Aalto University
School of Business

Macroeconomic forecasting and data revisions
Linear univariate GDP forecasting when revision process is characterized by news or noise: theory and evidence for Finland

Economics

Master's thesis

Roope Kolehmainen

2016

Department of Economics
Aalto University
School of Business

Aalto-yliopisto
Kauppakorkeakoulu

Macroeconomic forecasting and data revisions

Linear univariate GDP forecasting when revision process is characterized by news or noise: theory and evidence for Finland

Master's thesis
Roope Kolehmainen
12.3.2016
MSc. in Economics

Hyväksytty kansantaloustieteen laitoksella xx.xx.2016 arvosanalla

1. tarkastajan nimi   2.tarkastajan nimi

Abstract

National accounting data is subject to revisions which can severely hamper the macroeconomic forecasting based on this data. Two forecast models which are both based on the same sample data but employ data that is revised to different degree (i.e. data is fetched in two distinct points in time) could produce considerably different forecasts. Revisions also impede forecast accuracy evaluation and might lead to spurious judgements regarding the predictive power of a particular forecasting model.

In this paper, I investigate the effect of data revisions on forecasting models, forecasting results, and forecasting accuracy when the forecasts are based on linear univariate models. In the theoretical analysis, I demonstrate the mechanisms through which the data revisions influence forecasting with these models. The empirical analysis is based on the Finnish real-time quarterly real GDP.

The empirical analysis is comprised of three parts. First, I investigate the randomness of the data revisions (known as news vs. noise characterization). Next, I analyze how the moment in time in which the forecasts are generated, i.e. a data vintages, influences GDP forecasting outcomes, models, and accuracy. Lastly, I devise a Monte Carlo experiment in order to study the interconnection between the randomness of the revisions, forecasting model data sample, and realizations. The comparison of the forecasting models is done by contrasting real-time data based models with the latest available data models. As realizations I use the initial publications, the first anniversary publications and latest available values.

The revision process of the Finnish real-time quarterly GDP is characterized by almost pure news. Inter alia, the variance of revisions increases in time, the revisions are correlated with later publications, and the relative proportion of the systematic dependencies between the first and latter publication is negligible. Moreover, the revisions do not exhibit statistically significant biases. The revisions exhibit significant influence on forecasting simulations as a whole. However, statistically significant deviations between forecasting models were observed only using sub-sample data with early observations. Monte Carlo analysis suggested that by measuring the characterization of the revision process it could be possible to select the model vintage depending on the realization to be forecasted. The simulations produced consistent and exclusive dependencies between the revision process, forecasting models, and realizations for individual forecasting horizons as well as for horizon averages with all the parameter scenarios.

Tiivistelmä

Kansantaloudellisen tilinpidon data tarkentuu ajan kuluessa, mikä voi huomattavasti vaikeuttaa makrotaloudellisten ennusteiden laadintaa. Kaksi ennustusmallia, jotka perustuvat samaan otosdataan, mutta joissa otokset ovat eriasteisesti tarkentuneita (ts. otosdatat ovat kerätty eri ajankohtina), voivat tuottaa merkittävästi erilaisia ennusteita. Datan tarkentuminen (datarevisiot) hankaloittaa myös ennustus-tarkkuuksien arviointia ja saattaa johtaa harhaanjohtaviin tulkintoihin ennustusmallien tehokkuudesta.

Tässä tutkimuksessa tutkin datarevisioiden vaikutusta ennustusmalleihin, -tuloksiin ja -tarkkuuksiin lineaaristen yhden muuttujan mallien avulla. Esitän teoreettisessa analyysissä, kuinka datarevisiot vaikuttavat ennustamiseen ko. mallien kautta. Empiiriset analyysit perustuvat suomalaiseen reaaliaikaiseen, reaaliarvoiseen bruttokansantuotteen kvartaalidataan.

Empiirinen analyysi koostuu kolmesta kokonaisuudesta. Tarkastelen ensin datarevisioiden satunnaisuutta (tunnetaan kirjallisuudessa news- ja noise-erotteluna). Seuraavaksi analysoin ennusteajankohdan eli datavuosikertojen vaikutusta BKT:n ennustustuloksiin, -malleihin ja -tarkkuuksiin. Viimeiseksi laadin Monte Carlo -simulaation, jossa tarkastelen revisioiden satunnaisuuden, ennustusmallin dataotosten ja realisaatioiden vuorovaikutuksia. Ennustusmalleja vertaillessa verrataan keskenään reaaliaikaiseen dataotokseen ja viimeisimpään dataan perustuvia malleja. Realisaatioina käytetään ensimmäisiä julkaisuja, vuosi ensimmäisen julkaisun jälkeen julkaistuja arvoja sekä viimeisimmän datavuosikerran arvoja.

Suomalaisen reaaliaikaisen kvartaali-BKT:n tarkentuminen on lähes puhdas news-prosessi. Mm. revisioiden varianssi kasvaa suhteessa aikaan, revisiot korreloivat myöhempien julkaisuiden kanssa ja aiempien ja myöhempien julkaisuiden välinen systemaattisten poikkeavuuksien suhteellinen osuus on lähes olematon. Revisioissa ei myöskään ole havaittavissa tilastollisesti merkitseviä vinoumia. Ennustussimulaatioissa revisioilla havaittiin olevan kokonaisuudessaan merkittävä vaikutus ennustuksiin. Ennustustarkkuus-vertailuissa tilastollisesti merkitseviä poikkeavuuksia mallien välille syntyi kuitenkin vain käyttämällä viimeisimpään datavuosikertaan nähden selvästi aiempaa osaotosta. Monte Carlo -analyysi antoi viitteitä siitä, että revisioiden satunnaisuutta arvioimalla voidaan valita ennustusmallin otosdata riippuen ennustettavasta arvosta. Simulaatiot tuottivat johdonmukaisia ja poissulkevia riippuvuussuhteita revisioprosessien, ennustusmallien ja realisaatioiden välille sekä yksittäisille ennustushorisonteille ja keskiarvoisesti kaikille kokeessa käytetyille parametreille.

Table of Contents

1. Introduction

“A basic requirement for successful economic forecasting is accurate data” (Cole, 1969, p. 1). However, the economic datasets at hand for the forecasters are hardly ever complete or definitely accurate. The data from national accounting is subject to indefinite years of refinements and the finalized data could be unavailable even for decades. Even though the vista of future developments in key national accounting variables such as growth of real GDP, interest rates, inflation, or unemployment can be perceived to be of interest to all economic agents varying from a central banker deciding on the conduct of monetary policy to a house buyer contemplating on withdrawing a mortgage, the economists have had difficulties forecasting these variables with eminent precision. Real-time forecasts are, especially, easily impaired by the data revisions which can be relatively large and completely random.

In presence of data revisions when selecting the most appropriate data for construction of a forecasting model and forecast accuracy evaluation, the forecasters are faced with several dilemmas. Is the new forecasting model more optimal compared to the preceding models or is the predictive performance solely a merit of better data? Is a particular forecast an attempt to predict the ultimate value of a given variable (which may or may not be available after years to come) or are we interested in what the real-time perception of the economic activity is? The forecasts which are based on today’s best available data, i.e. the current vintage data, are inclined to have an edge on the forecasts based on real-time data due to the series of refinements which inherently reduce the measurement errors and contain information on the definitional and benchmark changes. But, in real life the decision makers in many cases need a model that performs best in real-time and the handicap arising from the unknown future refinements should be accounted for in order to avoid judging forecasts based on spurious predictive power. Diebold and Rudebusch (1991) provide a tailor-made example: the US index of leading economic indicators had been perceived to be an exquisite predictor for recessions. However, it turns out the index had only ex ante predictive power and, after the elimination of the data revisions, which would not have been available to the forecaster in real-time, the index had no forecasting value.

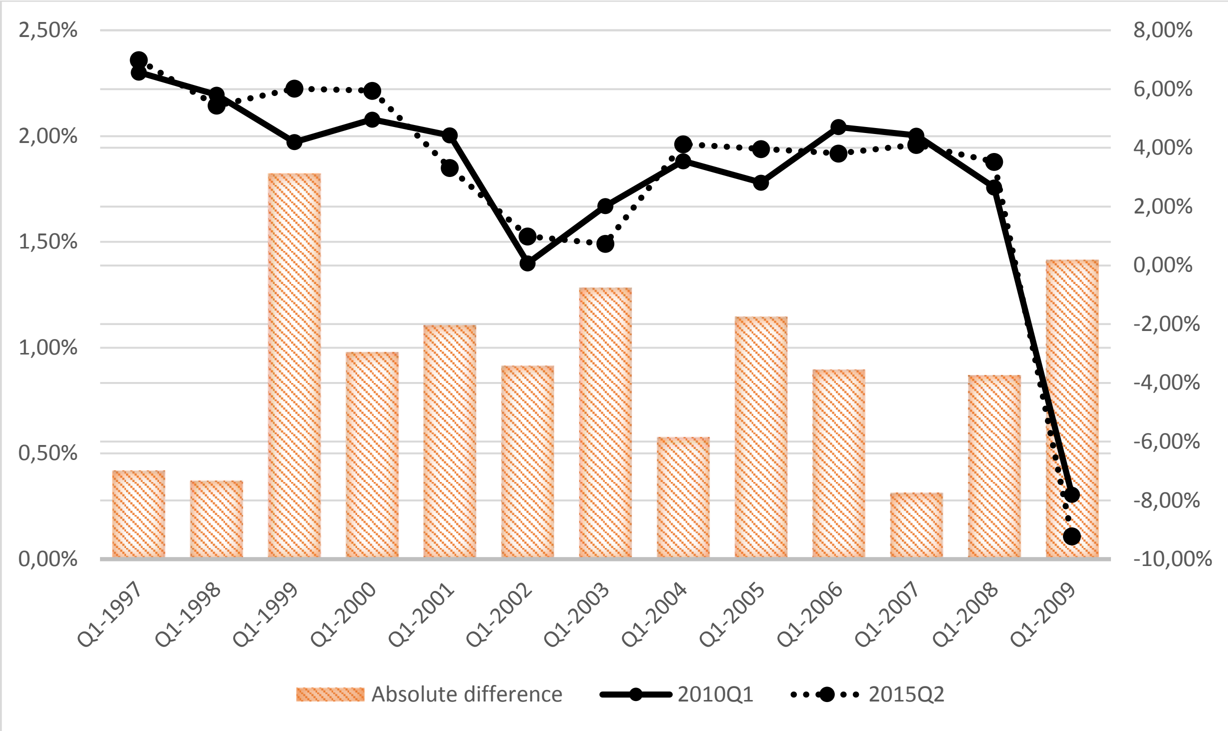


Figure 1.1: Visualization how the development of the year-on-year Finnish real GDP growth rates during 1997Q1-2009Q1 would have looked to an observer in 2010Q1 and 2015Q2.

The GDP forecasts are also impaired by the ambiguity of the data generating process. The economic theory does not define the functional form of data generating process for the key macroeconomic variables, and hence the choice for the forecasting model is somewhat discretionary. In this thesis, I will examine macroeconomic forecasting in the presence of data revision within linear univariate model, i.e. $ARIMA$-model, context. Linear univariate forecasting models have in multiple studies exhibited relatively unsurpassable predictive power and have been used as a benchmark in many of the forecasting evaluations.

The choice of the forecasting model data sample should also be evaluated from the statistical perspective. Theoretically, knowledge of the revision process could facilitate the model and data selection. If the revisions process is mainly comprised of unpredictable, variance increasing "news" the forecaster makes a mistake by constructing their real-time model with best available data. This data imposes an implicit contradictory assumption that unpredictable events would be predictable. On the contrary, revisions characterized by "noise" can be modelled a priori and should be accounted for in the data which the forecasting model is based on.

The purpose of this thesis is to examine, both theoretically and empirically, how the forecasting outcomes are affected by the selections of different data vintages for the forecasting model construction. Moreover, I will study the performance of these models with respect to different actuals of varying real-time degree. Additionally, I will analyze whether news and noise separation can assist in selecting the most appropriate data vintage for the forecasting model construction. The central research question for this thesis reads:

What is the impact of data revisions on forecasting outcomes when forecasting the Finnish GDP with linear univariate models? What are the implications of using different data vintages when revisions are characterized by news or noise for forecast model construction?

The central question is analyzed with a series of sub-research questions to facilitate the understanding of the conclusions of this thesis:

1. What are the data revisions and vintages? What type of characteristic are observed in the revision process and between the data vintages?
2. What are the theoretical and empirical effects of data revisions and their properties on forecasting?
3. How do different data vintages affect the forecasts within $ARIMA(p, d, q)$ framework?
4. What type of statistical properties do the Finnish real GDP growth data revisions exhibit?

I will first introduce the key concepts, discuss the relevant literature, and study $ARIMA(p, d, q)$ forecasting in the presence of data revisions theoretically in Chapters 2 and 3. The theoretical analysis is universal and in many aspects applicable for different macroeconomic time series variables such as inflation and interest rates which are also typical real-time forecasting objectives. In order to gain empirical relevance, I will study the Finnish data revision process and construct a series of forecasting simulations for the Finnish real GDP growth in Chapters 4 and 5. These results cannot necessarily be generalized as such due to varying properties of different time series and the differences in revision processes between the statistical institutions collecting the data.

2. Theoretical framework

Data revisions are typically unavoidable adjustments to the data that aim to improve the quality of the figures compared to the previous publication. A trade-off between timeliness and the quality of the high frequency data entails initial figures at the time of the publication often to be incomplete in many ways. For instance, in the absence of available data the numbers could be interpolated or estimated from another non-complete data series. Then, the subsequent publications, or vintages, of the same observed variable incorporate more information than the previous ones until the particular national account item is complete. (DiFonzo, 2005, pp. 6-7). Hence, a data revision can be measured as a difference between any earlier and subsequent publication. In economics, depending on the variable of concern and publication or measurement frequency, it is not unlikely, for example, that the size of the revision is larger than earlier figure itself or that the revision alters the sign of the previous estimate.

Data revisions can have major impacts on both empirical economics and economic forecasting. For an economist, the key question regarding the data revisions is whether the data revisions have the potency to influence the empirical macroeconomic research and the quantitative inferences the economists make. Corollary to the economist’s problem, forecasters are ultimately interested whether two forecasts, where the model is based on different editions of the same data set, lead to significantly different predictions of the future. Therefore, the data revision research on empirical macroeconomics and economic forecasting have many theoretical overlaps and also real-time economic theory can be utilized when analyzing the effects on forecasting.

Before the recent emergence of real-time databases, economic researchers and forecasters have conventionally based their models on the latest available data, i.e. the current vintage, which is easily accessible in any typical national accounting database. However, resorting solely to the current vintage data might considerably impair the theoretical and quantitative accuracy of a particular economic research or forecast. Data revisions can significantly affect conclusions drawn from quantitative research, and conversely, a better understanding of data revisions can consequentially improve the robustness of the empirical research and accuracy of the forecasts. Perhaps, therefore, the data revision research has predominantly committed on analyzing data revisions themselves. The data revisions occur at least in seven different ways (McKenzie, 2006, p. 8):

1. Incorporation of source data with more complete or otherwise better reporting (e.g. including late respondents) in subsequent estimates
2. Correction of errors in source data (e.g. from editing) and computations (e.g. revised imputation)
3. Replacement of first estimates derived from incomplete samples (e.g. sub-samples) judgmental or statistical techniques when data become available
4. Incorporation of source data that more closely match the concepts and / or benchmarking to conceptually more correct but less frequent statistics
5. Incorporation of updated seasonal factors
6. Updating of the base period

7. Changes in statistical methodology, concepts, definitions, and classifications

Sourcing from the above factors, there are several important aspects of the data revisions pertaining to empirical macroeconomic research and forecasting. Statistical properties that have been extensively researched and are merit noting due to their significance on economic forecasting are: serial correlation of the particular releases and revisions by their size and direction, predictability of the revisions, or separation to news or noise process.

On a later note, it is critical to annotate that revised data does not necessarily imply anything about the *accuracy* of the data. Instead, the dimension of quality related to revisions is *reliability*. Therefore, non-revised data can be less accurate than data which has gone through revisions. It is also possible that the data is either accurate or reliable – or both simultaneously. By following the Data Quality Assessment Framework of IMF it is crucial to distinguish between reliable and accurate data. Throughout this paper the definitions are (Carson and Laliberté, 2002, pp. 5-6):

- *"Accuracy* refers to closeness between the estimated value and the (unknown) true value that the statistics were intended to measure. Assessing the accuracy of an estimate involves evaluating the error associated with an estimate." (Carson and Laliberté, 2002, p. 6)
- *"Reliability* referes to the closeness of the initial estimated value(s) to the subsequent estimated values. Assessing reliability involves comparing estimates over time. In other words assessing reliability refers to revisions. This feature is identified separately for two reasons. First, it is usually the initial estimate that captures attention, whence the importance of its accuracy. Second, the separation helps bring out the fact that data that are not revised are not necessarily the most accurate." (Carson and Laliberté, 2002, p. 6)

In forecasting, the revisions indeed affect the *forecasting accuracy* because the revisions contain information which reduces the measurement errors.

Despite several studies of the statistical properties of data revisions and the debate of the potential effect on the empirical research dating back almost seven decades to a descriptive analysis by Friedman (1947, pp. 11-12), there is no clear-cut methodology how to treat the problems arising from data revisions. Seminal analysis on the properties of data revision was conducted by Zellner (1958) who asserts that a provisional releases in various national accounting items in the US, measured either infrequently or derived with a residual estimation, are subject to significant revision. Morgenstern (1963) and Cole (1969) discover similar patterns with their studies on the US GNP. Stekler (1967), on the other hand, argues that the revisions to the US GNP are typically small, and the direction of change is measured with significant precision. Thus, the provisional data serve as a good approximation of the economic movements.

The separation between news and noise and the predictability of data revisions is critical for quantitative analyses, as improved knowledge on the revision process facilitates selecting the forecasting model and the proper data vintage. If the revisions are fully comprised of unpredictable news, revisions are orthogonal with the earlier data vintages, and the initial values are efficient estimates of the final data. The variance of the data is, hence, expected to increase in later vintages. On the other hand, if data revisions reduce the noise contained in the early releases, the revisions are uncorrelated with the later releases but correlated with earlier vintages as the early release comprises of the final value and a measurement error (Croushore & Stark, 2003). In their seminal paper, Mankiw and Shapiro (1986) assert that the GDP revisions are unbiased, zero mean news and, thus the early releases are efficient estimates of the final value. Rudebusch (2001), Faust & al (2005), and Aruoba (2008), among several others, conclude that revisions reduce noise[1]. The application of news and noise separation to forecasting, however, has been scarce. Most notably, Clements and Galvao (2013), find that forecast accuracy statistics are worse for the news process than for noisy revisions.

Due to ambiguous, time-series specific, and statistical agency depended characteristics of data revisions and their serial correlations and contingent forecastability, the economist and forecasters need to carefully assess the underlying implications these properties have. Resorting solely to the refined data relies on the assumption that the agents are capable of anticipating the revisions completely. Put differently, estimating a forecasting model with the current vintage data incorporates an assumption that out-of-sample forecasts account for not only the non-occurred regular revisions but also all the future definitional and benchmark changes (which may in reality continue indefinitely into the future). On the contrary, relying on the preliminary data implicitly assumes that every revision or change is unpredictable. Additionally, the latest available data potentially ranges from reliable early observations in the sample to non-revised erroneous data as noted by Howrey (1978) which effectively asserts that the early observations contributing to the resulting forecast account for all the future revisions and the later estimates assume that revisions are unpredictable.

[1] All the above papers investigate the US output growth (real and/or nominal GDP or GNP). Faust & al. (2005) also study the real GDP growth of G7 countries concluding predictability in data revisions for all the other countries besides the US.

To carter the fact, the initial release is potentially a biased estimate of the final value several researchers have incorporated the use of real-time data vintages in their empirical studies and forecasts. Whereas, the latest available data and initial data are static periodic observations of the variables Y_{t-v-1}^{t-v} made either at present date t or at the first publication date, of each variable, the real-time vintage is the $(t-v)^{\text{th}}$ release of the data associated with observation $t-v-1$. A real-time database or a revision triangle, a collection of real-time vintages, is therefore comprised of vectors of observations such as:

$$\begin{bmatrix} Y_0^1 & \cdots & Y_0^{t-v} & \cdots & Y_0^t \\ & \ddots & \vdots & & \vdots \\ & & Y_{t-v-1}^{t-v} & \cdots & Y_{t-v-1}^t \\ & & & \ddots & \vdots \\ & & & & Y_{t-1}^t \end{bmatrix}$$

The revision triangle depicts the observed values as if today was a date of a particular vintage. Real-time vintage thus incorporates the best available data at the given point in time.

The seminal and the standard strand of forecasting studies the relationship between the data vintages and the forecasting outcomes. Customary methodology is based on the idea of comparing the forecasts utilizing the same structure where the other forecast is based on a particular data vintage and the other one on a different vintage. Before the wider accessibility to the real-time data bases, the forecasters have predominantly committed on studying the effects of using the preliminary data. Denton and Kuiper (1965) and Cole (1969) find converging results for the Canadian and the US GNP, respectively, that the initial data impairs forecasting accuracy substantially. On the contrary, Trivellato and Rettore (1986) show that for the Italian economy the impairment is much more minor.

Only recently, the researchers have started extensively studying the effect of real-time vintages on forecasting. Swanson (1996), Koening & al., and Croushore and Stark (2001) have been the advocates of the recent emergence in real-time analysis and forecasting. Croushore and Stark (1999, 2001, and 2003) have been the front-line researchers of the data revision analysis, real-time empirics and forecasting since the introduction of their real-time data set in 1999. They have starkly advocated the verification of the empirical results against real-time data. Moreover, they have extensively illustrated the applicability of real-time data in various forecasting exercises. Work of

Stark and Croushore and Clements and Galvao (2013) is the main inspiration to the findings in this thesis.

3. Theoretical analysis

In this section I will provide an a priori analysis of the effects of data revisions on the macroeconomic forecasting within linear univariate time-series framework. I will commence by providing the general properties of $ARIMA(p, d, q)$ by imposing the conceptual and notational framework for linear time-series models to facilitate the further analysis. Subsequently, I will introduce the potential mechanisms through which the data revisions might influence forecasting with linear stochastic time-series models.

Popularized by Box and Jenkins (1970), linear time-series models for stationary processes are a benchmark model set-up for stochastic processes and their forecasts. The linear univariate models are relatively simple theoretically and perform well forecasting key macroeconomic variables compared to more elaborate models. Therefore, the $ARIMA$-models provide a convenient platform for studying the effect of data revisions on GDP. Moreover, concentrating solely on the univariate models allows, for instance, disregarding a critical and tedious problem of cross-correlations between the early and final releases of different variables occurring in the multivariate models (Aruoba, 2008, p. 22). The economic hypothesis behind the stationary time-series model is that the majority of current behavior of a particular variable is explained by the persistence in its past observations. Autoregressive, AR, models relate the weighted average of its past observations and a disturbance term with the current observation and the moving average, MA, models describe the stationary process as a weighted sum of the current and lagged disturbances. The mixed autoregressive-moving average models, $ARMA(p, q)$ models, describe the stationary stochastic processes as p^{th} and q^{th} order autoregressive and moving average process as they display properties from both types.

The research on economic forecasting solely with the univariate models has been extensive since the introduction of $ARIMA(p, d, q)$ framework. The linear stationary models prove to be in many experiments well suited to forecast the dynamic economic dependencies and offer a relatively well performing benchmark when compared against the alternative, more complex univariate or multivariate methods[2]. Even in pure $AR(p)$ form, the linear stationary models have a relatively good forecasting performance for national accounting time series as shown by Meese and Geweke

[2] For a thorough discussion on a linear benchmark method selection for GDP and inflation modeling, see Marcellino (2008).

(1984). Similarly, Clements (2014) with the US GDP data appraise the predicting power of $AR(p)$ models concluding that multivariate models prevail only during occasional points of the business cycle. Moreover, for instance, Macellino & al (2003) and Marcellino (2008) conclude that linear univariate model based forecasts for Euro-area and the U.S. GDP, also when estimated with real-time vintages, perform better than the alternative linear models.

3.1 General properties of $ARIMA(p,d,q)$ processes

Prior to proceeding with the a priori analysis of data revisions in linear univariate model forecasting, I will introduce a selection of general properties of linear univariate models which have a vital role when forecasting in the presence of data revisions[3]. The classical $ARMA(p,q)$ is given as

$$Y_t = \delta + \sum_{i=1}^{p} \phi_i Y_{t-i} + \varepsilon_t + \sum_{i=1}^{q} \theta_i \varepsilon_{t-i} \quad \varepsilon \sim i.i.d\,(0,\sigma_\varepsilon^2). \qquad (3.1)$$

$ARMA(p,q)$ can be transformed to the pure $AR(p)$ or $MA(q)$ process by setting the order of the redundant process to 0.

For convenience, an $ARMA(p,q)$ process in 3.1 can be reformulated with the lag operator notation as

$$(1-\phi_1 L - \cdots - L^p)\tilde{Y}_t = \left(1-\theta_1 L - \cdots - \theta_q L^q\right)\varepsilon_t$$

or

$$\phi(L)\tilde{Y}_t = \theta(L)\varepsilon_t$$

and equivalently

$$\tilde{Y}_t = \phi^{-1}(L)\theta(L)\varepsilon_t \Leftrightarrow \theta^{-1}(L)\phi(L)\tilde{Y}_t = \varepsilon_t. \qquad (3.2)$$

where L is the lag operator and $\tilde{Y}_t = Y_t - \mu$ is a deviation of the process from its fixed reference or from its mean in case the process is stationary. (Box & Jenkins, 1970).

From equation 3.2 we can study the two critical properties of the mixed time series processes, stationarity and invertibility, which are also focal points studying the effects of data revisions. A time-series process is stationary if its joint probability distribution remains constant over time. Hence, the mean and the variance are time independent and the process does not drift away from its fixed reference point. The stationarity of an $ARMA(p,q)$ process is obviously dependent only on

[3] Theoretical analysis in section 3. concerns mixed linear univariate models in order to discuss all the potential sources through which the data revisions could potentially influence the forecasts. In the empirical experiments in Chapters 4 & 5 only pure autoregressive models are considered.

the stationarity of an $AR(p)$ process (as $MA(q)$ component formed of weighed sum of $i.i.d$ disturbances is always stationary). The time-series Y_t is a stationary process and the autoregressive operator $\phi^{-1}(L)$ in equation 3.2 converges if the roots of the characteristic equation $\phi(L) = 0$ are all outside of the unit circle. Therefore, we must have that the modulus of the roots $|L_i| > 1 \; \forall \; i = 1,2, \dots, p$. Moreover, as the fixed reference point for an $AR(p)$ process can be written as

$$\mu = \frac{\delta}{\sum_{i=1}^{p} \phi_i} \tag{3.3}$$

and we can derive the necessary condition for stationarity that

$$\left|\sum_{i=1}^{p} \phi_i\right| < 1. \tag{3.4}$$

Invertibility is a necessary property to relate the present variables with past observations (Box & Jenkins, 1970 p. 52). Invertibility property allows for an unambiguous determination of lag order via the autocorrelation functions. Whereas stationarity is determined by the $AR(p)$ component of an $ARMA(p,q)$ process, invertibility is solely determined by $MA(q)$. If $MA(q)$ is invertible it has an $AR(\infty)$ presentation. Corollary to the stationarity condition derived above, the moving average operator $\theta^{-1}(L)$ in equation 3.2 converges if all the roots of the characteristic equation $\theta(L) = 0$ lie outside the unit circle. Put differently, if $|L_i| > 1 \; \forall \; i = 1,2, \dots, p$, the process can be inverted to $AR(\infty)$.

In economics, many of the empirical time series (e.g. national accounts and security prices) do not have a fixed mean or variance and are, hence, non-stationary. In case the necessary stationarity condition in equation 3.4 is violated (i.e. the roots of the characteristic equation, $\phi(L) = 0$, of the $AR(p)$ have modulus less than unity) the process exhibits non-stationary behavior with $ARMA(p,q)$ modeling. However, in some cases it is possible to transform non-stationary time series into stationary series by differencing the data in particular order d.

Suppose we have a $ARMA(p,q)$ model such that in equation 3.2. $\tilde{Y}$ is a non-stationary $AR(p)$ process such that d moduli of the roots of the characteristic equation $\varphi(L) = 0$ are exactly unity and the remainder are outside of the unit circle. Then we can write the equation as

$$\varphi(L)\tilde{Y}_t = \phi(L)(1-L)^d\tilde{Y}_t = \theta(L)\varepsilon_t \tag{3.5}$$

where $\varphi(L)$ represents a stationary autoregressive operator of order $p - q$. Utilizing differencing operator notation $\Delta = 1 - L$ we can reformulate the equation 3.5 as

$$\varphi(L)\tilde{Y}_t = \phi(L)\Delta^d Y_t = \theta(L)\varepsilon_t \iff \phi(L)w_t = \theta(L)\varepsilon_t$$

where, $w_t = \Delta^d Y_t = \Delta^d \tilde{Y}_t$, since $\Delta^d Y_t = \Delta^d \tilde{Y}_t$ for $d \geq 1$. Hence, the time series is homogenous non-stationary with order d and which can be represented by a stationary, invertible $ARMA(p, q)$ process where the moduli of the roots of $\varphi(L)$ are all greater than unity. (Box & Jenkins, 1970).

The above homogenous non-stationary processes are referred to as integrated autoregressive-moving average models of order p, d, and q, or $ARIMA(p, d, q)$ models. Many of the economic time series, especially in context of national accounting, exhibit non-stationary properties by growing with a deterministic trend – either with a linear or exponential trend. Noting that, the mean of stationary process $\Delta^d Y_t$ is

$$\mu_w = \frac{\delta}{\sum_{i=1}^{p} \phi_i}$$

hence instead of a stochastic trend when $\delta = 0$, the series includes a deterministic polynomial trend of degree d. If $d > 1$, that is the time series displays an exponential trend, or the interest is in the relative changes of the time series rather than the absolute levels, we need to apply non-linear transformation. Converting the time series into logarithmic form will linearize the exponential trend and standardize the variance. Data is transformed into logarithmic form by (Box & al., 1970, pp. 100-103)

$$Y_t = \left(1 + \frac{Y_t - Y_{t-1}}{Y_{t-1}}\right) Y_{t-1} = (1 + g_t)Y_{t-1}$$

$$\ln Y_t = \ln\big((1 + g_t)Y_{t-1}\big)$$

$$g_t \cong \ln \frac{Y_t}{Y_{t-1}}. \qquad (3.6)$$

3.2 $ARIMA(p, d, q)$ and forecasting

The performance of a certain forecast is most often evaluated based on the accuracy of the aggregated forecasted values relative to the observed values. Hence, forecasters most commonly wish to minimize the mean square error, MSE (or alternatively the root mean square error or $RMSE$) of the forecasted value and actual observation. These two are most typical forecast assessment measures utilized in forecasting literature. As the forecast deviation is a random variable the forecaster's problem is to choose the lead time l forecast at origin-t, or $\hat{Y}_{t+l|t}$, such that $E\left(e_{t+l|t}^2\right) = E[(Y_{t+l} - \hat{Y}_{t+l|t})^2]$ is minimized (Box & al., 1970, p. 138).

In $ARIMA(p,d,q)$ forecasting model specification the optimal MSE minimizing forecast is given by the conditional expectation of Y_{t+l} that is by $\hat{Y}_{t+l|t} = E(Y_{t+l}|Y_t, Y_{t-1}, \dots, Y_1)$. Lead time l $ARIMA(p,d,q)$ generated observation may be expressed as an infinite weighted sum of current and previous shocks as:

$$Y_{t+l} = \sum_{j=0}^{\infty} \psi_j \varepsilon_{t+l-j}.$$

Naturally, an out-of-sample forecast can only be based in the data available until time t so we may define the forecastable value as $t+l$. Our best forecast is then a linear function of error terms

$$\hat{Y}_{t+l|t} = \sum_{j=0}^{\infty} \psi_j^* \varepsilon_{t-j}$$

where weights ψ_j^* are to be determined to minimize MSE. As by assumption $E(\varepsilon_i \varepsilon_j) = 0$ for $i \neq j$, we have that

$$E\left(e_{t+l|t}^2\right) = E[(Y_{t+l} - \hat{Y}_{t+l|t})^2] = (1 + \psi_1^2 + \dots + \psi_{l-1}^2)\sigma_\varepsilon^2 + \sum_{j=0}^{\infty} (\psi_{l+j} - \psi_{l+j}^*)^2 \, \sigma_\varepsilon^2.$$

Expression is minimized by setting $\psi_{l+j}^* = \psi_{l+j}$ for $j = 0{,}1$ ... and we have

$$Y_{t+l} = (\varepsilon_{t+l} + \psi_1 \varepsilon_{t+l-1} + \dots + \psi_{l-1}\varepsilon_{t+1}) + (\psi_l \varepsilon_t + \psi_{l+1}\varepsilon_{t-1} + \cdots) = \hat{\varepsilon}_{t+l|t} + \hat{Y}_{t+l|t}$$

where $\hat{\varepsilon}_{t+l|t}$ is the forecast error. The optimum forecast $\hat{Y}_{t+l|t}$ is merely a conditional expectation of Y_{t+l}

$$\hat{Y}_{t+l|t} = \psi_l \varepsilon_t + \psi_{l+1}\varepsilon_{t-1} + \cdots = E(Y_{t+l}|Y_t, Y_{t-1}, \dots, Y_1)$$

since the expected values of $\hat{\varepsilon}_{t+l|t}$ for an unbiased forecast are equal to 0.

Moreover, important to note is that one step ahead forecast error, i.e. lead time $l = 1$ forecast for any $ARIMA$ model specification, is

$$\hat{\varepsilon}_{t+1|t} = Y_{t+1} - \hat{Y}_{t+1|t} = \varepsilon_{t+1} \, .$$

Therefore, the residuals ε_t which in the above specification generate the process as a set of independent random variables also are one step ahead forecast errors and the forecast error variance is the variance of the error term. Moreover, in a typical forecasting exercise the true parameters of the $ARIMA(p,d,q)$ process are unknown and estimated from the data. Therefore, the actual forecast error will be larger than $var(\hat{\varepsilon}_{t+1|t}) = (1 + \psi_1^2 + \cdots + \psi_{l-1}^2)\sigma_\varepsilon^2$ which assumes that the model parameters are known.

In addition to MSE or $RMSE$ there are a variety of other forecasting error statistics. In order to validate the results derived from $RMSE$, I will also calculate the mean absolute error and mean error statistics:

$$MAE = n^{-1} \sum_{t=1}^{n} |\varepsilon_t|$$

$$ME = n^{-1} \sum_{t=1}^{n} \varepsilon_t \,.$$

Both MAE and ME also reveal the characteristics of the forecast errors. ME offsets negative and positive forecasts errors and hence could possibly disclose information on disproportional biases in either direction. Relatively large $RMSE$ error statistics compared to MAE indicate higher variance in the forecast errors. The smaller the difference between $RMSE$ and MAE the lower the variance of individual forecast errors.

Real-time forecasting brings an additional dimension to forecast evaluation. Due to variation arising from the revision process in the predictand and dependent variables, out-of-sample forecasts become susceptible to the changes in the correlation structure of the data. Calculating the forecasting error statistics is also more cumbersome when we acknowledge the existence of data revisions. The key question regarding the evaluation of the real-time data based forecasts is which data we should use as "actual" for computing the forecasting error, as the data is liable for indefinite revisions. This is especially the case if we want to forecast the true value of the GDP, and there is no consensus in the forecasting literature which realization should be used as the actual value. For example, Croushore (2006) argues that latest available data should be used as the approximation for the truth. However, defining latest available as "actual" effectively hampers the forecasting accuracy of the real-time data based models compared to the latest available data based models due to the forecasting error reducing effect of regular and benchmark revisions contained in the available data. Hence, the forecasts using real-time data might be significantly sensitive to the selection of the realization as shown by Stark and Croushore (2002).

Because the aim of this thesis is to evaluate whether the forecasts based on different vintages really are different from the each other it is vital to test the differences in forecasting accuracy. I will evaluate forecasts based on Harvey & al. (1997) modification of the Diebold and Mariano (1995) predictive accuracy test which allows more precise testing of longer forecast horizons l and smaller sized samples.

Harvey & al. (1997) employ the squared errors as loss differential so that $d_t = \varepsilon_{1t}^2 - \varepsilon_{2t}^2$. Forecasts are equal under the null hypothesis $H_0 : E(d_t) = 0 \; \forall t$. The test is based on the observed sample mean

$$\bar{d} = n^{-1} \sum_{t=1}^{n} d_t$$

where the variance of $\bar{d}$ is consistently estimated by

$$var(\bar{d}) \approx n^{-1}\left[\gamma_0 + 2\sum_{k=1}^{l-1} \gamma_k\right]$$

and γ_k is the k^{th} autocovariance of the d_t. The autocovariance in turn can be estimated by

$$\hat{\gamma}_k = n^{-1} \sum_{t=k+1}^{n} (d_t - \bar{d})(d_{t-k} - \bar{d}).$$

By substitution to the variance can be reformulated as:

$$S_1 = \frac{\bar{d}}{\sqrt{[\widehat{var}(\bar{d})]}}$$

which is the Diebold and Mariano test statistic. The modified, bias corrected test statistic[4] is

$$S_1^* = \sqrt{\frac{n+1-2l+n^{-1}l(l-1)}{n}}\; S_1 \,.$$

After the bias correction, the test statistic is compared with critical values of Student's t-distribution with $n - 1$ degrees of freedom.

3.2.1 Channel effects

Forecasts in $ARIMA$-framework are influenced by the data revisions through three different channels. Data revisions change the data input used to make the forecast (direct channel), data revisions alter the estimated coefficients (indirect channel), and the revisions could impact the elementary construction of the model or the forecast results are effected through deliberated ex-ante model specifications (specification channel).

In fashion of Croushore & Stark (2003), to analyze how data revisions could have a significant influence on the forecast through the direct and indirect channels, I consider a pure autoregressive p^{th} order, $AR(p)$, data generating process for modeling the gross domestic product measured in levels. Despite the simplicity of the example, the conclusions drawn from this set-up are evidently present in the real-time forecasting. Moreover, we can see from the below illustration that despite

[4] For the thorough modification, see Harvey & al. (1997, p. 3).

the stochasticity of the data revisions, the revisions will have a considerable effect if certain circumstances are fulfilled. Setting $q = 0$ in equation 3.1 yields a pure $AR(p)$ process

$$Y_t = \delta + \sum_{i=1}^{p} \phi_i Y_{t-i} + \varepsilon_t \quad \varepsilon \sim i.i.d\,(0, \sigma_\varepsilon^2) \tag{3.7}$$

where gross domestic product Y at time t is generated by its previous realizations weighted by coefficient ϕ up to period p. Subsequently, I consider a forecast utilizing the above data generating process in (3.7). The data could be subject to revisions. Let $Y_{t,v}$ represent the realized v vintage value for Y in the period t. Hence, we can express the revision to the data by $Y_{t,v} - Y_{t,v-1}$.

To turn to forecasting, I let $Y_{t|t-1,v}$ stand for a forecast for period t made based on the information in $t-1$ as reported in the vintage v. This set-up allows for examining the forecast made on the basis of any data vintage. For instance, forecasting current economic activity with real-time model, the data vintage v would be given so that it contains the data observations from the $t-1$ vintage. The general one step ahead forecast with $AR(p)$ model is then given by

$$Y_{t|t-1,v} = \hat{\delta}_v + \sum_{i=1}^{p} \hat{\phi}_{i,v} Y_{t-i,v} \tag{3.8}$$

where the estimated autoregressive parameters depend on the utilized data vintage and are thus indexed by v.

To examine how the forecasts are impacted by data revisions, suppose we have made a forecast based on another data vintage $w \neq v$ subject to the same $AR(p)$ data generating process as in (3.7) so that we would utilize the same forecast rule as in 3.8: $Y_{t|t-1,w} = \hat{\delta}_w + \sum_{i=1}^{p} \hat{\phi}_{1,w} Y_{t-1,w}$. We can express the changes in our t-period forecast caused by the data revisions by subtracting the forecast estimated based on the data vintage v from the forecast based on vintage w:

$$Y_{t|t-1,w} - Y_{t|t-1,v} = \hat{\delta}_w - \hat{\delta}_v + \sum_{i=1}^{p} (\hat{\phi}_{i,w} Y_{t-i,w} - \hat{\phi}_{i,v} Y_{t-i,v}). \tag{3.9}$$

From equation 3.9 we can quite easily study the impacts of the data revisions through the direct and indirect channels, ceteris paribus. However, deriving the joint effects of these channels algebraically is inconceivable and, de facto, an empirical matter which I explore more in chapter 4. The model specification channel cannot be directly examined with the equation 3.8 but will be investigated below.

3.2.1.1 Direct channel

Suppose the data revisions affect only the data input for the forecasting model and the revisions do not affect the estimated coefficients. Thus, we have that $(Y_{t-i,w} - Y_{t-i,v})_{\, i\in\{1,p\}} \neq 0$ and $\hat{\delta}_v = \hat{\delta}_w$ and $\hat{\phi}_{i,v} = \hat{\phi}_{i,w} = \hat{\phi}_i,\ i \in \{1,p\}$ our forecast change in $AR(p)$ specification, expressed in equation 3.8, becomes

$$Y_{t|t-1,w} - Y_{t|t-1,v} = \sum_{i=1}^{p} \hat{\phi}_i \left(Y_{t-i,w} - Y_{t-i,v}\right). \tag{3.10}$$

The above equation 3.10 immediately reveals how the forecast revision depends on the size of the revisions to the data $(Y_{t-i,w} - Y_{t-i,v})_{\, i\in\{1,p\}}$ and its sensitivity to the magnitude of the estimation coefficients. However, the sensitivity to the data revisions depends largely on characteristics of the data series. It is easy to see directly from the equation 3.10 that the data revisions do not have an impact on the forecast, for instance in case of a stationary white noise process for which $\forall \hat{\phi}_i \approx 0$, $i \in \{1,p\}$. However, they have a significant impact in case the process exhibits tendency for persistence and $\exists\, \hat{\phi}_i \neq 0$, $i \in \{1,p\}$. The larger the coefficient the larger the deviation is between the forecasts using different data vintages.

It is possible to derive a more general result for the direct channel effect of the data revisions on the forecast change in case of the stationary process. For stationary processes $E(Y_t) = E(Y_{t-1}) = E(Y_{t-2}) = \cdots = \mu$ yields the formula for the mean in equation 3.3 ($\mu = \frac{\delta}{1-\sum_{i=1}^{p} \phi_i}$)

Applying this to the forecast change in equation 3.10 and remembering the conditions for data vintages and estimated coefficients that we defined above for the direct channel we get

$$E(Y_{t|t-1,w} - Y_{t|t-1,v}) = \frac{\hat{\delta}_w}{1-\sum_{i=1}^{p} \hat{\phi}_{i,w}} - \frac{\hat{\delta}_v}{1-\sum_{i=1}^{p} \hat{\phi}_{i,v}} = 0.$$

This allows us to conclude that the expected forecast deviation for pure stationary processes through the direct channel is zero.

3.2.1.2 Indirect channel

Now we make a contrary assumption and suppose that the data revisions do not have an impact on the variables used in the forecast estimation but affect the estimated coefficients. An exemplary set-up in which this type of model behavior is possible is that the data is not revised for the early lags but revisions occur for the succeeding lags. For a concrete instance, suppose that the revisions occur due to both noise and news. During the Q1's only noise affects the revisions and the national accounts preliminary data for the preceding fiscal year is only made more reliable by recalculations.

During Q2 the true economic developments are observable for the first time, and the news caused larger revisions to the data. Hence, we could have $\forall\, Y_{t-i},\ Y_{t-i,v} \approx Y_{t-i,w} \equiv Y_{t-i}$, and, but for $Y_{t-(p+k)},\ Y_{t-(p+k),v} \neq Y_{t-(p+k),w}$, where $i \in \{1, p\}$ and $k > p$, meaning that the data is potentially revised only for the lags occurring after the pth lag. In this case we then set $Y_{t-i,w} \cap Y_{t-i,v} \equiv Y_{t-i,v\ i\in\{1,p\}}$ but keep the assumption that $\hat{\delta}_w \neq \hat{\delta}$ and $\forall\, \hat{\phi}_{i,w} - \hat{\phi}_{i,v} \neq 0$ all the estimated coefficients differ depending on the data vintage. Hence, our $AR(p)$ process in the equation 3.9 becomes

$$Y_{t|t-1,w} - Y_{t|t-1,v} \approx \hat{\delta}_w - \hat{\delta}_v + \sum_{i=1}^{p} (\hat{\phi}_{i,w} - \hat{\phi}_{i,v}) Y_{t-i} \qquad (3.11)$$

We see directly from 3.11 that our forecast revision is now influenced merely by the magnitude of the effect the data revisions have on the estimated coefficients.

3.2.1.3 Specification channel

The data revisions do not only shift the forecasting results through the changes in data input and the coefficient estimation but they can also alter the functional form of the forecasting model. The specification channel is not directly visible in our above example but the changes in model specification caused by revisions to the data can shift the forecasting results. The two most essential causes for the deviations in the forecasting results when utilizing different data vintages are order selection in $ARIMA$-models and stationarizing operations for non-stationary data. The order selection problem occurs when the differing data vintages affect the optimal order of the time series model that we utilize for forecasting. For instance, in our above autoregressive forecasting example it could be possible that for data vintage v the optimal order for the AR process would be 2 and for w the optimal would be the 3rd order. Hence, with vintage v we would forecast with $AR(2)$ process whereas using revised vintage w we would utilize $AR(3)$ process. Moreover, in order to get meaningful forecasts with the time series models we first need to stationarize the data if the time series is non-stationary, and it is possible that the data revisions alter the results in case stationarizing operations are applied.

The changes of the lag specification due to data revisions can be analyzed through the duality of autoregressive and moving average models. To start the analysis, assume that the revised data follows a $AR(1)$ data generating process

$$Y_t = \phi Y_{t-1} + \epsilon_t \qquad (3.12)$$

in which Y_t is revised data for GDP growth rate and $\epsilon_t \sim WN\ (0, \sigma_\epsilon^2)$. Then assume that the real-time data for GDP growth, X_t, is merely a stochastic description of the revised data measured with error, u_t, so that

$$X_t = Y_t + u_t$$

where $u_t \sim (0, \sigma_u^2)$ and $\rho(X_t, u_t) = 0$ implying that the revisions are adding news[5]. Applying Wold's decomposition theorem, the linear representation for the real-time data is $ARMA(1,1)$

$$X_t = \phi X_{t-1} + \varepsilon_t + \theta \varepsilon_{t-1} \qquad (3.13)$$

where

$$\theta = -\frac{\phi\, \sigma_u^2}{\sigma_\varepsilon^2}$$

with

$$\sigma_\varepsilon^2(1+\theta^2) = \sigma_\epsilon^2 + \sigma_u^2(1+\ \phi^2).$$

From the above set-up we can make two crucial conclusions concerning the changes in model specifications. First, the lag structure for models using revised data are likely to differ from the models using real-time data input. This is due to the dynamic structures of the models. The linear representation for the real-time variable X_t in the equation 3.13 can be expressed with lag operator notation as

$$\phi(L)X_t = \theta(L)\varepsilon_t$$

where $\phi(L) = 1 - \phi L$ and $\theta(L) = 1 + \theta L$. With the help of this notation, the $ARMA(1,1)$ process can easily be expressed as an $AR(\infty)$ process $iff\ |\theta| \cap\ |\phi| < 1$ as follows

$$\varepsilon_t = \frac{\phi(L)X_t}{\theta(L)} = (1+\theta L)^{-1}(1-\phi L)X_t$$

$$X_t = (\phi - \theta)\sum_{i=0}^{\infty}\theta^i X_{t-i-1} + \varepsilon_t$$

Hence, whereas revised data followed strictly an $AR(1)$ data generating process in equation 3.12, the model structure for the real-time data is potentially subject to infinite $AR(\infty)$ process. Therefore, the impact to the lag specification depends on the magnitude of the coefficient ϕ estimated with the revised data and the size of the revision, or measurement error, u_t. (Elliot, 2002, pp. 2-3).

---

[5] It is possible to study the effects under noise assumption also which leads to similar results with more complicated computations. See Elliot (2002, p. 2)

Second, the persistence of the revised time series, in addition to the direct channel, also attributes to the model specification. The larger the ϕ coefficient of our $AR(1)$ process for the revised data set the more likely is that lag structures will differ between the models utilizing different data sets. The explanation is this: the past observations are included in the model until their explanatory power becomes insignificant for explaining the current observation. With $AR(p)$ model specification the explanatory power sources directly from the size of the coefficients and thus the absolute value of the parameter ϕ directly contributes to the selected lag structure. Consider again the above infinite lag operator specification for the real-time data X_t. For the same sized measurement error u_t in the real-time data, the larger the absolute value of the $AR(1)$ estimation coefficient ϕ the more likely is that the additional past observations are included in the model and differences in the lag specifications will occur. (Elliot, 2002, p. 3).

Additional aspect to model specification is the effect of data revisions on order selection criterion. Typically, the economists and forecasters are aided by different selection criteria such as Akaike information criterion (AIC) or Schwarz-Bayesian information criterion (SIC or SBIC) to construct the univariate model. Stark and Croushore (2002) show that forecasts where lag selection is governed by SIC are much less sensitive to the data revisions than forecasts where specification is selected by AIC. This is because SIC penalizes for additional lags and hence creates insulation to the forecasts against the data revisions compared to AIC. AIC, on the contrary, displays greater variability and longer lag lengths that enhances the sensitivity of the forecasts to data revisions.

3.2.1.4 Stationarizing operations

As discussed above, non-stationary data series need to be transformed to stationary to produce meaningfully fitted models with $ARIMA(p, d, q)$. Logarithmizing is applicable if the data series has an exponential trend, its variance increases or decreases in time, or in the particular experiment the interest is in growth rates rather than in levels. The data revisions typically will not have an ex-ante effect for the necessity to stationarize the data as it should be extremely unlikely that the data revisions would alter the necessary conditions for stationarity. However, data revisions could potentially have ex-post effect on the forecasting model.

Consider logarithmic transformation as described in 3.6 for same observations t of the gross domestic product utilizing different data vintages v and w. Logarithmic transformation for Y_t given the vintage is thus

$$g_{t,v} \cong \ln \frac{Y_{t,v}}{Y_{t-1,v}}$$

and similarly the growth rate for using the vintage w is $g_{t,w} = \ln \frac{Y_{t,w}}{Y_{t,w-1}}$. Naturally, a nonrecurring revision to level data will also shift the growth rate. But, consider a recurring and systematic revision which occurs so that $Y_{t,w} = (1+\alpha)Y_{t,v}$ and $Y_{t-1,w} = (1+\alpha)Y_{t-1,v}$. Now the level variables are affected once again but $g_{t,w} = \ln \frac{Y_{t,w}}{Y_{t,w-1}} = \ln \frac{(1+\alpha)Y_{t,v}}{(1+\alpha)Y_{t,v-1}} = g_{t,v}$ and growth rate remains unaffected between the data vintages.

This finding is consistent with the empirical research which unequivocally supports that the national accounting variables measured in levels are far more sensitive to the data revisions than growth rates. For instance, Howrey (1996) shows that data revisions affect US real GNP forecasts much more when data is measured in levels instead of percentage growth rates.

4. Empirical analysis

This section is comprised of three main components. The first part describes the data employed throughout the chapter. In the second part I evaluate the statistical properties of the revision process in order to uncover potential patterns between the data vintages. Guided by the second part, I construct multiple forecasting experiments in the third section in order to evaluate the optimal vintage employment in forecasting the Finnish quarterly growth rates in different scenarios. These empirical forecasting applications are followed by Monte Carlo forecasting experiment simulating both news and noise revision processes.

4.1 Data

The input data sample used throughout this chapter is seasonally adjusted quarterly Finnish real GDP real time-series data from the *Main Economic Indicators Original Release Data and Revisions Database* (MEIORDRD) available at http://stats.oecd.org/mei/default.asp?rev=1. MEIORDRD is updated by OECD. The database contains real-time time series data for key macroeconomic variables for OECD and Euro countries in addition to a few major economies in the world. The data is available based on the publishing frequency of the national statistical institutes. The Statistics Finland publishes national accounts on a quarterly basis such that a new observation and an edition is published during Q+1 (see table 4.1; the Statistics Finland publishes the accounts in 65 days after the end of the previous quarter, and the data is available at MEIORDRD typically in a month from the publication).

The sample of observations used for statistical analysis and model construction cover a period of 1996Q1-2014Q4 and the editions 2000Q2-2015Q1. The GDP data is transformed to quarter-on-quarter (q-o-q) growth rates, $\frac{Q_{t,v}}{Q_{t-1,v}} - 1$ which are percentage rates expressed with respect to previous quarter within a particular vintage, v. Hence, for the latest data vintage 2015Q1, there are a total of 76 q-o-q observations, and for the first observation, 1996Q1, there are 60 vintages. All in all there are total 2790 observations pertaining to different vintages.

Table 4.1. Data release schedule and release notation

Quarter of year Y	***YQ1***	***YQ2***	***YQ3***	***YQ4***
1st release	July Y	October Y	January Y+1	April Y+1
2nd release	October Y	January Y+1	April Y+1	July Y+1
3rd release	January Y+1	April Y+1	July Y+1	October Y+1
4th release	April Y+1	July Y+1	October Y+1	January Y+2
Y1 (Q+4 release)	July Y+1	October Y+1	January Y+2	April Y+2
Y2 (Q+8 release)	July Y+2	October Y+2	January Y+3	April Y+3
Y3 (Q+12 release)	July Y+3	October Y+3	January Y+4	April Y+4
Y5 (Q+20 release)	July Y+5	October Y+5	January Y+5	April Y+5
L (Final value)	2015Q1	2015Q1	2015Q1	2015Q1

The above table describes the data release schedule including notation used to refer to the data throughout sections 4.1 and 4.2.

4.2 Descriptive statistics

In order to obtain a hint on the optimal vintage treatment in forecasting, I will first perform a series of evaluations on the characteristics of the data. As provided in the precedent sections, if data revisions can be characterized as reducing noise the forecaster ought not necessarily to rely on the real-time data due to predictability of the revisions. Vice versa, if the revisions add news to data, the revisions themselves are unforecastable, and in order to avoid effectively making a spurious implicit assumption that unpredictable is predicable, revised data should not necessarily be used.

In forecasting, one has to pay particular attention on bias and dispersion in the revision data. Uncovering potential systematic tendencies for the revised estimates to increase or decrease as we move forward in vintages could indicate predictability in the revisions. In the below summary statistics table[6] the bias can be studied by analyzing the 12 measures starting from mean absolute revision. Bias can be also detected by studying the normality of the data (by concluding biasedness in case normality applies) but I will directly evaluate biasedness of the revisions through the values calculated in the summary statistics table 4.2.

[6] Pre-programmed revision analysis spreadsheets created by OECD task-force and amended by the author are used in this thesis and are freely available at: http://www.oecd.org/std/automatedprogramstoperformrevisionsanalysis.htm.

Table 4.2. Revision summary statistics

Summary statistics	2nd-1st	Y1-1st	Y1-2nd	Y2-1st	L-1st	Y2-Y1	Y3-1st	Y5-1st	Y5-Y3	Y3-Y1	Y3-Y2	L-Y3	L-Y5	L-Y1	L-Y2
Sample	96Q1-14Q2	96Q1-13Q4	96Q1-13Q4	96Q1-12Q4	96Q1-14Q4	96Q1-12Q4	96Q1-11Q4	96Q1-09Q4	96Q1-09Q4	96Q1-11Q4	96Q1-11Q4	96Q1-11Q4	96Q1-09Q4	96Q1-13Q4	96Q1-12Q4
n	74	72	72	68	76	68	64	56	56	64	64	64	56	72	68
Mean absolute revision (MAR)	0,32	0,47	0,35	0,58	0,69	0,29	0,58	0,76	0,40	0,44	0,29	0,47	0,42	0,54	0,54
Mean revision ($\bar{R}$)	0,03	0,04	0,01	0,00	0,03	-0,03	0,04	0,04	0,00	-0,02	0,01	-0,01	0,01	-0,03	0,00
St. dev($\bar{R}$) - HAC formula	0,05	0,08	0,06	0,09	0,09	0,04	0,09	0,11	0,06	0,04	0,04	0,06	0,05	0,06	0,07
Mean squared revision (MSR)	0,29	0,49	0,27	0,65	0,90	0,14	0,68	1,06	0,33	0,32	0,18	0,38	0,30	0,54	0,52
Root mean squared revision (RMSR)	0,54	0,70	0,52	0,81	0,95	0,37	0,83	1,03	0,57	0,57	0,43	0,61	0,55	0,74	0,72
Relative mean absolute revision (RMAR)	0,31	0,45	0,34	0,52	0,71	0,26	0,52	0,68	0,36	0,38	0,26	0,43	0,37	0,53	0,51
t-stat	0,47	0,54	0,24	0,05	0,32	-0,83	0,44	0,38	-0,06	-0,37	0,21	-0,10	0,14	-0,50	-0,03
t-crit 1%	3,43	3,43	3,43	3,44	3,43	3,44	3,45	3,48	3,48	3,45	3,45	3,45	3,48	3,43	3,44
t-crit 5%	1,99	1,99	1,99	2,00	1,99	2,00	2,00	2,00	2,00	2,00	2,00	2,00	2,00	1,99	2,00
t-crit 10%	1,67	1,67	1,67	1,67	1,67	1,67	1,67	1,67	1,67	1,67	1,67	1,67	1,67	1,67	1,67
1% significant?	NO	NO	NO	NO	NO	NO	NO	NO	NO	NO	NO	NO	NO	NO	NO
5% significant?	NO	NO	NO	NO	NO	NO	NO	NO	NO	NO	NO	NO	NO	NO	NO
10% significant?	NO	NO	NO	NO	NO	NO	NO	NO	NO	NO	NO	NO	NO	NO	NO
Correlation(First, Later)	0,90	0,83	0,92	0,78	0,68	0,96	0,79	0,66	0,91	0,91	0,95	0,89	0,92	0,83	0,85
Correlation(Revision,First)	0,07	-0,07	-0,2*	-0,13	-0,2*	-0,14	-0,07	-0,28**	-0,3**	-0,11	0,00	-0,27**	-0,13	-0,24**	-0,23*
Correlation(Revision,Later)	0,49***	0,5***	0,21*	0,51***	0,58***	0,15	0,56***	0,53***	0,13	0,32**	0,31**	0,19	0,28**	0,33***	0,32***
Min Revision	-2,22	-2,59	-1,81	-2,74	-4,11	-1,09	-3,71	-3,72	-2,01	-1,47	-1,21	-1,39	-1,34	-2,08	-1,57
Max Revision	2,32	1,79	1,45	2,26	2,38	0,83	1,84	2,38	1,93	1,47	1,10	1,84	1,41	2,17	2,43
Range	4,54	4,38	3,26	5,00	6,49	1,91	5,54	6,10	3,94	2,95	2,31	3,23	2,75	4,26	4,00
% Later > Earlier	45,95	55,56	47,22	42,65	52,63	39,71	46,88	58,93	48,21	43,75	48,44	50,00	48,21	43,06	48,53
% Sign(Later) = Sign(Earlier)	87,84	81,94	81,94	82,35	77,63	91,18	82,81	80,36	91,07	90,63	90,63	81,25	83,93	81,94	79,41
Variance of Later estimate	1,55	1,59	1,59	1,66	1,61	1,66	1,80	1,73	1,73	1,80	1,80	1,73	1,81	1,68	1,74
Variance of Earlier estimate	1,19	1,20	1,58	1,24	1,12	1,64	1,23	1,34	1,84	1,61	1,60	1,80	1,73	1,59	1,66
UM %	0,23	0,37	0,07	0,00	0,09	0,64	0,23	0,17	0,00	0,09	0,04	0,01	0,02	0,20	0,00
UR %	0,47	0,45	3,99	1,55	4,03	1,43	0,36	7,88	7,82	0,83	0,03	7,43	2,19	5,80	5,67
UD %	99,30	99,18	95,94	98,34	95,88	97,43	99,28	91,87	91,08	98,75	99,96	92,56	96,66	94,03	93,81
% of times indicating Acceleration or Deceleration	90,41	78,87	85,92	79,10	68,00	88,06	76,19	65,45	76,36	82,54	88,89	76,19	78,18	80,28	83,58

KEY:
1st: First published estimate
L: Latest published estimate
Y1: Estimate published 1 year later
Y2: Estimate published 2 years later
Y3: Estimate published 3 years later
Y5: Estimate published 5 years later
2nd: Second published estimate

Significance:
1% level: ***
5% level: **
10% level: *

The other important factor impairing the usefulness of the initial data is significant dispersion that translates into uncertainty (Öller & Hansson, 2005, p. 8). Dispersion can be concluded by studying multiple measures jointly and, for instance, by detecting abnormally large ranges for the revisions.

4.2.1 Unbiasedness and efficiency

By analyzing Mean revision ($\bar{R}$) and Mean absolute revision (MAR) first in table 4.2 we can obtain a general idea whether the revisions are zero mean news or variance reducing noise correction. Quick glance at the $\bar{R}$ measure in the summary statistics table reveals that there are no systematic over- or under-estimation in initial estimations[7], and that the revisions are centered around zero for all the period ranges under comparison. Moreover, directly from the summary we can observe that for all the comparison periods, excluding two, the variance increases for the later estimates suggesting that revisions occur due to news.

To obtain more robust results, we can study the statistical significance of the mean revisions. Revisions are biased if the mean revision is statistically different from zero (either negative or positive). Instead of the standard t-test, I will use a modified t-test in order to account for the potential association between the revisions made in different periods, i.e. the serial correlation in the data. $T\text{-}stat$ values are calculated as:

$$t = \frac{\bar{R}}{st.dev(\bar{R}) - HAC}$$

where heteroscedasticity and autocorrelation consistent standard deviation of mean revision or HAC is given as square root of:

$$\widehat{var(\bar{R})} = \frac{1}{n(n-1)}\left(\sum_{t=1}^{n}\hat{\varepsilon}_t^2 + \frac{4}{3}\sum_{t=2}^{n}\hat{\varepsilon}_t\hat{\varepsilon}_{t-1} + \frac{2}{3}\sum_{t=3}^{n}\hat{\varepsilon}_t\hat{\varepsilon}_{t-2}\right),\ \hat{\varepsilon}_t = R_t - \bar{R}\ ^{8}. \quad (4.1)$$

$T\text{-}stat$ values are then compared to the critical values corresponding 1%, 5%, and 10% significance levels. For none of the values are we capable of finding statistical significance for any of these levels implying that we cannot refute the hypothesis that the mean revisions are zero.

Mean squared revision (MSR) and root mean squared revision $RMSR$ measure the variance of a revision based on a symmetric and quadratic loss function for an unbiased initial estimate. MSR is defined as

$$MSR = \frac{1}{n}\sum_{t=1}^{n}(R_t)^2$$

[7] Moreover "% Later > Earlier" also indicates no over- or under-estimation suggesting that revision is, on n-weighted average, positive in 47,89 % of the cases.

[8] Appendix 1. discloses the derivation of the HAC formula.

and the $RMSR$ is merely a square root of the statistic. We can further decompose the MSR to smaller proportions to gain insight on which effects are driving forces behind the revisions. I use the standard Theil (1961) composition following Granger & Newbold (1973):

$$MSR = \bar{R}^2 + (\sigma_P - \rho\sigma_L)^2 + (1 - \rho^2)\sigma_L^2$$

where σ_L and σ_P are standard deviations of the latter and the previous estimates in the comparison periods, and ρ is the correlation of these two estimates (L and P denote latter and previous estimates, respectively. For other notation, see table 4.1 and key in the table 4.2). Dividing by MSR yields

$$1 = UM + UR + UD$$

with $UM = \frac{\bar{R}^2}{MSR}$, $UR = \frac{(\sigma_P - \rho\sigma_L)^2}{MSR}$, and $UD = \frac{(1-\rho^2)\sigma_L^2}{MSR}$. (DiFonzo, 2005).

UM, or the mean error, measures the proportion of MSR which is not equal to zero. UR can be interpreted as a slope error by considering a regression model for later estimate which is explained by the earlier estimate: $L_t = \alpha + \beta P_t + \varepsilon_t$. If the earlier estimates are unbiased we have $\alpha = 0$ and $\beta = 1$[9]. Hence, the statistic measures the degree that β is different from 0. On the contrary, if the regression would fit the data perfectly, $\beta = 1$, and $UD = 0$. Therefore, UD can be interpreted as a disturbance term which is not caused by systematic deviations of the later estimations from the earlier estimations. (DiFonzo, 2005, pp. 18-19)

Given the above, if the first estimates are characterized with high proportions of UD, the interpretation is that the estimates are not systematically dependent, or that the revisions are due to news. In all of the comparison periods the proportion of UD is extremely high. Investigating the periods with initial estimates as data origin the disturbance proportions are ranging from 91,08% to 99,30%. This gives us a solid indication that the revision process is characterized by news instead of noise.

To reinforce the above assessment that the revisions to the Finnish q-o-q growth rates is mainly comprised of unpredictable news we can estimate the significance of the correlation first between the earlier estimate and the revision and second between the later estimate and the revision. As discussed in Chapter 2, if the revisions are correlated with the later estimates the information that is made available between the estimation points (i.e. unpredictable news) is incorporated in the estimation process of the later estimate. This implies that the early estimates are efficient estimates

[9] Note that least squares estimators for this regression are $\hat{\alpha} = \bar{L} - \hat{\beta}\bar{P}$ and $\hat{\beta} = \frac{\sigma_{LP}}{\sigma_P^2}$, and that $\sigma_P - \rho\sigma_L = \sigma_L(1 - \hat{\beta})$.

for the later estimates. On the contrary, if the revision process is characterized by noise, the revision and the early estimates exhibit significant correlation implying the early estimates do not efficiently incorporate all the available information.

Indeed, we discover corroborating results that the revisions are significantly correlated with the later estimates, i.e. the revision process is characterized by news, when compared against the alternative of zero correlation. For all the comparison periods which utilize the initial release as the origin, the correlation is significant on 1% level, whereas of these periods, only 'Y5-1st' exhibits correlation between early estimate and revision on a 5% significance level. Hence, it becomes quite safe to conclude that the early estimates seem to be non-biased estimates for the later estimates.

4.2.2 Dispersion

Dispersion can hamper the use of the initial data as it describes the volatility of the data revisions (Öller & Hansson, 2005, p. 8). Especially during verges of economic downturns and when business cycles reverse upward from troughs one might expect to discover dispersion in revision data. Hence, as the sample in this study contains the financial crisis of 2008 the results indeed indicate moderate to significant dispersion. Moreover, as we widen the sample range, definitional changes account toward the spread of dispersion[10]. Unavoidably, the degree of dispersion stems from the quality of data and the forecaster has only limited possibilities account for this quantitatively.

For instance, revision ranges are up to 6,49 percentage points (p.p.) and overall relatively high when comparison period includes the initial data. For a comparison, 'Y2-Y1' range, in which the Y1 data has been already revised four times, has a range of 1,91 p.p. MAR itself is a good indicator of the dispersion, avoiding offsetting effect of negative and positive revisions. Also, by comparing $\bar{R}$ with MAR it can be concluded that there is an evident spread between the revisions from first estimate to the last one within the comparison period. Similarly to MAR, $RMSR$ is an excellent measure for dispersion by joining the degree of bias and the variance of a revision. Correlation between $RMSR$ and range is 0,92, MAR and range is 0,85, and $RMSR$ and MAR is 0,98 corroborating the implicit proposition that large ranges and higher non-offsetting revision measures indeed indicate for large revision spread.

[10] Definitional changes can be deduced to have a negligible effect on the data as we observed no biasedness in section 4.2.1, see McKenzie (2007, p. 10)

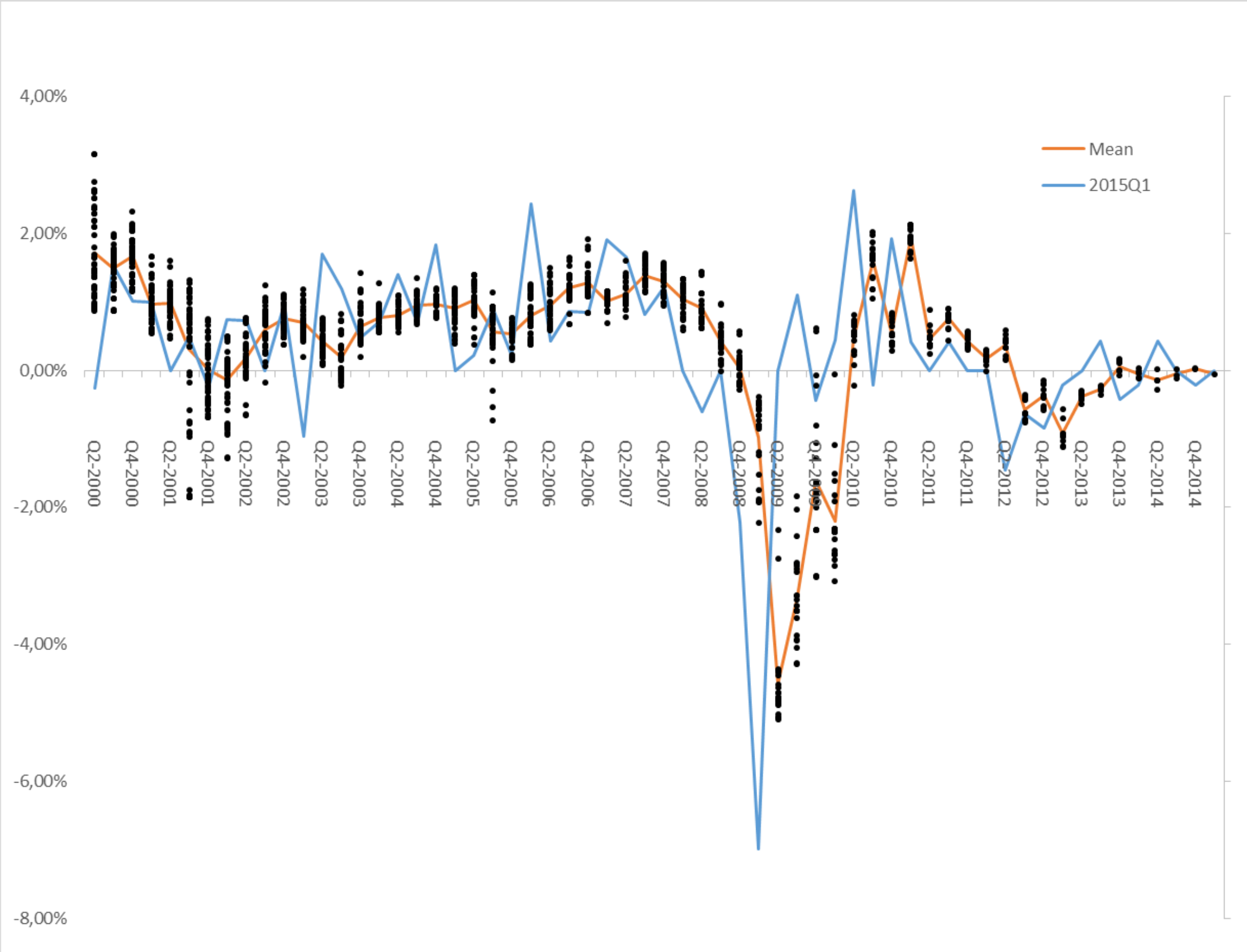
4.3 Forecasting

4.3.1 Repeated forecasting analysis

To start the empirical forecasting analysis, I will evaluate the overall influence data revisions can induce on forecasting. I construct a series of forecasts starting from the first non-observed value in the first vintage of the data sample, 2000Q2, to the last vintage 2015Q1 for which the forecastable observation is the ongoing quarter. With a similar "repeated forecasting" approach to Stark and Croushore (2002), I estimate the q-o-q growth rates with an $ARIMA(p, 1,0)$ model using each available edition to construct a separate one-step out-of-sample forecast for a particular date t. For the first quarter to be forecasted, 2000Q2, there are 60 vintages, and for 2007Q4 there are 30 vintages. Hence, for all the 60 quarters for which forecasts are generated with all the available vintages, there are 1830 forecasts in total. Effectively, we mimic a scenario in which we would have, for example, during 2000Q2 built our first forecasting model for the ongoing quarter. Then, during each subsequent quarter our forecasting model for 2000Q2 would have been updated based on new available data edition.

I limit the $AR(p)$ model specification to allow for individual p^{th} order selection for each of the time series and vintage used to generate the forecasting model. The autoregressive order is selected with Akaike Information Criterion (AIC). The maximum p^{th} order is 6.

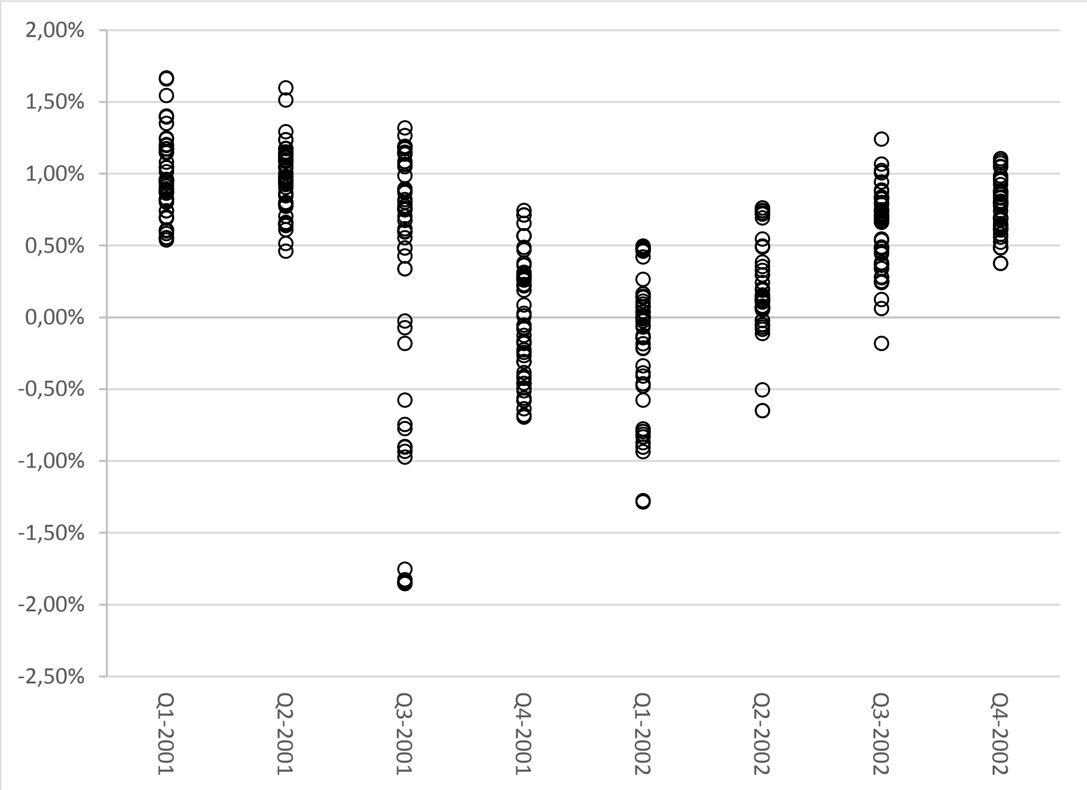
Figure 4.1. Repeated forecasts, mean forecast, and 2015Q1 vintage realization



The figure 4.1 summarizes all the generated forecasts for 2000Q2-2015Q1. The vertical pillars of dots in the figure are different forecasts made for the same quarter, and each dot represents a forecast that would have been made based on the information available in a given vintage. The two lines plot the mean for the forecasted values between the vintages and observed values of the latest available data (2015Q1) time series. The pillar of dots for the first forecasted quarter 2000Q2 thus contains 60 dots, and the number of dots reduces as we move along the vertical axis.

We can directly observe that the forecasts based on the different vintages exhibit a great disparity. The largest spread between the generated forecasts for a single period is 3,631 p.p. (2009Q4) and even the average spread is 0,976 p.p. The forecast ranges are most condensed when the economy is in a state of steady growth such that q-o-q growth rates are positive but relatively low. These periods also exhibit the smallest average absolute revisions. However, the forecast ranges become much more volatile once the quarterly growth rates deviate from the typical low positive growth rates. Especially during the downward turning points the forecasts exhibit greatest disparity. These periods are also associated with the highest absolute average revisions.

Figure 4.2. Repeated forecasts, 2001Q1-2002Q4



Because the number of generated forecasts is tediously large and the number of generated forecasts is reduced as we move forward in time, it is convenient to look at a smaller slice of the graph pertaining to earlier quarters in the sample for which higher number of forecasts are generated. In figure 4.2. the graph shows eight sets of different forecasts made for 2001Q1-20002Q4. In the first pillar there are 57 dots and for the last quarter there are 50 dots (the number of the dots may seem smaller as they are overlapping).

Despite the fact that many of the dots are overlapping, it is immediately evident that there is substantial variance between the forecasts made for the same period but in different time. As noted above, the disparity between the forecasts is modest for the periods with relatively steady growth (for instance, 2002Q2 and 2002Q4). However, once the economy hits a point of downturn the forecasts immediately get dispersed. In 2001Q3, most pessimistic forecasts are based on earlier publications of the data which drastically overestimated the severity of the economic downturn in the aftermath of the dot-com bubble. Because almost 70% of the forecasting models for 2001Q3 are based on autoregressive order 1, much of the variation in the forecasts between different vintages can be credited on the revisions of the precedent quarter. The early estimates for 2001Q2 q-o-q contraction ranged from -1,73% to -2,43% whereas the later vintages suggested zero growth thus producing much more optimistic forecasts. Therefore, the forecasts ranging from optimistic to pessimistic changes sign and range from 1,266% to -1,854%. As a contrast, for 2002Q2 (neglecting the outlier forecasts), and Q42002 the forecasts are most condensed due to a more undispersed estimates between the editions for the previous quarters.

The above figures 4.1 and 4.2 illustrate the effect of all the channels through which the forecasts are affected by using different vintages. I will eliminate the specification channel effect by selecting the most appropriate $ARIMA(p,1,0)$ model based on the data in the first vintage. The model is selected again by using Akaike Information Criterion. Investigating again only a slice of the possible periods of which the forecasts can be made, I can evaluate the effect that different model selection has on the collection of forecasts made for 2001Q1-20002Q4.

Figure 4.3. Repeated forecast comparison, free AR(p) forecasts contrasted with AR(p) selected based on the first vintage

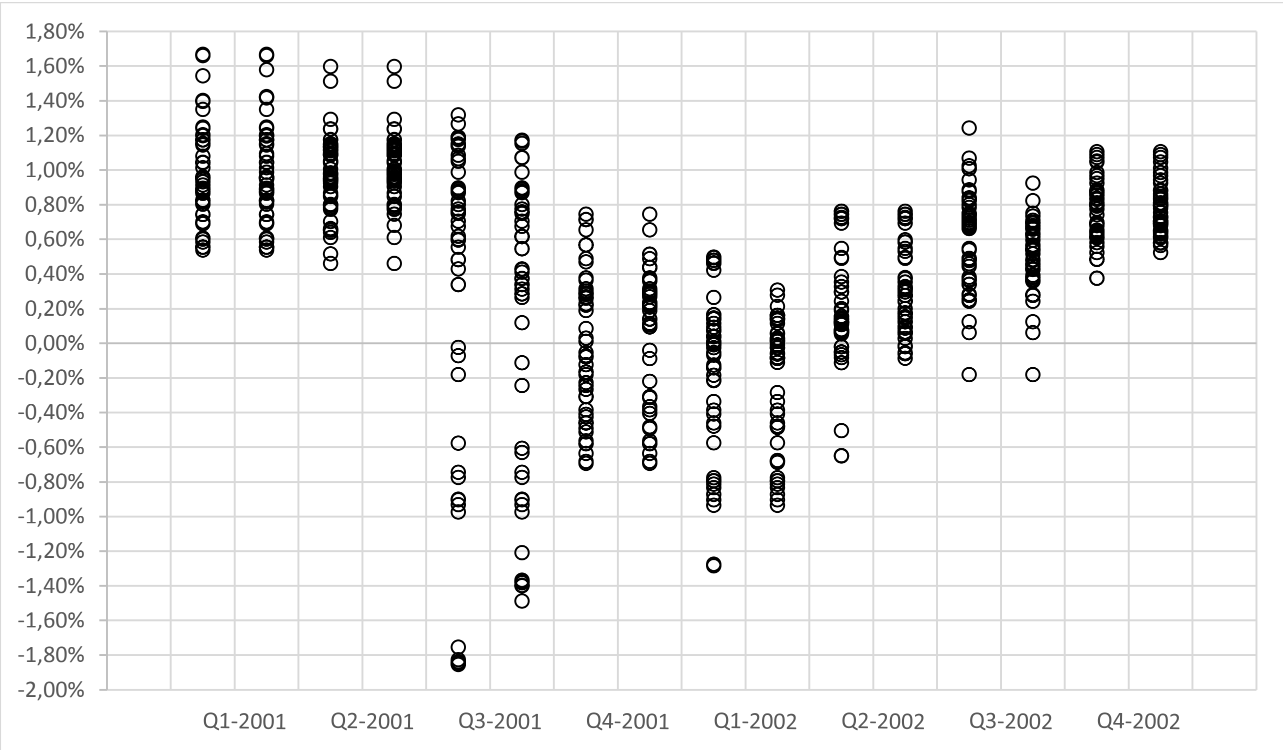


Figure 4.3 depicts the elimination of the specification channel. The pillars are in pairs such that the left set of pillars in each pair repeats the results in figure 4.2. The right pillars include the elimination of the specification channel by fixing the functional form of the forecasting model based on the first vintage. In all but one of the models the autoregressive order p was 2 (for 2001Q3 it was 1). We can see that the ranges become slightly more condensed but all in all the forecasts seem to be very unanimous to each other when the model specification channel effect is eliminated against the model based on the first vintage. However, despite arguably a negligible influence on the forecasting outcomes, the specification channel effect seems to have an underlying impact on the above results. Table 4.3 summarizes the functional forms of the forecasting models for each

quarter after the elimination of specification channel. *N* depicts the number of generated forecasts – or dots in the pillar graph. *N(different)* specifies the number of cases in which free autoregressive order selection resulted in different forecasting model (percentage value in brackets is the proportional amount to *N*). As we see in table 4.3, for 2001Q4-2002Q2, the specification channel could change the functional form of the forecasting model up to almost 60 % of the forecasts made for a particular vintage. In total of 186 forecasts, representing 43,46 % generated data points in the right pillars for each quarter, the vintage-dependent forecasting model specification was different from the functional form based on the first available vintage data.

Table 4.3. Summary of specification and direct channel elimination

	Functional form	**AR-parameters**	**N**	**N(different)**	**Average differential**
2001Q1	$ARIMA(2,1,0)$	(-0,9936; -0,6139)	57	19 (33,33 %)	0,08
2001Q2	$ARIMA(2,1,0)$	(-0,9443; -0,6719)	56	13 (23,21 %)	-0,24
2001Q3	$ARIMA(1,1,0)$	(-0,5104)	55	17 (31,91 %)	0,09
2001Q4	$ARIMA(2,1,0)$	(-0,8609; -0,4160)	54	31 (57,41%)	0,14
2002Q1	$ARIMA(2,1,0)$	(-0,8376; -0,4649)	53	31 (58,49%)	0,01
2002Q2	$ARIMA(2,1,0)$	(-0,9019; -0,4712)	52	31 (59,62%)	0,00
2002Q3	$ARIMA(2,1,0)$	(-0,8876; -0,5389)	51	26 (50,98 %)	0,18
2002Q4	$ARIMA(2,1,0)$	(-0,8175; -0,5601)	50	18 (36,0%)	0,149
TOTAL			428	186 (43,46%)	

To investigate the direct channel effect, I fix the coefficient with respect to the model suggested by the first vintage. Forecasting model parameters are summarized in table 4.3. In only two of the 420 forecasts (forecasts based on first vintage data will obviously be identical in the comparison and hence are excluded from the analysis) two individual one-step ahead forecasts for the same period forecasts deviated by more than 1 p.p. from each other and in 31 forecasts the deviations exceeded the average q-o-q growth percentage of 0,5071 %. Furthermore, the (geometric) average differentials between the one-step-ahead forecasts for the same period generated with fixed and non-fixed coefficient models are relatively small ranging from 0,00 to 0,24 percentage p.p. in absolute terms.

We can deduce that the indirect channel has the greatest impact on the forecast deviation. The differences in coefficient estimations then magnify the direct channel effect. Interestingly, the mean one-step-ahead forecast performs relatively well both in real-time and predicting the latest available values. The $RMSE$ of the mean forecast is 1,097 for real-time forecast and 1,479 for the latest available data. The mean is different on 5 percent significance level compared to the rolling forecasts made with latest available data (see next section and table 4.4 for corresponding forecast error statistics for the rolling latest available data based forecasts). This is in harmony with Elliot (2002) who argues that in practice the whole real-time dataset should be employed in order to minimize the forecast error.

4.3.2 Forecast simulation

To further analyze how the revision process influences forecasting, I run a series of empirical exercises comparing forecasts based on real-time vintages to those based on latest available data. I forecast the Finnish q-o-q real output growth with $ARIMA(p, d, q)$ model in order to compare the forecasts generated with models based on real-time data to forecasts generated with latest available data.

I follow the methodology of Stark and Croushore (2002) with deliberate modifications by proceeding in the following way. Firstly, I use the data from 1996Q1 to 2000Q1 which would have been available to the forecaster in beginning of 2000Q2 to estimate a model for real GDP growth rate. Secondly, I forecast q-o-q GDP growth rates for the subsequent four quarters starting from 2000Q2, the first quarter for which data would not have been available, to 2000Q3, 2000Q4, and 2001Q1. Thirdly, I continue by moving forward quarter by quarter in the sample repeating steps 1 and 2 always adding one more observation from the diagonal of the revision triangle. However, I deviate from Stark and Croushore's experiment by allowing a free $ARIMA(p, 1{,}0)$ specification (with maximum lag length of 6) and coefficient selection at each vintage in order to fully capture the real-time effect (therefore, effectively I have as many forecasting models as there are computed forecasts). I then repeat the exact same forecasting procedure using the latest available vintage as forecasting model data. The forecasts made with real-time data are referred to as **RT** and forecasts based on latest available data are denoted by **LA**. Fourthly, I compute the forecast errors by using three alternative "actuals": latest available data, first anniversary edition after the fourth-step-ahead forecast (for 2000Q2 the actual is 2002Q1 edition for 2000Q2), and the first available publication for a particular forecast (for 2000Q2 the actual is 2000Q3 edition for 2000Q2).

There is no definite answer in the forecasting literature to which values should be regarded as actuals and the definition of the actuals is driven by the purpose of a particular forecast as discussed in section 3.1. Central bankers conducting monetary policy decisions would definitely have more use on the knowledge over near future GDP publications and would define first publications as actuals. However, if the future 'true' value is of the interest to a forecaster, the selection for actual is not clear cut. Croushore (2006, p. 17) suggests using latest available figures as a proxy for actuals.

Working with real-time databases enables researchers to estimate variety of actuals. My selection of actuals is targeted to discover potential links in forecasting results and revision process studied in section 4.2.

The below tables each display three different sets of results pertaining to different values I use as actuals. I compare the root mean squared error ($RMSE$), mean absolute error (MAE) and mean error (ME) for all four forecast horizons. Number of generated forecasts is represented by N. The amount of generated forecasts for which error statistics can be calculated decrease while moving further in the sample as the availability to actuals decrease. Furthermore, I test the statistical significance of the forecasting differentials between vintages by constructing Diebold and Mariano test statistic based on a quadratic loss function and scaling it according to modifications suggested by Harvey & al (1997). P-value column reports the p-values of the Harvey forecast difference test.

Table 4.4 Forecast error statistics, model: RGPD ~ ARIMA(p,1,0), full sample

Forecast step	***N***	***RMSE***		***MAE***		***ME***		***p-***
		RT	**LA**	**RT**	**LA**	**RT**	**LA**	***value***
Actual value: Latest available data								
1-step forecast	60	1,450	1,633	0,975	1,060	-0,044	-0,026	0,196
2-step forecast	59	1,511	1,720	1,015	1,079	-0,037	0,003	0,401
3-step forecast	58	1,605	1,617	1,035	1,048	-0,071	-0,096	0,927
4-step forecast	57	1,686	1,864	1,114	1,151	-0,095	-0,104	0,386
Actual value: First publication								
1-step forecast	60	1,192	1,231	0,839	0,883	-0,039	-0,021	0,747
2-step forecast	59	1,233	1,435	0,883	0,945	-0,041	-0,001	0,446
3-step forecast	58	1,295	1,383	0,938	0,955	-0,078	-0,103	0,569
4-step forecast	57	1,483	1,602	1,073	1,092	-0,097	-0,107	0,547
Actual value: Four quarters later								
1-step forecast	58	1,502	1,611	1,042	1,102	-0,126	-0,108	0,416
2-step forecast	57	1,446	1,578	1,007	1,032	-0,104	-0,037	0,237
3-step forecast	56	1,425	1,425	0,991	0,989	-0,135	-0,134	0,152
4-step forecast	55	1,503	1,514	1,049	1,025	-0,123	-0,089	0,413

In table 4.4 the autoregressive order p in $ARIMA(p,1,0)$ is selected by AIC for each vintage. Similarly, the model employing last available data is reconstructed as more observations are utilized. Hence, for each forecast, the forecasting model is uniquely constructed based on the data available during that period. Somewhat strikingly, the forecast error measures are quite similar across the evaluated error statistic, actuals, and forecast horizons. The $RMSE$s are actually lower for the real-time data models for all the actuals and forecast horizons evaluated. Also, interestingly, the LA models perform better in real-time relative to their performance for forecasting revised values. This is in contrast with the presumption that the latest available data would have an advantage (especially if more revised data is used as actual data) over the real-time data based forecasts. However, we cannot reject the null hypothesis that the forecast differentials are different from zero for any of the generated forecasts. The results are identical if I fix the functional form of the latest available data forecasts and estimate the LA models with $ARIMA(2,1,0)$ specification with fixed coefficients which is the most appropriate model for the full observation range in the last vintage. Recalling Howrey's (1978) remark, however, as I am employing full sample data which has observations revised to very different degrees, the models based on the latest available data contain the forecast error reducing information only partly.

The above results are in harmony with Stark and Croushore (2002). Corollary to their experiment, I will split my sample to sub-samples to investigate the effects of the real-time forecasting and the trade-off between the sample observations used for model building and forecast evaluation. Taking only earlier vintages of the sample as the basis of the model also enables the investigation to the complete real-time forecasting impact as the observations in current vintage model have already gone through a sequence of revisions.

Table 4.5 Forecast error statistics, model: RGPD ~ ARIMA(p,1,0), 2001Q1-2005Q4

Forecast step	***N***	***RMSE***		***MAE***		***ME***		***p-***
		RT	**LA**	**RT**	**LA**	**RT**	**LA**	***value***
Actual value: Latest available data								
1-step forecast	20	0,999	0,786	0,778	0,651	0,271	-0,077	0,104
2-step forecast	20	1,102	0,890	0,909	0,730	0,338	0,003	0,057 *
3-step forecast	20	1,169	0,876	0,888	0,694	0,338	0,043	0,123
4-step forecast	20	1,075	0,819	0,900	0,684	0,385	0,048	0,064 *
Actual value: First publication								
1-step forecast	20	1,522	1,245	1,083	0,918	0,044	-0,304	0,197
2-step forecast	20	1,468	1,237	1,092	0,889	0,098	-0,237	0,168
3-step forecast	20	1,177	1,038	0,902	0,773	0,256	-0,039	0,346
4-step forecast	20	1,410	1,075	1,083	0,824	0,268	-0,069	0,262
Actual value: Four quarters later								
1-step forecast	20	1,235	0,938	0,880	0,684	0,106	-0,242	0,091 *
2-step forecast	20	1,217	0,976	0,924	0,710	0,195	-0,140	0,100
3-step forecast	20	0,946	0,706	0,732	0,585	0,325	0,030	0,286
4-step forecast	20	1,132	0,720	0,875	0,603	0,323	-0,014	0,183

The results of the subsample 2001Q1-2005Q4 in table 4.5 exhibit widening spread in the forecasting error statistics. When the latest data is used as actuals, even the last observations in the forecasting model sample has been republished 38 times. All the forecast error statistics are smaller for the models utilizing the latest available data. Using latest available data as actuals 2- and 4-step forecasts differences are statistically significant on a 10% level (***, **, and * denote 1%, 5%, and 10% statistical significance, respectively). Now, in accordance with our presumptions, the performance of the LA models increase as the forecastable value gets more revised. However, the same is true for the RT models indicating a poor real-time forecasting ability.

Interestingly, the real-time data based forecasts tend to generate positively biased forecasts for 'true values' with MEs up to 0,385 %. In section 4.2 we concluded that the direction of revisions were quite evenly distributed with 48% of the revisions increasing the previous estimate. However, we indeed observed that previous estimates for some revision ranges were exceptionally larger proportionally than latter estimates (for instance Y2-Y1 60,29% of the revisions corrected the estimates downwards). Also 11/15 of the estimated revision periods in table 4.2 had an aggregate proportional downward correction. Hence, the models based on real-time data could for this reason suffer from a positive bias even using real-time data as actuals with prolonged forecast horizon l.

A peculiar fact is that the forecast models based on latest available data also perform better in real-time given $RMSE$ and MAE compared to the real-time models. However, we observe a negative bias with similar conclusions as described above for the real-time forecasts.

Table 4.6 Forecast error statistics, model: RGPD ~ ARIMA(p,1,0), 2006Q1-2010Q4

Forecast step	***N***	***RMSE***		***MAE***		***ME***		***p-***
		RT	**LA**	**RT**	**LA**	**RT**	**LA**	***value***
Actual value: Latest available data								
1-step forecast	20	2,175	2,524	1,587	1,883	-0,060	0,376	0,174
2-step forecast	20	2,233	2,714	1,551	1,867	-0,142	0,256	0,332
3-step forecast	20	2,320	2,451	1,505	1,707	-0,181	-0,025	0,622
4-step forecast	20	2,479	2,868	1,633	1,841	-0,230	0,075	0,314
Actual value: First publication								
1-step forecast	20	1,262	1,495	0,956	1,122	0,441	-0,004	0,217
2-step forecast	20	1,383	2,009	1,007	1,360	0,420	-0,004	0,313
3-step forecast	20	1,565	1,813	1,158	1,276	0,099	-0,001	0,504
4-step forecast	20	1,769	2,131	1,288	1,507	0,225	-0,002	0,369
Actual value: Four quarters later								
1-step forecast	20	2,115	2,415	1,614	1,883	-0,196	0,240	0,236
2-step forecast	20	2,202	2,707	1,566	1,945	-0,259	0,139	0,336
3-step forecast	20	2,306	2,531	1,701	1,811	-0,335	-0,179	0,459
4-step forecast	20	2,453	2,841	1,771	1,980	-0,347	-0,042	0,294

The differences between real-time and latest available models again diminish as the sample data approaches more recent periods in the subsample 2006Q1-2010Q4 (table 4.6). Again the real-time data based models exhibit smaller error statistics compared to those based on the latest available vintage. In this setup, the degree of the revisions between the vintages is presumably not very different as I find no statistically significant differences between models utilizing real-time data and the models based on the current vintage.

With this subsample, the RT models forecast the early realizations better relative to the more recent vintages of actuals. Interestingly, we observe quite large directional biases for both model types and for all the realizations (excluding first publication forecasts with LA models). Peculiarly, the forecast bias is now negative for RT models forecasting the revised realizations which is exactly opposite to the conclusions drawn from table 4.5. Hence, the results seem to be mixed and sub-sample dependent.

Table 4.7 Forecast error statistics, model: RGPD ~ ARIMA(2,1,0), full sample

Forecast step	***N***	***RMSE***		***MAE***		***ME***		***p-***
		RT	**LA**	**RT**	**LA**	**RT**	**LA**	***value***
Actual value: Latest available data								
1-step forecast	60	1,441	1,491	0,962	0,978	-0,060	-0,056	0,576
2-step forecast	59	1,494	1,526	0,996	0,973	-0,056	-0,039	0,844
3-step forecast	58	1,592	1,530	1,025	1,007	-0,080	-0,093	0,561
4-step forecast	57	1,669	1,757	1,098	1,106	-0,111	-0,100	0,538
Actual value: First publication								
1-step forecast	60	1,177	1,147	0,835	0,824	-0,055	-0,050	0,745
2-step forecast	59	1,223	1,264	0,873	0,859	-0,060	-0,043	0,810
3-step forecast	58	1,284	1,284	0,928	0,912	-0,088	-0,101	0,999
4-step forecast	57	1,455	1,498	1,057	1,048	-0,114	-0,102	0,764
Actual value: Four quarters later								
1-step forecast	58	1,487	1,496	1,029	1,029	-0,142	-0,138	0,924
2-step forecast	57	1,434	1,436	0,997	0,951	-0,123	-0,098	0,391
3-step forecast	56	1,425	1,369	0,987	0,955	-0,145	-0,145	0,146
4-step forecast	55	1,485	1,465	1,035	1,012	-0,140	-0,102	0,764

To mimic a more realistic real life forecasting exercise I do not allow for a free autoregressive order selection for each vintage. Instead, I estimate the model based on the data which would have been available for a forecaster at each particular quarter. The most optimal model specification is $ARIMA(2,1,0)$ which I use again to first construct forecasts with the full sample range data and then with the first sub-sample.

For the full sample forecasts we again have very similar error statistic results to the free selection setup (table 4.7). The statistics are not significantly different from each other when the model is based on real-time data and current vintage. Moreover, the comparison between the error statistic results between the free autoregressive order selection and fixed model specification suggest that the model specification does not contribute significantly to the forecasting errors. All the error statistics are strikingly identical for all the actuals and both vintages. Hence, it seems that the specification channel does not necessarily have a major impact on forecasting accuracy.

Table 4.8 Forecast error statistics, model: RGPD ~ ARIMA(2,1,0), 2001Q1-2005Q4

Forecast step	***N***	***RMSE***		***MAE***		***ME***		***p-***
		RT	**LA**	**RT**	**LA**	**RT**	**LA**	***value***
Actual value: Latest available data								
1-step forecast	20	0,962	0,783	0,750	0,641	0,243	-0,044	0,117
2-step forecast	20	1,070	0,866	0,873	0,698	0,302	0,011	0,085 *
3-step forecast	20	1,149	0,862	0,878	0,682	0,328	0,050	0,149
4-step forecast	20	1,025	0,788	0,868	0,665	0,352	0,063	0,009 ***
Actual value: First publication								
1-step forecast	20	1,477	1,250	1,055	0,939	0,016	-0,270	0,205
2-step forecast	20	1,442	1,248	1,055	0,921	0,062	-0,228	0,223
3-step forecast	20	1,169	1,017	0,892	0,769	0,246	-0,032	0,352
4-step forecast	20	1,341	1,052	1,051	0,806	0,236	-0,054	0,206
Actual value: Four quarters later								
1-step forecast	20	1,187	0,948	0,852	0,705	0,078	-0,208	0,091 *
2-step forecast	20	1,183	0,986	0,887	0,742	0,159	-0,131	0,161
3-step forecast	20	0,931	0,695	0,722	0,578	0,315	0,037	0,295
4-step forecast	20	1,061	0,693	0,843	0,585	0,290	0,001	0,132

In table 4.8, turning back to the same subsample as in table 4.5 with fixed forecasting model order, I find almost identical result to the free autoregressive selection set-up. The magnitudes of the forecast error statistics are extremely close to the errors in table 4.5 again suggesting that the specification channel contributes relatively little to the forecast deviations. The models based on the latest available data perform better relative to the real-time data models for all the actuals and horizons. Again, we observe the directional biases when attempting to forecast revised values with real-time models as well as forecasting real-time values with latest available models.

4.3.3 Monte Carlo simulation

The previous chapters and the sections in this chapter demonstrated, both empirically and theoretically, that forecasting accuracy of the linear univariate models is significantly affected by the data vintage selection for the forecasting model. In the above section it was found that the selection of actuals contributes to the forecast error statistics obtained using different vintages as a basis of the forecasting model. However, in section 4 it was shown that the Finnish quarterly real GDP revision process was almost purely characterized by news and hence the above forecasting experiments were based on a news process. Therefore, since the focus of this study is to find potential effects of news and noise revision processes on the forecasting accuracy, I will evaluate forecasting performance of models based on different data vintages with Monte Carlo simulations.

I have calibrated the parameters based on the findings in section 4. Following Clements and Galvao (2013) I simulate $AR(1)$ process for a quarterly GDP percentage change data. By altering the intercept term δ and the $AR(1)$ coefficient ϕ, I investigate three different time invariant scenarios (Scenario 1-3). The parameters are set such that the true process has a mean of 0,5 corresponding approximately to 0,51 average change q-o-q GDP in Finland. The standard deviation of the data generating process is set to unity (the average standard deviation across the vintages is 1,0123).

I set $w = 20$ such that the data goes through 20 revisions (or there are 20 different vintages w) before reaching its 'true' value $\tilde{Y}$. The data is constructed such that the series in each w vintage is comprised of news and noise components $v_{t,t+s}$ and $e_{t,t+s}$, such that $Y_{t,t+s} = \tilde{Y}_t + v_{t,t+s} + e_{t,t+s}$. I consider only pure news and noise processes by setting $\sigma_{e_l} = 0$ and $\sigma_{v_l} \neq 0$ for news process and $\sigma_{e_l} \neq 0$ and $\sigma_{v_l} = 0$ for noise process. Following Clements and Galvao (2013) I assume first and fourth revisions to be non-zero mean and the means are set to eight percent of the first release data. Also, the standard deviations of the revisions are set to diminish so that revisions 2-19 have a standard deviations of 50 % and the final revision 25 % of that of the first revision. Table. 4.9 summarizes the parameters for each of the scenarios for both noise and news processes. The methodology for the calibration of the parameters is disclosed in Appendix 2.

Table 4.9 Summary of the calibrated parameters for the Monte Carlo simulation

Noise	δ	ϕ	σ	μ_{e_1}	$\mu_{e_2} = \mu_{e_3} = \mu_{e_4}$	σ_{e_1}	$\sigma_{e_{2,4,\dots,20}}$	$\sigma_{e_{3,5,\dots,19}}$
Scenario 1	0,25	0,5	1	0,0741	0,0345	0,5590	0,1443	0,2500
Scenario 2	0,175	0,65	1	0,0741	0,0345	0,6370	0,1645	0,2849
Scenario 3	0,325	0,35	1	0,0741	0,0345	0,5168	0,1334	0,2311
News	δ	ϕ	σ	$\mu_{v_1} = \mu_{v_4}$	σ_{v_1}	$\sigma_{v_{2,\dots,10}}$	$\sigma_{v_{20}}$	
Scenario 1	0,25	0,5	1	0,0426	0,6575	0,3288	0,1644	
Scenario 2	0,175	0,65	1	0,0450	1,4118	0,7059	0,3530	
Scenario 3	0,325	0,35	1	0,0413	0,5233	0,2616	0,1308	

I use a similar approach with forecasting as in sections 4.3.1 and 4.3.2. The sample size of the first vintage forecasting model is 50 and due to expanding window method the sample size increases by one as $t + s$ increases by one. $l = 4$ again such that the forecasts are calculated recursively for four periods ahead. The RT forecasts are again created for each vintage and collectively compared to the LA forecasts made based on true value data $\tilde{Y}_t$. The forecast error evaluation to the actual values is made using the $\tilde{Y}_t$ data as latest available data and considering the vector comprised of $Y_{t+50,t}$, $t =$

1 ... 21 as the first publication data. Therefore, all the forecast comparisons made against the latest available data as actual comprise of 20 forecasts accuracy evaluations for each forecast step. When the first available data is used as actuals, the number of forecasts is reduced by one as the forecast horizon increases by one due to limited amount of actuals towards the end of the sample. The data generating process simulation and the above forecasting procedure is repeated 500 times.

Table 4.10 Monte Carlo forecast error statistics, average figures for all the forecast horizons

Scenario/ Revision process	Actual value: Latest available						Actual value: First publication					
	RMSE		**MAE**		**ME**		**RMSE**		**MAE**		**ME**	
	RT	**LA**	**RT**	**LA**	**RT**	**LA**	**RT**	**LA**	**RT**	**LA**	**RT**	**LA**
Scenario 1												
Noise	1,140	1,119	0,928	0,911	-0,019	-0,006	1,375	1,342	1,102	1,077	-0,011	-0,002
News	2,258	2,271	1,828	1,839	-0,009	-0,026	2,018	2,037	1,621	1,640	-0,024	-0,035
Scenario 2												
Noise	1,258	1,208	1,026	0,986	0,012	0,014	1,450	1,368	1,166	1,102	0,022	0,020
News	5,001	5,054	4,038	4,081	-0,017	-0,027	4,384	4,471	3,531	3,612	-0,028	-0,034
Scenario 3												
Noise	1,052	1,039	0,854	0,844	-0,028	-0,019	1,288	1,273	1,037	1,024	-0,015	-0,008
News	1,745	1,751	1,412	1,417	0,018	0,012	1,556	1,564	1,259	1,267	0,009	0,006

Table 4.10 summarizes the forecast error statistics as averages for all $l = 1, \dots, 4$ forecasts. The results reveal that the error statistics, when the revisions are reducing noise, are much smaller compared to the news process regardless of which data is used as actuals and for model construction. For instance, the $RMSE$ statistic for the model using latest available data as the basis of the model and actuals is 5,054 when the revision process adds news but only 1,208 when revisions reduce noise. The result is in harmony with the presumptions as the expected standard deviations for the news process are larger than for the noise reducing revision process. ME statistics are not extensively different from zero, and thus we can interpret that the relatively large standard deviation produces high volatility for the forecasts when revision process is characterized by news.

Furthermore, we see that the forecast errors increase with the AR parameter ϕ. Firstly, a larger autoregressive coefficient implies higher standard deviations for the revisions and consequently for the data series in the observed vintages. Hence, again we see higher volatility of the generated forecasts contributing to larger non-offsetting error statistics (also the magnitude of the differences between $RMSE$ and MAE indicate a higher variance between individual forecast errors). Secondly, in section 3.2.1 we concluded that the persistence in the data generating process contributes directly to the forecast deviations through the direct channel. This is indeed the case: the absolute differences in both $RMSE$ and MAE statistics between the models based on different data diminish as the autoregressive coefficient is reduced. These findings also match with Clements and Galvao (2013) who conclude that forecast accuracy spread is larger for news revisions and increases with higher persistence in autoregressive coefficients.

The differences between the forecasting models based on different vintages are much more subtle. However, there seems to be a systematic pattern in the relative size of the error statistics based on the model data. The average $RMSE$ and MAE statistics as well as the individual h-step averages (see Appendix 3.) are always lower for the models based on the latest available data when the revision process is characterized by reducing noise. The exact opposite is true for when the revisions are adding news. For the news process, the real-time models always (for both averages and individual horizons) perform better forecasting the real-time values relative to forecasting the true values. Similarly, the LA models performed better forecasting for the true values compared to real-time values when revision process is characterized by noise. Interestingly the pattern is opposite in a reverse scenario and, for instance, the real-time models perform relatively better forecasting the true value as opposed to the real-time values when revision process is characterized by noise. In table 4.5 we found that the LA models performed better at predicting the final value relative to real-time value but, for instance, 4.4 and 4.6 we found the opposite. This could again, with the directional biases observed in table 4.5 and 4.6, hint that our sub-samples were to some extent characterized differently, and hence, produce mixed results. The results in tables 4.4 and 4.6 are aligned with findings in Monte Carlo simulation.

5. Conclusions

Macroeconomic forecasting based on the national accounts data has proved to be cumbersome. The data is published with lag and it is subject for series of revisions therefore hampering the forecasting model construction and evaluation. Just recently, with emergence of real-time databases and expanding research, the forecasters have begun to understand more thoroughly the potential pitfalls when forecasting in the presence of data revisions. I have chosen to study the utilization of real-time vintages relative to latest available data. Real-time data is defined as the best available data at hand for forecaster during a particular period and latest available data is "today's" data which contains all the past revisions and can be found in any conventional macroeconomic database. Additionally, I investigate whether the separation of the revision process to news (and variance) adding and noise (variance) reducing revisions influences the selection of the data vintage for the forecasting model. I carry out a data analysis to uncover the characteristics of the revision process in Finland, perform an empirical forecasting study, and construct a Monte Carlo experiment with parameters calibrated based on the Finnish data.

I discover that the revisions to the Finnish real GDP have a significant influence on forecasting when forecasts are compiled with linear univariate models, i.e. $ARIMA(p, d, q)$ models. All three potential channels (directly through the data input, indirectly through the coefficients, and through the functional form specification) through which the forecasts are affected were present with indirect channel being the most influential for the highlighted sub-sample. Generating one-step-ahead forecasts produced forecast spreads of up to 3,631 p.p. for a single forecast period with an average spread of almost 1% for the whole sample. Forecasting deviations are most condensed during the periods of low positive growth which also exhibited smallest mean absolute revisions.

Despite the extensive research over the separation to news and noise, the application of this separation to forecasting is meager. However, uncovering the patterns in revisions could possibly facilitate the selection of the forecasting model data. If revisions are adding news, the real-time observations contain all the data available at time. Hence, the observations in the real-time vintage are efficient estimates for the true value. I find that the data revision process for Finnish q-o-q real GDP data is dominated by adding news characterization. Specifically for the early revisions, indicators such as significant correlations between the revision and the latter release, and high relative proportions of disturbance terms, UD's, in mean squared revision decomposition point to news revisions dominance.

I find significant deviations between the real-time and current vintage forecasts only when forecasts are based on early sample subsets and are contrasted on the latest available data and first anniversary revision data. Only a handful of significant forecast differentials between the models are found by deploying sub-samples. In the full sample evaluations, the forecast error statistics from forecasts based on real-time vintages tend to be smaller compared to the forecasts based on latest available data even when compared to latest available data as actuals. This is corroborated with the findings in the Monte Carlo experiment where the real-time based models produce smaller forecast error statistic when the revision process simulates adding news. On the contrary, when the revision process is characterized by noise, the latest available data models generate smaller forecast errors than the real-time data based models. Moreover, the Monte Carlo simulation suggests that the real-time models have the best real-time forecasting performance when the revision process is characterized by news and the latest available data models perform best forecasting the actual values when revisions are characterized by noise.

Appendix 1. Derivation of heteroscedasticity and autocorrelation consistent (HAC) variance formula

The HAC variance formula in equation 4.1 is utilized to accommodate potential serial correlations in revisions. To start the derivation, consider a simple regression model $R_t = \mu_t + \varepsilon_t$[11], where $t = 1, \ldots, n$. The error term may be serially correlated and (or) heteroscedastic. In that case, the ordinary least squares (OLS) coefficient estimators are consistent but the usual OLS standard errors generally are not. Instead, the standard errors should be computed by using a heteroscedasticity and autocorrelation consistent (HAC) estimator for the variance.

To derive the Newey & West (1987) HAC estimator consider a multiple regression model

$$\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\varepsilon}$$

where $\mathbf{y}$ is a $n \times 1$ dimensional vector of observations on the dependent variable, $\mathbf{X}$ is a $n \times (k + 1)$ dimensional matrix of regressors, $\boldsymbol{\beta}$ is a $(k + 1) \times 1$ dimensional vector of unknown regression coefficients, and $\boldsymbol{\varepsilon}$ is $n \times 1$ dimensional random vector with zero mean and unknown covariance matrix $\boldsymbol{\Omega}$. The OLS estimator is $\widehat{\boldsymbol{\beta}} = (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{y}$ and the Newey-West estimator of the covariance matrix for $\widehat{\boldsymbol{\beta}}$ is

$$\widehat{Var(\boldsymbol{\beta})} = (\mathbf{X}'\mathbf{X})^{-1}\widehat{\boldsymbol{\Omega}}(\mathbf{X}'\mathbf{X})^{-1}$$

where

$$\widehat{\boldsymbol{\Omega}} = \frac{n}{n-k}\left\{\sum_{t=1}^{n} \hat{\varepsilon}_t \boldsymbol{X}_t \boldsymbol{X}_t{}' + \sum_{v=1}^{q}\left(\left(1 - \frac{v}{q+1}\right)\sum_{t=v+1}^{n} \boldsymbol{X}_t\ \hat{\varepsilon}_t \hat{\varepsilon}_{t-v} \boldsymbol{X}_{t-v}{}' + \boldsymbol{X}_{t-v} \hat{\varepsilon}_{t-v}\ \hat{\varepsilon}_t\ \boldsymbol{X}_t{}'\right)\right\}.$$

The parameter q is the truncation parameter of the HAC estimator which represents the number of autocorrelations that is used to estimate the dynamic causal effects of the OLS residuals $\hat{\varepsilon}_t = \hat{y}_t - \mathbf{x}_t'\widehat{\boldsymbol{\beta}}$, $t = 1, \ldots, n$. As we are interested in the mean, i.e. the intercept of the regression model $R_t = \mu_t + \varepsilon_t$ the formula gets simplified. The OLS estimate of the mean is $\bar{R} = \frac{1}{n}\sum_{t=1}^{n} R_t$, matrix $\mathbf{X}$ is a $n \times 1$ dimensional vector of one, and $\mathbf{X}'\mathbf{X} = n$. Therefore we have that

[11] The reasoning follows OECD revision analysis interpretation guide (DiFonzo, 2005, pp. 25-26) influenced by Jenkinson & Stuttard (2004).

$$\widehat{var(\bar{R})} = \frac{1}{n(n-k)}\left\{\sum_{t=1}^{n}\hat{\varepsilon}_t^2 + 2\sum_{v=1}^{q}\left(\left(1-\frac{v}{q+1}\right)\sum_{t=v+1}^{n}\hat{\varepsilon}_t\hat{\varepsilon}_{t-v}\right)\right\}.$$

By adjusting the parameters to $k = 1$ and $q = 2$ (which is a suitable common selection for the sample size for all the comparison periods[12]) we arrive to $\widehat{var(\bar{R})}$ in equation 4.1.

[12] For the discussion regarding the selection of the truncation parameter, see for example Stock & Watson (2003, p. 641)

Appendix 2. Calibration and derivation of Monte Carlo parameters

In this appendix I derive the formulas used to calibrate the Monte Carlo parameters in section 4.3.3. The derivations are based on the statistical framework provided by Jacobs and Van Norden (2011), Clements and Galvao (2013) and Hännikäinen (2015). Both Clements and Galvao and Hännikäinen calibrate the variables to represent conditions in the US, hence the derivations shown here are essentially the same but the results differ at times due to differences in statistical characteristics.

The $t+s$ vintage estimate of Y for period t, denoted $Y_{t,t+s}$, $s=1,\ldots,w$, is considered to consist of true value $\tilde{Y}_t$ and news and noise components, $v_{t,t+s}$ and $e_{t,t+s}$ such that $Y_{t,t+s}=\tilde{Y}_t+v_{t,t+s}+e_{t,t+s}$. Recalling the earlier discussion we have that for pure news process $Cov\left(Y_{t,t+s},v_{t,t+s}\right)=0$ and for pure noise $Cov\left(Y_{t,t+s},e_{t,t+s}\right)=0$. Following Jacobs and Van Norden (2011) and Clements and Galvao (2013), I combine all w different vintages of Y_t, v_t, and e_t into a vector such that

$$\mathbf{Y}_t=\mathbf{i}\tilde{Y}_t+\mathbf{v}_t+\mathbf{e}_{\mathrm{t}}$$

where $\mathbf{i}$ is a w-vector of ones. The true process is assumed to follow $AR(1)$ data generating process such that

$$\tilde{Y}_t=\delta+\phi\tilde{Y}_{t-1}+\sum_{i=1}^{w}\mu_{v_i}+\sigma\eta_{1,t}+\sum_{i=1}^{w}\sigma_{v_i}\eta_{2t,i}\quad \text{(A1.1)}$$

where $\eta_{1,t}$ and $\eta_{2t,i}$ are $niid$ (0,1) ($i=1,\ldots,w$).

I set $w=20$ so that my dataset consists of initial vintage and twenty revised estimations with the last vintage corresponding to the final value. Both first vintage data, $Y_{t,t+1}$, and final data, $\tilde{Y}_t$, are covariance stationary processes, i.e. invariant in time. Hence, the expected value for $\tilde{Y}_t$ is simply the fixed reference point for a stationary process (see equation 3.3):

$$\mu_{\tilde{Y}}=\frac{\delta+\sum_{i=1}^{l}\mu_{v_i}}{1-\phi}.$$

The expected value for $Y_{t,t+1}$ is then $E\left(Y_{t,t+1}\right)=\mu_{\tilde{Y}}-\sum_{i=1}^{w}\mu_{v_i}-\mu_{e_1}$ or

$$\mu_{Y_{t,t+1}}=\frac{\delta+\phi\sum_{i=1}^{w}\mu_{v_i}}{1-\phi}-\mu_{e_1}.$$

Therefore, for the pure *noise* process we have

$$\mu_{\tilde{Y}} = \frac{\delta}{1-\phi}$$

and

$$\mu_{Y_{t,t+1}} = \frac{\delta}{1-\phi} - \mu_{e_1}.$$

For the pure *news* process we have

$$\mu_{\tilde{Y}} = \frac{\delta + \sum_{i=1}^{w} \mu_{v_i}}{1-\phi}$$

and

$$\mu_{Y_{t,t+1}} = \frac{\delta + \phi \sum_{i=1}^{w} \mu_{v_i}}{1-\phi}.$$

I follow the methodology of Clements and Galvao (2013) and assume that the first and fourth revision have non-zero mean. This assumption is supported by statistical analysis in section 4.2 (the authors actually consider the first and fifth revisions as non-zero). I assume that these revisions are both λ times the mean of the first release data. λ is set to 0,08 which is an approximate multiplier for the mean real GDP q-o-q growth in Finland, 0,5071 , resulting to the mean revision of 0,04.

An i^{th} revision r for observation t is defined as $r_{t,i} = Y_{t,t+1+i} - Y_{t,t+i} = \mu_{v_i} + \sigma_{v_i}\eta_{2t,i} - \mu_{e_i}$, for $i = 1, \dots, w$.

When revisions are comprised of *noise* we have that

$$E(r_{t,1}) = -\mu_{e_2} + \mu_{e_1} = \lambda E(Y_{t,t+1})$$

$$E(r_{t,2}) = -\mu_{e_3} + \mu_{e_2} = 0$$

$$E(r_{t,3}) = -\mu_{e_4} + \mu_{e_3} = 0$$

$$E(r_{t,4}) = -\mu_{e_5} + \mu_{e_4} = \lambda E(Y_{t,t+1})$$

$$E(r_{t,5}) = -\mu_{e_6} + \mu_{e_5} = 0$$

$$\vdots$$

$$E(r_{t,19}) = -\mu_{e_{20}} + \mu_{e_{19}}$$

$$E(r_{t,20}) = \mu_{e_{20}} = 0$$

$E(r_{t,20}) = \mu_{e_{20}} = 0$ also implies that $\mu_{e_{19}} = \mu_{e_{18}} \dots = \mu_{e_5} = 0$. Therefore $\mu_{e_4} = \mu_{e_3} = \mu_{e_2} = \lambda E(Y_{t,t+1})$ and $\mu_{e_1} = 2\lambda E(Y_{t,t+1})$. Plugging in the result for $E(Y_{t,t+1}) = \mu_{Y_{t,t+1}}$ yields:

$$\mu_{e_1} = \frac{\delta}{1-\phi}\frac{2\lambda}{1+2\lambda}$$

$$\mu_{e_2} = \mu_{e_3} = \mu_{e_4} = \frac{\delta}{1-\phi}\frac{\lambda}{1+2\lambda}.$$

The expected revisions for pure *news* process are

$$E(r_{1,t}) = \mu_{v_1}$$

$$\vdots$$

$$E(r_{20,t}) = \mu_{v_{20}}$$

such that $\mu_{v_2} = \mu_{v_3} = \mu_{v_5} = \cdots = \mu_{v_{20}} = 0$ and $E(r_{1,t}) = \mu_{v_1} = E(r_{4,t}) = \mu_{v_4}$. Noting $E(r_{1,t}) = \lambda E(Y_{t,t+1})$ results in

$$\mu_{v_1} = \lambda\frac{\delta+\phi\sum_{i=1}^{w}\mu_{v_i}}{1-\phi} = \frac{\lambda\delta}{1-(1+2\lambda)\phi} = \mu_{v_4}.$$

Next, I derive the standard deviations for the data revisions and for the first and final data vintages. First, to find the variance for final value $\tilde{Y}_t$ I modify the equation A1.1 yielding:

$$\tilde{Y}_t - \mu_{\tilde{Y}} = \phi(\tilde{Y}_{t-1} - \mu_{\tilde{Y}}) + \sigma\eta_{1,t} + \sum_{i=1}^{w}\sigma_{v_i}\eta_{2,t,i}. \quad \text{(A1.2)}$$

Variance is then found by multiplying A1.2 by $\tilde{Y}_t - \mu_{\tilde{Y}}$ and taking expectations of a modified equation for $\tilde{Y}_t$ gives

$$E(\tilde{Y}_t - \mu_{\tilde{Y}})^2 = \phi E[(\tilde{Y}_t - \mu_{\tilde{Y}})(\tilde{Y}_{t-1} - \mu_{\tilde{Y}})] + E[(\tilde{Y}_t - \mu_{\tilde{Y}})\sigma\eta_{1t}] + E\left[(\tilde{Y}_t - \mu_{\tilde{Y}})\sum_{i=1}^{w}\sigma_{v_i}\eta_{2t,i}\right].$$

Since, $E[(\tilde{Y}_t - \mu_{\tilde{Y}})\sigma\eta_{1t}] = \sigma^2 E(\eta_{1t}^2) = \sigma^2$ and $E[(\tilde{Y}_t - \mu_{\tilde{Y}})\sum_{i=1}^{w}\sigma_{v_i}\eta_{2t,i}] = \sum_{i=1}^{w}\sigma_{v_i}^2 E(\eta_{2t,i}) = \sum_{i=1}^{w}\sigma_{v_i}^2$ we can reformulate the variance for $\tilde{Y}_t$ as

$$Var(\tilde{Y}_t) = \phi\rho_1 Var(\tilde{Y}_t) + \sigma^2 + \sum_{i=1}^{w}\sigma_{v_i}^2$$

where ρ_1 is the first autocorrelation coefficient. Because for $AR(1)$ $\rho_1 = \phi$ (Box & Jenkins, 1970, p. 57) we have that for *noise* process

$$\sigma_{\tilde{Y}_t} = \sqrt{\frac{\sigma^2}{1-\phi^2}}$$

and for *news* process

$$\sigma_{\tilde{Y}_t} = \sqrt{\frac{\sigma^2+\sum_{i=1}^{w}\sigma_{v_i}^2}{1-\phi^2}}.$$

The variance for $Y_{t,t+1}$ is hence

$$Var(Y_{t,t+1}) = Var(\tilde{Y}_t - \sum_{i=1}^{w}\mu_{v_i} - \sum_{i=1}^{w}\sigma_{v_i}\eta_{2t,i} - \mu_{e_1} + \sigma_{e_1}\eta_{3t,i})$$

$$= Var(\tilde{Y}_t) + \sum_{i=1}^{w}\sigma_{v_i}^2 Var(\eta_{2t,i}) + \sigma_{e_1}^2 Var(\eta_{3t,i}) - 2\sum_{i=1}^{w}\sigma_{v_i} Cov(\tilde{Y}_t,\eta_{2t,i})$$

$$+ 2\sigma_{e_1} Cov(\tilde{Y}_t,\eta_{3t,i}) - 2\sum_{i=1}^{w}\sigma_{v_i}\sigma_{e_1} Cov(\eta_{2t,i},\eta_{3t,i}).$$

Since, $Cov(\tilde{Y}_t,\eta_{3t,i}) = Cov(\eta_{2t,i},\eta_{3t,i}) = 0$ and $Cov(\tilde{Y}_t,\eta_{2t,i}) = \sum_{i=1}^{l}\sigma_{v_i}$ we have that

$$Var(Y_{t,t+1}) = Var(\tilde{Y}_t) - \sum_{i=1}^{l}\sigma_{v_i}^2 + \sigma_{e_1}^2.$$

By combining this result with $\sigma_{\tilde{Y}_t}$ we have that for *noise* process

$$\sigma_{Y_{t,t+1}} = \sqrt{\frac{\sigma^2}{1-\phi^2} + \sigma_{e_1}^2}$$

and for *news* process

$$\sigma_{Y_{t,t+1}} = \sqrt{\frac{\sigma^2+\sum_{i=1}^{l}\sigma_{v_i}^2}{1-\phi^2} - \sum_{i=1}^{l}\sigma_{v_i}^2} = \sqrt{\frac{\sigma^2+\phi^2\sum_{i=1}^{l}\sigma_{v_i}^2}{1-\phi^2}}.$$

Finally, I derive the standard deviations for the revisions. Again following Clements and Galvao (2013) and the notation of Hännikäinen (2015) I set $\sigma_r = \alpha\sigma_{Y_{t,t+1}}$, where α denotes the ratio of the standard deviation of the first revision to the standard deviation of first real. Based on the Finnish data sample in section 4.2 α is set to 0.45. For simplicity, like Clements and Galvao, I assume that standard deviations for the revisions 2-19 are 50 of the σ_{r_1}so that we have $\alpha/2$, and for revision 20 standard deviations are 25 so that we have $\alpha/4$.

To derive the standard deviations for the revisions consider the variance for the first revision

$$\sigma_{r_1}^2 = Var(\mu_{v_1} + \sigma_{v_1}\eta_{2t,1} - \mu_{e_2} + \sigma_{e_2}\eta_{3t,2} + \mu_{e_1} - \sigma_{e_1}\eta_{3t,1}).$$

We see directly that if the revisions are characterized by *noise*, $\sigma_{r_1}^2 = \sigma_{e_2}^2 + \sigma_{e_1}^2$ and for *news* we have $\sigma_{r_1} = \sigma_{v_1}$. Iterating forward the standard deviations for *noise* process we have

$$\sigma_{r_1}^2 = \sigma_{e_2}^2 + \sigma_{e_1}^2$$

$$\sigma_{r_2}^2 = \sigma_{e_3}^2 + \sigma_{e_2}^2$$

$$\vdots$$

$$\sigma_{r_{19}}^2 = \sigma_{e_{20}}^2 + \sigma_{e_{19}}^2$$

$$\sigma_{r_{20}}^2 = \sigma_{e_{20}}^2.$$

Since we assumed equal standard deviations for revision 2-19 we can deduce that $\sigma_{e_2}^2 = \sigma_{e_4}^2 = \cdots = \sigma_{e_{18}}^2 = \sigma_{e_{20}}^2$ and $\sigma_{e_3}^2 = \sigma_{e_5}^2 = \cdots = \sigma_{e_{17}}^2 = \sigma_{e_{19}}^2$. Because we also imposed that $\sigma_{r_{20}} = \frac{\alpha}{4}\sigma_{Y_{t,t+1}}$ and $\sigma_{r_{2,..19}} = \frac{\alpha}{2}\sigma_{Y_{t,t+1}}$ which implies $\frac{\sigma_{r_{20}}^2}{\sigma_{r_{19}}^2} = \frac{1}{4}$. By inserting these findings to the identity $\alpha\sigma_{Y_{t,t+1}} = \sigma_{e_2} + \sigma_{e_1}$ we find that

$$\sigma_{e_1} = \sqrt{\frac{15\alpha^2}{16-15\alpha^2}\frac{\sigma^2}{1-\phi^2}}$$

$$\sigma_{e_2} = \sigma_{e_4} = \cdots = \sigma_{e_{18}} = \sigma_{e_{20}} = \sqrt{\frac{\alpha^2}{16}\frac{16}{16-15\alpha^2}\frac{\sigma^2}{1-\phi^2}}$$

$$\sigma_{e_3} = \sigma_{e_5} = \cdots = \sigma_{e_{17}} = \sigma_{e_{19}} = \sqrt{\frac{3\alpha^2}{16}\frac{16}{16-15\alpha^2}\frac{\sigma^2}{1-\phi^2}}$$

For the *news* process we utilize the fact that $\sum_{i=1}^{l}\sigma_{v_i}^2 = 5.5625\sigma_{v_1}^2$. Plugging this to $\sigma_{r_1} = \sigma_{v_1} = \alpha\sigma_{Y_{t,t+1}}$ yields

$$\sigma_{v_1} = \sqrt{\frac{\alpha^2\sigma^2}{1-(1+5.5625\alpha^2)\phi^2}}$$

and $\sigma_{v_{2,..,19}} = \frac{\sigma_{v_1}}{2}$ and $\sigma_{v_{20}} = \frac{\sigma_{v_1}}{4}$.

Appendix 3. Full Monte Carlo simulation results

Table A2.1 Forecasting error statistics from Monte Carlo simulation, noise revisions

Forecast step	Actual value: Latest available						Actual value: First publication					
	RMSD		***MAE***		***ME***		***RMSD***		***MAE***		***ME***	
	RT	**LA**	**RT**	**LA**	**RT**	**LA**	**RT**	**LA**	**RT**	**LA**	**RT**	**LA**
Scenario 1												
1-step forecast	1,081	1,017	0,874	0,823	-0,017	-0,004	1,379	1,330	1,102	1,064	-0,011	0,002
2-step forecast	1,149	1,136	0,934	0,924	-0,017	-0,004	1,372	1,335	1,099	1,070	-0,010	0,000
3-step forecast	1,166	1,160	0,951	0,946	-0,020	-0,006	1,375	1,349	1,104	1,083	-0,011	-0,003
4-step forecast	1,166	1,164	0,952	0,950	-0,022	-0,008	1,374	1,356	1,105	1,090	-0,012	-0,006
Scenario 2												
1-step forecast	1,124	1,011	0,908	0,817	0,001	0,006	1,417	1,328	1,131	1,062	0,015	0,021
2-step forecast	1,262	1,215	1,031	0,990	0,009	0,011	1,442	1,351	1,158	1,085	0,022	0,020
3-step forecast	1,311	1,286	1,073	1,053	0,017	0,017	1,466	1,384	1,183	1,118	0,028	0,021
4-step forecast	1,333	1,318	1,093	1,081	0,022	0,022	1,476	1,410	1,193	1,141	0,024	0,016
Scenario 3												
1-step forecast	1,028	0,993	0,834	0,806	-0,028	-0,017	1,301	1,275	1,045	1,023	-0,016	-0,004
2-step forecast	1,056	1,049	0,858	0,851	-0,033	-0,023	1,290	1,274	1,039	1,025	-0,016	-0,009
3-step forecast	1,059	1,056	0,861	0,857	-0,028	-0,019	1,285	1,275	1,036	1,027	-0,016	-0,012
4-step forecast	1,063	1,059	0,864	0,860	-0,025	-0,017	1,275	1,268	1,029	1,022	-0,010	-0,008

Table A2.2 Forecasting error statistics from Monte Carlo simulation, news revisions

	Actual value: Latest available						Actual value: First publication					
Forecast step	***RMSD***		***MAE***		***ME***		***RMSD***		***MAE***		***ME***	
	RT	**LA**	**RT**	**LA**	**RT**	**LA**	**RT**	**LA**	**RT**	**LA**	**RT**	**LA**
Scenario 1												
1-step forecast	2,190	2,224	1,772	1,800	-0,007	-0,026	1,950	1,986	1,562	1,597	-0,023	-0,042
2-step forecast	2,264	2,275	1,832	1,843	-0,017	-0,035	2,000	2,021	1,606	1,626	-0,026	-0,039
3-step forecast	2,285	2,290	1,851	1,856	-0,006	-0,021	2,045	2,058	1,644	1,656	-0,024	-0,032
4-step forecast	2,291	2,294	1,855	1,858	-0,006	-0,020	2,078	2,085	1,672	1,680	-0,025	-0,030
Scenario 2												
1-step forecast	4,710	4,820	3,782	3,878	-0,017	-0,023	4,146	4,271	3,315	3,432	-0,019	-0,024
2-step forecast	5,001	5,059	4,043	4,083	-0,011	-0,022	4,328	4,427	3,481	3,574	-0,030	-0,036
3-step forecast	5,118	5,146	4,140	4,164	-0,017	-0,030	4,479	4,552	3,617	3,686	-0,035	-0,041
4-step forecast	5,177	5,192	4,188	4,200	-0,021	-0,034	4,583	4,634	3,710	3,757	-0,030	-0,034
Scenario 3												
1-step forecast	1,730	1,742	1,403	1,415	0,013	0,006	1,534	1,549	1,238	1,252	0,009	0,001
2-step forecast	1,747	1,752	1,413	1,417	0,016	0,010	1,550	1,558	1,253	1,261	0,009	0,006
3-step forecast	1,750	1,754	1,415	1,418	0,019	0,013	1,563	1,569	1,266	1,272	0,005	0,004
4-step forecast	1,754	1,758	1,417	1,420	0,024	0,018	1,577	1,582	1,279	1,284	0,010	0,011

Bibliography

Aruoba, S. Borağan. "Data revisions are not well behaved." Journal of money, credit and banking 40.2- 3 (2008): 319-340.

Box, George EP, and Gwilym Jenkins. "M. (1970)." Time Series Analysis: Forecasting and Control." Holden-D. iv, San Francisco.

Carson, Carol S., and Lucie Laliberté. "Assessing accuracy and reliability: a note based on approaches used in national accounts and balance of payments statistics." (2002): 1-21.

Clements, Michael P. "Real-time factor model forecasting and the effects of instability." Computational Statistics & Data Analysis (2015).

Clements, Michael P., and Ana Beatriz Galvão. "Real-Time Forecasting Of Inflation And Output Growth With Autoregressive Models In The Presence Of Data Revisions." *Journal of Applied Econometrics* 28.3 (2013): 458-477.

Cole, Rosanne. "Data errors and forecasting accuracy." Economic forecasts and expectations: analysis of forecasting behavior and performance. NBER, 1969. 47-82.

Croushore, Dean. "Forecasting with real-time macroeconomic data." Handbook of economic forecasting 1 (2006): 961-982.

Croushore, Dean, and Tom Stark. "A real-time data set for macroeconomists: Does the data vintage matter?." Review of Economics and Statistics 85.3 (2003): 605-617.

Denton, Frank T., and John Kuiper. "The effect of measurement errors on parameter estimates and forecasts: A case study based on the Canadian preliminary national accounts." The Review of Economics and Statistics (1965): 198-206.

Diebold, Francis X., and Glenn D. Rudebusch. "Forecasting output with the composite leading index: A real-time analysis." Journal of the American Statistical Association 86.415 (1991): 603-610.

Diebold, Francis X., and Roberto S. Mariano. "Comparing Predictive Accuracy." Journal of Business & Economic Statistics 13.3 (1995): 253-63.

DiFonzo, T. "The OECD project on revisions analysis: First elements for discussion." Ponencia presentada en la OECD steseg Meeting. 2005.

Elliott, Graham. "Comments on 'Forecasting with a real-time data set for macroeconomists'." Journal of Macroeconomics 24.4 (2002): 533-539.

Faust, Jon, John H. Rogers, and Jonathan H. Wright. "News and noise in G-7 GDP announcements." Journal of Money, Credit and Banking (2005): 403-419.

Friedman, Milton. "Lerner on the Economics of Control." The Journal of Political Economy (1947): 405-416.

Granger, Clive WJ, and Paul Newbold. "Some comments on the evaluation of economic forecasts." Applied Economics 5.1 (1973): 35-47.

Howrey, E. Philip. "The use of preliminary data in econometric forecasting." The Review of Economics and Statistics (1978): 193-200.

Hännikäinen, Jari. "Essays on Real-Time Macroeconomic Forecasting." (2015).

Jenkinson, G., and N. Stuttard. "Revisions information in ONS first releases." Economic Trends 604 (2004): 70-72.

Koenig, Evan F., Sheila Dolmas, and Jeremy Piger. "The use and abuse of real-time data in economic forecasting." Review of Economics and Statistics 85.3 (2003): 618-628.

Mankiw, N. Gregory, and Matthew D. Shapiro. News or noise? An analysis of GNP revisions. No. w1939. National Bureau of Economic Research, 1986.

Marcellino, Massimiliano. "A linear benchmark for forecasting GDP growth and inflation?." Journal of Forecasting 27.4 (2008): 305-340.

Marcellino, Massimiliano, James H. Stock, and Mark W. Watson. "Macroeconomic forecasting in the euro area: Country specific versus area-wide information." European Economic Review 47.1 (2003): 1-18.

McKenzie, Richard. "Undertaking revisions and real-time data analysis using the OECD main economic indicators original release data and revisions database." (2006).

McKenzie, R., and Z. Adam. "Revisions in Quarterly GDP of OECD countries: an Update." *Working Party of National Accounts* (2007).

Meese, Richard, and John Geweke. "A comparison of autoregressive univariate forecasting procedures for macroeconomic time series." Journal of Business & Economic Statistics 2.3 (1984): 191-200.

Morgenstern, Oskar, and Princeton University Press. On the accuracy of economic observations. Vol. 2. Princeton: Princeton University Press, 1963.

Newey, Whitney K., and Kenneth D. West. "Hypothesis testing with efficient method of moments estimation." International Economic Review (1987): 777-787.

Rudebusch, Glenn D. "Is the Fed too timid? Monetary policy in an uncertain world." Review of Economics and Statistics 83.2 (2001): 203-217.

Stark, Tom, and Dean Croushore. "Forecasting with a real-time data set for macroeconomists." Journal of Macroeconomics 24.4 (2002): 507-531.

Stekler, H. O. "Data revisions and economic forecasting." Journal of the American Statistical Association 62.318 (1967): 470-483.

Stock, James H. "Forecasting economic time series." A Companion to Theoretical Econometrics, Blackwell Publishers (2001): 562-84.

Stock, James H., and Mark W. Watson. Introduction to econometrics. Vol. 104. Boston: Addison Wesley, 2003.

Swanson, Norman. "Forecasting using first-available versus fully revised economic time-series data." *Studies in Nonlinear Dynamics & Econometrics* 1.1 (1996).

Trivellato, Ugo, and Enrico Rettore. "Preliminary data errors and their impact on the forecast error of simultaneous-equations models." Journal of Business & Economic Statistics 4.4 (1986): 445-453.

Zellner, Arnold. "A statistical analysis of provisional estimates of gross national product and its components, of selected national income components, and of personal saving." Journal of the American Statistical Association 53.281 (1958): 54-65.

Öller, Lars-Erik, and Karl-Gustav Hansson. "Revision of national accounts." Journal of Business Cycle Measurement and Analysis 2004.3 (2005): 363-385.