

Evaluating Real-Time Forecasts in Real-Time*

Dick van Dijk^{1,†} Philip Hans Franses¹

Francesco Ravazzolo^{1,2}

¹*Econometric Institute
Erasmus University Rotterdam*

²*Norges Bank*

ECONOMETRIC INSTITUTE REPORT EI2007-33

August 27, 2007

Abstract

The accuracy of real-time forecasts of macroeconomic variables that are subject to revisions may crucially depend on the choice of data used to compare the forecasts against. We put forward a flexible time-varying parameter regression framework to obtain early estimates of the final value of macroeconomic variables based upon the initial data release that may be used as actuals in current forecast evaluation. We allow for structural changes in the regression parameters to accommodate benchmark revisions and definitional changes, which fundamentally change the statistical properties of the variable of interest, including the relationship between the final value and the initial release. The usefulness of our approach is demonstrated through an empirical application comparing the accuracy of forecasts of US GDP growth rates from the Survey of Professional Forecasters and the Greenbook.

Keywords: Data revision, forecast evaluation, parameter uncertainty, Bayesian estimation, structural breaks.

JEL Classification: C11, C22, C53, C82, E01, E27

*We are grateful to Sharon Koziicki and Norman Swanson for detailed suggestions. We thank Paolo Giordani, Robert Kohn and Richard Paap for helpful discussions concerning the methodology. Any remaining errors are our own.

[†]Corresponding author: Dick van Dijk, Econometric Institute, Erasmus University Rotterdam, P.O. Box 1738, NL-3000 DR Rotterdam, The Netherlands. Tel.: +31-10-40891263, Fax: +31-10-4089162. *E-mail address:* djvandijk@few.eur.nl

1 Introduction

This paper deals with the evaluation of forecasts for macroeconomic variables such as quarterly GDP growth. The interest lies in comparing the current quality of two or more forecasts, provided by, for example, (members of) the Survey of Professional Forecasters (SPF), the Greenbook of the Federal Reserve, or other institutions. A complication that arises when aiming to compare such forecasts in real-time is that GDP is subject to revisions after its initial release, such that it is not obvious which data to use as “actuals” in the forecast evaluation. As discussed by Croushore (2006), the appropriate choice largely depends on the purpose of the forecast user. For example, given that financial markets mostly respond to the initial release of macroeconomic variables such as employment and GDP, an investor or market analyst might be most interested in obtaining an accurate forecast of the first release of these variables, which therefore constitute the relevant actuals. On the other hand, and this is the perspective taken in this paper, a policymaker may be more interested in forecasting the “true” current value of the variables that play a role in her policy decisions. The problem then is that, due to the occurrence of revisions, the true value of current quarterly GDP growth only becomes available after a few years, if ever.

Revisions come in a variety of forms. Following the first release of current GDP shortly after the end of the current quarter, “regular” revisions are provided due to additional information becoming available to the statistical agency, updating of seasonal adjustment factors, and other reasons. It is only after three years or so that GDP for the current quarter gets a fixed value, which may be treated as the “true” or final value. However, once every five to ten years, GDP and other macroeconomic indicators undergo further and sometimes quite dramatic “benchmark” revisions due to, for example, changes in the data-collection procedures and measurement methods at the statistical agency and to definitional changes. These benchmark revisions

often change the entire history of GDP data and thereby further complicate the task of forecast evaluation for two reasons. First, the forecaster cannot reasonably be expected to predict the substantial changes in a variable due to benchmark revisions. Second, as a forecast user the policymaker is most likely concerned about the current definition of GDP. This implies that we cannot simply take the most recent available data as actuals for a historical forecast evaluation exercise, as they no longer represent the relevant economic variable, at least not for observations in the more distant past. In sum, the question we address in this paper is how we should evaluate and compare forecasts of the current value of macroeconomic variables such as GDP, when due to regular revisions we know the true value only much later in time and when benchmark revisions may fundamentally alter the definition of the variable.

The solution that we propose makes use of a forecasting model for the final value of GDP¹ based on currently available information. This includes the current release of GDP, revisions to growth rates of previous quarters, and other macroeconomic and financial variables that may bear relevant information concerning future (regular) revisions. The estimates or forecasts that can be obtained from this model are treated as realizations of current GDP, and these are in turn used as actuals to evaluate rival forecasts made by the SPF, the Greenbook, and so on.

The novel feature of our forecasting model is that it allows for instability in the relationship between the final value of quarterly GDP and the various regressors, such as the current GDP release. This is achieved by allowing for occasional structural breaks of random magnitude in the regression parameters. We believe that this flexibility is mandatory due to the occurrence of benchmark revisions. As discussed

¹As noted above, the “true” or final value of current GDP may never become known due to the occurrence of benchmark revisions. Hence, it seems difficult to define the final value in the first place. It seems most common to use the data release after approximately three years or the last release before a benchmark revision as an estimate of the final value. We follow this practice in the empirical application in Section 3.

in Corradi *et al.* (2007), these benchmark revisions imply that a given real-time dataset should be considered as a collection of different economic variables. That is, first (and other) releases and final values for different calendar quarters essentially refer to different variables. One of the consequences of such benchmark revisions then is that the relationship between early releases and the final value may change over time, which gets reflected in structural breaks in the regression coefficients in a model relating the final value (or the revision) to the early release. Swanson and van Dijk (2006) provide empirical evidence for the presence such structural changes in models for real-time industrial production and inflation data, see also Aruoba (in press).

The interest in the analysis of real-time data and revisions has been growing steadily over the last decade, stimulated by the increased availability of large real-time data sets for key macroeconomic variables such as employment, money and output measures. A large part of the literature has been concerned with the statistical properties of first release data. In particular, following Mankiw *et al.* (1984) and Mankiw and Shapiro (1986), much attention has been devoted to examining whether the initial release is a rational forecast or a noisy measure of the final value. In the first case, the revision constitutes “news” unrelated to the initial release, while in the second case, the revision constitutes (the removal of) “noise” unrelated to the final value. We refer to Faust *et al.* (2005) for an extensive recent discussion of the news and noise hypotheses. The empirical evidence on the nature of the revisions is mixed, and it seems fair to say that real-time data for most macroeconomic variables contains elements of both news and noise. Mork (1987), Neftçi and Theodossiou (1991), Brodsky and Newbold (1994), Rathjens and Robins (1995), and Croushore and Stark (2001, 2003), among others, analyse these properties for real-time GDP data, which is the variable studied in our empirical application below. Several filtering methods have been suggested to remove the measurement error and extract

an accurate estimate of the true value based on possibly noisy initial releases, see Jacobs and van Norden (2007) for an overview.

The consequences of revisions for forecasting and forecast evaluation have also been studied, see Stekler (1978) and Zarnowitz (1978) for early accounts. The seminal paper by Diebold and Rudebusch (1991) has been very influential, demonstrating that while revised data for the Conference Board’s composite leading index improve forecasts for industrial production, real-time data do not. Other, more recent empirical applications to macroeconomic forecasting with real-time include Swanson (1996), Swanson and White (1997), Amato and Swanson (2001), Orphanides (2001), Orphanides and van Norden (2002, 2005), Bernanke and Boivin (2003), Koenig *et al.* (2003), Kozicki (2004), McGuckin *et al.* (2007), and Chauvet and Piger (in press), see Croushore (2006) for a survey. Several of these studies also provide links to, in particular, monetary policy based on real-time data. Although related to our paper, we do observe that most of the cited and other studies are concerned with the question how to construct forecasts in the presence of data revisions. By contrast, here we take the forecasts as given and focus on their evaluation, that is, on the appropriate choice of “actuals”. In that sense, our paper is most closely related to Robertson and Tallman (1998), who show empirically that the accuracy of GDP forecasts may differ considerably depending upon whether the first or a later release is used to compare the forecasts against.

The outline of our paper is as follows. In Section 2 we outline the forecasting model for the true value of current GDP that we will use to evaluate and compare professional forecasts. In Section 3 we implement this model for quarterly growth rates of US GDP and find that differences across forecasters can appear or disappear depending on the “true” value of GDP that is used. In Section 4 we conclude and give suggestions for further research.

2 Methodology

First we introduce some notation. Let y_t^{t+j} , $j = 1, 2, \dots$, denote the value of quarterly GDP growth (or another macroeconomic variable of interest) for calendar quarter t as it is made available during quarter $t + j$. Hence, for $j = 1$, y_t^{t+1} concerns the initial release of GDP, y_t^{t+2} concerns the first revised value released one quarter later, etc. We use y_t^f to denote the final value of GDP growth, which would be obtained when $j \rightarrow \infty$ if benchmark revisions did not occur.² Furthermore, let r_t^{t+j} , $j = 1, 2, \dots$ denote the magnitude of the revision between the j -th release of y_t and the final value, that is,

$$r_t^{t+j} = y_t^f - y_t^{t+j},$$

and to simplify notation define $r_t \equiv r_t^{t+1} = y_t^f - y_t^{t+1}$ as the full revision between the first release and the final value.

The aim of this paper is to consider models for forecasting the revision r_t^{t+j} based on information available at time $t + j$. We remark that this could be generalized in a straightforward manner to $r_t^{t+j,h}$, $h = 1, 2, \dots$, being the h -period revision between the releases of y_t in quarter $t + j$ and quarter $t + j + h$, that is, $r_t^{t+j,h} = y_t^{t+j+h} - y_t^{t+j}$, but we do not explore this possibility here.

Traditionally, attempts to forecast revisions between the release in period $t + j$ and the final value of y have been made using of linear regression models with constant parameters. In particular, for the complete revision r_t , this renders

$$r_t = x_t' \beta + \varepsilon_t, \tag{1}$$

where $x_t = (1, y_t^{t+1}, w_t)'$ is assumed to include at least an intercept and the first release y_t^{t+1} , and possibly a $(k \times 1)$ vector w_t of other variables observed at the time of the initial release in period $t + 1$. The error term ε_t is assumed to be uncorrelated

²Note that we do not consider the possibility of multiple releases within a single quarter, but it is straightforward to generalize the model set-up described below to incorporate this.

with x_t and to have a constant variance denoted as σ^2 . This framework has often been used for testing whether the initial release y_t^{t+1} is a rational forecast of the final value y_t^f , see Faust *et al.* (2005), Swanson and van Dijk (2006), Aruoba (in press) and references cited therein. This hypothesis implies that the revision r_t^f is uncorrelated with information available at time $t+1$ or, put differently, $\beta = 0$ in (1). In the terminology of Mankiw *et al.* (1984), this constitutes the “news” hypothesis.

2.1 A flexible model for revisions

As discussed in the Introduction, the assumption that the relationship between the initial release and the final value is constant over time may be unrealistic due to the occurrence of benchmark revisions involving redefinitions of the variable y and changes in data-collection and processing techniques. Of course, there are multiple ways to extend the linear regression model in order to capture parameter instability. Here we propose an attractive flexible specification that allows for occasional structural breaks in the regression parameters, that is

$$r_t^f = x_t' \beta_t + \varepsilon_t, \quad (2)$$

where $\beta_t = (\beta_{1,t}, \beta_{2,t}, \dots, \beta_{k+2,t})$ is a vector of time-dependent regression coefficients, which evolve over time according to

$$\beta_{i,t} = \beta_{i,t-1} + \kappa_{i,t} \eta_{i,t}, \quad i = 1, \dots, k+2, \quad (3)$$

where $\eta_{i,t} \sim N(0, q_i^2)$ for $i = 1, \dots, k+2$, and $\kappa_{i,t}$ is an unobserved uncorrelated 0-1 process with $\Pr[\kappa_{i,t} = 1] = \pi_i$ for $i = 1, \dots, k+2$. Hence, the i -th regression parameter $\beta_{i,t}$ remains the same as its previous value $\beta_{i,t-1}$ unless $\kappa_{i,t} = 1$ in which case it changes with $\eta_{i,t}$, see Giordani *et al.* (2007) and Ravazzolo *et al.* (in press) for other applications of this approach.

The flexibility of the specification in (3) stems from the fact that the parameters $\beta_{i,t}$, $i = 1, \dots, k+2$, are allowed to change every time period, but they need not

change at any point in time. The occurrence of a change is described by the latent binary random variable $\kappa_{i,t}$, while the magnitude of the change is determined by $\eta_{i,t}$, which is assumed to be normally distributed with mean zero. Another attractive property of (3) is that the changes in the individual regression parameters are not restricted to coincide but rather are allowed to occur at different points in time.

For estimation and inference in the model (2) with (3) we advocate a Bayesian approach. This provides the posterior distribution of the parameters in the model, that is, the probability and magnitude of the structural changes in the regression parameters, π_1, \dots, π_{k+2} and q_1^2, \dots, q_{k+2}^2 , respectively, plus the variance of the error term, σ^2 . In addition, we thus obtain the posterior distribution of the latent $\kappa_{i,t}$ processes for $i = 1, \dots, k + 2$ and $t = 1, \dots, T$, with T denoting the sample size. This allows to establish *ex post* when breaks in the relationship between the revision r_t and the initial release y_t^{t+1} or the other variables in w_t have occurred. From an *ex ante* perspective, in the construction of revision forecasts the model and the Bayesian approach to inference enables us to account for the fact that such breaks may also happen in future periods.

In the remainder of this section we briefly discuss specification of the prior distributions and the posterior simulation algorithm. The interested reader is referred to Giordani *et al.* (2007) and Ravazzolo *et al.* (in press) for more detailed discussion.

2.2 Prior specification and posterior simulation

The model parameters are collected in the vector $\theta = (\pi_1, \dots, \pi_{k+2}, q_1^2, \dots, q_{k+2}^2, \sigma^2)'$. To facilitate the posterior simulation we make use of independent conjugate priors. For the parameters governing the structural break probabilities we take Beta distributions

$$\pi_i \sim \text{Beta}(a_i, b_i) \quad \text{for } i = 1, \dots, k + 2, \quad (4)$$

where the hyperparameters a_i and b_i can be set according to our prior belief about the frequency of occurrence of structural breaks. For the variance parameters we take the inverted Gamma-2 prior

$$q_i^2 \sim \text{IG-2}(\nu_i, \delta_i) \quad \text{for } i = 1, \dots, k + 2, \quad (5)$$

and

$$\sigma^2 \sim \text{IG-2}(\nu_s, \delta_s), \quad (6)$$

where ν_i , δ_i , $i = 1, \dots, k + 2$, ν_s , and δ_s are parameters which can be chosen to reflect the prior beliefs about the variances. Realistic values of the parameters in the different prior distributions obviously depend on the problem at hand. In Section 3 below we discuss the prior settings for our empirical application to US quarterly GDP growth.

The posterior distributions of the model parameters can be obtained by making use of the Gibbs sampler of Geman and Geman (1984) coupled with the data augmentation method of Tanner and Wong (1987). In addition, we thus obtain the posterior distributions for $B = \{\beta_t\}_{t=1}^T$ and $K = \{\kappa_t\}_{t=1}^T$ with $\kappa_t = (\kappa_{1,t}, \kappa_{2,t}, \dots, \kappa_{k+2,t})'$, as these latent variables are simulated alongside the model parameters θ in the sampling procedure.

The complete data likelihood function, which is needed to derive the posterior distribution of θ , is given by

$$p(r, B, K | x, \theta) = \prod_{t=1}^T p(r_t | x_t, \beta_t, \sigma^2) p(\beta_t | \beta_{t-1}, \kappa_t, q_1^2, \dots, q_{k+2}^2) \prod_{i=1}^{k+2} \pi_i^{\kappa_{i,t}} (1 - \pi_i)^{1 - \kappa_{i,t}}, \quad (7)$$

where $r = (r_1, \dots, r_T)$ and $x = (x'_1, \dots, x'_T)'$. It follows directly from (2) and (3), respectively, that the terms $p(r_t | x_t, \beta_t, \sigma^2)$ and $p(\beta_t | \beta_{t-1}, \kappa_t, q_1^2, \dots, q_{k+2}^2)$ are normal density functions. Combining (7) with the prior density $p(\theta)$, which follows from (4)–(6), we obtain the posterior density

$$p(\theta, B, K | r, x) \propto p(\theta) p(r, B, K | x, \theta). \quad (8)$$

For the Gibbs sampling procedure we employ the efficient sampling algorithm of Gerlach *et al.* (2000) to handle the (occasional) structural breaks. The sampling scheme can be summarized as follows:

1. Draw K conditional on θ , r and x .
2. Draw B conditional on K , θ , r and x .
3. Draw θ conditional B , K , r and x .

In step 1 the latent variables $\kappa_{i,t}$, measuring the (occasional) structural breaks, are drawn using the algorithm of Gerlach *et al.* (2000), which derives its efficiency from generating $\kappa_{i,t}$ without conditioning on the states $\beta_{i,t}$. The conditional posterior density for κ_t , $t = 1, \dots, T$ unconditional on B is

$$\begin{aligned} p(\kappa_t | K_{-t}, \theta, r, x) &\propto p(r | K, \theta, x) p(\kappa_t | K_{-t}, \theta, x) \\ &\propto p(r_{t+1}, \dots, r_T | r_1, \dots, r_t, K, \theta, x) \\ &\quad p(r_t | r_1, \dots, r_{t-1}, \kappa_1, \dots, \kappa_t, \theta, x) p(\kappa_t | K_{-t}, \theta, x), \end{aligned} \tag{9}$$

where $K_{-t} = \{\kappa_s\}_{s=1, s \neq t}^T$. Note that the term $p(\kappa_t | K_{-t}, \theta, x)$ in (9) is simply given by $\prod_{i=1}^{k+2} \pi_i^{\kappa_{i,t}} (1 - \pi_i)^{1 - \kappa_{i,t}}$. The two remaining densities $p(r_{t+1}, \dots, r_T | r_1, \dots, r_t, K, \theta, x)$ and $p(r_t | r_1, \dots, r_{t-1}, \kappa_1, \dots, \kappa_t, \theta, x)$, can easily be evaluated as demonstrated in Gerlach *et al.* (2000). As κ_t can take only a finite number of values, the integrating constant can easily be computed by normalization.

In step 2 of the above algorithm the simulation smoother of Carter and Kohn (1994) is used to compute the full conditional posterior density for the latent regression parameters B . The Kalman smoother is applied to derive the conditional mean and variance of the latent factors. For the initial value β_0 we choose a multivariate normal prior with mean zero.

Finally, in step 3 we can use standard results in Bayesian inference to sample the parameters θ . Hence, the probabilities π_i are sampled from Beta distributions and the variance parameters q_i^2 and σ^2 are sampled from inverted Gamma-2 distributions.

3 Evaluating US GDP forecasts

We employ the methodology laid out in the previous section to evaluate the accuracy of forecasts for US quarterly real GDP growth rates. In particular, we compare the one-quarter ahead consensus forecast provided by the Survey of Professional Forecasters (SPF) and the so-called Greenbook forecast produced by the Board of Governors of the Federal Reserve. Both sets of forecasts are obtained from the Federal Reserve Bank of Philadelphia’s real-time data website.³ The SPF has included real GDP since the survey that was held in 1981Q3 and are available until 2006Q4. The Greenbook forecasts date back earlier, to 1968Q3, but are published with a five-year delay such that the most recent forecast available concerns the GDP growth rate in 2001Q4. This means that the common sample period runs from 1981Q3 until 2001Q4. We obtain actual real-time GDP data also from the Federal Reserve Bank of Philadelphia’s real-time data website. This concerns quarterly vintages starting in 1965Q4, where the time series in each vintage starts in 1947Q1. We refer to the studies mentioned in the introduction for analyses of the (regular) revisions in real GDP. Aruoba (in press) provides a recent account, showing that revisions are “not well-behaved” and are predictable to some extent. Benchmark revisions in real GDP occurred in the first quarter of 1966, 1971, 1976, 1981, 1986, 1992, 1996, and 2001, see Siklos (2007) for an analysis of these revisions.

We evaluate the SPF and Greenbook forecasts for the sample period 1984Q1-2001Q4, for a total of 72 quarters. In order to obtain an impression of the impact that the choice of actuals has on the perceived forecast accuracy, we compare the forecasts for the growth rate in quarter t against (a) the first release, which becomes available in quarter $t + 1$, (b) the second release, which becomes available in quarter $t + 2$, (c) the “final” value, which is defined as the value that becomes available three years after quarter t , or the last value prior to a benchmark revision, in case one

³<http://www.philadelphiafed.org/econ/forecast/reaindex.html>

occurs in the intermediate period between quarters t and $t + 12$, and (d) estimates of the final value obtained from a regression model relating the revision to the initial release and other relevant information that is available at the time the forecast becomes available. We employ both models with fixed parameters of the form (1) and models with time-varying parameters as in (2) with (3). In addition to these models for the full revision between the first release and the final value, we also consider their equivalents for the revision between the second release y_t^{t+2} and the final value y_t^f , that is, using r_t^{t+2} as dependent variable in (1) and (2). Accordingly, we include the second release in x_t instead of the initial release y_t^{t+1} .

Concerning the additional regressors w_t in the revision forecasting models we follow Aruoba (in press) by including revisions to the GDP growth rates for previous quarters and several business cycle proxies. This choice is motivated by the observations that often successive revisions are correlated and that revisions seem to vary systematically with the state of the business cycle, see also Swanson and van Dijk (2006). To be precise, in the models for the full revision r_t^f we use the revisions between the first release of the growth rates for quarters $t - 1, \dots, t - 5$ and the release in quarter $t + 1$, that is, $r_{t-1}^{t,1}, \dots, r_{t-5}^{t-4,5}$. In addition, as business cycle proxies we use the S&P500 index return in quarter t , the 3-month T-bill rate at the end of quarter t , as well as the term spread, defined as the difference between the 10-year T-bond rate and the 3-month T-bill rate measured at the same time. The same variables are included in the model for the revision between the second release and the final value, with the difference that we now include the revisions between the first release of the growth rates for quarters $t, \dots, t - 5$ and the release in quarter $t + 2$, plus the first revision $r_t^{t+1,1} = y_t^{t+2} - y_t^{t+1}$ of the growth rate in quarter t itself.

We set the prior for the hyperparameters a_i and b_i in the Beta distribution for the prior probability of breaks in the regression parameters π_i equal to $a_i = 0.8$ and $b_i = 18$ for all i . This implies that the prior mean duration between breaks

in a particular regression parameter is equal to 24 quarters. For the variances of the break magnitudes, that is q_i^2 , $i = 1, \dots, k$, we take very peaked priors with mode near zero to limit the number of potential breaks. Finally, the Gamma-2 prior parameters for σ^2 are 2 with 20 degrees of freedom.

Before we turn to the evaluation of the SPF and Greenbook forecasts, we provide some insight into the relevance of allowing for structural changes in the parameters of the forecasting models for the revisions by considering in-sample results, which are obtained by estimating the models using the revisions over the sample period 1966Q4-2001Q4. Figures 1 and 2 show the posterior mean for the latent binary variable $\kappa_{i,t}$ governing the occurrence of changes in the regression parameters, together with the associated posterior mean for $\beta_{i,t}$ in the models for the revisions following the first and second release, respectively. For the coefficients, the 25th and 75th percentiles of the posterior distributions are also shown. Several interesting findings emerge from these graphs. First, the posterior means of $\kappa_{i,t}$ show quite erratic, ‘spiky’ behavior, suggesting that the probabilities of structural breaks in the parameters vary considerably from one period to the next. Previously we argued that changes in the relationship between the initial data release and the final value are mainly due to the occurrence of benchmark revisions, and the results in Figures 1 and 2 suggest that this is indeed the case for US real GDP. For example, in Figure 1 we observe a high probability of structural change in several parameters around the benchmark revision in 1976Q1. In addition, substantial changes in the regression coefficients seem to have occurred following the widely documented volatility reduction of GDP growth around 1984, see also Aruoba (in press). Second, the posterior mean of the break probability for the different past revision variables bear great resemblance. This is partly caused by the fact that these revisions often are (positively) correlated. Third, the magnitude of the change in parameters can be substantial. For example, the graph on the righthand side of panel (d) in Figure 1 shows that the

posterior mean of the coefficient of the second revision of the growth rate in quarter $t - 2$ changes from -0.3 at the start of the sample period to approximately zero at in 2001, see also panel (e) of Figure 2.

- insert Figures 1 and 2 about here -

For the real-time forecast evaluation exercise, the regression models in (1) and (2) are estimated recursively using an expanding window of data, starting with 1966Q4-1981Q1. Given that we need the data release from 1984Q2 to obtain the full revision for the GDP growth rate in 1981Q1, the resulting parameter estimates can be used to obtain an estimate of the final value in 1984Q1 for which the first release is observed in that quarter.

Although we emphasize the use of the final value estimates obtained from the forecasting models for the revisions as actuals in forecast evaluation, we may of course also use these estimates themselves as alternative forecasts of the final values. In that sense, we obtain additional evidence for the relevance of allowing for structural change in the regression coefficients. Comparing the forecasts of the final value obtained from the model with constant parameters in (1) with the actual final values, we find a mean squared prediction error (MSPE) of 2.38, while for the model with occasional structural change in (2) with (3) this is equal to 1.62. Hence, allowing for structural breaks in the parameters leads to considerably more accurate forecasts of the final value of quarterly real GDP growth rates.

We now turn to the evaluation and comparison of the SPF consensus forecasts and the Greenbook forecasts. Table 1 reports MSPE values using different data releases as actuals. We observe a marked discrepancy in the relative accuracy of these forecasts depending on which data are used as actuals. when the first or second release of GDP is used in the forecast evaluation, the Greenbook forecasts appear more accurate as they lead to smaller MSPE values. By contrast, when the actual final value is used, the SPF consensus forecast achieves an MSPE that is

about 5% lower than the Greenbook forecast (3.96 against 4.25). Using forecasts of the final value from the revision forecasting models, we find a similar ranking, as the SPF consensus forecast is more accurate than the Greenbook forecast, although admittedly the differences in MSPE are not very large. Nevertheless, if the aim of the forecast user is to obtain an accurate prediction of the final value, it seems better to use the revision forecasts to construct a preliminary estimate of the final value than to use the first or second release.

- insert Table 1 about here -

4 Conclusions

The accuracy of real-time forecasts of macroeconomic variables that are subject to revisions may crucially depend on the choice of data used to compare the forecasts against. In case the aim is to forecast the “true” value, the initial release is not suitable for this purpose due to measurement error. The most recent available data also should not be used due to benchmark revisions and definitional changes, which may fundamentally change the properties of the variable of interest. In this paper we have put forward a flexible time-varying parameter regression framework to obtain early estimates of the final value of macroeconomic variables based upon the initial release and other information available at the time the forecasts are produced. These estimates may be used as actuals in forecast evaluation. Allowing for structural changes in the regression parameters is needed as benchmark revisions may also alter the relationship between the final value and the initial release.

The empirical application for quarterly US real GDP growth rates demonstrates the relevance of allowing for structural change in the parameters of the revision forecasting models, as these seem to occur fairly often but especially around benchmark revisions, while the magnitude of the change in parameters can be substantial. Using the forecasts of the final value obtained from our model leads to judgements of the

(relative) quality of the accuracy of SPF and Greenbook forecasts that are more in line with an evaluation using the actual final value than when using the initial data release.

Interesting topics for future research include generalizing our time-varying parameter regression model to a multivariate context, in order to exploit the correlation between revisions of different macroeconomic variables. Another possible extension concerns the distinction between news and noise components in the revisions, which is not accounted for in the current setup.

References

- Amato, J., and Swanson, N.R. (2001), “The Real-Time Predictive Content of Money for Output,” *Journal of Monetary Economics*, 48, 3–24.
- Aruoba, D.S. (in press), “Data Revisions Are Not Well-Behaved,” *Journal of Money, Credit, and Banking*.
- Bernanke, B.S., and Boivin, J. (2003), “Monetary Policy in a Data-Rich Environment,” *Journal of Monetary Economics*, 50, 525–546.
- Brodsky, N., and Newbold, P. (1994), “Late Forecasts and Early Revisions of United States GNP,” *International Journal of Forecasting*, 10, 455–460.
- Carter, C. and Kohn, R. (1994), “On Gibbs Sampling for State-Space Models,” *Biometrika*, 81, 541–553.
- Chauvet, M. and Piger, J. (2007), “A Comparison of the Real-Time Performance of Business Cycle Dating Methods,” *Journal of Business and Economic Statistics*, to appear.
- Corradi, V., Fernandez, A. and Swanson, N.R. (2007), “Information in the Revision Process of Real-Time Datasets,” Rutgers University, working paper.
- Croushore, D. (2006), “Forecasting with Real-Time Macroeconomic Data,” in Elliott, G., Granger, C.W.J. and Timmermann, A. (eds.), *Handbook of Economic Forecasting*, Elsevier, North-Holland, pp. 961–982.
- Croushore, D., and Stark, T. (2001), “A Real-Time Dataset for Macroeconomists,” *Journal of Econometrics*, 105, 111–130.
- (2003), “A Real-Time Dataset for Macroeconomists: Does Data Vintage Matter?,” *Review of Economics and Statistics*, 85, 605–617.
- Diebold, F.X., and Rudebusch, G.D. (1991), “Forecasting Output with the Composite Leading Index: A Real-Time Analysis,” *Journal of the American Statistical Association*, 86, 603–610.
- Faust, J., Rogers, J.H., and Wright, J.H. (2005), “News and Noise in G-7 GDP Announcements,” *Journal of Money, Credit and Banking*, 37, 403–419.
- Geman, S. and Geman, D. (1984), “Stochastic Relaxation, Gibbs Distributions and the Bayesian Restoration of Images,” *IEEE Transaction on Pattern Analysis and Machine Intelligence*, 6, 721–741.
- Gerlach, R., Carter, C. and Kohn, R. (2000), “Efficient Bayesian Inference for Dynamic Mixture Models,” *Journal of the American Statistical Association*, 95, 819–828.
- Giordani, P., Kohn, R. and van Dijk, D. (2007), “A Unified Approach to Nonlinearity, Outliers and Structural Breaks,” *Journal of Econometrics*, 137, 112–133.
- Jacobs, J.P.A.M. and van Norden, S. (2007), “Modelling Data Revisions: Measurement Error and Dynamics of True Values,” Les Cahiers de CREF 07-09.
- Koenig, E., Dolmas, S. and Piger, J. (2003), “The Use and Abuse of Real-Time Data in Economic Forecasting,” *Review of Economics and Statistics*, 85, 618–628.
- Kozicki, S. (2004), “How Do Data Revisions Affect the Evaluation and Conduct of Monetary Policy?,” Federal Reserve Bank of Kansas City *Economic Review*, 5–38.

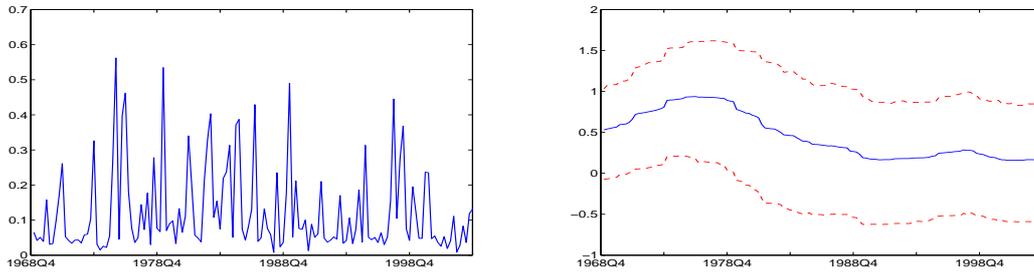
- Mankiw, N.G., Runkle, D.E., and Shapiro, M.D. (1984), "Are Preliminary Announcements of the Money Stock Rational Forecasts?," *Journal of Monetary Economics*, 14, 15–27.
- Mankiw, N.G., and Shapiro, M.D. (1986), "News or Noise: an Analysis of GNP Revisions," *Survey of Current Business*, 66, 20–25.
- McGuckin, R.H., Ozyildirim, A. and Zarnowitz, V. (2007), "A More Timely and Useful Index of Leading Indicators," *Journal of Business and Economic Statistics*, 25, 110–120.
- Mork, K.A. (1987), "Ain't Behavin': Forecast Errors and Measurement Errors in Early GNP Estimates," *Journal of Business and Economic Statistics*, 5, 165–175.
- Neftçi, S.N., and Theodossiou, P. (1991), "Properties and Stochastic Nature of BEA's Early Estimates of GNP," *Journal of Economics and Business*, 43, 231–239.
- Orphanides, A. (2001), "Monetary Policy Rules Based on Real-Time Data," *American Economic Review*, 91, 964–985.
- Orphanides, A. and van Norden, S. (2002), "The Unreliability of Output Gap Estimates in Real Time," *Review of Economics and Statistics*, 84, 569–583.
- (2005), "The Reliability of Inflation Forecasts Based on Output Gap Estimates in Real Time," *Journal of Money, Credit, and Banking*, 37, 583–601.
- Rathjens, P., and Robins, R.P. (1995), "Do Government Agencies Use Public Data? The Case of GNP," *Review of Economics and Statistics*, 77, 170–172.
- Ravazzolo, F., Paap, R., van Dijk, D. and Franses, P.H. (in press), "Bayesian Model Averaging in the Presence of Structural Breaks," in Wohar, M. and Rapach, D.E. (eds.), *Forecasting in the Presence of Structural Breaks and Model Uncertainty - Frontiers of Economics and Globalization Series*, Amsterdam: Elsevier.
- Robertson, J.C., and Tallman, E.W. (1998), "Data Vintages and Measuring Forecast Performance," *Federal Reserve Bank of Atlanta Economic Review*, 83 (Fourth Quarter), 4–20.
- Siklos, P.L. (2007), "What Can We Learn From Comprehensive Data Revisions for Forecasting Inflation? Some U.S. Evidence," Wilfred Laurier University, working paper.
- Stekler, H.O. (1967), "Data Revisions and Economic Forecasting," *Journal of the American Statistical Association*, 62, 470–483.
- Swanson, N.R., and van Dijk, D. (2006), "Are Statistical Reporting Agencies Getting It Right? Data Rationality and Business Cycle Asymmetry," *Journal of Business and Economic Statistics*, 24, 24–42.
- Swanson, N.R. (1996), "Forecasting Using First-Available Versus Fully Revised Economic Time-Series Data," *Studies in Nonlinear Dynamics and Econometrics*, 1, 47–64.
- Swanson, N.R. and White, H. (1997), "A Model Selection Approach to Real-Time Macroeconomic Forecasting Using Linear Models and Artificial Neural Networks," *Review of Economics and Statistics*, 79, 540–550.
- Zarnowitz, V. (1978), "Accuracy and Properties of Recent Macroeconomic Forecasts," *American Economic Review*, 68, 313–319.

Table 1: Mean squared prediction errors for one-step ahead forecasts of quarterly US real GDP growth rates

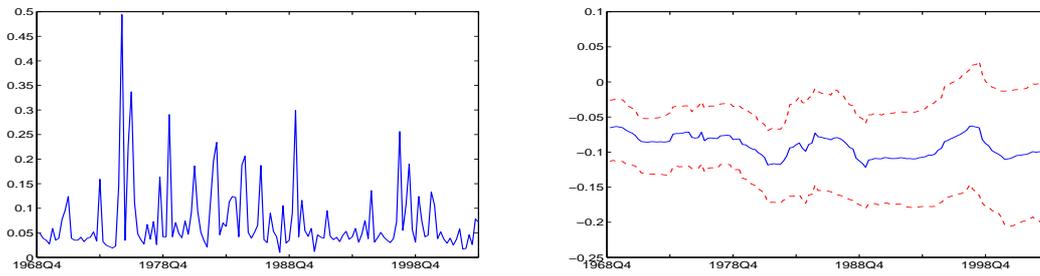
	First release	Second release	Fully revised	Final CP, 1	Final TVP, 1	Final CP, 2	Final TVP, 2
Greenbook	2.01	2.00	4.25	3.11	2.32	3.70	2.99
SPF mean	2.05	2.22	3.96	3.02	2.24	3.55	2.96

Note: The table reports mean squared prediction errors (MSPE) for SPF and Greenbook forecasts of quarterly real GDP, evaluated against different actuals. CP indicates forecasts of the final value obtained from the model with constant parameters in (1), TVP indicates forecasts of the final value obtained from the model with time-varying parameters in (2) with (3). The addition 1 and 2 indicates whether the revision following the first or second release is predicted.

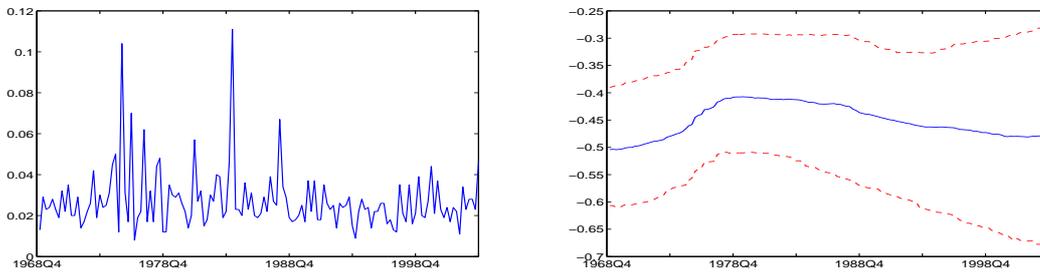
Figure 1: Marginal posterior densities of the breaks and β parameters



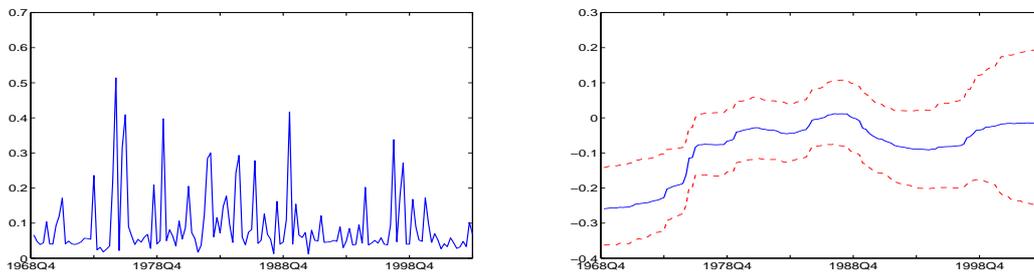
(a) Intercept



(b) First release y_t^{t+1}

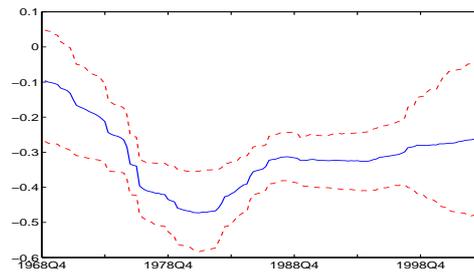
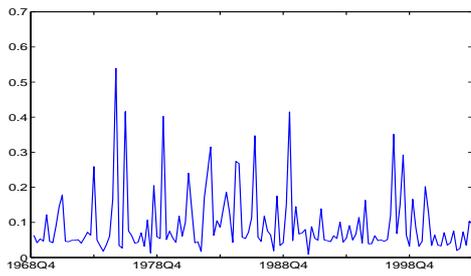


(c) Revision $r_{t-1}^{t,1}$

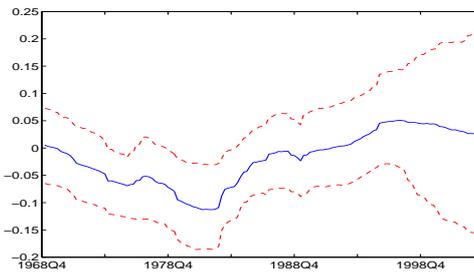
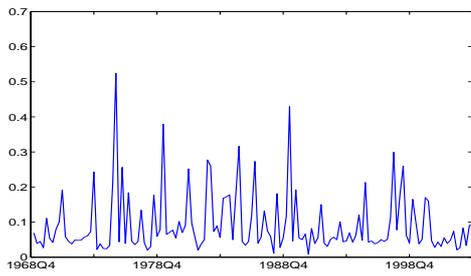


(d) Revision $r_{t-2}^{t-1,2}$

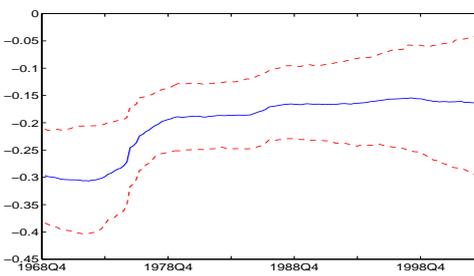
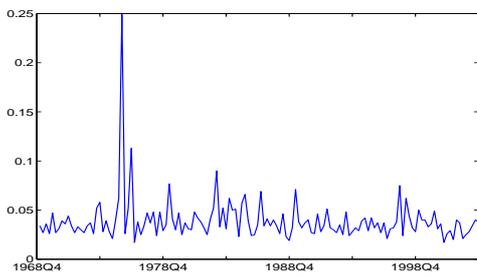
(continued on next page)



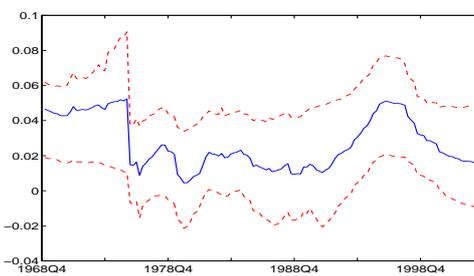
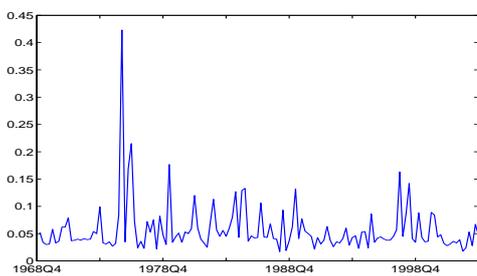
(e) Revision $r_{t-3}^{t-2,3}$



(f) Revision $r_{t-4}^{t-3,4}$

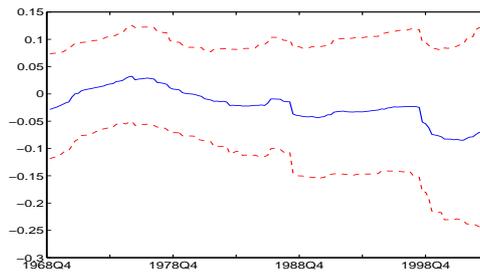
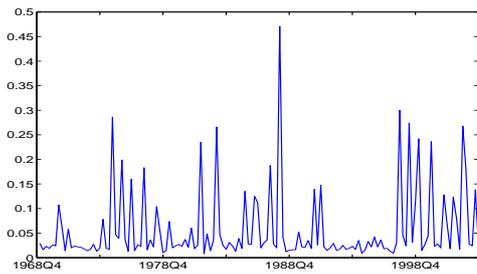


(g) Revision $r_{t-5}^{t-4,5}$

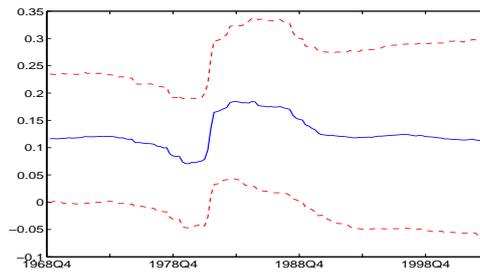
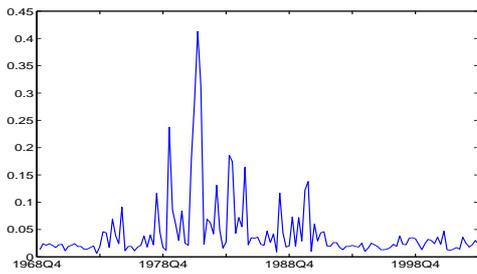


(h) S&P 500, 3-month return

(continued on next page)



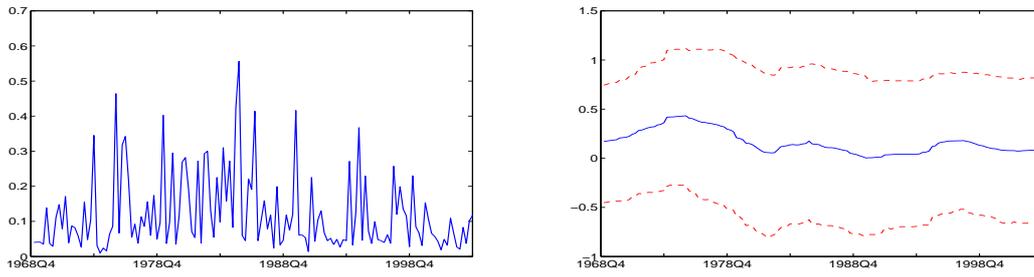
(i) 3-month T-bill rate



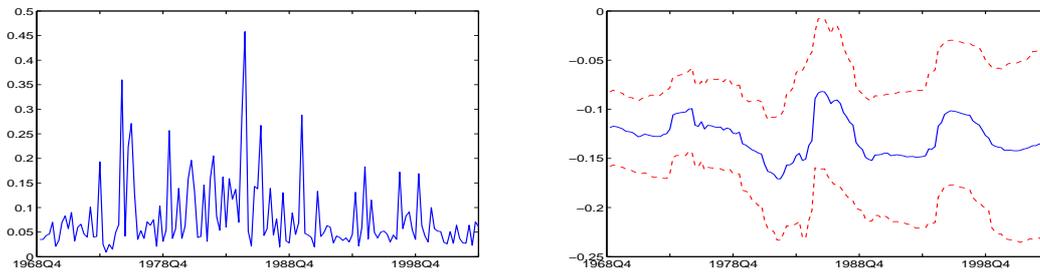
(j) Term spread

Note: The graphs in this figure show the posterior means (solid line) of $\kappa_{i,t}$ on the left side and $\beta_{i,t}$ on the right side in the regression model (2) for the complete revisions in quarterly GDP growth rates, estimated over the period 1966Q4-2001Q4. The dashed lines in the graphs for the coefficients are the 25th and 75th percentiles of the posterior densities.

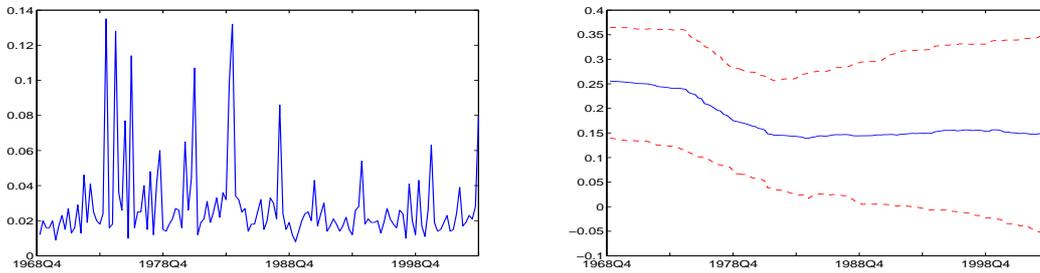
Figure 2: Marginal posterior densities of the breaks and β parameters



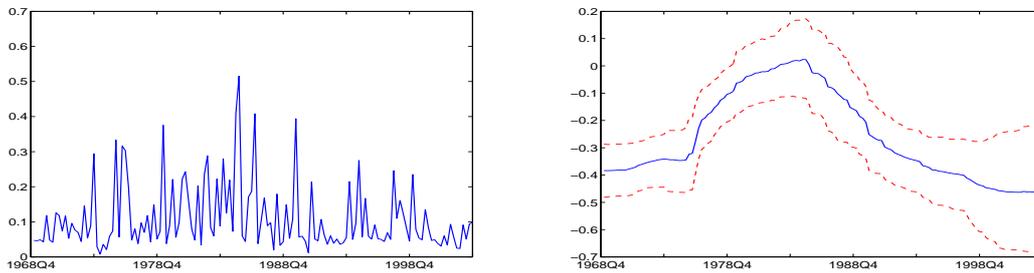
(a) Intercept



(b) Second release y_t^{t+2}

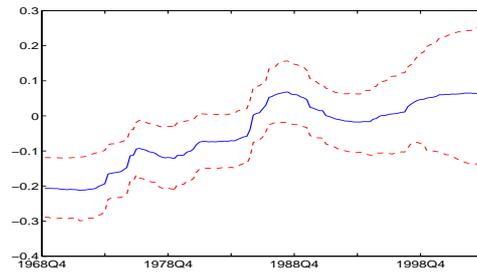
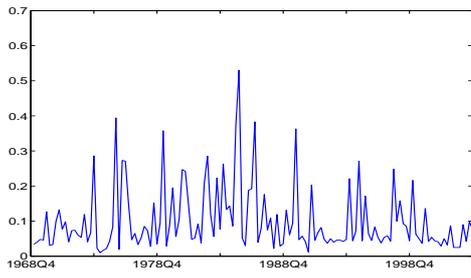


(c) Revision $r_t^{t+1,1}$

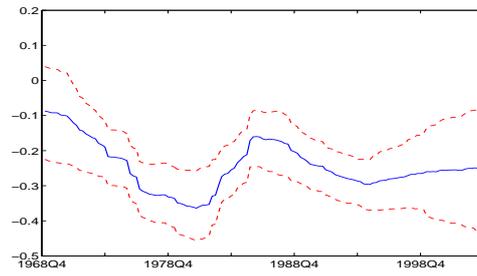
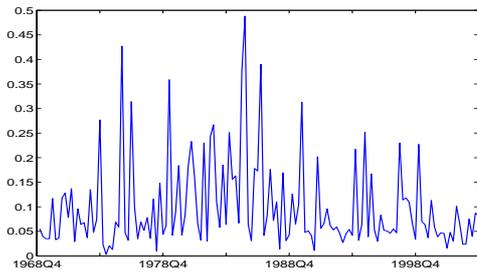


(d) Revision $r_{t-1}^{t,2}$

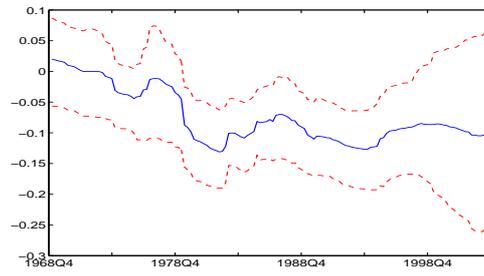
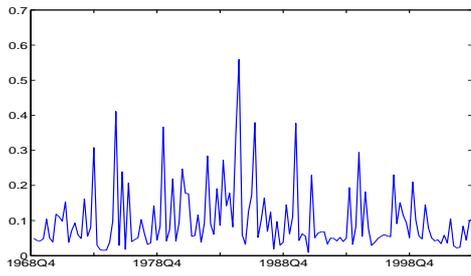
(continued on next page)



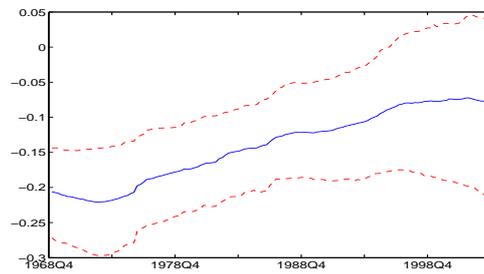
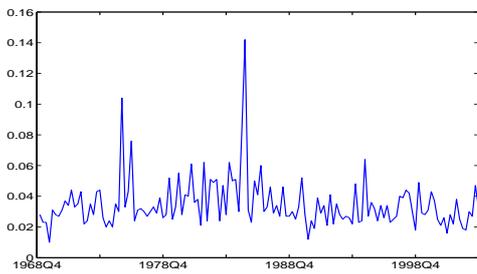
(e) Revision $r_{t-2}^{t-1,3}$



(f) Revision $r_{t-3}^{t-2,4}$

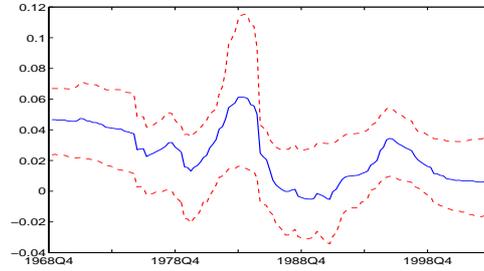
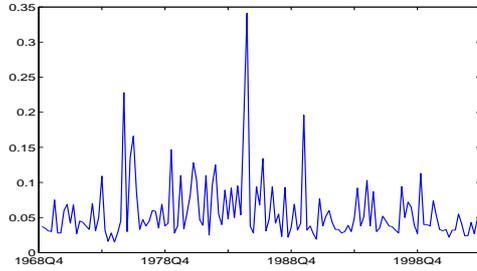


(g) Revision $r_{t-4}^{t-3,5}$

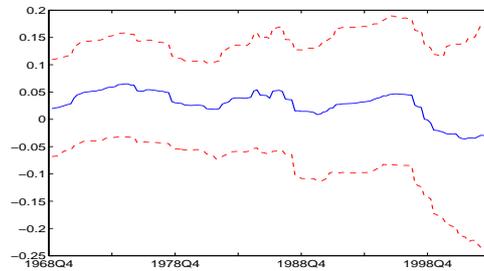
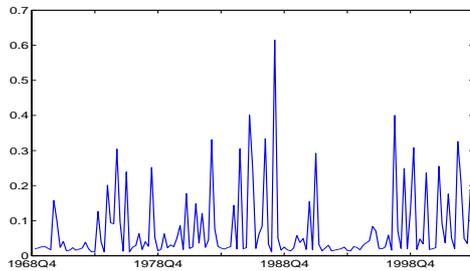


(h) Revision $r_{t-5}^{t-4,6}$

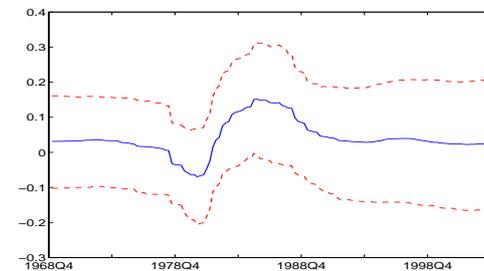
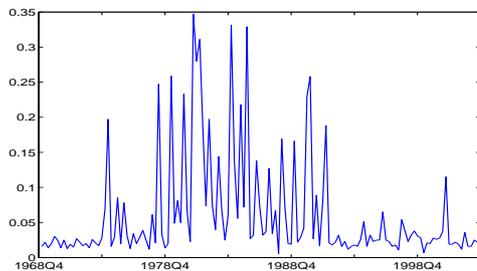
(continued on next page)



(i) S&P 500, 3-month return



(j) 3-month T-bill rate



(k) Term spread

Note: The graphs in this figure show the posterior means (solid line) of $\kappa_{i,t}$ on the left side and $\beta_{i,t}$ on the right side in the regression model (2) for the revisions between the second release and the final value of quarterly GDP growth rates, estimated over the period 1966Q4-2001Q4. The dashed lines in the graphs for the coefficients are the 25th and 75th percentiles of the posterior densities.