CESifo Working Papers

www.cesifo.org/wp

Monetary Policy Analysis in Real-Time. Vintage Combination from a Real-Time Dataset.

Carlo Altavilla
Matteo Ciccarelli

CESIFO WORKING PAPER NO. 3372
CATEGORY 7: MONETARY POLICY AND INTERNATIONAL FINANCE
MARCH 2011

PRESENTED AT CESIFO AREA CONFERENCE ON MACRO, MONEY & INTERNATIONAL FINANCE, FEBRUARY 2011

CESifo *Working Paper No. 3372*

Monetary Policy Analysis in Real-Time. Vintage Combination from a Real-Time Dataset.

Abstract

This paper provides a general strategy for analyzing monetary policy in real time which accounts for data uncertainty without explicitly modelling the revision process. The strategy makes use of all the data available from a real-time data matrix and averages model estimates across all data releases. Using standard forecasting and policy models to analyze monetary authorities' reaction functions, we show that this simple method can improve forecasting performance and provide reliable estimates of the policy model coefficients associated with small central bank losses, in particular during periods of high macroeconomic uncertainty.

JEL-Code: E520, E580, C320, C530, C820.

Keywords: monetary policy, Taylor rule, real-time data, great moderation, forecasting.

Carlo Altavilla
University of Naples 'Parthenope'
Via Medina, 40
80133 Naples
Italy
altavilla@uniparthenope.it

Matteo Ciccarelli
European Central Bank
Kaiserstrasse 29
60311 Frankfurt am Main
Germany
matteo.ciccarelli@ecb.int

We thank Christopher Carroll, Domenico Giannone, Roberto Golinelli, Kajal Lahiri, and Benoit Mojon for helpful comments and suggestions and conference participants at the CESifo Area Conference on Macro, Money and International Finance and the Fourth Italian Congress of Econometrics and Empirical Economics for comments. This paper should not be reported as representing the views of the European Central Bank (ECB), or ECB policy. Remaining errors are our own responsibility.

1 Introduction

Data revisions matter for forecasting and policy analysis (e.g. Croushore 2010). Adjusting for the existence of data revisions requires a forecaster or a policymaker to explicitly model the data measurement process. It has been broadly accepted in the literature that data revisions may contain news – so that when data are initially released they are optimal forecasts of later data and revisions are unpredictable – or noise – in which case the revision is correlated with the initial data and revisions are predictable (e.g. Aruoba, 2008). However, as Jacobs and Van Norden (2010) point out, revisions need not be either pure news or pure noise but can be a mixture. In this case, modelling efforts are greatly complicated and the gains in terms of, e.g., forecast accuracy might not be so big.

In this paper we aim to provide a general estimation strategy for performing monetary analyses in real time with forecasting and policy models. The proposed strategy needs to be robust to the revision process, regardless of its exact nature and without the need to model it. The paper poses two main questions: (1) How should the policymaker handle data uncertainty? and (2) How much information contained in a real-time data set would be convenient or optimal to use?

Our starting point is that, in general, the existence of multiple vintages of data for a given variable might render incorrect the use of a single vintage when evaluating a model because the stochastic relationship between vintages is not taken into account and therefore the estimated uncertainty is distorted.

To avoid vintage dependency of policy analyses and forecast accuracy, we propose combining the information contained in all available vintages. Our strategy (i) considers the available data vintages from a real-time data matrix as different units of a kind of panel dataset; (ii) applies mean group estimation techniques to average estimates across data releases in order to average out part of the measurement error, and (iii) suggests that a model should be estimated by using all vintages of a real-time data matrix. As revisions seem to be more sizeable and volatile in periods of high macroeconomic volatility, our prior is that this strategy could work better precisely in these periods.

Our approach shares the similar views of previous studies. Guerrero (1993), for instance, proposes combining historical and preliminary information to obtain timely time series data, using simple regression models that link preliminary and final data. Translated into the language of a real time data matrix, the approach relates the final column of the data (the latest vintage available) with the diagonal (real-time data), but disregards the revision process incorporated in all other vintages.

Patterson (2003) combines the data generating process and the data measurement process with a nesting model that comprises the links amongst generic variables and the links within data vintages on the same variable and across variables, extracts a common trend for each variable and then checks whether the common trends cointegrate. The analysis is, therefore, performed on levels and becomes unfeasible when the number of vintages or variables increases. Moreover, being based on levels, the approach is more subject to contamination due to benchmark revisions, i.e. those changes that statistical agencies make to their methodologies or statistical changes such as changes of base year or seasonal weights.

Pesaran and Timmermann (2007) evaluate forecast combination methods which weigh or pool estimates based on estimation windows of different lengths. The analysis, however, is not performed in real time and evaluates the combination strategies based only on one sample (the latest vintage available).

Finally, in a previous work (Altavilla and Ciccarelli 2007) we explored the role of both model and vintage combination in forecasting, using a mean-group estimation approach to exploit the information contained in all vintages of the variables of interest. Unlike that paper, however, the present one considers only vintage combinations and performs a pure real-time analysis, testing the validity of the approach in an economic model.

Using standard forecasting and policy models to analyse monetary authorities' reaction functions, we test our averaging strategy to estimate the parameters of a policy rule and to forecast the interest rate. The choice of the latter as our target variable is motivated by the fact that this variable, though being a function of possibly revised variables in standard models, is not subject to revision itself, which makes the definition of the actual value needed to evaluate, e.g., a forecast straightforward. In forecasting variables with a revisable target, instead, the latter would not be known with certainty as these variables are always subject to (possibly large) revisions. As a consequence, real-time policy evaluations and forecasts would crucially depend on the actual value used to compute the various statistics and tests.

We perform our analysis over the vintages from 1965Q4-2010Q4, checking in particular whether results differ over the two sub-samples 1970-1984 (pre Great Moderation) and 1984-2010 (Great Moderation and beyond). Our results show that our approach can not only improve forecasting performance but also provides reliable and stable estimates of policy rule parameters and central bank losses, in particular over parts of the sample characterized by high volatility.

Exponential smoothing strategies are also employed to average over past vintages and choose the smoothing parameter (and therefore the number of vintages) which minimizes a loss function in an optimal choice of past available information. It turns out that the best strategy – in terms of central bank losses and mean squared prediction error – would require using all past data vintages and averaging the estimates with equal weights.

The rationale for using all vintages instead of simply the last few years – after which data releases become rational in the sense of Swanson and van Dijk (2006) – and also including benchmark revisions, is the same as the rationale for combining different models with a possibly naïve weighting scheme (Stock and Watson 2004; Timmermann 2006; Clark and McCracken 2010). Most revisions after the first few years mainly reflect differences in the statistical methods used to construct the data. Although constructed somewhat differently, these data measure the same economic concept (e.g. GDP). It seems, therefore, natural to estimate a model for each of these measures and then average the estimates. Our results suggest that attaching an equal weight to the estimates based on all data vintages constructed with different statistical methods is a valid risk diversification strategy for coping with data uncertainty.

The road map of the paper is the following. We first describe the data set used in the analysis and show that some features of the revisions are time-varying and can be difficult to model. This implies the need for an estimation strategy which accounts for data revision in a general manner and is robust to the choice of the appropriate data set without necessarily modelling the revision process (Section 2). Next, we introduce the idea of combining estimates from different data vintages to help us cope with these implications (Section 3). We then illustrate the appropriateness of our strategy in real-time monetary policy analyses (Section 4), and finally we conclude and discuss the implications of our findings (Section 5).

2 Selected features of revisions

Macroeconomic time series are revised for two main reasons. First, statistical agencies regularly update their initial estimates when additional information becomes available. Second, comprehensive or benchmark revisions to time series are introduced roughly every five years because new aggregation methods, measurement concepts or survey methods are introduced in statistical accounting systems leading to a significant change in the structure of the data.

While the nature and the importance of regular revisions has been extensively studied in the literature for most variables commonly used in policy and forecasting models, the effect of benchmark revisions on forecasting and policy analysis has been discussed to a lesser extent.[1] In their pioneering study, Croushore and Stark (2001) warned that benchmark revisions are not so easy to characterize. More recently, using a standard backward- and forward-looking specification of the Phillips curve, Siklos (2008) analyses the statistical significance of benchmark revisions to key US macroeconomic variables and finds that benchmark revisions entail some important information content which could be used to improve the forecasting performance of models.

Starting from these findings, we do not follow the usual practice of pulling benchmark vintages out from the data and focusing on the time period between successive benchmark revisions, or of adjusting the data to somehow homogenize vintages following benchmark revisions. On the contrary, all revisions play an important role in the rationalization of our combination approach, which aims precisely at averaging model estimates across different measurements of the same economic concepts (i.e. GDP growth or inflation) taken at different vintages.

Note that the literature has often monitored with particular care the benchmark revision introduced in 1996, which entailed a shift from fixed-weighted to chain-weighted measures of national income and product account (NIPA) data.[2] As in our combination approach we average the estimates based on different vintages and not the data itself, we do not treat any benchmark revision with special care. Moreover, our results do not seem to be affected by the change to chain weighting any more than they are by other previous revisions (see below).

In the remaining part of this section we will briefly revisit some main features of revisions already discussed in the previous literature and characterize in particular two time-varying aspects: (1) the magnitude and volatility of revisions and (2) the news/noise features. We focus only on two macroeconomic variables – real GDP growth, and GDP deflator inflation – which will be used in our subsequent analysis. In this respect, therefore, our analysis is partial and tailored to our purposes.

We use the data of the real-time data set for macroeconomists developed at the Federal Reserve Bank of Philadelphia and described in great detail, for instance, in Croushore and Stark (2001).

[1]For studies on regular revisions, see e.g. Mankiw et al. (1984), Mankiw and Shapiro (1986), and Aruoba (2008).

[2]In a fixed-weight measure of GDP everything produced today contributes to GDP according to what it was worth in a past benchmark year. On the contrary, chain weights move the benchmark along with time. Under chain weights, therefore, a given good entering the GDP is worth the average, or chain, of what it costs to buy it in the current and previous year.

2.1 Size and volatility

First, we examine the magnitude and volatility of the revisions of the growth rate of real output and the inflation rate based on the output deflator. Quarterly observations of quarterly vintages are used. Percentages are expressed in annual terms.

We denote with y_t^v the realization of the v-th vintage of the generic variable y at time t. A revision from a vintage v to a vintage $v+k$ is therefore defined as

$$r_t^{v+k} = y_t^{v+k} - y_t^v. \tag{1}$$

Consistently with the literature, we define the *final* revision by choosing $k = 12$ quarters (see e.g. Aruoba, 2008). Note, however, that this is only a standard convention, as there is never a final revision as such.

Our available data is schematically represented by the following table:

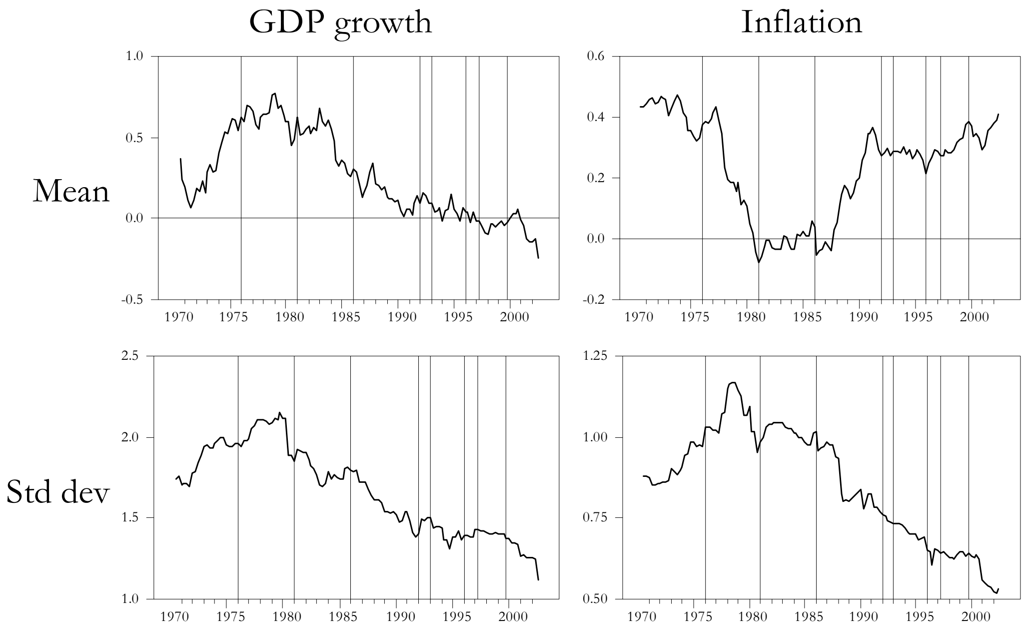
	$1965Q4$	$1966Q1$	$\cdots$	$2010Q4$	
$1947Q1$	$y_{47:1}^{65:4}$	$y_{47:1}^{66:1}$	$\cdots$	$y_{47:1}^{10:4}$	$1947Q1$
$1947Q2$	$y_{47:2}^{65:4}$	$y_{47:2}^{66:1}$	$\cdots$	$y_{47:2}^{10:4}$	$1947Q2$
$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$
$1965Q3$	$y_{65:3}^{65:4}$	$y_{65:3}^{66:1}$	$\cdots$	$y_{65:3}^{10:4}$	$1965Q3$
		$y_{65:4}^{66:1}$	$\cdots$	$y_{65:4}^{10:4}$	$1965Q4$
			$\ddots$	$\vdots$	$\vdots$
				$y_{10:3}^{10:4}$	$2010Q3$

where vintages are in the columns and time series observations in the rows. Therefore, the first available total revision is $r_{65:3}^{68:4} = y_{65:3}^{68:4} - y_{65:3}^{65:4}$; and the last one is $r_{07:3}^{10:4} = y_{07:3}^{10:4} - y_{07:3}^{07:4}$. The time series of revisions, therefore, goes from 1965Q3 to 2007Q3.

Time-varying descriptive statistics of the final revisions are reported in Figure 1, which shows the (centred) rolling sample average and standard deviation of the revisions for real GDP growth (left panel) and inflation (right panel), using a 10-year moving window. Vertical gridlines represent benchmark vintage dates.[3]

[3] For both real GDP and the GDP deflator, the vintages that incorporate benchmark revisions are the following: 1965:4, 1976:1, 1981:1, 1986:1, 1992:1, 1993:1, 1996:1, 1997:2, 1999:4, 2000:1, 2000:2, 2004:1, and 2009:3.

Figure 1: Descriptive Statistics – 10-year moving window

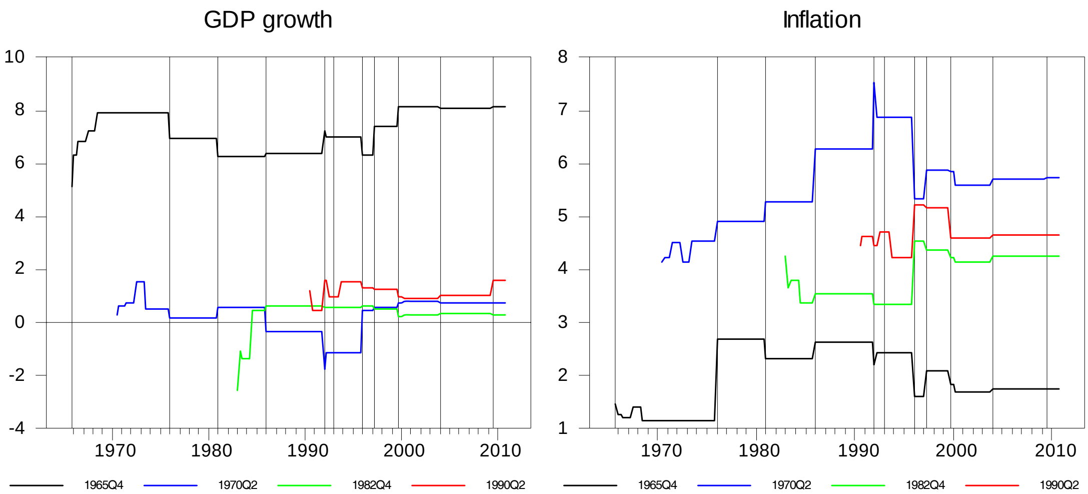


Note: the chart reports the centred rolling sample average (first row) and standard deviation (second row) of the revision for real GDP growth (left panel) and inflation (right panel), using a 10-year moving window. Vertical gridlines represent benchmark vintage dates.

That revisions are not well behaved has largely been discussed in the literature (see e.g. Aruoba, 2008). The rolling statistics, however, also reveal that the revision process has been far from stationary and its characteristics are time-varying. The charts show a strong positive and significant autocorrelation structure of both mean and standard deviation for GDP growth and inflation, meaning that the size of the revisions and the uncertainty surrounding them have not only been relatively high but also substantially persistent. Since approximately 1984, the volatility and – to some extent – the positive mean of the revisions display the same negative trend, which has been used to describe the Great Moderation period (Stock and Watson, 2003), possibly implying that the two phenomena (the macroeconomic uncertainty and a biased and volatile revision process) are strictly interrelated.

The size and variability of the revisions can be also illustrated from a different perspective, as is done e.g. in Croushore (2010, Fig. 1). For any observation date, one can track GDP growth and the inflation rates as they were successively revised across all the available vintages since their initial releases. This is done in Figure 2 for both variables as measured, for instance, in 1965Q4, 1970Q2, 1982Q4, and 1990Q2.

Figure 2: Values of various quarters as measured at all vintages



Note: the figure shows for four representative time periods (i.e. 1965Q4, 1970Q2, 1982Q4, and 1990Q2) the value of the GDP growth and the inflation rates for each observation date as they were successively revised across all the available vintages since their initial releases. Vertical gridlines indicate benchmark revision dates.

It is clear that both variables have undergone several revisions of different magnitude with many redefinitions, which have made them fluctuate in a non-monotonic manner. Among other things, for instance, the charts show that after the more recent revisions the values of both variables are sometimes not greatly dissimilar from what they had been at some point in the past or even at their first release. Subsequent revisions have, therefore, introduced different measurements of the same economic concepts, but there is no reason to believe that empirical tests based on the more recent measures would be more valid than those based on older data.

In fact, with so many different ups and downs in the two variables, the result of any estimation method, model validation or empirical test of an economic theory is going to depend significantly on the choice of the vintage. This implication has already been illustrated by Croushore and Stark (2003) for various economic examples, where testing a certain hypothesis on data sets of different vintages might lead to conflicting results, thus casting doubt on the robustness of given economic theories. In our view, Figure 2 indicates that, rather than checking a given model vintage by vintage, it can be more natural to estimate or test the model on each vintage and then average the results across vintages.

2.2 Testing news and noise

The next issue is the characterization of a revision in terms of news or noise. The literature has classified final revisions into two categories (see e.g. Aruoba, 2008). If revisions are *noise*, the initial releases are simply equal to the final series (the truth) measured with errors and revisions are predictable; if revisions are *news*, the initial releases are optimal forecasts of the later data and revisions are unpredictable. The empirical evidence of the news/noise feature of revisions is mixed and has been shown to be variable-, country- and sample-dependent. Moreover, Jacobs and van Norden (2010) have recently argued that revisions, rather than being pure news or pure noise, can be a mixture. In this case, efforts at modelling to incorporate the structure of the revision process are greatly complicated and the gains in terms of e.g. forecast accuracy might not be significant.

The difficulty in characterizing revisions as pure news or pure noise can also stem from the fact that they might add news or reduce noise in a time-varying manner. In Figure 3 we report the rolling p-values of the F-test for checking the news/noise hypotheses with Mincer-Zarnowitz (1969) type regressions. To classify revisions as news or noise we consider the regressions:

$$r_t^{v+k} = \alpha_1 + \beta_1 y_t^v + \varepsilon_t^1 \tag{2}$$

$$r_t^{v+k} = \alpha_2 + \beta_2 y_t^{v+k} + \varepsilon_t^2 \tag{3}$$

where the null $H_0 : \alpha_1 = \beta_1 = 0$ would test the news hypothesis, and the null $H_0 : \alpha_2 = \beta_2 = 0$ would test the noise hypothesis. As Aruoba (2008) suggests, both hypotheses should be tested as these hypotheses are mutually exclusive but not collectively exhaustive, which implies that both hypotheses could in principle be rejected.[4]

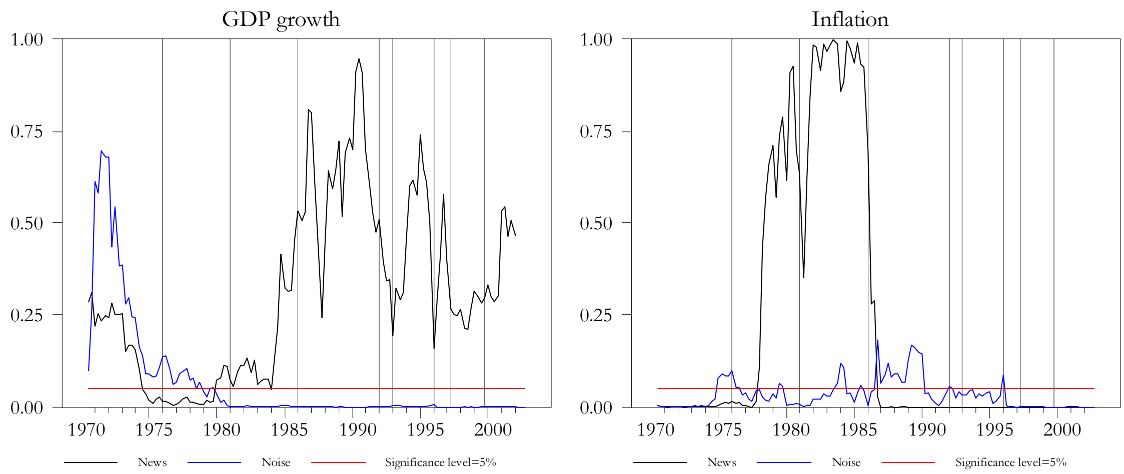
The rolling p-values using a 10-year moving window indicate that these characteristics are indeed time-varying. Interestingly, while for inflation both hypothesis are rejected over the whole sample except for the disinflation period (1980-1986), where revisions were news, for GDP growth the test rejects the noise hypothesis in favour of news over part of the disinflation and in the great moderation period. In other words, the test seems to indicate that not only revisions have been less sizeable and less volatile (as shown above) since approximately the mid-1980s, but also that government data agencies might have released data in a more efficient manner over the same period.

Note that the marked time variation of the testing results is another proof that a given model

[4]Note that most studies uses linear regression to test for the rationality of early release data. For a test of data rationality which accounts for possible nonlinear dependence, see the recent work by Corradi et al. (2009).

tested over different vintages may give rise to potentially different results. Finally, it is also worth noting that none of the empirical regularities discussed above seem to show much dependence on chain/fixed weighting of the NIPA data, or at least no more than on other benchmark revisions.

Figure 3: Testing news vs. noise. Rolling p-values – 10-year moving window



Note: the figure reports for real GDP growth (left panel) and inflation (right panel) the rolling p-values of the F-test for checking the news/noise hypotheses with Mincer-Zarnowitz (1969) type regressions using a 10-year moving window. The red line represents the 5% significance level threshold. Vertical gridlines indicate benchmark revision dates.

2.3 Implications

Three facts seem to emerge from the empirical analysis carried out in this section. First, revisions are sizeable and volatile in periods of high macroeconomic volatility. Second, government data agencies add news or reduce noise in a time-varying manner. Third, revisions may influence the stability of models or the empirical implications of economic theories tested over different vintages of the same variables.

The first two facts imply that measurement errors associated with data releases were bigger and more volatile before the disinflation and the great moderation period than afterwards. The third fact implies that data uncertainty is also responsible for a form of model instability which has to do with the inability of a given model to replicate results using data sets of different vintages. Consequently, all measurements of the same economic concept between the real-time value of the initial vintage and the revised value of the latest available vintage can be of use in the estimation of a model.

The time-varying features of the revision process make it difficult to adjust macroeconomic and forecasting models, as the direction of the revisions are not easily predicted and modelling revision characteristics might be greatly complicated by these facts. Therefore, an estimation strategy is needed which generally accounts for data revision, is robust to the choice of the most appropriate data set, and alleviates the related measurement error problem without necessarily modelling the revision process. In the next section we introduce the idea of combining estimates from different data vintages, which should help us cope with all these issues.

3 Estimation combination using vintages of a real-time data matrix

How should researchers and policymakers respond to imperfect data and which estimation strategy would make results robust to the revision process, regardless of its nature? Which data should be used to evaluate or estimate models? Should we use all the data available or just a subset of revised data in estimating models? The literature has only partially answered these questions, mainly by reporting comparisons between real-time and latest available data (Denton and Kuiper, 1965; Molodtsova et al., 2008), or by checking the robustness of model theories using only a few alternative data vintages (Boschen and Grossman, 1982; Dewald et al, 1986; Croushore and Stark, 2003).

Starting from the conclusion of the previous literature that a given theory (or an identification scheme in a structural model) may give rise to different results for different vintages, so that modellers and empirical researchers should check their results for robustness across different vintages of data, we argue that a robust set of results can be obtained by borrowing from the literature on model combination. Consequently, instead of checking results vintage by vintage, and in order to cope with the presence of measurement errors in the data revision process in a general form, we suggest a simple combination strategy which averages model estimates across all the available vintages.[5]

We answer our initial question on how to handle data uncertainty by starting from the idea that the existence of multiple vintages of data for a given variable might render incorrect or simply incomplete the use of a single vintage when evaluating or estimating a model, because the stochastic relationship between vintages is not taken into account and the estimated uncertainty is distorted. Hence, to avoid a vintage dependency of policy or forecasting models, we propose combining the

[5]It follows that in the Blanchard and Quah example discussed by Croushore and Stark (2003), for instance, the output and unemployment responses to a demand shock (CS 2003, Fig.6 p. 615) would be better represented by an average across vintages than by a single impulse response function based on one of the vintages.

information contained in all vintages.

Suppose that the problem is to explain the dynamics of a variable (or a vector of variables) y_{t+h}^{v} at a given vintage v. We start by specifying a model or a data generating process (DGP):

$$y_{t+h}^{v} = m\left[y_t^v, x_t^v; \beta_v; \varepsilon_{t+h}^v | \Omega_t(v')\right] \tag{4}$$

where a model m links the variable y_{t+h}^{v} to past observations of it and other variables, denoted by x_t^v. Note that both sets of variables (y and x) might be subject to revision. The vector β_v collects all parameters of the model to be estimated, and ε_{t+h}^v is an error term whose properties remain to be specified.

The choice of the information set Ω_t to be used in the construction or the estimation of the model is crucial to our approach. When $v' = v_t$, the set includes information up to time t relative only to the current vintage v. Alternatively, with $v' = (v_1, v_2, ..., v_t)$, the set might include information up to time t of both *current* and *past* vintages. The idea here is not so much that of modelling the revision process and using its systematic properties to improve the forecast (see e.g. Swanson and van Dijk, 2006), but rather to capture a general stochastic relationship between the vintages – without necessarily specifying their DGP – and to use all the available vintages to estimate the parameters of the model. In other words, the approach implies that if, for example, we want to forecast in real time the variable y_{T+h}^v, instead of using only the historical values $y_{t=1,..,T}^v$ of the same vintage v, we would fit a model on the whole dataset $y_{t=1,...,T}^{v=v_1,...,v_T}$ and estimate the parameters by averaging estimates over all past vintages. In the case of a general linear regression model

$$y_{t+h}^{v} = X_t^v \beta_v + \varepsilon_{t+h}^v \qquad t = 1, 2, ..., T_v \tag{5}$$

one would estimate the model for each vintage up to the current one (v), take an average of the estimates $\left(\hat{\beta}_1, ..., \hat{\beta}_v\right)$, and forecast y_{T+h}^v as

$$\tilde{y}_{t+h}^{v} = X_t^v \bar{\beta}\left(v\right) \tag{6}$$

where $\bar{\beta}\left(v\right)$ is a (possibly weighted) average of the single $\hat{\beta}_j$

$$\bar{\beta}\left(v\right) = \sum_{j=1}^{v} \omega_j \hat{\beta}_j. \tag{7}$$

This strategy has the advantage that empirical results become less dependent on the current vintage of the data when, say, evaluating a model or forecasting accuracy. In this respect, we share the view that [...] if empirical results do not hold up across alternative vintages of the data, then those results are of limited value (Croushore and Stark 2003), but only to the extent that the results obtained with different vintages are treated as if they stemmed from different models.

In fact, our approach can be rationalized in the general framework of model combination, and empirical regressions subject to data revisions can be viewed as results from different models. A strategy which accounts for all vintages and averages estimation results across them is similar to a model combination approach which averages results across different methodologies. Therefore, the idea of using only one data set (e.g. the latest available or the real-time data) to check the results of a model can be as misleading as the idea of using only one model to check a theory. Similarly, estimation combination (averaging over all available vintages) is as valid when dealing with data uncertainty as model combination (averaging over all available models) is when dealing with model uncertainty.

This explains why, as mentioned in Section 2, we do not pull out benchmark vintages from the data and focus on the time elapsed between successive benchmark revisions, and why all data revisions are essential to our approach – a potentially useful risk diversification strategy for coping with data uncertainty.

The choice of the weights in the average (7) depends on the empirical problem. In this paper we check two weighting schemes: (i) equal weights and (ii) exponential-smoothing weights. In the first case, the weighting will simply be:

$$\omega_j = 1/v$$

In the second, we assume

$$\omega_j = \lambda (1-\lambda)^j \qquad \lambda \in (0,1) \tag{8}$$

where the smoothing parameter λ balances the importance of the past vintages with that of the most recent ones. By replacing (8) in (7), iterating, and taking the limit as $j \to \infty$ we have the usual adaptive formula

$$\bar{\beta}(v) = (1-\lambda)\bar{\beta}(v-1) + \lambda\widehat{\beta}_v \tag{9}$$

Hence, when λ is close to 0 the contribution of recent vintages in the estimation (and therefore in the forecasting) is similar to that of the older observations, and the weighting scheme is closer to the equal weight approach. With values of λ close to one, instead, we attribute more weight to the most recent vintages. In fact, with $\lambda = 1$ only the latest available vintage is considered in the estimation.

It is important to stress that our approach averages across model estimates. A different approach to using current and past vintages would directly average the data over vintages in a factor-model style and then use the factors to estimate a model. However, apart from the difficulty of averaging data which have been measured in different ways in the course of the various benchmark revisions, it can be shown that the degree of model instability is high, in that the estimation of a single model over different data vintages gives rise to different parameter estimates (Altavilla and Ciccarelli, 2007). Given this heterogeneity, it is therefore better to estimate the model over each vintage and then aggregate the estimates, instead of aggregating the data and then estimating the model (Pesaran and Smith, 1995).

Finally, note that the approach is similar in spirit to that suggested by Pesaran and Timmermann (2007) of selecting estimation windows to cope with structural breaks. They extend the concepts of model uncertainty and model combination to a wider class of models where regression equations subject to breaks are viewed as different models. Instead of averaging model estimates across different samples of the same vintage (the latest available), our approach averages over the estimates obtained from different vintages, and therefore considers regression equations subject to revisions as different models. In this sense, our combination approach can be seen as a risk diversification strategy in the face of uncertainty regarding not only data measurement but also possible structural changes.

In the next sections we will apply our combination strategy to estimating and forecasting a Taylor rule in real time, and compare the results with those of a single-vintage approach, which only uses the latest available vintage.

4 Real-time monetary policy with vintage combination

4.1 Policy Analysis

There is a growing body of literature suggesting that policy analysis based on real-time data often reaches substantially different conclusions from analysis concentrating on the latest data available.[6] For instance, Orphanides (2001, 2003) finds that real-time monetary policy recommendations may differ considerably from those obtained using ex post revised data, and that the estimated policy reaction functions based on revised data yield misleading descriptions of historical policy. The solutions proposed in the literature for handling data uncertainty are either using additional variables

[6]See Croushore (2010) for a comprehensive review of existing evidence.

that are not subject to revision (Coenen et al., 2005, for example, suggest the use of money supply as indicator variable when output data is uncertain) or reducing the reaction of policymakers to real-time data by implementing a less aggressive monetary policy strategy (Aoki, 2003).

In this section we analyze whether and how the combination strategy described in Section 3 for handling data uncertainty may influence the design of optimal monetary policy. In particular, we analyze the performance of simple monetary rules that are consistent with inflation targeting. We concentrate on the stabilizing properties of the rules, on the size of the response coefficients implied by the estimated reaction functions, and on the associated losses.

4.1.1 The model

The monetary authority minimizes expected losses of a social loss function subject to the economy, and sets up a policy rule. The economy is summarized by the Rudebusch and Svensson's (1999) macroeconomic model (RS henceforth).

This model has been widely employed in empirical monetary policy analysis and it will therefore be easy to compare our results with those of previous studies. The use of a purely backward-looking model might be seen as a limitation of our analysis, but Estrella and Fuhrer (2003) test both forward- and backward-looking monetary policy models and find that the RS model does not suffer from the claimed instability of backward-looking models, whereas forward-looking models do. Moreover, they demonstrate that the relevance of the Lucas critique to the RS model is empirically limited.

The model consists of an aggregate supply equation that relates inflation (π) to the output gap ($\widetilde{y}$) and inflation lags, and an aggregate demand equation that relates the output gap to the interest rate (i):

$$\pi_{t+1} = \sum_{j=0}^{3} \alpha_j \pi_{t-j} + \alpha_4 \widetilde{y}_{t-1} + \varepsilon_t^{\pi} \tag{10}$$

$$\widetilde{y}_{t+1} = \sum_{j=0}^{1} \beta_j \widetilde{y}_{t-j} - \beta_2 \left(\bar{i}_t - \overline{\pi}_t\right) + \varepsilon_t^{\widetilde{y}} \tag{11}$$

Here $\widetilde{y}$ is the difference between actual real GDP and potential GDP in percentage points;[7] $\overline{\pi}_t$ is

[7] Potential output is estimated by applying a standard Hodrick-Prescott (HP) filter with the penalty parameter set to 1600. The inclusion of an analytical concept, such as an output gap, in the model greatly complicates the design of the monetary rule. Orphanides and van Norden (2002) show that several univariate methods for estimating current output gaps are considerably inaccurate because of the unreliability of the models in estimating end-of-sample values and to a lower extent because of data revisions. Numerous studies (e.g. Orphanides, 2001; and Rudebusch, 2001) have shown that real-time measurement problems associated with the use of an output gap in standard monetary policy rules may lead to severe policy mistakes.

four-quarter inflation in the GDP chain-weighted price index, i.e. $\overline{\pi}_t = \frac{1}{4}\sum_{j=0}^{3}\pi_{t-j}$; and $\overline{i}_t$ is the four-quarter average federal funds rate, i.e. $\overline{i}_t = \frac{1}{4}\sum_{j=0}^{3} i_{t-j}$.

The central bank minimizes an intertemporal loss function that has a positive relationship with the deviation between the goal variables and their target levels:

$$\min E_t \sum_{\tau=0}^{\infty} \delta^{\tau} L_{t+\tau} \tag{12}$$

where the period loss function is:

$$L_t = \vartheta\pi_t^2 + \phi\widetilde{y}_t^2 + \gamma\left(i_t - i_{t-1}\right)^2 \tag{13}$$

and E_t denotes expectations conditional upon the information set available at time t, $\Omega_t(v')$; δ is a discount factor, $0 < \delta < 1$; and ϑ, ϕ and γ are non-negative weights. We set $\gamma = 0.5$ and $\vartheta = 1$ and use several values of ϕ varying between 0 and 10. We consider $\phi = 1$ as a benchmark for our analysis.

As shown in Rudebush and Svensson (1999), for $\delta = 1$, the loss function can be written as the weighted sum of the unconditional variances of the target variables:

$$E_t\left[L_t\right] = \vartheta Var\left[\pi_t\right] + \phi Var\left[\widetilde{y}_t\right] + \gamma Var\left[i_t - i_{t-1}\right] \tag{14}$$

For the policy rules, we consider a linear feedback instrument rule:

$$i_{t|\Omega_t(v')} = fX_t = \sum_{j=0}^{3} f_j^{\pi}\pi_{t-j} + \sum_{j=0}^{1} f_j^{\widetilde{y}}\widetilde{y}_{t-j} + \sum_{j=0}^{2} f_j^{i} i_{t-j} \tag{15}$$

where f is a conformable row vector.[8]

The problem of minimizing the loss function in each period subject to the model presented above is standard and results in an optimal linear feedback rule which, under the limit assumption of $\delta = 1$, converges to a closed-form solution for the vector f (see RS *p.240*).

As remarked above (section 3), the choice of the information set Ω_t is crucial to our analysis. When $v' = v_t$, the set only includes the latest information available. In this case, we sequentially estimate the model and solve the optimization problem with the single vintage approach (henceforth, SV). Alternatively, with $v' = (v_1, v_2, ..., v_t)$ the set also includes information contained in past vintages. In each period, we use all the available vintages to estimate the state of the economy following the (mean-group) vintage combination strategy outlined above (henceforth, VC).

[8]McCallum and Nelson (2004) found that also in the context of forward-looking models optimal monetary rules can be well approximated by simple feedback rules based on an interest rate instrument of the type presented in the paper.

4.1.2 Results

We first check how sensitive the estimated long-run response coefficients might be to the information set used in the estimation. These coefficients are computed as:

$$\tilde{f}_{\pi} = \frac{\sum_{j=0}^{3} f_{\pi}^{j} \pi_{t-j}}{1 - \sum_{j=0}^{2} f_{i}^{j} i_{t-j}} \qquad \text{and} \qquad \tilde{f}_{\widetilde{y}} = \frac{\sum_{j=0}^{1} f_{\widetilde{y}}^{j} \widetilde{y}_{t-j}}{1 - \sum_{j=0}^{2} f_{i}^{j} i_{t-j}}.$$

A common result when analyzing the optimal policy that a central bank should follow in response to developments in the economy is a considerable difference between the reaction coefficients implied by the optimal policy rules and those implied by the historical evidence. In particular, the historical behaviour of central banks is usually less aggressive than that implied by optimal rules. Several authors, such as Rudebusch (2001) and Tetlow and von zur Muehlen (2001), relate this less aggressive policy to the uncertainty that policymakers face when setting interest rates.

In order to assess whether vintage combination might overcome this empirical evidence, table 1 presents an overview of the results of previous studies using the same RS model compared with our benchmark estimates. The first column indicates the authors and the sample period over which they obtained their results. The second, third and fourth columns report the weights attached to the different targets in the loss function. The next two columns show the long-term response coefficients for each study. Finally, the last four columns report our results obtained by using the single vintage and the vintage combination (last two columns) methods on the same sample analyzed by each of the previous studies.

It clearly emerges that during the last decades the FED has become more responsive to changes in the target variables. As expected, results for the single vintage method are very close to those obtained by previous studies. The small difference depends on the exact sample period used in the various analyses. More interestingly, table 1 clearly indicates that, although also increasing over time, the size of the response coefficients obtained with our VC method is significantly smaller than those obtained in previous studies based on single vintages.

A more comprehensive representation of the results is reported in Figure 4, where we compare the long-run response coefficients retrieved from the optimal feedback rule obtained from both approaches (SV and VC) over the whole sample period and for several values of the weights.

In particular, for each quarter the chart shows the value of the response coefficients across the weights the monetary authorities attach to inflation and the output gap. While the weights on inflation and interest rate stabilization are set at 1 and 0.5 respectively, the weights on output ϕ

Table 1: Evidence from other studies

Authors	Weights			Long-term Response Coefficients					
				Literature		This paper			
						Single Vintage		Vintage Combination	
	ϕ	$1-\phi$	γ	$\tilde{f}_\pi$	$\tilde{f}_y$	$\tilde{f}_\pi$	$\tilde{f}_y$	$\tilde{f}_\pi$	$\tilde{f}_y$
Rudebusch-Svensson ('99) Sample: 1961:1-1996:2	1	1	0.5	2.72	1.56	2.5	1.56	1.695	0.89
Rudebusch-Svensson ('02) Sample: 1961:1-1996:4	0.4	0.4	0.2	2.81	1.68	2.6	1.58	1.701	0.904
Dennis ('06) Sample: 1982:1–2000:2	1	1	0.25	2.63*	1.75	2.83	2.01	1.87	1.1
Brock et al. ('07) Sample: 1970:2 - 2002:4	1	1	0.1	3.2	2.1	3.32	2.49	2.08	1.32
Cateau ('07) Sample: 1970:1–2003:2	1	1	0.5	2.88	1.57	2.7	1.8	1.78	1

Note: the table compares the results of previous studies using the same RS model with our benchmark estimates. The first column indicates the authors and the sample period over which they have obtained their results. The second, third and fourth columns report the weights attached to the different targets in the loss function. The next two columns show for each study the long-term response coefficients. Finally, the last four columns report our results obtained by using the single vintage and the vintage combination (last two columns) methods over the same sample analyzed by each of the previous studies. * indicates that the monetary rule reacts to future inflation.

vary between 0 and 10 in each time period, in increments of 0.1. We therefore, have 100 possible combinations of weights for each of the 141 vintages from 1975Q4 to 2010Q4.

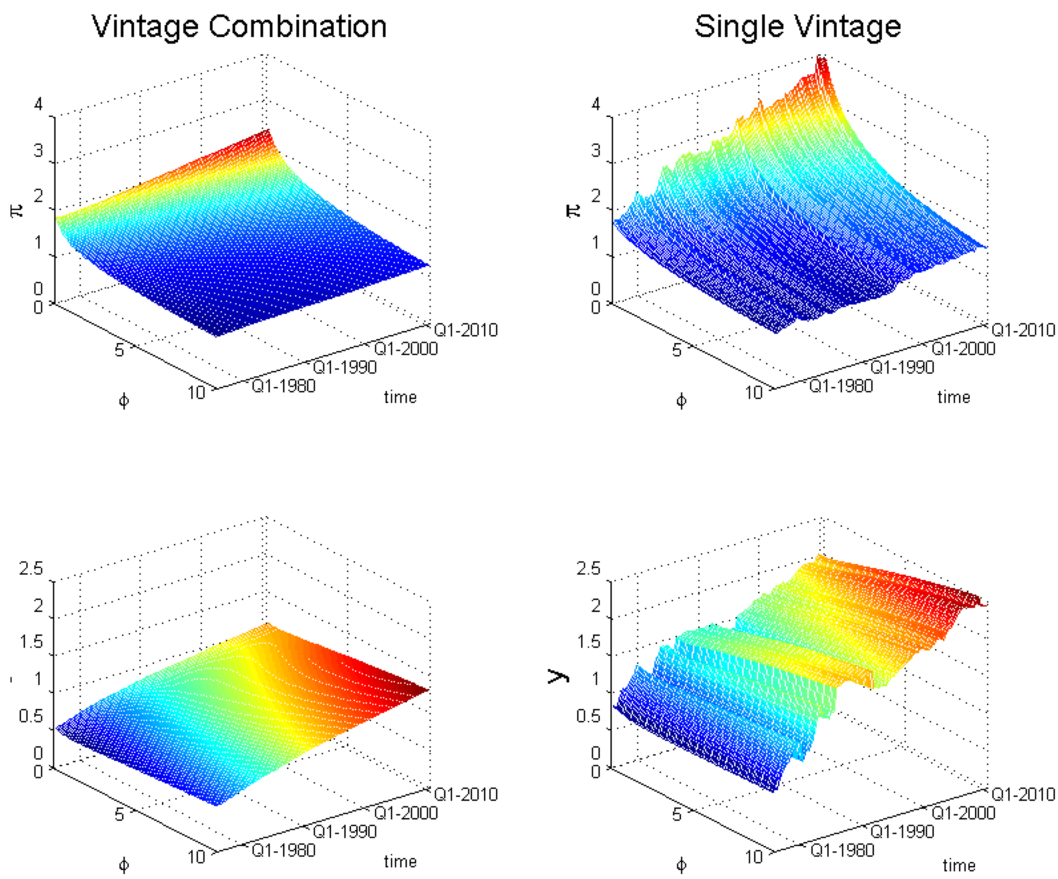
Note that for both the SV and the VC approaches the estimated response coefficients for inflation always satisfy the so-called Taylor principle and, in line with the literature on monetary policy rules, they are larger than unity, thus exerting a stabilizing effect on inflation. Moreover, for both approaches the response coefficients $\tilde{f}_\pi$ and $\tilde{f}_{\tilde{y}}$ show a positive trend over time, meaning that the FED has become more responsive to changes in inflation and to output gap.

This increasing trend in output gap coefficients has also been documented by Tetlow and Ironside (2007). Estimating the FRB/US macroeconomic model over 30 different vintages, they observe a steady increase in the reaction coefficients during the period 1996-2003. More recently, Taylor and Williams (2010) also observe that since the onset of the great moderation the response coefficients of output gap and inflation changes have increased significantly.

Although the VC method also presents the same increasing pattern in both coefficients, two main differences are worth emphasizing: 1) the size of the coefficients is smaller; and 2) the responses are less volatile over time. Both features are desirable for a monetary policy implementation strategy.

The first is in line with the literature finding that when considering data uncertainty reaction coefficients are lower. Rudebusch (2001), for example, shows that taking into account the real-time data uncertainty in the RS model leads to an attenuation of the optimal policy rule.

Figure 4: Response Coefficient Dispersion



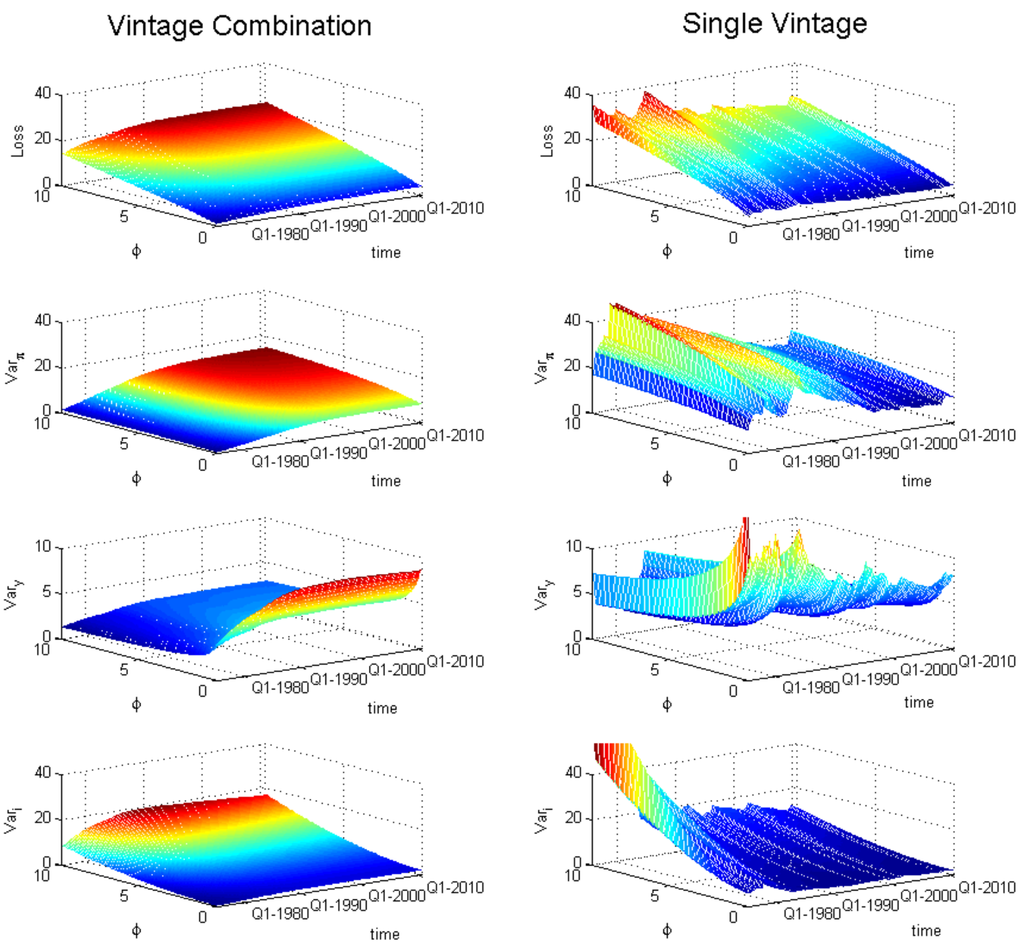
Note: for each quarter the chart shows the value of the inflation (first row) and output gap (second row) long-term response coefficients estimated using the vintage combination method (first column) and the single vintage method (second column). While the weights on inflation and interest rate stabilization are set at 1 and 0.5 respectively, the weights on output gap stabilization ϕ vary between 0 and 10, in increments of 0.1.

The second difference implies conservative behaviour by the central bank, which does not immediately react to out-of-target developments in inflation and output. In line with a medium-term orientation of policy strategy, the reaction of the central bank is much more stable over time. This stability is a consequence of the estimation method, as VC reaction coefficients are based on estimated coefficients which have been averaged with the previous estimates. In turn, this implies that when we minimize the central bank losses with our combination method only persistent changes in the estimates will gradually and smoothly appear, whereas a vintage-by-vintage approach will typically imply more volatile and erratic behaviour.

One may, therefore, wonder whether the VC approach is also optimal from the central bank perspective compared to a SV one. Figure 5 reports the expected losses and the variance of the three target variables, and shows how sensitive the results are to alternative weights the central bank puts on inflation and output stabilization. Again, for each time period $\gamma = 0.5$, $\vartheta = 1$, and ϕ varies from 0 to 10 in increments of 0.1.

The charts highlight the trade-off between inflation and output stabilization. In general, the inflation variance is more sensitive to the weights given in the loss function, compared to the variance of the output gap. For most of the sample, when estimating the model with our combination strategy the variances of the goal variables are lower than those obtained when using the latest available data. In other words, the stabilization of inflation and output that would have been obtained had the authorities relied on a vintage combination strategy is higher than that implied by the single vintage strategy. Moreover, the losses associated with the SV approach are higher than those estimated with the VC approach. These findings suggest that taking into account the revision history by means of simple averages of the estimates not only produces lower and more stable central bank reactions but also increases the stabilizing properties of the rule and is associated with lower losses.

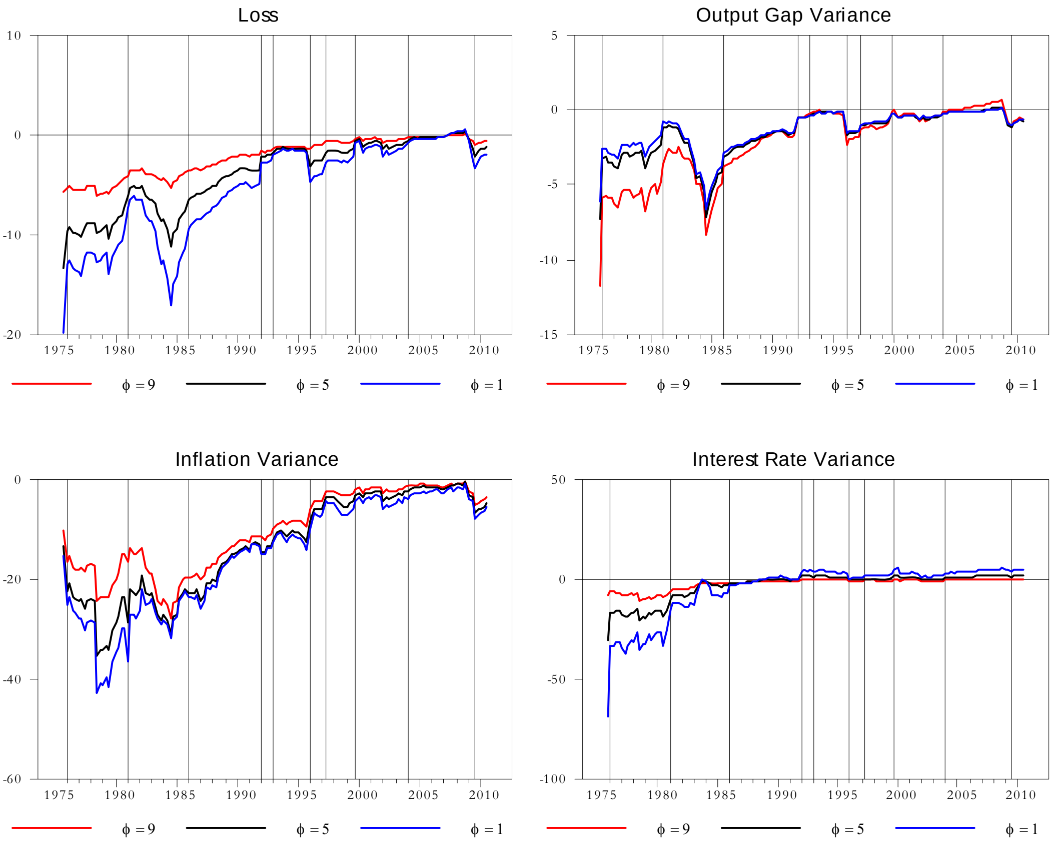
Figure 5: Loss and Variance Surface



Note: for each quarter the chart shows the value of the central bank loss (first row), inflation variance (second row), output gap variance (third row), and interest rate variance (fourth row) obtained using the vintage combination method (first column) and the single vintage method (second column). While the weights on inflation and interest rate stabilization are set at 1 and 0.5 respectively, the weights on output gap stabilization ϕ vary between 0 and 10, in increments of 0.1.

These results are not homogeneously valid over the whole sample and there are periods where simply using the latest available data gives rise to results similar to those based on the VC strategy. Figure 6 clearly identifies these periods. For three different values of ϕ (1.0, 5.0 and 9.0), the chart plots the differences between the losses and variances of the target variables estimated with the SV and VC approaches. Vertical gridlines represent benchmark vintage dates. Negative values indicate a lower loss and a lower variance of the monetary rule estimated with our approach.

Figure 6: Differences between Vintage Combination and Single Vintage



Note: Vertical gridlines indicate benchmark revision dates. Each line in the figure expresses, for three different values of output gap stabilization weight (i.e. 1.0, 5.0 and 9.0), the differences between the losses and variances of the target variables estimated with the SV and VC approaches. Therefore, negative values indicate a lower loss and a lower variance of the monetary rule estimated with the VC approach.

The figure shows that a general estimation approach which averages estimates over all current and past data vintages outperforms the single-vintage strategy in terms of stabilizing properties of the estimated monetary policy rules for most of the sample under analysis. Interestingly, the advantage of the combination strategy substantially decreases from the beginning of the great moderation period, disappears over the more stable period until 2007 and becomes somewhat significant again in recent years, in particular during the crisis and with the end of the great moderation. In other words, consistently with what we anticipated in the descriptive analysis of the revisions (section 2), the findings are particularly valid over the first and the last part of the sample – two periods characterized by relatively high volatility.

4.1.3 How much information in data combination?

One may wonder why at a given point in time we should estimate our model using all previous history (as summarized by all data vintages available) instead of just taking, say, the last five years of vintages. Or, in other words, why should we weigh all vintages equally instead of weighing the most recent ones more? Here we check the issue by modifying the VC approach (equal weight) with the exponential smoothing strategy described in section which averages over past vintages with a system of decreasing weights.

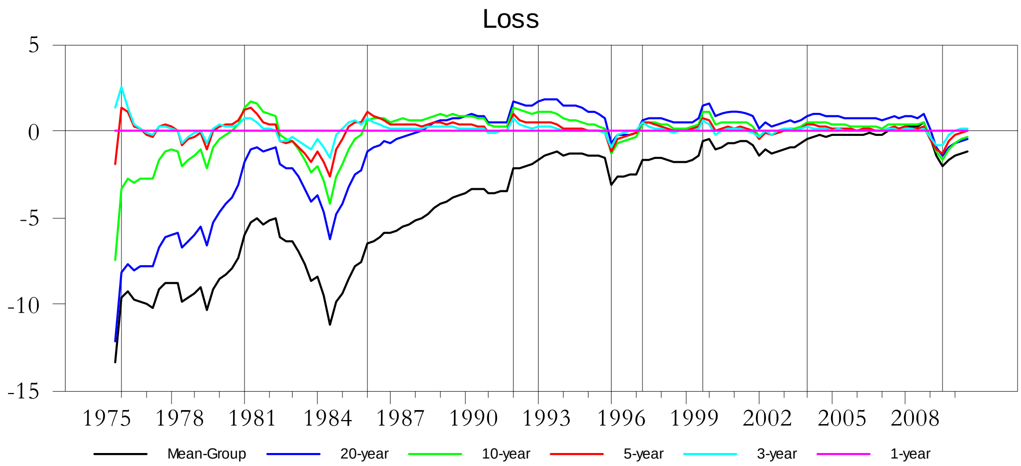
We present results for five representative values of the smoothing parameter λ. Given the quarterly frequency, we choose the following values of λ: 0.057, 0.13, 0.28, 0.45, and 0.9 which correspond to averages over the last twenty, ten, five, three and one year of data, respectively. Results (Figure 7) are reported only for our benchmark parameterization, where $\phi = \vartheta = 1$ and $\gamma = 0.5$. Colours are explained in the chart legend. Note that the black line in figure 7 is the same as the black line in figure 6.

The chart shows that the advantage in terms of expected loss increases almost monotonically with the number of vintages used in the average, and that using all vintages is always the best strategy.

Note that the pattern of the differential losses is similar across values of λ and that the use of additional information from previous vintages is particularly helpful during periods of high macroeconomic volatility. In fact, all lines approach zero during the mid 1980s, meaning that, from the start of the great moderation, the advantage of using more years of vintages instead of estimating the model with a single vintage significantly decreased. Moreover, while it is almost always better to average across all vintages, it would not have been optimal to average using only a few years of data over the period 1985-2007.

The end of the great moderation and the possible return of high and volatile revisions implied by the recent crisis have pushed all lines back into negative territory as of 2008Q4. We take the latter as clear evidence that in periods of high uncertainty the use of past vintages in the estimation process is always optimal. Note, however, that the last benchmark revision in 2009Q3 implied a new change in the single-vintage estimation as shown by the fact that the differential losses seem to go back to zero after that date. If this effect is maintained by the next revisions or reverted by a structural change due to the crisis remains to be seen. In the face of this uncertainty, the results seem to recommend our combination strategy in any case.

Figure 7: Difference between Single Vintage and Vintage Combination Models



Note: Vertical gridlines indicate benchmark revision dates. Each line in the figure is expressed in deviation from the SV strategy using only the latest available quarter (λ=1). So, for example, the blue line represents the difference between the vintage combination model, which decreasingly weights the last 20 years of vintages, and the single vintage model. All models are estimated using $\phi = \vartheta = 1$ and $\gamma = 0.5$

4.1.4 Discussion

An interpretation of the previous results can be proposed in the context of the discussion on rule-based monetary policy in practice. Although no central bank will be bound to the prescription of any simple rule (or any optimal control algorithm) our estimated rule should be interpreted as a normative guide for monetary policy decisions.

Simple instrument rules, such as the Taylor-rule, have been frequently criticized for being based on information sets which are more limited than the ones effectively accounted for by, monetary authorities. In this sense, the vintage-after-vintage re-optimization of our VC approach exploits one important dimension for broadening the information set of the rule. In fact, although it does not prescribe to the monetary authority the selection of instruments according to a more comprehensive setting, it includes in the analysis past economic developments by means of an efficient use of all data vintages while accounting for data uncertainty without modelling the revision process.

Moreover, the observed stability (associated with relatively low losses) of the policy rule response coefficients given by our VC strategy seems to be supportive of a continuation-based policy making of the kind recently proposed by Jensen and McCallum (2010). On the contrary, an approach based on a single vintage estimation is consistent with an optimal vintage-dependent rule, where the data revision process is not accounted for when planning the optimal monetary strategy.[9]

Our VC approach also suggests a cautious monetary policy strategy. This evidence is in line with a monetary strategy that does not react strongly to out-of-target developments in inflation and output attenuating the response coefficients in an optimal policy rule. The result is a consequence of the fact that the measurement error in early vintages of data significantly influences the results obtained with real-time data when using the SV approach and less so when using a VC strategy.

Finally, our results indicate that from a policy perspective it is always recommendable to use a combination approach in periods of high macroeconomic volatility. During such periods, it is likely that data revisions become more sizeable and more volatile than in normal periods and the associated measurement error affects model estimation and evaluation more heavily. Our combination approach, by assigning only a small weight to the new information, works as a useful risk diversification strategy, as it minimizes the risk of taking policy decisions based on data subject to potentially large revisions. Moreover, by equally weighing all vintages, the approach assigns the same probability to the estimates coming from data measured in different ways and, at the same time, it is more cautious in the face of uncertainty stemming from structural breaks. The cost is that possible structural changes will be recognized with some delay – when it becomes clear that the changes come from the economy and not from the revision process. However, the imperfect knowledge of the nature of the structural changes in the economy might render difficult to benefit from incorporating these features in the model.[10]

[9] In a forward-looking environment, Jensen and McCallum (2010) have shown that an optimal unconditional continuation policy (where policymakers commit to implementing the time-invariant policy that minimizes the unconditional expected value of their targets) is preferable to a timeless-perspective policy. As Woodford (2003) also stresses, a lack of continuity (i.e. that a policy continues the optimal plan from an earlier period) also substantially reduces the credibility of a policy rule.

[10] In a forecasting experiment, for example, Clark and McCracken (2008) have shown that when the date of the break is unknown, ignoring evidence of structural change and depending on its type and magnitude, can even lead to more accurate forecasts.

4.2 Forecasting

Data revision also affects the forecasting performance of models (e.g. Croushore, 2010). The availability of real-time data sets has allowed the examination of several related forecasting issues, including the relative forecasting performance of latest available vs. real-time data, and the amount of information required to improve the forecasting accuracy of a given model. As discussed in section 3, our approach is potentially helpful in dealing with these issues. In particular, the use of our combination strategy accounts for data revision without necessarily modelling an assumed pattern of the revisions themselves, and avoids the choice between real-time (the diagonal of the real-time data set) and latest-available (the last column) data.

4.2.1 Model and experiment

To illustrate the potential advantages of the approach in a forecasting exercise, we use it to forecast the interest rate in real time with a standard VAR model that contains a Taylor-type interest rate rule similar to the one employed in the previous sections. In particular, we now generalize the two-equation model of the previous subsection given by (10) and (11) with a trivariate VAR where the interest rate is the third endogenous variable.

The model is represented by the following general specification:

$$z_t^v = C^v + A^v (L) z_{t-1}^v + \varepsilon_t^v \tag{16}$$

where $A^v (L)$ is a lag polynomial, and the parameters potentially vary with the vintage v. The vector z_t^v contains inflation, output growth, and the nominal interest rate. The inflation rate and the output growth are the s-quarter growth rates of the GDP deflator and real GDP respectively in percentage points in annual rates, i.e. $\pi_{t,s} = \frac{400}{s} (\ln P_t - \ln P_{t-s})$ and $q_{t,s} = \frac{400}{s} (\ln GDP_t - \ln GDP_{t-s})$. The nominal interest rate is the federal funds rate. We report results for $s = 1$ and $s = 4$, i.e. for quarter-on-quarter (q-o-q) and year-on-year (y-o-y) transformations of prices and GDP. No assumptions are required about the structure of the variance-covariance matrix, and the VAR is estimated in standard Bayesian fashion *á la* Litterman with a symmetric tightness function, a general tightness of 0.2, a lag decay equal to 1.5, and with mean prior coefficients of its own lag equal to 0 for q, and to 1 for π and i.

As in the previous subsection, the equation for the fed funds in the BVAR represents a backward-looking Taylor rule. Although our simple rule does not contain forward-looking variables, recent papers (e.g., Molodtsova et al., 2008; Orphanides, 2003) find that Taylor rules for the U.S. using inflation forecasts are nearly identical to those using lagged inflation rates. In addition, as non-linear models do not seem to improve forecast accuracy over our sample (see Altavilla and Ciccarelli, 2010) we use a linear specification of the VAR, which is also potentially usable under different prior assumptions (see Altavilla and Ciccarelli, 2009). The chosen model specification is among the top performing models used by Clark and McCracken (2010) in a forecasting competition under possible structural breaks.[11]

Finally, note that the choice of the forecasting object depends on the fact that the interest rate is not itself subject to data revisions and therefore the analysis is not complicated by the choice of the actual values for the computation of the loss functions. This choice is consistent with our general idea that a final revision or an actual value are meaningless concepts.

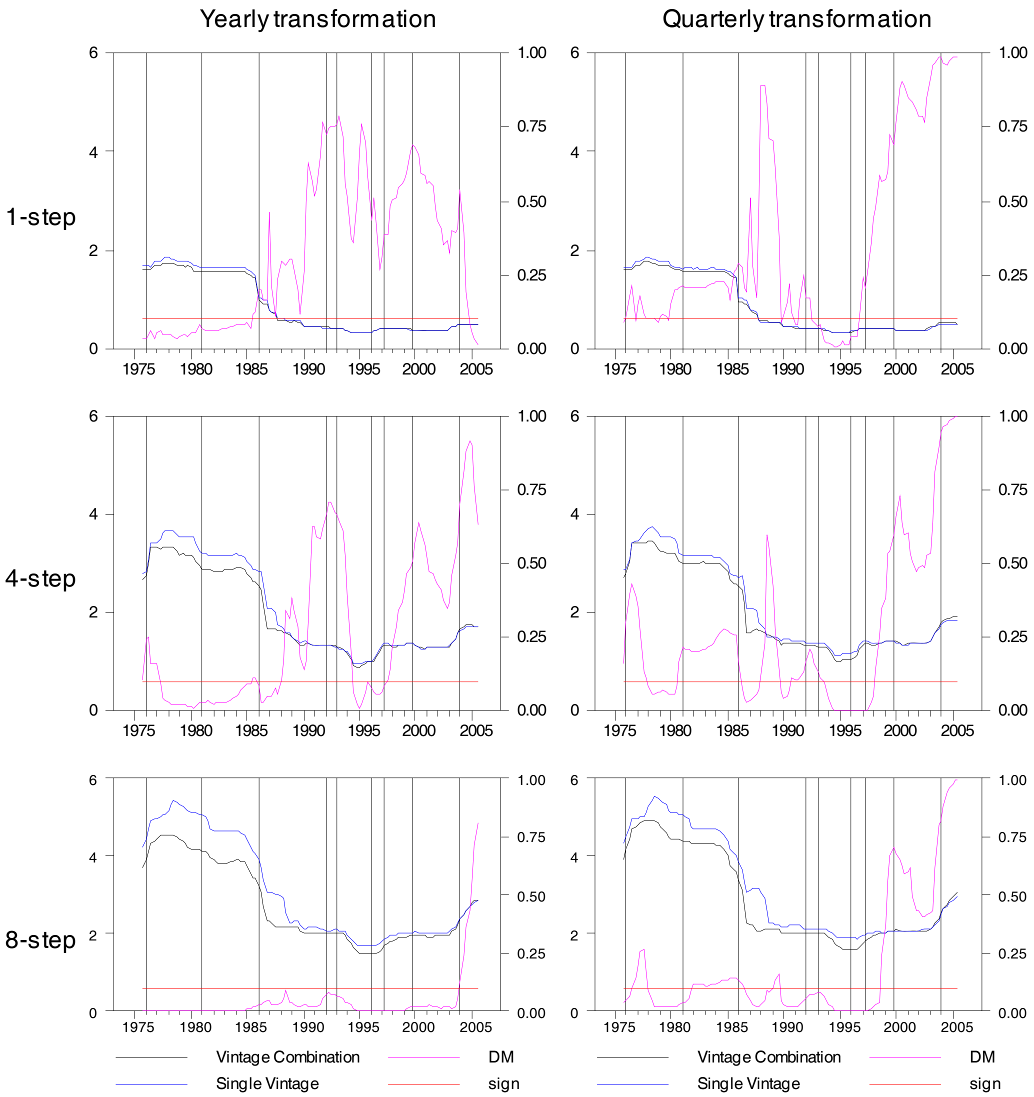
In the benchmark experiment we use (16) to forecast the interest rate i in real time at $h = 1, 4, 8$ quarters ahead, and compare the VC against the SV strategy. A comparison with a "real-time" strategy – which considers only the data on the diagonal of the real-time data matrix – was performed for completeness but is not reported for two main reasons: first, because the results are never better than those of the other two strategies; and second, we believe that policymakers and forecasters when performing a real-time exercise not only do not ignore the time series information contained in the latest available vintage, but might even be willing to consider data revisions taking place from one vintage to another. Both elements are clearly missing if we perform the exercise using only the diagonal of a real-time data matrix.

4.2.2 Results

We run the experiment over the period 1959:1 through 2010:4, and estimate a VAR(4) for both the quarterly and the yearly transformation using the training sample 1959:1 to 1970:4 for the first mean-group (VC) estimation. The accuracy is measured by a rolling Root Mean Squared Error centred on the period 1975:1 to 2005:4, with a ten-year rolling window.

[11]In that paper, the authors also show that model combination techniques are valid tools for coping with structural instabilities whose exact forms are difficult to identify. The same idea is also in Pesaran and Timmermann (2007), who consider estimation based on different windows of the same sample as different models and show that equal-weight averages have a good forecasting performance in the presence of structural changes. In some sense, our modelling strategy combines these two views without explicitly modelling the revision process or the possible structural instability.

Figure 8: Rolling RMSE and DM test - 10-year moving window



Note: the figure reports for each forecasting horizon (1-, 4-, and 8-steps ahead), and for each data transformation (the left panel refers to yearly transformation while the right panel to quarterly reformation) the rolling RMSE (using a ten-year rolling window) for the VC and SV methods (left scale in each graph) together with the p-value of a rolling Diebold-Mariano test (right scale in each graph). The red line represents the 10% significance level threshold. Vertical gridlines indicate benchmark revision dates.

Figure 8 reports a rolling RMSE for the two approaches (left scale) together with the p-value of a rolling Diebold-Mariano (DM) test (right scale). As before, vertical gridlines represent benchmark vintage dates. Our combination strategy almost always provides the lowest RMSE at all forecast-horizons and across different data transformations, the difference from the simpler strategy being increasingly higher and more significant with the forecast horizon. The difference between the two RMSEs is also more significant on average over the pre-Great moderation period (1975-1985) than afterwards, as shown in Table 2, where the average p-values of the same test are reported over the full sample and over two sub-samples for the two data transformations.

Table 2: Diebold-Mariano forecast comparison test. P-values

	Year-on-Year			Quarter-on-Quarter		
Horizon	1970-2010	1970-1984	1985-2010	1970-2010	1970-1984	1984-2010
1-step	0.034	0.036	0.16	0.12	0.12	0.68
4-step	0.029	0.030	0.34	0.090	0.083	0.59
8-step	0.001	0.000	0.15	0.022	0.024	0.25

Note: the table reports the average p-values of the DM test over different sub-samples for the two data transformations. The null hypothesis of equal accuracy (equal RMSE) is tested against the alternative that the combination approach has a smaller RMSE than the SV one. Newey and West autocorrelation consistent standard errors have been used.

In the DM test, we check the null hypothesis of equal accuracy (equal RMSE) against the alternative that the combination approach has a smaller RMSE than the SV one. Newey and West autocorrelation-consistent standard errors have been used. Consistently with the results obtained with the policy analysis reported above, while in the first part of the sample the RMSEs are different at between 3 and 10 percent significance level, in the so-called Great moderation period the forecast accuracy of the vintage combination method is not statistically different from that of the single vintage strategy. The result is somewhat consistent with the expected idea that a combination method could perform better when the measurement errors of the revised variables are greater. In particular, this also seems consistent with what was seen in Section 2, where we showed that the revisions of GDP and inflation are more news than noise during the great moderation and the disinflation periods.

The results are qualitatively unchanged when various robustness checks are performed, including, e.g., other prior assumptions and different numbers of lags in the VAR.[12] We have also examined the issue of how much information would be optimal to include in the estimation average by means

[12] All additional results are available upon request.

of the same exponential weighting scheme used in the policy subsection:

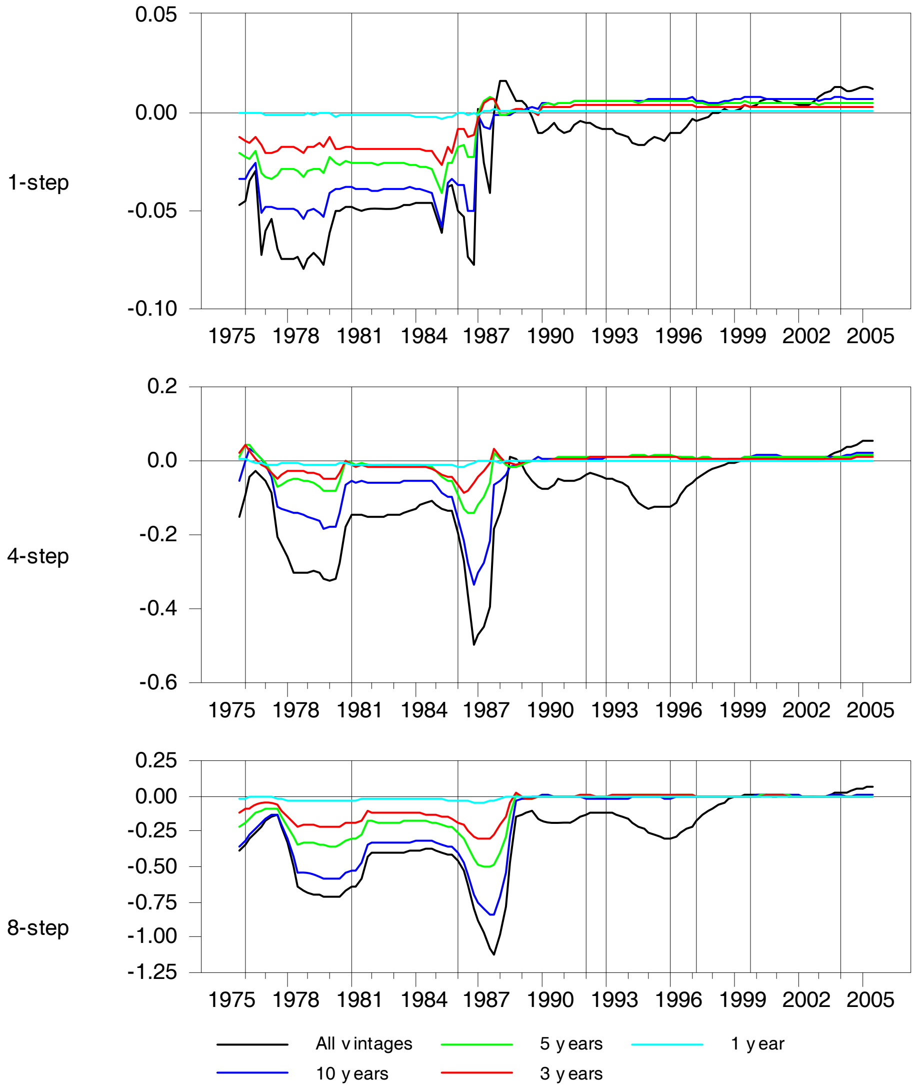
$$\bar{\beta}_v = (1-\lambda)\,\bar{\beta}_{v-1} + \lambda\widehat{\beta}_v \tag{17}$$

where the smoothing parameter λ balances the importance of the past vintages with that of the most recent ones. Ideally one could choose the smoothing parameter λ (and therefore the number of vintages) which minimizes the RMSE. Therefore, we compute a rolling RMSE for different values of λ. Figure 9 reports the results for the quarterly transformation of GDP and inflation and for $\lambda = 0.13$, 0.28, 0.45, and 0.9 – which correspond to averages with decreasing weights over the last ten, five, three and one year of data respectively. For the sake of comparison, we also report results for the average over all vintages ($\lambda \rightarrow 0$). Each line in the figure is expressed in deviation from the SV strategy using only the latest available quarter ($\lambda = 1$).

The results are very much in line with those discussed in the policy analysis of the previous subsection. Note that, as seen above, it is still advantageous in terms of accuracy in accounting for past vintages from the late 70s to the mid 80s, whereas this advantage shrinks when comparing the performance of the alternative models during the great moderation period.

More interestingly, our preferred model (the solid black line in figure 9), where all vintages are equally weighted, is the top performing model at all horizons and in particular during the pre-Great Moderation period. On the other hand, Figure 9 also highlights the fact that the forecasting performance monotonically decreases as we assign less and less weight to past vintages, with the SV model (where only the most recent data are considered in the estimation, i.e. $\lambda = 1$) always being the worst performer.

Figure 9: Rolling RMSE - Difference between Single Vintage and Vintage Combination Models



Note: Vertical gridlines indicate benchmark revision dates. Each line in the figure is expressed in deviation from the SV strategy using only the latest available quarter (λ=1). So, for example, the blue line represents for each forecasting horizon (i.e. 1, 4 and 8) the difference between the vintage combination model, which decreasingly weights the last 10 years of vintages, and the single vintage model.

4.2.3 Discussion

From this limited analysis it emerges that our combination approach, which estimates a forecasting model using all current and past vintages, seems to be a promising avenue in what Croushore (2010) calls prescriptive forecast analysis, i.e. the analysis of how forecasts should be made when we know that the data will be revised. Our findings suggest that a forecasting method which uses the entire real-time data matrix can perform better than a method that uses just one column or the diagonal. Consequently, when forecasting in real-time or comparing the performance of alternative models we should estimate our models on all the available data instead of simply basing the analysis on a single vintage, be it the latest available or the real-time diagonal.

One of the main reasons for a forecast being affected by data revisions is the fact that revisions change the data used to estimate the model and this change affects the coefficients estimated. Our procedure is more stable to such a change for it averages the estimates over all vintages, with the latest vintage receiving the same weight as any other vintage in the data matrix. This, in turn, implies that newly-released data of the latest available vintage – which is the last available observation on the real-time diagonal and, as such, is subject to a possibly sizeable revision – receives only a tiny weight in the total average. At the same time, all revisions of the historical data are treated as equally important for the estimation of the model.

Finally, our results can be rationalized in the general framework of forecast or model combination, where a model should be understood in a general sense as said at the end of section 3 (Pesaran and Timmermann, 2007), and where estimates obtained by the same model are pooled across different data vintages. In this respect, the findings are also consistent with the existing evidence that combining forecasts with a naïve weighting scheme that assigns an equal weight to each model is very often the best strategy in terms of RMSE (see e.g. Stock and Watson 2004; Timmermann 2006; and Clark and McCracken 2010).

5 Conclusions

In this paper we have proposed and discussed a general strategy for conducting monetary policy which accounts for data uncertainty without explicitly modelling the revision process. The method considers the available data vintages of a real-time data matrix as different units of a kind of panel dataset and employs simple combination techniques to average estimates across all data vintages. This is a novel approach with important implications for real-time analysis, which is typically based on methods which sequentially estimate a forecasting or policy model using only the latest information available – the last vintage of data – or the real-time data – the diagonal of the real-time data matrix. The approach has been tested with two empirical exercises based on a standard Taylor rule.

First, we have analyzed whether combining the information coming from the entire revision history of the selected variables may influence the design of monetary policy in a standard macroeconomic model. The results indicate that the optimal monetary rule obtained by estimating the model with the latest data available is more aggressive than that obtained when estimating the model with the combination approach, which instead uses all available data vintages of the real-time data matrix. Lower values of long-term response coefficients were consistently found compared to previous studies. Moreover, the vintage combination approach outperforms the single vintage strategy in terms of the stabilizing properties of the estimated monetary policy rules, especially during periods characterized by high volatility.

Second, we have investigated the forecasting performance of Taylor-type monetary rules. The results suggest that the combination method outperforms the simpler approach based on one data vintage for most of the sample period, with a statistically significant difference over the pre-Great moderation period.

In both types of exercise, we have also analyzed whether at a given point in time it is optimal to estimate the model using all previous history (as summarized by all the available data vintages) instead of taking only a few years of vintages (for instance between two benchmark revisions). We have addressed this issue by modifying the equal-weight approach with an exponential smoothing strategy which averages over past vintages with a system of decaying weights. The results confirm that our combination strategy, which assigns an equal weight to all vintages, is always optimal in the respective metric (loss function in the policy analysis and RMSE in the forecasting exercise).

Finally, we have shown that our combination approach might have a comparative advantage in particular over periods characterized by high volatility. These results, therefore, would suggest the proposed approach as a general strategy for coping with the higher volatility of the economic variables (and, presumably, of their revisions) after the recent financial crisis and with the end of the great moderation.

Our approach can be rationalized in the framework of model combination as a valid risk diversification strategy for coping with data uncertainty. An equal-weight strategy which uses all data vintages, by giving relatively more weight to the past than to the current vintage is more conservative and less subject to vintage dependency than the usual approaches based on the comparison of single vintages. The latter strategy might render a monetary policy decision awkwardly unstable or introduce a time-varying inconsistency into a given economic model. Our findings imply that the sensitivity of a model to variations in the data or the consistency of a theory should be checked on an average of the whole data history rather than on single vintages haphazardly chosen, or simply on the latest vintage available. The latter, in particular, is still subject to future (e.g. benchmark) revisions which could determine further and controversial changes in the interpretation of model results or in the action chosen by a policymaker.

References

[1] Altavilla, C. and M. Ciccarelli (2007), Information combination and forecast (st)ability: Evidence from vintages of time-series data, ECB WP no. 1580.

[2] Altavilla, C. and M. Ciccarelli (2009), The Effects of Monetary Policy on Unemployment Dynamics under Model Uncertainty: Evidence from the US and the Euro Area, *Journal of Money, Credit and Banking*, 41:1265-1300.

[3] Altavilla C. and M. Ciccarelli (2010), Evaluating the Effect of Monetary Policy on Unemployment with Alternative Inflation Forecasts, *Economic Modelling*, 27:237-253.

[4] Aoki, K. (2003), On the optimal monetary policy response to noisy indicators, *Journal of Monetary Economics*, 50:501-523.

[5] Aruoba, S.B. (2008), Data Revisions Are Not Well Behaved, *Journal of Money, Credit and Banking*, 40:319-340.

[6] Boschen, J. F. and H.I. Grossman (1982), Tests of Equilibrium Macroeconomics Using Contemporaneous Monetary Data. *Journal of Monetary Economics,* 10:309–333.

[7] Brock, W.A., S. N. Durlauf, and K.D. West (2007), Model uncertainty and policy evaluation: Some theory and empirics, *Journal of Econometrics*, 136:629-664.

[8] Cateau, G. (2007), Monetary policy under model and data-parameter uncertainty, *Journal of Monetary Economics*, 54:2083-2101.

[9] Clark T.E. and M.W. McCracken (2008), Forecasting with small macroeconomic VARs in the presence of instability. In *Forecasting in the Presence of Structural Breaks and Model Uncertainty*, Rapach DE, Wohar ME (eds). Emerald Group Publishing: Bingley, UK, pp. 93–147.

[10] Clark T.E. and M.W. McCracken (2010) Averaging Forecasts from VARs with Uncertain Instabilities, *Journal of Applied Econometrics*, 25: 5–29.

[11] Coenen, G., A. Levin and V. Wieland (2005), Data uncertainty and the role of money as an information variable for monetary policy, *European Economic Review*, 49(4): 975-1006.

[12] Corradi, V., A. Fernandez, and N.R. Swanson (2009), Information in the Revision Process of Real-Time Datasets, *Journal of Business & Economic Statistics*, 27:455-467.

[13] Croushore, D. (2010), Frontiers of Real-Time Data Analysis, *Journal of Economic Literature,* forthcoming.

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

[15] Croushore, D., and T. Stark (2003), A real-time data set for macroeconomists: Does the data vintage matter? *The Review of Economics and Statistics*, 85:605–617.

[16] Dennis R. (2006), The policy preferences of the US Federal Reserve, *Journal of Applied Econometrics,* 21:55-77

[17] Denton, F.T., and J. Kuiper (1965), The Effect of Measurement Errors on Parameter Estimates and Forecasts: A Case Study Based on the Canadian Preliminary National Accounts, *Review of Economics and Statistics,* 47:198-206.

[18] Dewald, W.G., J.G. Thursby, and R.G. Anderson, Replication in Empirical Economics: The Journal of Money, Credit, and Banking Project, *American Economic Review* 76:587–603.

[19] Estrella, A. and J.C. Fuhrer (2003), Monetary Policy Shifts and the Stability of Monetary Policy Models, *The Review of Economics and Statistics*, 85:94-104.

[20] Guerrero, V.M. (1993), Combining historical and preliminary information to obtain timely time series data, *International Journal of Forecasting*, 9:477-485.

[21] Jacobs J.P.A.M. and S. van Norden (2010), Modeling Data Revisions: Measurement Error and Dynamics of "True" Values, *Journal of Econometrics,* forthcoming.

[22] Jensen, C., and B.T. McCallum (2010), Optimal Continuation versus the Timeless Perspective in Monetary Policy *Journal of Money, Credit and Banking*, 42:1093–1107.

[23] Mankiw, N. G., and M.D. Shapiro (1986), News or Noise: An Analysis of GNP Revisions, *Survey of Current Business*, pp. 20-25.

[24] Mankiw, N. G., D.E. Runkle, and M.D. Shapiro (1984), Are Preliminary Announcements of the Money Stock Rational Forecasts? *Journal of Monetary Economics,* 14:15-27.

[25] Mincer, J., and V. Zarnowitz (1969), "The evaluation of economic forecasts", in J. Mincer, editor, *Economic Forecasts and Expectations: Analyses of Forecasting Behavior and Performance*, National Bureau of Economic Research, New York, chapter 1, pp.3–46.

[26] Molodtsova T., A. Nikolsko-Rzhevskyy, and D.H. Papell (2008), Taylor rules with real-time data: A tale of two countries and one exchange rate, *Journal of Monetary Economics,* 55:S63–S79

[27] Orphanides, A. (2001), Monetary policy rules based on real-time data. *The American Economic Review*, 91:964-985.

[28] Orphanides, A. (2003), Historical monetary policy analysis and the Taylor rule. *Journal of Monetary Economics*, 50:983-1022.

[29] Orphanides, A. and S. van Norden (2002), The Unreliability of Output-Gap Estimates in Real Time, *The Review of Economics and Statistics*, 84:569-583.

[30] Patterson K.D. (2003), Exploiting information in vintages of time-series data, *International Journal of Forecasting*, 19:177–197.

[31] Pesaran, M. H. and R. Smith (1995), Estimating long-run relationships from dynamic heterogeneous panels, *Journal of Econometrics*, 68:79-113.

[32] Pesaran, M. H., and A. Timmermann (2007), Selection of estimation window in the presence of breaks, *Journal of Econometrics*, 137:134-161.

[33] Rudebusch, G.D. (2001), Is the Fed Too Timid? Monetary Policy in an Uncertain World, *Review of Economics and Statistics* 83:203-17.

[34] Rudebusch, G.D., and L.E.O. Svensson (1999), Policy Rules for Inflation Targeting, in John B. Taylor (ed.), *Monetary Policy Rules*, Chicago: University of Chicago Press, pp.205-46.

[35] Rudebusch, G.D., and L.E.O. Svensson (2002), Eurosystem monetary targeting: Lessons from U.S. data, *European Economic Review*, 46(3): 417-442.

[36] Siklos, P. L. (2008), What Can We Learn from Comprehensive Data Revisions for Forecasting Inflation? Some US Evidence, in David E. Rapach and Mark E. Wohar (eds.), *Forecasting in*

the Presence of Structural Breaks and Model Uncertainty, Emerald Group Publishing Limited, Chapter 7, pp.271-299.

[37] Stock, J.H., and M.W. Watson (2003), Has the business cycle changed and why? in: NBER Macroeconomics Annual 2002, 17: 159-230.

[38] Stock, J.H., and M.W. Watson (2004), Combination forecasts of output growth in a seven country data set, *Journal of Forecasting* 23:405–430.

[39] Swanson, N.R. and D. van Dijk (2006), Are Statistical Reporting Agencies Getting It Right? Data Rationality and Business Cycle Asymmetry, *Journal of Business and Economic Statistics,* 24:24-42.

[40] Taylor J.B. and J.C. Williams (2010), Simple and Robust Rules for Monetary Policy, NBER Working Paper no. 15908.

[41] Tetlow R.J. and B. Ironside (2007), Real-Time Model Uncertainty in the United States: The Fed, 1996-2003, *Journal of Money, Credit and Banking*, 39:1533-1561.

[42] Tetlow, R.J. and P. von zur Muehlen (2001), Robust Monetary Policy with Misspecified Models: Does Model Uncertainty Always Call for Attenuated Policy?, *Journal of Economic Dynamics and Control* 25:911-49.

[43] Timmermann, A. (2006), Forecast combinations, in Elliott, G., C. Granger, and A. Timmermann (eds.), *Handbook of Economic Forecasting*, vol. 1. North-Holland, Amsterdam, pp. 135–196.

[44] Woodford, M. (1999), Commentary: How Should Monetary Policy Be Conducted in an Era of Price Stability? In *New Challenges for Monetary Policy,* A Symposium Sponsored by the Federal Reserve Bank of Kansas City, pp. 277–316.

[45] Woodford, M. (2003), Optimal interest-rate smoothing, *Review of Economic Studies,* 70:861–886.