DP2006/02

Forecasting Substantial Data Revisions in the Presence of Model Uncertainty

Anthony Garratt, Gary Koop and Shaun P. Vahey

February 2006

JEL classification: E01, C11, C32, C53

Discussion Paper Series

DP2006/02

Forecasting Substantial Data Revisions in the Presence of Model Uncertainty*

Anthony Garratt, Gary Koop and Shaun P. Vahey†

Abstract

A recent revision to the preliminary measurement of GDP(E) growth for 2003Q2 caused considerable press attention, provoked a public enquiry and prompted a number of reforms to UK statistical reporting procedures. In this paper, we compute the probability of "substantial revisions" that are greater (in absolute value) than the controversial 2003 revision. The predictive densities are derived from Bayesian model averaging over a wide set of forecasting models including linear, structural break and regime-switching models with and without heteroskedasticity. Ignoring the nonlinearities and model uncertainty yields misleading predictives and obscures the improvement in the quality of preliminary UK macroeconomic measurements relative to the early 1990s.

* The views expressed in this paper are those of the author(s) and do not necessarily reflect those of the Reserve Bank of New Zealand. We thank Dean Croushore, Simon van Norden, Athanasios Orphanides, Adrian Pagan, Simon Potter and James Yetman for helpful comments. We are also grateful to seminar participants at the Society for Computational Economics 2005 meetings, the CIRANO Data Revisions Workshop, University of New South Wales, University of Otago, and Norges Bank. Financial support from the ESRC (Research Grant No RES-000-22-1342) is acknowledged gratefully.

† Anthony Garratt (Birkbeck College), Gary Koop (University of Strathclyde), Shaun Vahey (Reserve Bank of New Zealand). Address: Reserve Bank of New Zealand, 2 The Terrace, Wellington, New Zealand. Tel.: +64 4 471 3670, Fax: +64 4 473 1209. *email address: vaheys@rbnz.govt.nz.*

©Reserve Bank of New Zealand

1 Introduction

It is widely understood that statistical agencies should revise macroeconomic data measurements. Delayed information flows ensure that initial measurements of economic variables routinely contain inaccuracies, and transparent statistical agencies seek to provide the most accurate measurements feasible, given their information set. Since data agencies aim to reduce data inaccuracies (among other considerations), the UK financial press often interpret unusually large revisions as preliminary indicators of statistical degradation.

The considerable controversy surrounding the preliminary expenditure measurement of GDP (known as GDP(E)) growth for 2003Q2 prompted the UK's Statistics Commission (2004) to instigate a wide-ranging and public review of statistical reporting procedures. The review (hereafter referred to as the "Mitchell Report"after principal investigator, James Mitchell) made a number of specific recommendations to enhance transparency and documented public concerns about statistical quality. Shortly after the Mitchell Report, a *Code of Practice* set out a new protocol for revisions (National Statistics, 2004). This specified that the incidence of "substantial revisions" would be used to monitor statistical performance.

Motivated by the aftermath of the 2003Q2 substantial revision, we outline an approach to predict the (conditional) probability of revisions for UK GDP(E) growth. For each observation in our evaluation period, we generate a predictive density by Bayesian model averaging (BMA) over a wide set of forecasting models for revisions. In addition to the standard linear specification, the set of models includes many nonlinear alternatives. We focus on the revision between the first and second measurements of the growth rates, where second release lags the first by one quarter. Our definition of a revision approximates that used by the financial press and the Office of National Statistics (ONS) to assess revisions. We report probabilities of revisions greater than x in absolute value conditional on the initial measurement, where x takes on a number of values. Since the Mitchell Report[1] emphasised that considerable public and financial market attention followed the revision of just over 0.3 percentage points to the preliminary 2003Q2 GDP(E) growth measurement, we define revisions greater than this threshold as "substantial".[2] A time series plot of the recursively estimated probabilities serves as an ocular tool to aid assessment of revision performance.

[1] See Statistics Commission, 2004, vol.1, p18 and vol.2 p4.

[2] The *Code of Practice* gives no guidance on the precise definition of "substantial" to be used for monitoring purposes (National Statistics, 2004).

Our BMA methodology differs from the standard approach to charactersing revisions adopted in the literature (see, for example, Mankiw, Runkle and Shapiro, 1984, and more recently Faust, Rogers and Wright, 2005). The classical approach typically uses a single linear regression model with the data revision as the dependent variable and the initial measurement as the explanatory variable. Although commonly used in ONS studies, such as Akritidis (2003a and 2003b) and George (2005), the potential for nonlinearities and model uncertainty are ignored. Recent papers by Swanson and van Dijk (2006) and Garratt and Vahey (2006) have found that structural breaks and regime switching affect the revisions processes using US and UK data respectively. But neither of these academic studies report predictive densities using (some) models that exhibit the multiple breaks in the error variance associated with sporadic structural reforms to data reporting procedures.

To illustrate the importance of nonlinearities, we report results for the "best" model (selected using the Bayesian Information Criterion) and the linear model, in addition to those from model averaging. We break our empirical work into two parts. In the first part, we examine the extent to which the various models are supported by the data. We find weak evidence for breaks and regime-switches in the regression coefficients and strong evidence in favor of breaks in the error variance, that occur mostly since the early 1990s.

In the second part of the empirical work, we focus on recursively estimated out of sample predictives. We show that the linear model yields misleading results because it misses structural breaks in the error variance picked up by the best model. Since models with variance breaks receive a great deal of support, they are weighted heavily in the BMA exercise.

Our recursive out of sample BMA predictions reveal that the probability of substantial revisions fell sharply in the mid-1990s to level out at less than five percent after 1998Q2. This confirms that the early 1990s reforms to UK statistical reporting procedures discussed by Wroe (1993) had beneficial impacts, primarily through the error variance, reducing the expected frequency of substantial revisions to roughly once every five years.

The remainder of the paper is organised as follows. Section 2 discusses the background and consequences of the 2003Q2 GDP(E) revision. Section 3 examines our models for UK data revisions. Section 4 discusses econometric methods and the subsequent section describes the data. Section 6 presents the results. The final section concludes.

2 A substantial revision: background and aftermath

In the absence of this one revision to quarterly GDP growth, we believe the press comment would not have become nearly as critical as it did.

Mitchell Report, Statistics Commission, 2004, Vol. 2, p32.

The extreme press reaction to the 2003Q2 revision was conditioned partly by the history of statistical reforms, by the institutional arrangements which govern the production of UK data and by expectations of future public scrutiny.

The history of British statistics, summarised in HM Treasury (1998, annex A), clarifies the key role of public reviews in the provision of UK data. Policymakers became concerned about the quality of macroeconomic statistics in the 1980s. Nigel Lawson (1992, p845), Chancellor of the Exchequer 1983-1989, described official UK macro data as "little more than a work of fiction". In 1989 the Pickford Review documented considerable downwards bias in the initial measurements of many macroeconomics indicators (see the discussion by Egginton, Pick and Vahey, 2002). To remedy this, the Central Statistical Office (CSO, forerunner of the ONS) expanded to take responsibility for a greater proportion of UK statistics and reformed many of the underlying surveys. Wroe (1993) discussed these reforms in detail, together with the two 'Chancellor's Initiatives' introduced in the early 1990s to further enhance statistical quality.

In the light of these structural reforms, the UK press took a close interest in monitoring statistical quality. For example, at least 10 newspapers and 15 financial commentators passed comment on national statistics in the 12 months prior to the controversial revision to 2003Q2 GDP growth (see Statistics Commission, 2004, vol.2 p28-33).

The GDP revision at the end of September 2003 sparked particularly strong press hostility. An initial measurement of 0.3 percent for quarterly GDP(E) was revised up by just over 0.3 percentage points.[3] Concern about the press reaction and the threat to public confidence led the Statistics Commission to

[3] Len Cook, National Statistician 2000-2005, noted in oral testimony to a Treasury Select Committee of UK politicians that first measurements of GDP(E) are released nearly a month earlier than in other European Union countries. The transcript can be downloaded from http://www.publications.parliament.uk/. We return to this timeliness issue and the definition of revisions used in our econometric models in section 5.

instigate the review conducted by James Mitchell of the National Institute of Economic and Social Research.[4] The recommendations published in early 2004 focused on transparency and the use of forecast information in statistical reporting. The Mitchell Report identified forecast information included in construction output figures (by the Department of Trade and Industry) as the primary source of the notorious 2003Q2 revision (Statistics Commission, 2004, vol.2 p50-61).[5] A subsequent MORI survey of data users confirmed that many felt UK statistics had become inadequate (see Statistics Commission, 2005, p5). The Statistics Commission accepted that some reforms, including greater autonomy, would enhance statistical credibility.[6]

Conditioned by the extreme press hostility to the 2003Q2 revision, the "Code of Practice" for reporting revisions specified a protocol for the treatment of substantial revisions (see National Statistics, 2004). These are defined as:

> ...those which lie outside the range of revisions normally associated with the statistics in question and which tend, therefore, to have a more significant impact."
>
> National Statistics (2004, p7)

Decisions to make substantial revisions now require the authority of the relevant Chief Statistician (p10), must be accompanied by a public explanation (p13) and will be used as "diagnostic tools to monitor and improve quality" (p14) (see National Statistics, 2004).[7] More routine revisions are monitored in detail too, through "revision triangles" which record revisions through time (see Jenkinson and George, 2005) and frequent revisions analyses (see, for example, George, 2005).[8]

The statistical reforms in the aftermath of the 2003Q2 revision are ongoing. Gordon Brown, Chancellor of the Exchequer, confirmed on 28 November 2005 that the ONS would be made independent at a date yet to be announced.

[4] The Statistics Commission provides independent advice on UK national statistics; see http://www.statscom.org.uk/.

[5] Section 5 discusses the many potential causes of data revisions.

[6] Another recent public review, the Allsopp review in March 2004, argues for greater provision of UK regional data and larger surveys for macro data. See http://www.hm-treasury.gov.uk/allsopp.

[7] The Code of Practice also sets out the protocol for "unexpected" revisions (which might be caused by errors) and by "scheduled" revisions (which are not). Analysis based on these characteristics represents an interesting area for future work. This requires more detailed information on the causes of each revision than contained in the Castle and Ellis (2002) data used in this study.

[8] The triangles for quarterly growth rates are published on the National Statistics website, http://www.statistics.gov.uk/.

This effectively reversed the 1989 decision to make the Chancellor of the Exchequer responsible for the CSO.[9] Nigel Lawson (1992, p378), Chancellor at the time, noted the unpopularity of the CSO annexation within the statistical agency. Some staff felt that the Chancellor would be subject to accusations of "fiddling the figures".

3 Modelling the revision process

Given the ramifications of the substantial 2003Q2 revision, our empirical analysis aims to assess the probability of similar sized (and smaller) revisions. The standard approach to characterising data revisions adopted by, for example, Mankiw, Runkle and Shapiro (1984) and Faust, Rogers and Wright (2005) uses a single linear regression model:

$$Y_t^k = \alpha^k + \beta^k X_t^1 + \varepsilon_t^k, \tag{1}$$

where X_t^k is the k^{th} measurement of a variable and $Y_t^k = X_t^k - X_t^1$ is the revision between the k^{th} and the first measurement.

Since the press reacted strongly to the second quarterly measurement for 2003Q2, the revision of interest is defined as the second measurement minus the first; we set $k = 2$. Hereafter, we suppress the superscripts for simplicity. A more common treatment of revisions, see for example Mankiw, Runkle and Shapiro (1984), Diebold and Rudebusch (1991), Faust, Rogers and Wright (2005) and Garratt and Vahey (2006), compares preliminary measurements with those taken at a particular vintage date.[10] Among others, Aruoba (2005) and Croushore (2005) compare preliminary measurements with those taken just before a "benchmark" revision. Neither definition of revisions although common in the academic literature, matches that used by the UK financial press to monitor statistical quality.

It is straightforward to carry out Bayesian inference in this linear model. Using Bayesian methods, inference about the parameters (e.g. to test whether $\alpha = \beta = 0$) can be based on the posterior, $p(\alpha, \beta | Data)$ and forecasting can be carried out on the predictive $p(Y_{T+h}|Data)$ where Y_{T+h} is an out of sample data revision to be forecast. Since the Bayesian approach generates the entire predictive distribution, analysis can utilise point forecasts (eg

[9] The CSO was enlarged in 1996 and re-branded the ONS.

[10] We repeated the analysis of bias reported below using the same definition of revisions in Garratt and Vahey (2006). This analysis confirmed their characterisation of UK revisions, with similar break dates (assuming constant error variance).

$E(Y_{T+h}|Data)$), measures of forecast precision (eg $var(Y_{T+h}|Data)$), probabilities of forecast regions (eg $\Pr(Y_{T+h} > 0|Data)$), or credible intervals (the Bayesian variant of confidence intervals).[11] To assess the likelihood of revisions of a particular magnitude (which might attract press attention) requires the probability of forecast regions.

Recent papers by Swanson and van Dijk (2006) and Garratt and Vahey (2006) have found structural breaks and regime switching to affect the revisions processes and, hence, we study a more general class of models written as:

$$Y_t = \begin{cases} \alpha_1 + \beta_1 X_t + \sigma_1 \varepsilon_t \text{ if } s_t = 1 \\ \alpha_2 + \beta_2 X_t + \sigma_2 \varepsilon_t \text{ if } s_t = 2 \\ \dots \\ \dots \\ \alpha_N + \beta_N X_t + \sigma_N \varepsilon_t \text{ if } s_t = N \end{cases} \tag{2}$$

where ε_t is $N(0,1)$

This class of models allows for multiple breaks in the error variances (and other parameters) which would result from sporadic structural reforms to data reporting procedures. The N different regimes depend upon s_t and this can be defined in various ways. Structural break variants of (2) define:

$$s_t = \begin{cases} 1 \text{ if } t < \tau_1 \\ 2 \text{ if } \tau_1 \leq t < \tau_2 \\ \dots \\ \dots \\ N \text{ if } t > \tau_{N-1} \end{cases} \tag{3}$$

so that structural breaks occur at times $\tau = (\tau_1, .., \tau_{N-1})'$. The break dates can be treated as unknown parameters and estimated from the data. The posterior, $p(\alpha, \beta, \tau|Data)$, (and the predictive) reflect parameter uncertainty about τ, just as for any parameter. In this case, (2) and (3) define a single model: a linear regression model with $N-1$ breaks. Alternatively, we can interpret each configuration of the breakpoints as defining a particular model, in which case (2) and (3) defines a whole class of models. We discuss the implications of the different interpretations below.

Another possible definition of s_t defines a simple regime-switching model with:

[11] Koop (2003, Chapter 3) gives details of the relevant methods and formulae.

$$s_t = \begin{cases} 1 & \text{if } X_t < r \\ 2 & \text{if } X_t \geq r \end{cases} \tag{4}$$

where r is the threshold (treated as an unknown parameter). That is, the revisions process can have different properties depending on whether the first measurement of the variable is above or below a threshold. In this model, the regime shifting depends on the threshold trigger (the first measurement of the variable) and the estimated threshold itself (r). For this reason, we refer to the regime-shifting in this model as endogenous.

In addition, following Swanson and van Dijk (2006) and Castle and Ellis (2002), we investigate the possibility that the revision process varies over the business cycle. Like the NBER for the US, the Economic Cycle Research Institute produces a set of dates for peaks and troughs for UK growth cycles. These are commonly used for empirical research (eg Osborne and Sensier, 2002).[12]We consider a set of models defined by (4) with $s_t = 1$ for periods beginning at (but not including) the trough date through (and including) the peak date, and $s_t = 2$ otherwise. Thus, $s_t = 1$ can be interpreted as defining expansionary periods and $s_t = 2$ contractionary periods. In this case, the regime shifting depends on this business cycle dating variable and, thus, we refer to this sort of regime-shifting as exogenous.

We also experimented with (but do not report results) using the following variants of (1) and (2):

$$Y_t = \alpha + \beta X_t + W_t\gamma + \varepsilon_t,$$

$$Y_t = \begin{cases} \alpha_1 + \beta_1 X_t + W_t\gamma_1 + \sigma_1\varepsilon_t & \text{if } s_t = 1 \\ \alpha_2 + \beta_2 X_t + W_t\gamma_2 + \sigma_2\varepsilon_t & \text{if } s_t = 2 \\ \dots & \\ \dots & \\ \alpha_N + \beta_N X_t + W_t\gamma_N + \sigma_N\varepsilon_t & \text{if } s_t = N \end{cases}$$

where W_t is a vector of explanatory variables containing information available at the same date as the first measurement. Swanson and van Dijk (2006) found that US revisions can be forecast using macroeconomic indicators. For our UK GDP data, various choices for W_t including lags of X_t and first observations of GDP components gave qualitatively similar results to those obtained with (1) and (2) and so are not reported below.

[12] See http://www.businesscycle.com/.

4 Econometric methods

Bayesian methods use the rule of conditional probability to make inferences about unknown things (eg parameters, models) given known things (eg data). So, for instance, if $Data$ is the data and there are m competing models, $M_1, .., M_m$, each characterised by a vector of parameters θ^i for $i = 1, ..., m$, then a Bayesian would use the posterior distribution, $p\left(\theta^i|Data, M_i\right)$, to make inferences about the parameters in a particular model. If z is an unknown data point the researcher wishes to forecast, then the Bayesian would work with the predictive distribution, $p\left(z|Data\right)$. The posterior model probability, $p\left(M_i|Data\right)$, summarizes the information about which model generated the data. Precisely how $p\left(M_i|Data\right), p\left(z|Data\right)$ and $p\left(\theta^i|Data, M_i\right)$ are obtained depends on the empirical context. The logic of Bayesian inference suggests that prediction should involve averaging over both parameter and model space and hence:

$$p\left(z|Data\right) = \sum_{i=1}^{m} \int p\left(z, \theta^i, M_i|Data\right) d\theta^i. \tag{5}$$

Using the rules of probability, this can be written as:

$$\begin{aligned} p\left(z|Data\right) &= \sum_{i=1}^{m} \int p\left(z|Data, \theta^i, M_i\right) p\left(\theta^i|Data, M_i\right) p\left(M_i|Data\right) d\theta^i \quad (6) \\ &= \sum_{i=1}^{m} p\left(M_i|Data\right) \int p\left(z|Data, \theta^i, M_i\right) p\left(\theta^i|Data, M_i\right) d\theta^i. \end{aligned}$$

That is, the predictive density can be obtained using the predictive density in a particular model with given parameters (ie $p\left(z|Data, \theta^i, M_i\right)$), a posterior density for the particular model (ie $p\left(\theta^i|Data, M_i\right)$) and posterior model probabilities (ie $p\left(M_i|Data\right)$ for $i = 1, .., m$) and then integrating out both parameters and models. In this way, the Bayesian framework offers a logical way of treating parameter uncertainty and model uncertainty. The step where the models are integrated out is commonly referred to as Bayesian model averaging.

In order to carry out BMA procedures, we need to evaluate $p\left(M_i|Data\right)$. Using Bayes rule, we write this as:

$$p\left(M_i|Data\right) \propto p\left(Data|M_i\right) p\left(M_i\right), \tag{7}$$

where $p\left(Data|M_i\right)$ denotes the marginal likelihood and $p\left(M_i\right)$ the prior weight attached to this model (ie the prior model probability). For the

Bayesian, both of these quantities require prior information. Given the controversy attached to prior information, $p(M_i)$ is often simply set to the noninformative choice where, *a priori*, each model receives equal weight. Similarly, the Bayesian literature has proposed many benchmark or reference prior approximations to $p(Data|M_i)$ which do not require the researcher to subjectively elicit a prior (see, e.g., Fernandez, Ley and Steel, 2001). Here we use the Schwarz or Bayesian Information Criterion (BIC). Formally, Schwarz (1978) presents an asymptotic approximation to the marginal likelihood of the form:

$$\ln p(Data|M_i) \approx l - \frac{K \ln T}{2}. \tag{8}$$

where l denotes the log of the likelihood function evaluated at the MLE, K denotes the number of parameters in the model, and the sample is of length T. The exponential of (8) provides weights proportional to the posterior model probability used in BMA. The advantage of this choice is that (8) does not require the elicitation of an informative prior, it is familiar to non-Bayesians and it yields results which are closely related to those obtained using many of the benchmark priors used by Bayesians (see Fernandez, Ley and Steel, 2001).

With regards to the prior for the parameters (which enters $p(\theta^i|Data, M_i)$) we use the standard noninformative prior (see, e.g, Koop, 2003, page 38). For models with breakpoints (or thresholds), we also use a noninformative prior which attaches equal weight to every breakpoint (or threshold) value that implies that each regime contains at least 15% of the observations.

With *i.i.d.* Normal errors, it is straightforward to carry out Bayesian inference in all the models discussed in the previous section. That is, all of them are either directly Normal linear regression models or, conditional on breakpoints (thresholds) are Normal linear regression models.[13] Inference about the parameters (e.g. to test whether $\alpha_j = \beta_j = 0$, where $j = 1, ..., N$) is based on the posterior, $p(\alpha_j, \beta_j|Data)$ and forecasting is based on the predictive $p(Y_{T+h}|Data)$ where Y_{T+h} is the out of sample data revision to be forecast.

Although using the BIC allows us to approximate the marginal likelihood, $p(Data|M_i)$, without eliciting a prior for the parameters, the posterior model probability given in (7) requires the choice of a prior model probability, $p(M_i)$. We want to make a noninformative choice for this, but there is some ambiguity about this relating to our treatment of the breakpoints or

[13] For the breakpoint (threshold) models, we approximate the marginal likelihood using (8) for every possible breakpoint (threshold). When breakpoints (thresholds) are treated as parameters, their posteriors are proportional to these marginal likelihoods.

thresholds. To illustrate the basic issue, consider the case with either either zero or one break. Following Garratt and Vahey (2006), we consider every possible breakpoint, τ, that implies each regime contains at least 15% of the observations. If we interpret each possible value for τ as defining a different model then we have $.7T$ models with one break and one model with no breaks. This suggests a noninformative prior over model space is given by:

$$p\left(M_i\right) = \frac{1}{.7T+1}, \tag{9}$$

for $i = 1, .., .7T + 1$. However, one might want to interpret τ as a parameter and that there is only a single one-break model. The latter interpretation suggests the prior model probability should be:

$$p\left(M_i\right) = \frac{1}{2}, \tag{10}$$

for $i = 1, 2$. For BMA the differences between these two approaches can be substantial. The weights in BMA use posterior model probabilities, provided by equation (7). With the prior given in (9), the weights attached to one break models will be proportional to the *sum* of (the exponential of) (8) for every possible breakpoint. With the prior given in (10), the weight attached to the one break model with be proportional to the *average* of (the exponential of) (8) over every possible breakpoint. Arguably, both of these priors are reasonable; we present results using both. Some readers might prefer one approach over the other and can focus on empirical results using that approach. Other readers can be confident that, by covering two extreme approaches, intermediate cases lie somewhere between the two.

In our empirical work, we extend (9) and (10) to allow for many breaks. We consider $N = 1, 2, 3$ and 4 (i.e. allow for zero, one, two or three breaks) and every possible configuration of $\tau_1, .., \tau_{N-1}$ that implies each regime contains at least 15% of the observations. These choices are motivated by Garratt and Vahey (2006) who, using the approach of Bai and Perron (2003), never found more than three breaks and used trimming factors of 15% or larger using the same UK data source (discussed below). For our regime-switching models we consider the exogenous and endogenous threshold triggers specified above in the discussion after (4).

5 The data

The source for the revisions data used in this study is the Bank of England's real-time database for seasonally adjusted real quarterly GDP(E) growth,

from 1961Q3 through 2004Q2 (see Castle and Ellis, 2002). We also provide a comparative analysis of structural breaks, bias and nonlinearities for the GDP(E) components taken from the same database.[14]

The data were published initially by the CSO and its successor, the ONS, in *Economic Trends* and *Economic Trends: Annual Supplement*. Garratt and Vahey (2006) characterised the revision properties of the same GDP(E) series, together with a number of other macroeconomic variables and described this and other real-time UK databases in detail.

The Mitchell Report set out the 2004 timetable for revisions to UK National Accounts. By the end of our sample, revisions to an initial measurement for GDP occurred for the successive two months. So the preliminary release (M1), the second release (M2) and then the third (M3) typically differed (see Statistics Commission, 2004, vol.3 p23-24).[15] The substantial revision to the GDP measurement for 2003Q2, which attracted press hostility, took place with the M3 release. Unfortunately, the timetable of revisions evolved throughout our sample period. For example, prior to 1998 the first measurement for GDP occurs one month later than under the 2004 timetable in the Bank of England database.

For this study, we define the revision as the difference between the initial measurement of the quarterly growth rate of GDP available in the first month in a given quarter and its measurement occurring three months later.[16] This approach standardises the revisions timetable through our sample period and abstracts from the improved timeliness of preliminary GDP measurements. We treat the M1 and M3 releases for the notorious 2003Q2 GDP observation as the first and second measurements, respectively.

Figure 1 plots the first and second measurements of quarterly GDP growth used in our study between 1961Q3 and 2004Q2. The reduced volatility in the first and second measurements post-1989 reflected in part the unprecedented stability of recent economic growth.

Figure 2 plots revisions for the sub-sample 1980Q1 to 2004Q2 to help gauge the recent behaviour of revisions. These were much less volatile after 1990 and, in particular, the period 1998Q1 to 2001Q3 saw fluctuations within a

[14] These data are from http://www.bankofengland.co.uk/statistics/gdpdatabase.

[15] The ONS also revised earlier measurements as well as the most recent one, usually in the M3 release. Each year a "Blue Book" measure adjusted the quarterly data to match annual measure; see Robinson (2005).

[16] As discussed in Section 3, Garratt and Vahey's reported analysis compared preliminary measurements with March 2003 measurements.

tight band, less than 0.2 in absolute value. The six quarters preceding the 2003Q2 had three revisions greater than 0.2 in absolute value; the 2003Q2 revision was the largest since the 1980s.

Garratt and Vahey (2006) characterised UK revisions as typically biased. That is, the regression coefficients of the linear regression model are jointly non-zero. Although they found some support for lower revision variance after 1989, they found no breaks in the linear regression coefficients for GDP, although breaks were found for GDP components. Paterson and Heravi (1991), Symons (2001), Richardson (2003), Akritidis (2003a and 2003b) and George (2005) provide further real-time data analysis of various measures of UK GDP and its components. These studies often considered shorter samples of data than used by Garratt and Vahey (2006). In particular, the recent ONS studies used data from 1993 onwards but did not report tests for structural breaks based on longer samples. However, figure 2 provides little visual evidence of a break in 1993; the early 1990s saw a run of positive revisions from mid-1991 through to end-1994.

Castle and Ellis (2002) reviewed the causes of the UK revisions; more detailed discussion can be found in the Mitchell Report (see Statistics Commission, 2004, vol. 3, p21-27). Revisions occurred when new data arrive, the methodology changed and during re-basing of the National Accounts. The new data category sometimes involved the substitution of delayed survey information for earlier judgement. These revisions are fairly common. According to the Mitchell Report over 50 percent of the data used in the M1 release for 2003Q2 came from forecasts. Changes in methodology were rarer and recent changes of this type have known implementation dates (see Statistics Commission, 2004, vol. 1, p21). Wroe (1993) discussed a number of earlier methodological changes (with unknown implementation dates). Two recent changes (with known timing) stemmed from the switch to the European system of National Accounts in 1998 (for details see Castle and Ellis, 2002) and the switch to annual chain-linking in September 2003 (discussed by Charmokly and Soo, 2003). The (known) re-basing dates prior to that occurred approximately every five years. The impacts of these revisions should be relatively minor for our analysis since we consider quarterly growth rates.

6 Empirical results

We present our empirical results in two sections. The first examines the behaviour of GDP revisions over the period 1961Q3 to 1999Q2, focussing

on structural breaks, bias and regime switching in the revision process. As a diagnostic check on the modelling strategies, we also examine predictive distributions using the out of sample period from 1999Q3 to 2004Q2. The second section evaluates the predictives generated by recursive estimation over the evaluation period 1984Q3 to 2004Q2 and calculates the probability of substantial and smaller revisions.

6.1 Model

Evidence for structural breaks

Tables 1 and 2 present evidence on breaks in the regression coefficients for GDP (assuming no breaks in the error variance), together with analogous results for selected GDP components as benchmarks. Table 1 uses the extension of the prior given in (9) which treats each breakpoint as defining a model; but Table 2 uses the extension of (10) which treats breakpoints as parameters. In Table 1, the column labelled "Best" presents the model with highest BIC, based on a search over all possible breakpoints for $N = 1, 2, 3$ and 4. The remaining columns present the probability for each value of N. The probabilities are summed over all possible breakpoints, so the posterior model probability of the single model with the preferred values for $\tau_1, .., \tau_{N-1}$ will be lower (usually much lower) than the probabilities in these columns. For instance, in Table 1, $p(N = 2|Data)$ and $p(N = 2|Data, \tau_1 = 100)$ are different concepts with the latter being smaller than the former.

Table 1 reveals some uncertainty over how many breaks exist in the regression coefficients. The model without a break has the highest BIC for GDP, and for all the component series except Exports. But the probability of one break in GDP at around 0.4 is similar to that of no break. For Consumption, Government Expenditure, and Imports the probability of 1 break exceeds that of zero breaks. For Exports and Investment the probability of three breaks exceeds 0.2. For every variable there is at least a 0.6 chance that the single selected model is incorrect.

The evidence in favour of no breaks in the regression coefficients of GDP is much stronger in Table 2, where the model with no breaks receives over 0.9 posterior probability. Most of the components exhibit similar probabilities, except Exports where the no break model has a posterior probability of just 0.6.

We stress that the difference between Tables 1 and 2 is to be expected since

prior (10) attaches less weight to higher numbers of breaks. Another way of understanding this issue is to note that the models selected in Table 1 involve searching over every possible breakpoint(s) and choosing the breakpoint(s) which yield most evidence in support of a break(s). A researcher worried about data mining could feel that, by searching over many different models, these methods are bound to find evidence for a break somewhere. In Table 2 the results incorporate the uncertainty about the breakpoints and average over all possible breakpoints (with weights given by the posterior). However, only four different base models are being considered ($N = 1, 2, 3, 4$). Such an approach is less liable to the criticism that apparently significant results are simply due to data mining. Put yet another way, if M_i defines a model and $\tau = (\tau_1, .., \tau_{N-1})'$, then $p(M_i|Data, \tau_{\max})$ is used to select the models in Table 1 (where $\tau_{\max}$ is chosen to maximize the BIC), whereas $p(M_i|Data) = \int p(M_i|Data, \tau)\, p(\tau|Data)\, d\tau$ is used to select the model in Table 2. In our case, the number of values that τ can take on is hugely different for different values of N indicating that the possibilities for data mining varies hugely across $N = 1, 2, 3, 4$. Thus, $p(M_i|Data, \tau_{\max})$ and $p(M_i|Data)$ can be quite different. This issue has arisen in previous work involving nonlinear time series models where Bayesian approaches (which integrate out parameters analogous to τ) tend to find less evidence of nonlinearity than classical approaches (which search over all possible values of parameters analogous to τ). See, e.g., Koop and Potter (1998, 1999, 2001 and 2003).

Table 1: Probability of breaks (homoskedasticity, prior (9))

Variables	Best	No Break $N = 1$	1 Break $N = 2$	2 Breaks $N = 3$	3 Breaks $N = 4$
GDP	No breaks	0.401	0.404	0.164	0.030
Consump.	No breaks	0.134	0.483	0.311	0.072
Invest.	No breaks	0.010	0.085	0.381	0.525
Gov. exp.	No breaks	0.141	0.560	0.264	0.036
Exports	1 break 1991Q2	0.006	0.346	0.422	0.226
Imports	No breaks	0.068	0.294	0.286	0.353

Table 2: Probability of breaks (homoskedasticity, prior (10))

Variables	Best	No Break $N = 1$	1 Break $N = 2$	2 Breaks $N = 3$	3 Breaks $N = 4$
GDP	No breaks	0.991	0.009	0.000	0.000
Consump.	No breaks	0.967	0.033	0.001	0.000
Invest.	No breaks	0.914	0.075	0.010	0.001
Gov. exp.	No breaks	0.964	0.036	0.001	0.000
Exports	No breaks	0.621	0.365	0.013	0.001
Imports	No breaks	0.960	0.038	0.001	0.000

Tables 3 and 4 are comparable to Tables 1 and 2, except that they allow for breaks in variance as well as in the regression coefficients. There is now much stronger evidence for one or more breaks. Given the lack of evidence in favor of breaks in the conditional means provided by Tables 1 and 2, it is clear that the results in Tables 3 and 4 are being driven by breaks in the error variance.[17]

Tables 3 and 4 indicate that GDP growth probably has 3 structural breaks. For the single model with highest BIC (with the breaks dates given in the second column of Table 3), point estimates of the error variances in the four regimes are $0.17, 0.79, 0.11$ and 0.02. Thus, there are large shifts in the error variance and the final regime after 1988Q4 exhibits much lower error variance than earlier regimes. This pattern, of the final regime having an appreciably lower error variance, is repeated for most of the components of GDP. The measurements provided by the CSO/ONS became more accurate over time, particularly since the late 1980s.

Since Table 4 treats τ as an unknown parameter, we report its posterior mean and standard deviation. Since the posterior distribution of the break-points can be multi-modal and non-Normal, a standard deviation is not the best measure of dispersion. Nevertheless, we provide posterior standard deviations to give the reader of some idea for how precisely the breakpoints are estimated. It can be seen that, although many breakpoints are imprecisely estimated, the standard deviation for the third break in GDP growth is just two quarters. Hence the evidence suggest little support for the practice of discarding pre-1993 revisions common in recent ONS studies, discussed in the previous section.

[17] When we allow for breaks in the error variance, any difference in regression coefficients across regimes tends to become more substantial.

Table 3: Probability of breaks (heteroskedasticity, prior (9))

Variable	Best	No Break $N=1$	1 Break $N=2$	2 Breaks $N=3$	3 Breaks $N=4$
GDP	3 breaks 1977Q2 1983Q3 1988Q4	0.000	0.002	0.017	0.981
Consump.	2 breaks 1977Q3 1991Q2	0.000	0.008	0.853	0.139
Invest.	2 breaks 1971Q4 1991Q2	0.000	0.001	0.425	0.575
Gov. exp.	1 break 1992Q4	0.165	0.703	0.123	0.010
Exports	1 break 1972Q3	0.000	0.324	0.528	0.148
Imports	1 break 1972Q3	0.000	0.116	0.096	0.788

Table 4: Probability of breaks (heteroskedasticity, prior (10))

Variable	Best (St. dev.)	No Break $N=1$	1 Break $N=2$	2 Breaks $N=3$	3 Breaks $N=4$
GDP	3 breaks 1976Q4 (7) 1983Q2 (4) 1990Q3 (2)	0.000	0.365	0.105	0.531
Consump.	2 breaks 1979Q4 (16) 1991Q3 (2)	0.035	0.229	0.726	0.011
Invest.	2 breaks 1970Q4 (5) 1992Q3 (4)	0.001	0.035	0.860	0.105
Gov. exp.	No breaks	0.948	0.048	0.004	0.000
Exports	1 break 1971Q4 (6)	0.018	0.935	0.046	0.001
Imports	1 breaks 1972Q2 (5)	0.000	0.959	0.239	0.018

Evidence of bias

Researchers are often interested in whether the revision process is unbiased. That is, whether $\alpha_j = \beta_j = 0$ for $j = 1, .., N$. Our methods also allow us to address the question "what is the probability of unbiasedness at the end of the sample?" by calculating $p(\alpha_N = \beta_N = 0|Data)$ (for the linear model $\alpha_N = \alpha$ and $\beta_N = \beta$). Since this hypothesis does not make sense for the regime-switching models, we present results relating to unbiasedness in the present section only for the linear and structural break models.

Table 5 columns 2 and 4 present the probability of unbiasedness in all regimes, averaged over the linear and all the structural break models using BMA for the two prior specifications.[18] That is, column 2 presents results using the prior given in (9) which treats each breakpoint as a model and the fourth column presents results using the prior given in (10) which treats breakpoints as parameters. Overall, there is evidence of bias in the revisions process. Only Exports and Government Expenditure display weak evidence of unbiased revisions, but even for these variables the probability of unbiasedness never exceeds a half (and is well below ten percent when using the prior which treats breakpoints as unknown parameters).

Columns 3 and 5 display the probabilities of unbiased revision in only the final regime for the two prior specifications. In this case, the probability of unbiased GDP revisions is very close to one, providing further evidence of substantially improved statistical quality after the 1990. This pattern is shared by Consumption and, to a lesser extent, Investment; but final regime coefficients rarely exhibit unbiasedness for the other components of GDP growth.

The international evidence suggests that other countries also exhibit biased revisions. Faust, Rogers and Wright (2005) compared revisions for GDP across the G7 using OECD Main Economic Indicators data. They found significant downward bias for Germany, Italy, Japan and UK and less bias for Canada and the US.[19] Faust, Rogers and Wright (2005) reported strong evidence of unbiasedness for the UK when the sample was restricted to post-1988 data.

[18] As with all results in this paper, BICs are used to construct the posterior model probabilities appearing in this table (see equation 8).

[19] See also the study by Croushore and Stark (2001) based on US National Accounts.

Table 5: Probability of unbiased revisions

	Prior (9)		Prior (10)	
Variables	All	Final regime	All	Final regime
GDP	0.000	0.982	0.000	1.000
Consump.	0.043	0.987	0.044	0.998
Invest.	0.010	0.915	0.003	0.015
Gov. Exp.	0.272	0.003	0.031	0.013
Exports	0.416	0.000	0.057	0.000
Imports	0.000	0.000	0.000	0.000

Evidence for regime switching

Here we present the results examining exogenous and endogenous regime switching in the revision process. Recall that exogenous switching uses the ECRI growth cycle dates for the UK and endogenous switching follows the form outlined in equation (4) section 3.[20]

Table 6 presents Bayes factors comparing a regime switching model to the linear model. Values of a Bayes factor greater than one indicates the regime switching model receives more support than the linear model. For the exogenous regime switching models, the regime switching model receives less support than the linear model for GDP and its components, with and without heteroskedasticity. Therefore, we do not consider further any regime switching of this type.

For the endogenous regime-switching model, analogous to the discussion surrounding (9) and (10), we can treat various values for r as indicating individual models or r can be treated as a parameter integrated out when calculating the posterior model probability in (7). In table 6, the first number reported in each cell treats each threshold value as defining a model, the second averages over threshold values, and the estimated threshold value is shown in parentheses.[21] The evidence for regime switching continues to be weak, with Bayes factors below one for all variables except GDP. The Bayes factors are large for GDP (the exception being the homoskedastic case which averages over thresholds) providing strong evidence for regime switching in

[20] In particular, $s_t = 1$ for the periods 1962Q2-1963Q3, 1966Q3-1968Q1, 1971Q2-1973Q1, 1975Q3-1976Q3, 1977Q3-1979Q2, 1980Q3-1983Q4, 1984Q4-1985Q2, 1986Q1-1988Q1, 1991Q3-1994Q3, 1995Q3-1997Q3 and 1999Q2; elsewhere, $s_t = 2$.

[21] As with our structural break models, we use a flat prior over every possible threshold, r, such that each regime contains at least 15% of the observations.

the error variance. In particular, for very low first measurements of GDP growth (-1%), we get a high error variance; and for high values of the GDP growth, we get a low error variance. Deep recessions are associated with less accurate GDP growth data. Given this support for endogenous regime switching we include this form of model as one of the models in our BMA forecasting exercise.

Table 6: Bayes factors for linear and regime-switching models

Variables	Exogenous regimes		Endogenous regimes	
	Homo-skedastic	Hetero-skedastic	Homoskedastic	Heteroskedatic
GDP	0.013	0.013	34.95 0.870 (thrsh = −1.040)	45331 922.5 (thrsh = −1.040)
Consump.	0.008	0.001	0.219 0.014 (thrsh = −0.607)	0.181 0.009 (thrsh = 1.620)
Invest.	0.014	0.001	0.088 0.025 (thrsh = 1.230)	0.043 0.007 (thrsh = −2.830)
Gov. Exp.	0.018	0.002	0.371 0.105 (thrsh = 0.389)	0.066 0.014 (thrsh = 0.178)
Exports	0.019	0.491	0.185 0.024 (thrsh = −1.970)	0.008 0.004 (thrsh = 3.080)
Imports	0.013	0.025	0.088 0.041 (thrsh = 1.130)	0.077 0.011 (thrsh = 0.913)

Notes: For the endogenous regimes the first number reported treats each threshold value as defining a model, the second averages over threshold values and the estimated threshold value is shown in parentheses.

Predictive features of interest

So far we have focussed on presenting evidence on which (if any) of the models considered are supported by the data. The results suggest a large degree of model uncertainty which may be important for forecasting. Hence, as a precursor to our main empirical exercise of predicting substantial revisions (in section 6.2), we look at the basic properties of the predictives for the out of sample period 1999Q3 to 2004Q2, using models estimated on data 1961Q3 to 1999Q2. For the sake of brevity, we present only results using the model space prior given in (10). That is, we are treating breakpoints and thresholds as parameters. Furthermore, we will focus on GDP growth as it is the most important of our variables. Noting that data through 1999Q2 was used to

produce our previous results, we have 20 out of sample data points to use in our forecasting exercise.

We begin by presenting the actual value, predictive mean and standard deviation using Bayesian model averaging and, for comparison purposes, the traditional Linear model given in (1), together with BIC-selected "Best" model. In this case, the Best model has three structural breaks (timed at 1977Q2, 1983Q3 and 1988Q4) with breaks in the regression coefficients and the error variance. All predictive results are obtained using the standard non-informative prior (see for example Koop, 2003 p38 and p45-46). Of course, in models such as (1) and (2), prediction requires a value of the explanatory variable, for which we use the observed first measurement.

In Table 7, it can be seen clearly that BMA produces a more reasonable predictive distribution than the traditional Linear model. For instance, BMA predictive means are within two standard deviations of the actual revision in 19 out of 20 cases (exactly what a Normal approximation to the predictive suggests should happen). In contrast, the predictive standard deviation under the Linear model is approximately three and a half times as large. BMA gives a great deal of weight to models with structural breaks in the error variance that the Linear model misses. The smaller BMA error variance is then reflected in the much less dispersed predictive distribution. The Best model yields predictives similar to BMA using post-1999Q2 data. As the next section shows, however, differences arise at times when the model parameters are recursively estimated.

Table 7: Properties of the predictive distribution

	Actual revision	**BMA**		**Linear**		**Best**	
		Mean	St. dev.	Mean	St. dev.	Mean	St. dev.
1999Q3	−0.182	0.034	0.140	−0.012	0.483	0.050	0.145
1999Q4	−0.002	0.033	0.139	−0.001	0.483	0.045	0.145
2000Q1	0.086	0.029	0.137	0.042	0.483	0.027	0.143
2000Q2	0.088	0.037	0.139	−0.009	0.483	0.049	0.145
2000Q3	−0.002	0.032	0.138	0.013	0.483	0.040	0.144
2000Q4	0.086	0.027	0.138	0.064	0.483	0.018	0.143
2001Q1	0.086	0.028	0.137	0.054	0.483	0.022	0.143
2001Q2	0.170	0.027	0.138	0.064	0.483	0.017	0.143
2001Q3	−0.172	0.031	0.138	0.025	0.483	0.034	0.144
2001Q4	−0.255	0.027	0.138	0.065	0.483	0.017	0.143
2002Q1	0.045	0.025	0.139	0.083	0.483	0.010	0.143
2002Q2	−0.267	0.034	0.140	−0.011	0.483	0.050	0.145
2002Q3	0.240	0.032	0.138	0.013	0.483	0.040	0.144
2002Q4	−0.019	0.029	0.137	0.048	0.483	0.025	0.143
2003Q1	−0.111	0.026	0.138	0.071	0.483	0.015	0.143
2003Q2	0.305	0.027	0.138	0.059	0.483	0.020	0.143
2003Q3	0.210	0.031	0.138	0.024	0.483	0.035	0.144
2003Q4	0.014	0.034	0.140	−0.011	0.483	0.050	0.145
2004Q1	0.125	0.031	0.138	0.024	0.483	0.035	0.144
2004Q2	0.000	0.034	0.140	−0.011	0.483	0.050	0.145

6.2 Predicting substantial revisions

Since policymakers, statistical agencies and the press are interested in the probability of substantial GDP revisions in real time, we recursively estimate the models using data for 1961Q3 through to period t, where t=1984Q2, ..,2004Q1. For each of the 80 recursions, we calculate three one-step ahead probability events, $p(|Y_{T+1}| > a|\ \Omega)$ where Ω denotes information available at the time of the first release of GDP growth and $a = 0.05, 0.1$ and 0.3. The last of these thresholds matches our definition of a substantial revision. Since we average over all the models described (including linear, structural break and regime-switching models), and integrate out the parameters, we provide a formal treatment of model and parameter uncertainty.

Figures 3 to 5 display the probabilities of interest for BMA, the Best model

and the Linear model. Figure 3 gives the probability of revisions greater than 0.3 in absolute magnitude. Between 1986 and 1994, the three models predict probabilities between 0.5 and 0.7, with the BMA approach typically giving lower values than the Linear model or the Best model. Hence the BMA method highlights the risk that the other two models overstate the probability of substantial revisions early in the evaluation period. After 1998, the BMA and Best models indicate that substantial revisions are much less likely. The posterior probability of a substantial revision fell sharply between 1994Q1 and 1995Q2, before levelling out at around 0.05 but remained much higher for the Linear model, reaching around 0.5 by the end of the evaluation period. There is some evidence that the probability of substantial revisions has increased slightly since 2001Q2 for the BMA and Best models. Recall from figure 2 that a number of revisions greater than 0.2 in absolute value occurred just before the notorious 2003Q2 substantial revision.

Figures 4 and 5 provide information from the same posterior densities, evaluated at different revision events. Although the improvement in statistical quality is much less marked when the threshold value on the revision is 0.1 and 0.05, the timing of this transformation matches the previous plot.

The previous discussion compared three different approaches to forecasting revisions in GDP growth (i.e. using the Linear model, using a single Best model and BMA) to each other. However, one might also be interested in some general measure of forecast performance. Remember that we have calculated $p\left(|Y_{t+1}| > a|\Omega_t\right)$ for 80 periods, 1984Q3 to 2004Q2. Suppose we define a "correct forecast" as one where $p\left(|Y_{t+1}| \leq a|\Omega_t\right) > 0.5$ and the observed revision is less than a or $p\left(|Y_{t+1}| > a|\Omega_t\right) > 0.5$ and the observed revision is greater than a. For $a = 0.05$ and 0.1, our three approaches exhibit similar forecasting performance (i.e. roughly 79% of correct forecasts for $a = 0.05$ and roughly 60-61% of correct forecasts for $a = 0.10$). However, for the third probability event (i.e. $a = 0.30$), the substantial revision, we observe a high incidence of correct forecasts (a high "hit rate") using the BMA and Best model approach, of 74% and 69%, respectively. The Linear model, has a very low hit rate of only 19%.

A more formal measure of forecast performance is the Pesaran and Timmermann (1992) directional market timing statistic, PT. This also suggests differences between the BMA and Best model approaches as compared to the Linear model. The PT statistic allows a formal hypothesis test of directional forecasting performance. As shown in Granger and Pesaran (2000), this hypothesis test uses the same information as the Kuipers score which measures the proportion of above mean growth rates that were correctly fore-

cast minus the proportion of below mean growth rates that were incorrectly forecast. Under the null hypothesis that the forecasts and realisations are independently distributed the PT statistic has a standard normal distribution. For the substantial revision event, $p\left(|Y_{t+1}| > 0.3|\Omega_t\right)$, the data rejects the null of no ability to forecast observed changes, with values of 4.119 and 3.968 for the average and Best models, but is undefined for the Linear model suggesting a poor forecasting performance for the Linear model.

Thus, a strong message coming out of our analysis is that simply working with a Linear model yields misleading results. A second, slightly weaker message is that BMA offers some advantages over the strategy of simply searching over a wide set of models and choosing the single best model.

7 Conclusions

In this paper, we have shown that the probability of substantial revisions to UK GDP growth fell sharply during the mid 1990s, primarily as a result of structural breaks in the error variance of revisions. We calculate that the probability of a revision similar to (or more than) the absolute magnitude of the 2003Q2 revision was around 1:20 in 2003. Using a wide set of models, including linear and nonlinear regression models with and without heteroskedasticity, we adopted a noninformative-prior Bayesian approach to produce the predictive distributions and forecasts of interests. In contrast, earlier classical econometric studies of revisions neglected formal analysis of model uncertainty and structural breaks in the error variances. That approach yields misleading predictives for our UK revisions data.

References

Akritidis, L (2003a), 'Revisions to quarterly GDP growth', *Economic Trends*, 594, 94-101.

Akritidis, L (2003b), 'Revisions to quarterly GDP growth and expenditure components', *Economic Trends*, 601, 69-85.

Aruoba, B (2005), 'Data revisions are not well behaved', mimeo, CIRANO Workshop on Data Revisions, October 2005, Montreal.

Bai, J and P Perron (2003), 'Computation and analysis of multiple structural change models', *Journal of Applied Econometrics*, 18, 1-22.

Castle, J and C Ellis (2002), 'Building a real-time database for GDP(E)', *Bank of England Quarterly Bulletin*, February, 42-49.

Charmokly, Z and A Soo (2003), 'The application of annual chain-linking to the Gross National Income System', *Economic Trends*, 593, 41-47.

Croushore, D (2005), 'An evaluation of inflation forecasts from surveys using real-time data', mimeo, CIRANO Workshop on Data Revisions, October 2005, Montreal.

Croushore, D and T Stark (2001), 'A real-time data set for macroeconomists', *Journal of Econometrics*, 105, 111-130.

Diebold, F and G Rudebusch (1991), 'Forecasting output with the composite leading index: a real-time analysis', *Journal of the American Statistical Association*, 86, 603-610 .

Egginton, D, A Pick, and S Vahey (2002), ' 'Keep it real!' A real-time UK macro data set' , *Economics Letters*, 77, 15-20.

Faust, J, J Rogers, and J Wright (2005), 'News and noise in G7 GDP announcements', *Journal of Money, Credit and Banking*, 37, 403-420.

Fernandez, C, E Ley and M Steel (2001), 'Benchmark priors for Bayesian model averaging', *Journal of Econometrics*, 100, 381-427.

Garratt, A and S Vahey (2006), 'UK real-time data characteristics', *Economic Journal*, forthcoming.

George, E (2005), 'Revisions to quarterly GDP growth and its production and expenditure components, *Economic Trends*, 614, 30-39.

Granger, C and M H Pesaran (2000), 'Economic and statistical measures of forecast accuracy', *Journal of Forecasting*, 19, 537-560.

HM Treasury (1998), *Statistics A Matter of Trust: A Consultation Document*, The Stationery Office, London.

Jenkinson, G and E George (2005), 'Publications of revision triangles on the National Statistics website', *Economic Trends*, 614, 43-44.

Koop, G (2003), *Bayesian Econometrics*, John Wiley and Sons, Chichester.

Koop, G and S Potter (2003), 'Bayesian analysis of endogenous delay threshold models', *Journal of Business and Economic Statistics*, 21, 93-103.

Koop, G and S Potter (2001), 'Are parent findings of nonlinearity due to structural instability in economic time series?', *The Econometrics Journal*, 4, 37-55.

Koop, G and S Potter (1999), 'Dynamic asymmetries in U.S. unemployment', *Journal of Business and Economic Statistics*, 17, 298-312.

Koop, G and S Potter (1998), 'Bayes factors and nonlinearity: evidence from economic time series', *Journal of Econometrics*, 88, 251-281.

Lawson, N (1992), *The View From No. 11: Memoirs of a Tory Radical*, Bantam Press, London.

Mankiw, N G, D Runkle, and M Shapiro (1984), 'Are preliminary announcements of the money stock rational forecasts', *Journal of Monetary Economics*, 14, 15-27.

National Statistics (2005), *Code of Practice; Protocol on Revisions*, The Stationery Office, London.

Office for National Statistics. *Economic Trends*, various issues, The Stationery Office, London.

Office for National Statistics. *Economic Trends Annual Supplement*, various issues, The Stationery Office, London.

Osborne, D and M Sensier (2002), 'The prediction of business cycle phases: financial variables and international linkages' *National Institute Economic Review*, October.

Pesaran, M H and A Timmermann (1992), 'A simple nonparametric test of predictive performance', *Journal of Business and Economic Statistics*, 10, 461-465.

Patterson, K and S M Heravi (1991), 'Data revisions and the expenditure components of GDP', *Economic Journal*, 101, 887-901.

Richardson, C (2003), 'A time series approach to revisions', *Economic Trends*, 601, 86-89.

Robinson, H (2005), 'Revisions to quarterly GDP growth and its production (output), expenditure and income components' *Economic Trends,* 625, 34-49.

Schwarz, G (1978), 'Estimating the dimension of a model', *Annals of Statistics*, 6, 461-464.

Statistical Commission (2004), *Revisions to Economic Statistics*, Report no 17, Vol. 1,2 and 3, April. The Mitchell Report.

Statistical Commission (2005), *Official Statistics: Perceptions and Trust*, Report no. 24, London.

Swanson, N 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.

Symons, P (2001), 'Revisions analysis of initial estimates of annual constant price GDP and its components', *Economic Trends*, 568, 48-65.

Wroe, D (1993), 'Improving macro-economic statistics', *Economic Trends*, 471, 191-199.

Figure 1: First and Second GDP Quarterly Growth Measurements, 1961Q3 - 2004Q2

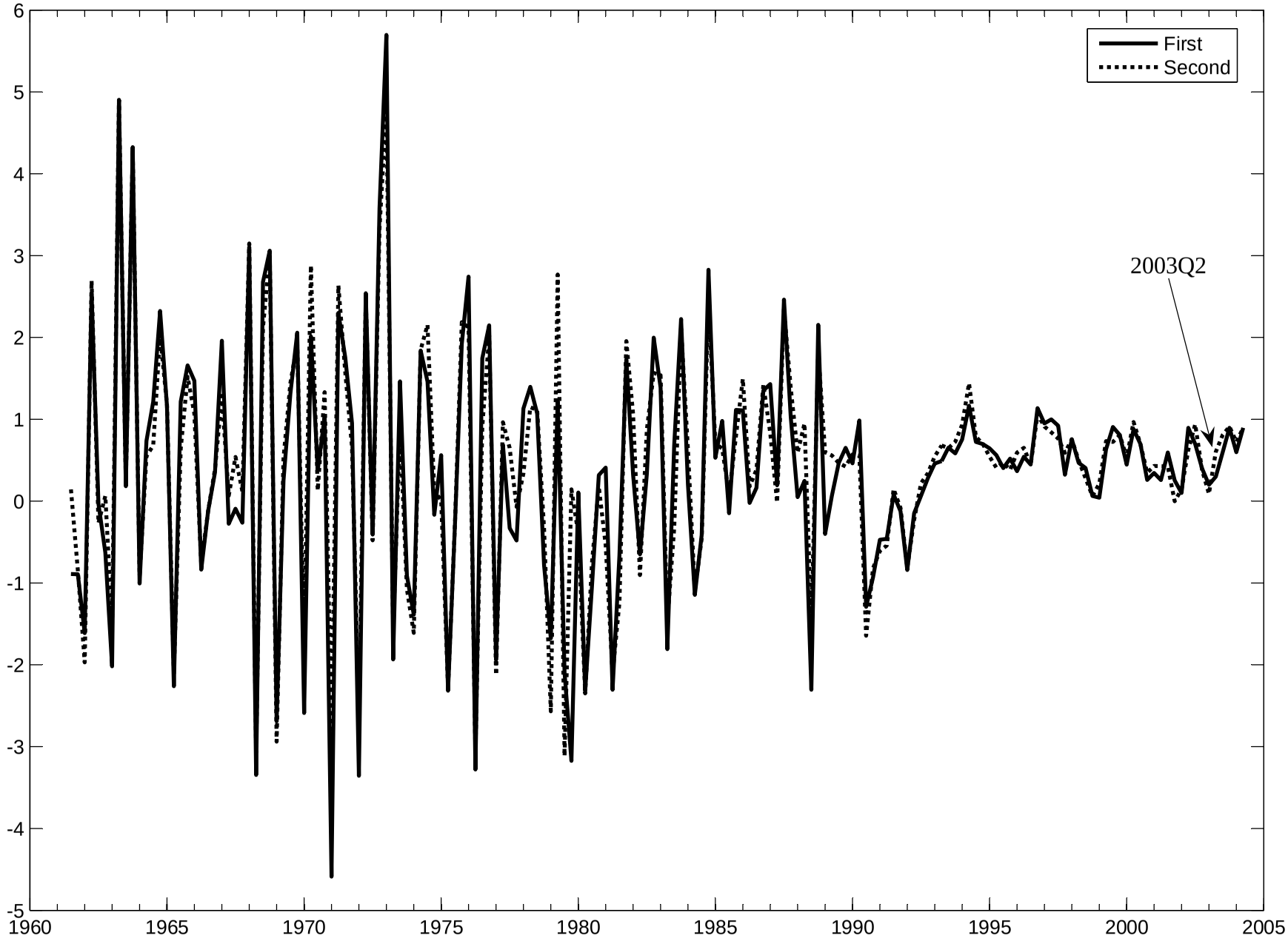

Figure 2: GDP Quarterly Growth Revisions, 1980Q1 - 2004Q2

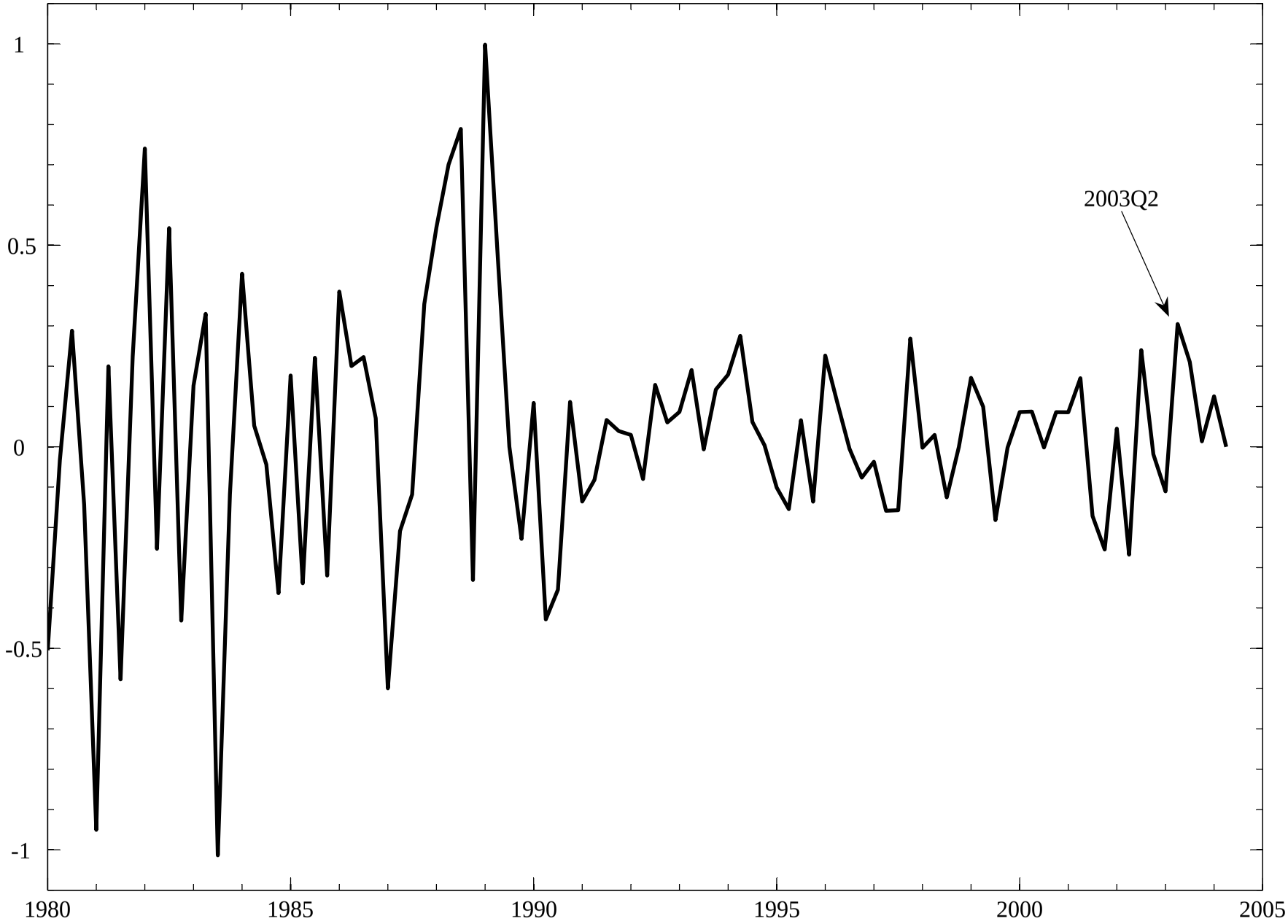

Figure 3: Probability of Absolute Revision Greater Than 0.3, $p(|Y_{T+1}|>0.3)$

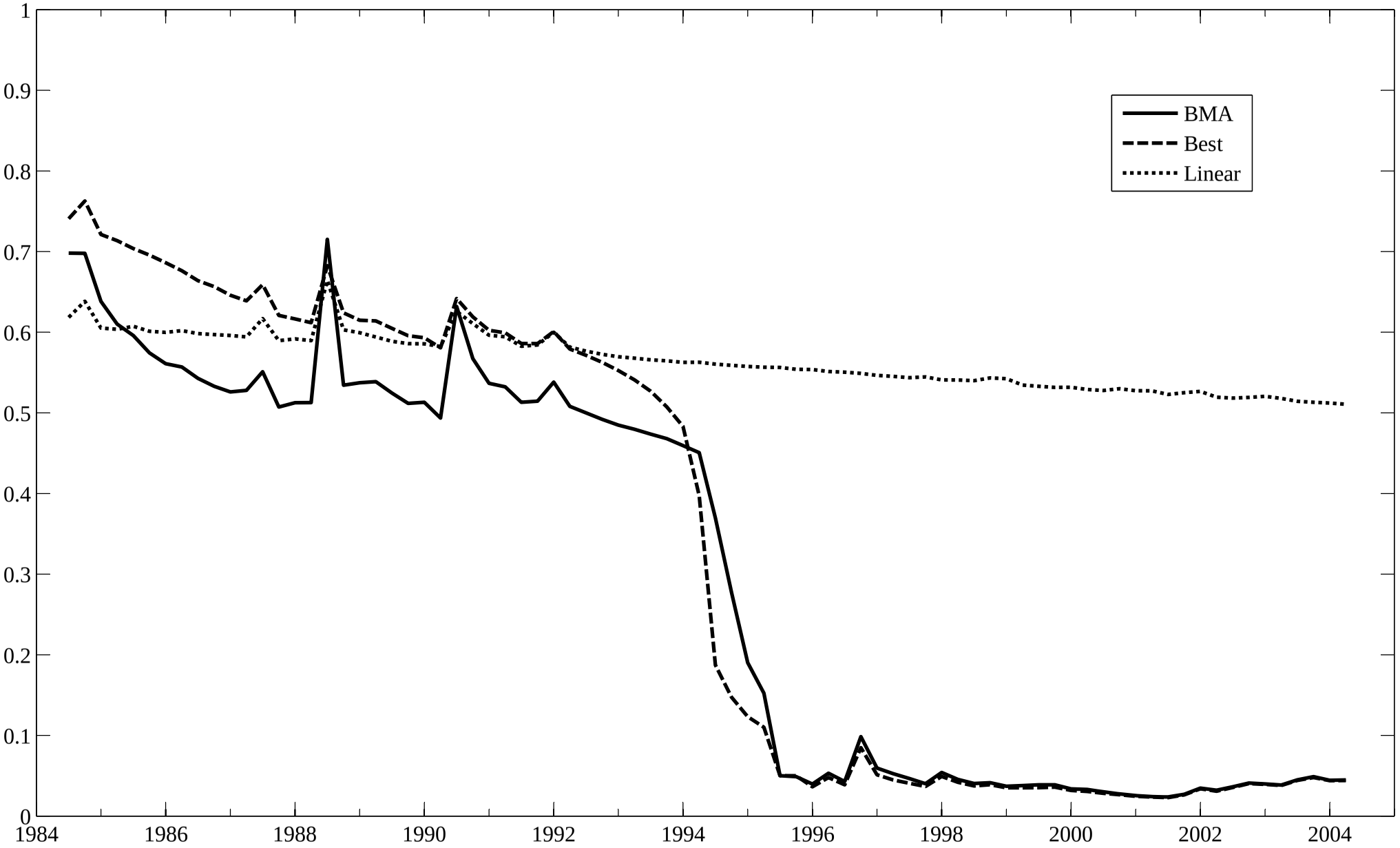

Figure 4: Probability of Absolute Revision Greater Than 0.1, $p(|Y_{T+1}|>0.1)$

Figure 5: Probability of Absolute Revision Greater Than 0.05, $p(|Y_{T+1}|>0.05)$