerlandica*, 63(2):125–130, 2009.

[8] A. Garratt and S. Vahey. Uk real-time macro data characteristics. *Economic Journal*, 116(509):119–135, 02 2006.

[9] R. Harrison, G. Kapetanios, and T. Yates. Forecasting with measurement errors in dynamic models. *International Journal of Forecasting*, 21(3):595–607, 2005.

[10] A. Harvey. *Forecasting, Structural Time Series Models and the Kalman Filter*. Cambridge University Press, Cambridge, 1989.

[11] A. Harvey, C. McKenzie, D. Blake, and M. Desai. : "Irregular Data Revisions". In A. Zellner, editor, *Applied Time Series Analysis of Economic Data, Economic Research Report Number ER-5*. Department of Commerce, Washigton D.C., 1983.

[12] E. Howrey. The use of preliminary data in econometric forecasting. *Review of Economics and Statistics*, 60:193–200, 1978.

[13] J. Jacobs and S. van Norden. Modeling data revisions: Measurement errors and dynamics of "true" values. *forthcoming Journal of Econometrics*, 2010. http://www.eco.rug.nl/ jacobs/jjdownload/jacobsvannordenfinalerfeb2010.pdf.

[14] N. Mankiw and M. Shapiro. News or noise: An analysis of gdp revisions. Technical report, Survey of Current Busines, May 1986.

[15] R. Mariano and H. Tanizaki. Prediction of final data with use of preliminary and/or revised data. *Journal of Forecasting*, 14:351–380, 1995.

[16] J. Mincer and V. Zarnowitz. 1: "The evaluation of economic forecasts". In J. Mincer, editor, *Economic Forecasts and Expectations: Analysis of Forecasting Behavior and Performance*. National Bureau of Economic Research, N. Y., 1969.

[17] K. Patterson. An integrated model of data measurement and data generation processes with an application to consumers expenditure. *The Economic Journal*, 105:54–76, 1995.

[18] P. Siklos. What can we learn from comprehensive data revisions for forecasting inflation: Some US evidence. Working Papers eg0049, Wilfrid Laurier University, Department of Economics, 2006.

[19] U. Trivellato and E. Rettore. Preliminary data errors and their impact on the forecast error of simultaneousequations models. *Journal of Business & Economic Statistics*, 4:445–453, 1986.