COUNTING BIASED FORECASTERS: AN APPLICATION OF MULTIPLE TESTING TECHNIQUES

Fabiana Gomez
David Pacini

Discussion Paper 15 / 661

27 May 2015

Department of Economics
University of Bristol
8 Woodland Road
Bristol BS8 1TN
United Kingdom

Counting Biased Forecasters: An Application of Multiple Testing Techniques

Fabiana Gomez,

University of Bristol

and

David Pacini[*],

University of Bristol

January 24, 2015

Abstract

We investigate the problem of counting biased forecasters among a group of unbiased and biased forecasters of macroeconomic variables. The innovation is to implement a procedure controlling for the expected proportion of unbiased forecasters that could be erroneously classified as biased (i.e., the false discovery rate). Monte Carlo exercises illustrate the relevance of controlling the false discovery rate in this context. Using data from the Survey of Professional Forecasters, we find that up to 7 out of 10 forecasters classified as biased by a procedure not controlling the false discovery rate may actually be unbiased.

JEL Classification: C12, C23, E17

Keywords: Biased Forecasters, Multiple Testing, False Discovery Rate.

[*]Corresponding Address: David.Pacini@bristol.ac.uk (E-mail); +44 (0) 11 79 28 84 37 (Phone); Department of Economics, University of Bristol, 8 Woodland Road, BS8 1TN, Bristol, United Kingdom.

1. Introduction

A large body of empirical literature evaluating forecasters' preferences focuses on counting biased forecasters among a group of biased and unbiased forecasters (see e.g., Figlewski and Wachtel, 1981; Brown and Maital, 1981; Zarnowitz, 1985; Keane and Runkle, 1990; Batchelor and Dua, 1991; Davies and Lahiri, 1995; Bonham and Cohen, 2001; Lim, 2001; Schuh, 2001; Elliot, Komunjer and Timmermann, 2008; Boero, Smith and Wallis, 2008; Capistran and Timmermann, 2009). The main difficulty when counting biased forecasters is that, since we do not observe the actual but the estimated bias of each forecaster, we can erroneously classify an unbiased forecaster as biased (and viceversa). One could attack this difficulty by using the estimated bias to calculate for each forecaster a p-value associated to the null hypothesis of unbiasedness and then, after choosing a significance level, to count the number of forecasters with a p-value below the pre-specified significance level. This procedure, called here for conciseness the *uncorrected procedure*, does not control however for the proportion of unbiased forecasters that can be erroneously classified as biased (i.e, the false discovery rate). This lack of control of the false discovery rate leads to wrong conclusions about the actual number of biased forecasters (see the Monte Carlo exercises in the text) because it overcounts the number of biased forecasters leading From the literature on multiple testing (see e.g., Efron, 2010), we know that an improvement in this sense can be obtained by applying multiple testing techniques controlling for the false discovery rate. Although there are potential gains from using these techniques, after reviewing the literature it appears that multiple testing has not been applied yet to the problem of counting the number of biased forecasters.

Motivated by the previous situation, the purpose of this paper is to implement and compare different multiple testing techniques controlling the false discovery rate to count the number of biased forecasters. These techniques offer the possibility to keep the expected proportion of unbiased forecasters that can be erroneously classified as biased controlled up to a convenient pre-specified threshold. The techniques, pioneered by Benjamini and Hochberg (1995), are easy to implement and they involve to compare ranked p-values associated to the null hypothesis of unbiasedness. It is worth mentioning that these techniques apply not only to the problem of counting biased forecasters but to any problem involving to count the number of rejected null hypotheses among a large number of multiple null hypotheses (see e.g., Stekler, 1987; Batchelor, 1990; D'Agostino, McQuinn and Whelan, 2012).

We illustrate the gains of controlling the false discovery rate in the context of counting biased forecasters by implementing multiple testing techniques to simulated and real data. The exercises on simulated data show that, when the number of forecasters being evaluated is large (say 50 or more), the multiple testing procedures have a better ability than the uncorrected procedure to detect biased forecasters while keeping the false discovery rate under control. The exercise with real data demonstrates the empirical relevance of controlling for false discoveries. We analyze data from the Survey of Professional Forecasters for real output forecasts. We find that, when controlling for the false discovery rate by employing the Benjamini-Hochberg procedure, up to 7 out of 10 forecasters classified as biased by the uncorrected procedure may be actually unbiased (i.e., false discoveries).

The rest of this paper is organized as follows. In the next section, we review the existing literature on counting biased forecasters. In Section 3, we describe the available procedures to count biased forecasters and discuss whether they control for false discoveries. In Section 4, we implement Monte Carlo exercises to illustrate the gains from using multiple testing procedures controlling the false discovery rate. In Section 5, we implement these procedures using data from the Survey of Professional Forecasters. Finally, Section 6 concludes.

2. Related Literature

The econometric literature on counting biased forecasters among a group of forecasters dates back to the work by Theil (1958) and Mincer and Zarnowitz (1969). Since these seminal papers, the literature has explored several dimensions of this problem: (i) testing procedure and error rate control; (ii) series, (iii) forecast horizon, (iv) data vintage, (v) subsample of forecasters. This paper intends to contribute in the first dimension. Following the literature, we implement a testing procedure based on a regression approach (see e.g., Zarnowitz, 1985; Schuh, 2001). In this context, the existing literature has either not controlled at all for false discoveries or has controlled for the probability of erroneously classifying as biased one or more genuine unbiased forecasters (i.e., the so-called family-wise error rate) using the so-called *Bonferroni correction procedure*.[1] Not controlling at all for false discoveries is not an attractive procedure because leads to overcount the number of biased forecasters. The Bonferroni correction procedure satisfactorily controls the family-wise error rate (see e.g., Efron, 2010; Romano, Shaikh and Wolf, 2010), but we can expect its ability to detect genuine biased forecasters to be very low when the number of forecasters being evaluated is large, say more than 50 (see the Monte Carlo exercises in the text). In this sense, controlling for the false discovery rate instead of the family-wise error rate, as we propose in this paper, it is of interest to avoid wiping out evidence in favor of biased forecasters.

The second and third dimensions of the problem of counting the number of biased forecasters are the series of forecasts to be explored and the forecast horizons. The literature finds some differences in the estimated number of biased forecasters across series, between univariate and multivariate forecasts, and forecast horizons (Brown and Maital, 1981; Zarnowitz, 1985; Komunjer and Owyang, 2012). Though the techniques in this paper applies to any series of forecasts and

[1] The Bonferroni correction procedure consists in counting the number of forecasters with a p-value below the threshold level resulting from dividing the pre-specified significance level by the total number of forecasters. In the text, we discuss this procedure in detail.

forecast horizons, for the sake of simplicity, we focus on the case of a univariate forecast. We choose the series for gross domestic product, which has been widely used in other papers (see Elliot et al., 2008).

Finally, the fourth and fifth dimensions of the problem are the data vintage and the subsample of forecasters. On one hand, many macroeconomic series are subject to revisions and forecasting performance can differ depending on whether real time or final data vintages are used (see Keane and Runkle, 1990). On other hand, there is evidence indicating that the number of biased forecasters may change over time, depending on the starting and ending date of the subsample (Croushore, 2010). To evaluate the sensitivity of our results to the choice of the data vintage, we measure "actual" values of the target variable using both the real time data (second revision) and the most recent vintage. To evaluate the sensitivity of our results to the choice of the subsample, we consider three subsamples of forecasters: forecasters observed during the period 1968:4-1979:1 (as in Zarnowitz, 1985), forecasters observed during the period 1981:3-2006:3 (as in Elliot et al., 2008), and forecasters observed during the period 1981:3-2012:4. We show that the relevance of our result does not hinge upon the data vintage or the subsample of forecasters.

3. Controlling for False Discoveries

In this section we formally describe the problem of counting biased forecasters. Next, we explain the problem generated by false discoveries. Finally, we describe different procedures to count biased forecasters, stress the importance of controlling for false discoveries and compare the accuracy of these procedures.

3.1 The Setup

Following the empirical literature evaluating forecasters' preferences (Zarnowitz, 1985; Elliot et al., 2008), our objective is to count the number of biased forecasters among a group of n biased and unbiased forecasters. Since we observe the target variable and the forecasts but we do not observe the bias of each forecaster, this objective raises two concerns. The first concern is to determine whether there are biased forecasters at all. If the answer to this question is negative, nothing more is required: the finding is that none of the forecasters is biased. If the answer is positive, a further concern comes up: How many are the biased forecasters? One possibility to address the first concern is to test the single null hypothesis that all the forecasters are unbiased. The second concern, however, requires more elaboration, as it is clear that testing a single hypothesis about all the forecasters is not going to answer how many the biased forecasters are. We next frame these concerns in the context of a multiple testing problem.

Formally, for $i = 1, .., n$ indexing forecasters and $t = 1, .., T$ indexing periods of time, let y_t denote the target value at period t of an economic variable, and let x_{it} denote the predicted value by forecaster i of that variable. The variable y_t might be, for instance, the gross domestic product (GDP) growth rate at quarter t, and x_{it} might the growth rate predicted by a Central Bank. If forecaster i is unbiased we must have:

$$\mathbb{E}(y_t) = \mathbb{E}(x_{it}) \tag{1}$$

For a given forecaster i, following the literature (see e.g., Zarnowitz, 1985; Elliot et al., 2008), a test of unbiasedness can be performed by testing in the regression model

$$y_t = \alpha_i + \beta_i x_{it} + u_{it}; \quad \mathbb{E}(u_{it}|x_{it}) = 0 \tag{2}$$

the null hypothesis $H_{0i} : \alpha_i = 0$ and $\beta_i = 1$ on the basis of a dataset $\{y_t, x_{it}\}_{i=1,t=1}^{n,T}$.[2] Following the recent literature (see e.g., Elliot et. al., 2008), we view the null hypothesis H_{0i} as a statement about unbiasedness but not necessarily for rationality.

[2]This regression-based approach is not the only approach to test unbiasedness. Schuh (2001) discusses an alternative procedure. As already suggested, the techniques reviewed below would apply to any test of unbiasedness (or efficiency) delivering a p-value for each forecaster.

To tackle our first concern, namely to determine whether there are biased forecasters at all, we may test the single null hypothesis

$$H_0^s : \alpha_i = 0, \beta_i = 1 \text{ for all } i \tag{3}$$

using the corresponding F-test. Implementing this test is straightforward, thus it is not be discussed.

If the single null hypothesis H_0^s is rejected (i.e. if we find that there are biased forecasters), the next step is to tackle the second concern, namely, to count the number of biased forecasters. To do so, we may test the multiple null hypotheses

$$H_{01} : \alpha_1 = 0, \beta_1 = 1 \ ; ... \ H_{0i} : \alpha_i = 0, \beta_i = 1 \ ; ...; \ H_{0n} : \alpha_n = 0, \beta_n = 1, \tag{4}$$

and count the number of rejected null hypotheses. For the sake of presentation, we assume that some decision rule produces a statistic z_i for each forecaster i, which will determine a decision of *unbiased* or *biased* for each of the n forecasters. For example, in the application below, z_i is the F-statistic associated to the null hypothesis H_{0i}. We denote the ordered p-values associated to the z_i statistics by

$$p_{(1)} \leq p_{(2)} \leq ... \leq p_{(n)} \tag{5}$$

with their corresponding null hypotheses labeled accordingly: $H_{(1)}, H_{(2)}, ..., H_{(n)}$.

The existing literature has treated the two concerns raised above using the same procedures. We next explain why tackling the problem of testing a single null hypothesis like H_0^s or H_{0i} and the problem testing the multiple null hypotheses H_{01},..,H_{0n} may benefit from using different procedures.

3.2 The Problem of Multiple Testing and False Discoveries

The standard procedure to test a single null hypothesis like H_{0i} is to compute a test statistic z_i and its associated p-value, to choose a significance level γ, and to reject H_{0i} if the p-value is below that significance level γ. The straightforward procedure to test the multiple null hypotheses H_{01},..,H_{0n} is to repeat the procedure for a single null hypothesis over the set of multiple hypotheses. This amount to test each single hypothesis in (4) at a chosen significant level γ, to reject the null hypothesis of unbiasedness for all forecaster i such that $p_{(i)} < \gamma$, and then to count the number of rejected null hypotheses. As already said, we call this procedure the *uncorrected procedure.* Although the uncorrected procedure is easy to implement, it does not control for the possibility of erroneously classifying an unbiased forecaster as biased. Such a problem does not arise when testing the single null hypothesis H_0^s, and it is a case of the so-called *multiple testing problem* (see Lehmann and Romano, 2005, Chapter 9).

To illustrate the multiple testing problem, Figure I presents a simplified Gaussian hypothetical example with $T = 52$ time periods. We suppose that the n forecasters are either unbiased (i.e. $\alpha_i = 0$ $\beta_i = 1$) or upward biased (i.e. $\alpha_i = 1$ and $\beta_i = 1$) with probability $\pi_0 = .7$ or $1 - \pi_0 = .3$, respectively. We wish to test simultaneously (not jointly) the null hypotheses in (4). In this hypothetical example, the F-statistic associated to the ordinary least squares estimator of (α_i, β_i) has either F-density with $(2, 50)$-degrees of freedom (when the null hypothesis of unbiasedness is valid) or a non-central F-density with $(2, 50)$-degrees of freedom and non-centrality parameter equal to 4 (when the null hypothesis of unbiasedness is not valid). These two densities are shown in Figure I, Panel A. The density shown in panel B is the cross-section density that (hypothetically) would be observed by a researcher. This density is a mixture of the two densities in Panel A, where the weight on each density is equal to the proportion of unbiased and biased forecasters. Suppose that we choose a significance level, γ, of 10%. With the uncorrected procedure described above, the researcher would expect to find 10% percent of the forecasters with a positive and significant F-statistic. This proportion is represented by the shaded region in the right tail of the cross-sectional density (Panel B). Does this area consist of only biased forecasters? Clearly not, because some forecasters are just unlucky; as shown in the shaded region of the right tail of Panel A, unbiased forecasters can exhibit positive and significant estimated F-statistics (the shaded region in the left tail of Panel B).

INSERT FIGURE I ABOUT HERE

The message conveyed by Figure I is that the uncorrected procedure does not control for unbiased forecasters that falsely exhibit significant statistics (i.e., a false discovery). Indeed, with the uncorrected procedure, the probability of erroneously classifying at least one biased forecaster rapidly increases with the number of forecasters in the sample. This is problematic because the uncorrected procedure may lead to reject the null hypothesis of unbiasedness more often than necessary. For instance, when the number of hypotheses is around 50 (as it is in our application) we shall be nearly certain to detect some biased forecaster even if all them are unbiased.

3.3 The Procedures

In this subsection, we review procedures to tackle the problem of counting biased forecasters while controlling for false discoveries.

3.3.1 Counting the Number of Biased Forecasters

Suppose that we reject the single hypothesis that all forecasters are unbiased. The following question is to determine how many biased forecasters there are. As already said, this question cannot be answered by testing the single null hypothesis (3). Moreover, the uncorrected procedure is a non-starter because it does not control for false discoveries. One possibility to improve upon the uncorrected procedure is to control the so-called *false discovery rate (FDR).*

Before defining the FDR, we need to introduce some notation. Let D denote the number of forecasters classified as biased (i.e., the number of discoveries) and let a denote the number of those that are actually unbiased (i.e., the number of false discoveries). The *false discovery rate*, is defined as the expected number of incorrectly rejected null hypotheses over the total number of rejections, $FDR := E(a/D)$ if $D > 0$ and 0 otherwise. Control of the FDR means that, for a given pre-specified level γ, the average realized proportion of incorrectly classified biased forecasters will be at most γ.

The following rule, pioneered by Benjamini and Hochberg (1995), is known to control the FDR under independence of the statistics $\{z_i\}_{i=1}^n$:

Lemma 1 (False Discovery Rate Control under Independence [Benjamini-Hochberg]) *Assume that the p-values* $(p_1, .., p_n)$ *are independent. For a fixed value* γ *in* $(0, 1)$*, let* k *be the largest index for which*

$$p_{(k)} \leq \frac{\gamma k}{n} \tag{6}$$

The decision rule that rejects $H_{(1)}, .., H_{(k-1)}$ *and does not reject* $H_{(k)}, .., H_{(n)}$ *controls the false discovery rate in the sense* $FDR \leq \gamma$*.*

The proof of Lemma 1 may be found in Efron (2010, Theorem 4.1).

The Benjamini-Hochberg procedure controls the FDR when the p-values are independent or exhibits positive regression dependence (see Benjamini and Yekutieli, 2001). In a given application, this may not be the case. A procedure that controls for FDR when p-values exhibit an unknown form of dependence is due to Benjamini and Yekutieli (2001):

Lemma 2 (False Discovery Rate Control under Dependence [Benjamini - Yekutieli]) *For a fixed value* γ *in* $(0, 1)$*, let* k *be the largest index for which*

$$p_{(k)} \leq \frac{\gamma k}{n \sum_{i=1}^{n} \frac{1}{i}} \tag{7}$$

The decision rule that rejects $H_{(1)}, .., H_{(k-1)}$ and does not reject $H_{(k)}, .., H_{(n)}$ controls the false discovery rate in the sense $FDR \leq \gamma$.

3.3.2 Is There any Biased Forecaster in the Sample?

We have assumed that the single hypothesis that all forecasters are unbiased can be rejected. We now review alternative procedures to test this null hypothesis.

As stated above, to decide whether there are biased forecaster at all, one possibility is to test the single null hypothesis (3). An alternative procedure would be to test the multiple null hypotheses (4) while controlling the so-called *family-wise error rate (FWER).* The FWER, is defined as the probability of one or more false discoveries, $FWER := P(a > 0)$. Control of the FWER means that, for a given significance level γ, the probability of classifying as biased at least one unbiased forecaster (rejecting at least one true null hypothesis) is less than γ. Controlling the FWER allows us to be $1 - \gamma$ confident that there are no unbiased forecasters among those classified as biased by our decision rule.

The most basic method to control the FWER is the Bonferroni correction procedure.

Lemma 3 (Family-wise Error Rate Control [Bonferroni]). *For a fixed value γ in $(0, 1)$, reject H_{0i} if $p_i < \gamma/n$. Then, the family-wise error rate is controlled in the sense $FWER \leq \gamma$.*

The proof of Lemma 3 may be found in Lehmann and Romano (2005, Theorem 9.1.1). While the Bonferroni correction procedure satisfactorily controls the FWER, its ability to detect cases in which the null H_{0i} is false (i.e., a genuine biased forecaster) will typically be very low for large n (see the Monte Carlo exercises below). An improvement in this sense is obtained by the method of Holm (1979):

Lemma 4 (Family-wise Error Rate Control [Holm]). *For a fixed value γ*

in $(0,1)$*, let* k *be the minimal index such that:*

$$p_{(k)} \leq \frac{\gamma}{n+1-k} \tag{8}$$

The decision rule that reject $H_{(1)},..,H_{(k-1)}$ *and does not reject* $H_{(k)},..,H_{(n)}$ *controls the family-wise error rate in the sense* $FWER \leq \gamma$*.*

The proof of Lemma 4 may be found in Lehmann and Romano (2005, Theorem 9.1.2). Both Bonferroni and Holm correction procedures can be applied when the p-values are dependent. The Holm correction procedure is more powerful than the Bonferroni's one because, by being less stringent in erroneously detecting a biased forecaster ($\gamma/n \leq \gamma/(n+1-k)$), it allows to detect more genuinely biased forecasters. Then, there seems to be no reason to use the Bonferroni correction procedure because it is dominated by the Holm correction procedure.

3.3.3 Comparing Procedures

We stress the fact that we propose to apply different procedures to tackle different concerns related to the problem of counting biased forecasters. This point has not been discussed in the literature. In particular, when the concern is to count the number of biased forecasters, we propose to use Benjamini-Hochberg or Benjamini-Yukutieli procedures. These procedures avoid to wipe out evidence favoring the classification of a forecaster as biased while keeping under control the proportion of incorrectly classified biased forecasters. When the concern relates to whether there are biased forecasters at all, we suggest to use the Bonferroni or the Holm correction procedures. The reason is that these procedures deliver a result close to the one obtained from testing the single null hypothesis that all the forecasters are unbiased. As theoretically stated above and further validated with data below, the Bonferroni or the Holm procedures are not convenient to count the number of biased forecasters because they miss the chances of detecting biased forecasters in the fear of incorrectly classifying them.

We can sort the different procedures according to the number of significant hypotheses they produce (i.e., discoveries). On one extreme, the less conservative procedure is the uncorrected procedure, which may incur a large proportion of incorrectly classified biased forecasters. On the other extreme, we have the Bonferroni and Holm correction procedures, that undercount the number of biased forecasters. The Benjamini-Hochberg and Benjamini-Yukutieli procedures, by allowing to keep under control the proportion of incorrectly classified biased forecasters, will deliver a number of biased forecasters between these two extremes.

4. Monte Carlo Exercises

In this section, we implement Monte Carlo exercises to illustrate the gains of controlling for the false discovery rate when counting the number of biased forecasters. In particular, we show that when the number of biased forecasters is large (say 50 or more): (i) The uncorrected procedure does not control for the false discovery rate; (ii) The Bonferroni correction procedure has a low ability to detect genuine biased forecasters; (iii) The Benjamini-Hochberg procedure, unlike the uncorrected procedure, it does control the false discovery rate and it has better ability than the Bonferroni correction procedure to detect biased forecasters.

4.1 Data Generating Process

We begin by describing the data generating process. For a given number of forecasters $n \in \{10, 50, 100\}$, we simulate $T \in \{50, 100, 150\}$ independent replications of the vector $(y_t, x_{1t}, .., x_{tn})$ from a multivariate normal distribution with mean and variance:

$$\mu := \begin{pmatrix} .5 \\ (.5-\alpha_1)/\beta_1 \\ \vdots \\ (.5-\alpha_n)/\beta_n \end{pmatrix} \quad ; \quad \Sigma_o^2 := \begin{pmatrix} 4 & 2 & 2 & ... & 2 \\ 2 & 2 & 1 & ... & 1 \\ 2 & 1 & 2 & ... & 1 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 2 & 1 & 1 & ... & 2 \end{pmatrix}$$

with $\alpha_i = 0$ and $\beta_i = 1$ for $i \in \{1, .., n \times .7\}$, $\alpha_i = .9$ and $\beta_i = 1$ for $i \in \{n \times .7 + 1, .., n \times .8\}$, $\alpha_i = 1.1$ and $\beta_i = 1$ for $i \in \{n \times .8 + 1, .., n \times .9\}$, and $\alpha_i = .3$ and $\beta_i = 1$ for $i \in \{n \times .9 + 1, .., n\}$. This design implies that, in each Monte Carlo experiment, 30% of the forecasters are biased. Moreover, the vector of disturbance terms $(u_{1t}, .., u_{nt})$ is independent of the vector of forecasts $(x_{1t}, .., x_{nt})$, and the variance of u_{it} is the same across forecasters i and periods of time t. The number of forecasters n and the number of periods T are similar to those in the empirical application below. We set the number of Monte Carlo replications equal to 1000.[3]

4.2 Test Statistic

We now describe the test statistic employed to count the number of forecasters. Since the variance of the disturbance terms in the design above does not depend either on the forecaster i or on the time period t, for each forecaster i, we calculate the homoskedastic-only F-statistic:

$$F_i := (R\hat{\theta}_i - r)'(R\hat{\Omega}R')^{-1}(R\hat{\theta} - r)/2$$

with R the 2-dimensional identity matrix, $\hat{\theta}_i$ the ordinary least squares (OLS) estimator of (α_i, β_i), $r := (0, 1)$, and $\hat{\Omega}$ the homoskcedastic-only estimator of

[3]The experiments were carried out in R using the package "mvtnorm" to simulate draws from the multivariate normal distribution. The seed was set equal to the number of the Monte Carlo replication.

the covariance matrix of the (OLS) estimator $\hat{\theta}_i$. Since the disturbance term u_{it} is independent of x_{it} and normally identically distributed over t, the statistic F_i follows a F-distribution with $(2, T-2)$ degrees of freedom for all i. Since the same target variable y_t is used to calculate all the elements of the vector of statistics $(F_1, .., F_i, .., F_n)$, the p-values are not independent across forecasters. Indeed, the vector $(F_1, .., F_i, .., F_n)$ follows a multivariate F-distribution. The multivariate F-distribution exhibits positive regression dependence (see e.g., Sarkar, 2002), which implies that we can expect the Benjamini-Hochberg procedure to control the false discovery rate (see Benjamini and Yekutieli, 2001) even if the statistics are not independent.

In this context, a discovery is a forecaster classified as biased for a given procedure. A false discovery is an unbiased forecaster classified as biased by a given procedure. To evaluate the performance of the different procedures, we calculate the Monte Carlo average of discoveries and the Monte Carlo false discovery rate for each of the procedures described in the previous section. From our discussion in the previous section, recall that implementing these procedures requires to choose a pre-specified level γ for the error rate under control. To evaluate the sensitivity of results to a particular choice of γ, we make γ vary in $\{.01, .05, .1\}$. Results are summarized in Table I below.

4.3 Results

We now comment on the results. They are summarize in Table I. By comparing the number of discoveries across procedures, Table I shows that the uncorrected procedure is the one that finds the largest number of biased forecasters. Under the uncorrected procedure the false discovery rate is not under control (see that the false discovery rate in column 14 associated to the uncorrected procedure is higher than the pre-specified level). This means that many of the discoveries made by the uncorrected procedure are actually false discoveries. In particular, when the number of forecasters being evaluated increases, the number false discoveries increases. This result highlights the importance of controlling the false discovery rate when the number of forecasters being evaluated is large (say more than 50). On the opposite extreme is the Bonferroni correction procedure. This is the procedure that finds the lowest number of biased forecasters. This procedure is too conservative, in the sense that it detects less biased forecasters than those in the sample.

As predicted by theory, the number of biased forecasters obtained by applying the Benjamini-Hochberg procedure is between these two extremes. Under this procedure the false discovery rate is lower than the pre-specified critical value (compare columns 1 and 12), which means that it is under control, while the number of discoveries is higher than the one obtained under the Bonferroni correction procedure (compare columns 5 and 7). This means that the Benjamini-Hochberg procedure improves over the uncorrected procedure by keeping the false discovery rate under control and over the Bonferroni Correction procedure by being less conservative. We see this result as evidence favoring the use of the Benjamini-Hochberg procedure.

Regarding to the Benjamini-Yekutieli procedure, its ability to detect biased forecasters is lower than the Benjamini-Hochberg procedure (compare columns 7 and 8). In all cases the Benjamini-Yekutieli procedure delivers a false discovery rate below the critical value.

Finally, under the Holm correction procedure, the number of detected biased forecasters is slightly higher than under the Bonferroni correction procedure and the rate of false discovery is below the pre-specified level in all cases. This shows the advantage of the Holm correction procedure over the Bonferroni's one. However, when compared with the Bejanmini-Hochberg or with the Benjamini-Yekutieli procedures, its ability to detect biased forecaster is lower (compare columns 6 and 7, and 6 and 8).

INSERT TABLE I ABOUT HERE

We conclude then that for the data generating process in the Monte Carlo exercise, the Benjamini-Hochberg procedure is the preferred procedure to count the number of biased forecasters.

5. Revisiting the Survey of Professional Forecasters

To demonstrate the empirical relevance of controlling for false discoveries when counting biased forecasters, in this section we revisit the Survey of Professional Forecasters (SPF) data on real output growth in the US - a series that has been the subject of many previous studies (e.g., Zarnowitz, 1985; Elliott et al., 2008).[4] From our understanding, these studies either employ the uncorrected procedure or the Bonferroni correction procedure. We next show that, when controlling for the false discovery rate by employing the Benjamini- Hochberg procedure, 7 out of 10 forecasters classified as biased by the uncorrected procedure may actually be unbiased (i.e., false discoveries). We also find that 1 out of 10 forecasters classified as unbiased by the Bonferroni correction procedure may be actually biased (i.e., true discoveries).

5.1 The Data

We begin the analysis by describing the data. Survey participants in the SPF are professional forecasters who provide point forecasts for several macroeconomic variables on a quarterly basis. As a target variable, we consider real output growth measured using either the real time data (second revision) and the most recent vintage. For the output growth forecasts, we use the one-quarter-ahead forecast of that variable (data label RGDP). So, from the perspective of this application, y_t

[4]For other sources of expectational data to which the methods described above could apply see the survey by Pesaran and Weale (2006).

and x_{it} are, respectively, the real output growth in quarter t and the one-period-ahead forecast made by forecaster i.[5] The SPF data set is an unbalanced panel. Each quarter some forecasters leave the sample and new ones are included. We follow the existing literature (e.g., Zarnowitz, 1985 and Elliot et al., 2008) and report results for forecasters participating for a minimum of 12, 20 or 30 quarters. We denote by T_i the number of periods we observe the forecaster i. We report results for three subsamples: forecasters observed during the period 1968:4-1979:1 (as in Zarnowitz, 1985), forecasters observed during the period 1981:3-2006:3 (as in Elliot et al., 2008), and forecasters observed during the period 1981:3-2012:4. Notice that we do not consider multivariate forecasts. Evaluating multivariate forecasts would require to modify the F-statistics implemented below along the lines discussed by Komunjer and Owyang (2012).

5.2 Counting Biased Forecasters

We now describe the procedures employed to count the number of biased forecasters. For each forecaster i, we first estimate the parameters α_i and β_i in Equation (2) by ordinary least squares, and then calculate the homoskedasticity-only F-statistic associated to the null hypothesis of unbiasedness $H_{0i} : \alpha_i = 0$ and $\beta_i = 1$.

Heteroskedasticity and serial correlation in the forecast errors may produce large values of the homoskedasticity-only F-statistic, wrongly causing them to reject the null hypothesis of unbiasedness. To address the issue of heteroskedasticity, we also calculate heteroskedasticity-robust F-statistics. The issue of autocorrelatation can be disregarded because, under the null hypothesis of unbiasedness, one-period-ahead forecast errors would optimally follow a white noise process (see Granger and Newbold, 1977, pp. 121-22). For a 1%, 5% and 10%-level, we employ the F-statistics to calculate p-values using the F distribution with $(2, T_i - 2)$ degrees

[5] The variable y_t is calculated as four hundreds times the difference in logs of the gross domestic product at period t and $t-1$. The variable x_{it} is calculated as four hundreds times the difference in logs of the one-quarter-head gross domestic product forecast at $t-1$ and the forecast for the current gross domestic product at quarter $t-1$.

of freedom for the homoskedasticity-only F-statistic and $(2, \infty)$ degrees of freedom the for heteroskedasticity-robust F-statistic. Since we calculate p-values only for those forecasters with non-negative F-statistics, the number of forecasters evaluated by the homoskedasticity-only F-statistic and the heteroskedasticity-robust F-statistic differ.

With the p-values at hand, we count the number of rejected null hypotheses associated to the uncorrected procedure (UC), the Benjamini-Hochberg procedure (BH), the Bonferroni correction procedure (BC), the Holm correction procedure (HP) and the the Benjamini-Yekutieli procedure (BY). To our knowledge, neither the BH nor the BY procedures have been applied before to the problem of counting biased forecasters. Before going on, it is worth to note that the BC, HC and BY procedures apply under any type of dependence but BH procedure may not control the FDR if the statistics are dependent across forecasters. In our case, the F-statistics are not independent because they are all calculated using the same target variable y_t, and the forecasts x_{it} are correlated across forecasters. Indeed, the vector of F-statistics follow a multivariate F-distribution. The multivariate F-distribution, however, exhibits positive regression dependence (see Sarkar, 2002). Since the BH procedure controls the FDR under this type of dependence (see Benjamini and Yekutieli, 2001), we expect the BH procedure to still produce reliable results.

5.3 Results

Table II summarizes the results about the number of biased forecasters along the dimensions discussed above. In particular, Panel A presents results for the same sample period considered by Zarnowitz (1985), Panel B for the same sample period considered by Elliot et al. (2008), and Panel C for the updated version of the subsample period considered by these latter authors. The first column in Table II, γ, represents the value of the error rate under control. For the BH and

BY procedures, it should be understood as the FDR. For the UC procedure, it represents the significance level. Finally, for the BC and HC procedures, γ should be interpreted as the $FWER$.

INSERT TABLE III ABOUT HERE

Our preferred results are highlighted in gray. In this case, the difference in the number of biased forecasters across procedures is the most pronounced.[6] They correspond to the case when γ is equal to .1, at least 12 observations per forecaster are requested, and the heteroskedasticity-robust F- statistic is used. There, we analyze 81 forecasters.

The BC and BH procedures find 2 biased forecasters. Thus, we can be 90% confident that there are biased forecasters in this subsample. As already discussed, to calculate the number of biased forecasters, the more convenient procedure is BH, which detects 9 biased forecasters while the FDR is below 10%. For the sake of completeness, we also present the results delivered by BY procedure. If instead of using the BH procedure, we used the UC procedure to calculate the number of biased forecasters, we would count 28 biased forecasters. These findings illustrate our message: 7 out of 10 biased forecasters detected by the UC procedure may actually be unbiased forecasters (i.e., false discoveries). Regarding the other results in Table I, the BC and the HC procedures find evidence of biased forecasters across all the subsamples, except for Panel B where no biased forecaster is detected according to the heteroskedasticity-robust F-statistic when at least 30 observations per forecaster are required. This suggests that biased forecasters might leave the sample. According to the heteroskedasticity-robust F-statistic, the number of biased forecasters detected by the BH procedure varies greatly across subsamples. There are between 3 and 13 biased forecasters in Panel A, between 1 and 9 in Panel B, and between 11 and 27 biased forecasters in Panel C. Our

[6] Obviously, in other results the difference is less pronounced. We recall that the objective of this application is to illustrate the potential gains of using multiple testing techniques.

findings do not qualitatively change and an even greater number of biased forecaster is detected when the homoskedasticity-only F-statistic is used instead of the heteroskedasticity-robust F-statistic. When comparing the number of biased forecasters in the periods 1981:3-2006:3 and 1981:3-2012:4, we find that the number of biased forecasters is greater in the last subsample. This suggests that unbiased forecasters during the period 1981:3-2006:3 may have become biased afterwards.

Table III summarizes the results about the number of biased forecasters when the target variable is the growth rate of the fully revised output.

INSERT TABLE III ABOUT HERE

Results are qualitatively similar to those presented in Table II, except for the subsample covering the period 1981:3-2006:3. For this subsample the BH and the BC procedures provide more evidence in favor to the conclusion that there are biased forecasters. Both procedures detect at least 5 biased forecasters for any value of γ and sample sizes. When comparing the number of biased forecasters in the periods 1981:3-2006:3 and 1981:3-2012:4, we find that the number of biased forecasters is lower in the last subsample (Panel B versus Panel C in Table III). This finding suggests that forecasters might be more concerned about forecasting the fully revised growth rate of real output.

6. Conclusions

This paper shows that applying the Benjamini-Hochberg multiple testing procedure to the problem of counting biased forecasters improves upon procedures implemented in the existing literature. We apply the Benjamini-Hochberg procedure to data from the Survey of Professional Forecasters and find that up to 7 out of 10 forecasters classified as biased by the uncorrected procedure may actually be unbiased (i.e., false discoveries). This evidence soften the asymmetric loss interpretation of judgments made by forecasters of macroeconomic variables and the perception that the mean squared error loss function may not be able to capture how some professional forecasters form expectations about future values of macroeconomic variables (see e.g., Granger, 1969).

References

BATCHELOR, R (1990): "All Forecasters Are Equal", *Journal of Business and Economic Statistics*, 8(2) pp. 143-44.

BATCHELOR, R. and P. DUA (1991): "Blue Chip Rationality Tests", *Journal of Money, Credit and Banking*, 23(4) pp. 692-705.

BENJAMINI, Y. and Y. HOCHBERG (1995): "Controlling the False Discovery Rate: A Practical and Powerful Approach to Multiple Testing", *Journal of the Royal Statistical Society, Series B*, 57(1) pp. 289-300.

BENJAMINI, Y. and D. YEKUTIELI (2001): "The Control of the False Discovery Rate in Multiple Testing under Dependency", *Annals of Statistics*, 29(4) pp. 1165-88.

BOERO, G., J. SMITH, and K. WALLIS (2008): "Evaluating a Three-dimensional Panel of Forecasts: The Bank of England Survey of External Forecasters", *International Journal of Forecasting*, 24(3) pp. 354-67.

BONHAM C. and R. COHEN (2001): "To Aggregate, Pool, or Neither: Testing the Rational Expectations Hypothesis Using Survey Data", *Journal of Business and Economic Statistics*, 19(3) pp. 278-91.

BROWN, B. W. and S. MAITAL (1981): "What Do Economists Know? An Empirical Study of Experts' Expectations", *Econometrica*, 49(2) pp. 491-504.

CAPISTRAN, C. and A. TIMMERMANN (2009): "Disagreement and Biases in Inflation Expectations", *Journal of Money, Credit and Banking*, 41(2) pp. 365-396.

CROUSHORE, D. (2010): "An Evaluation of Inflation Forecasts from Surveys Using Real-Time Data", *B.E. Journal of Macroeconomics: Contributions*, 10(1) article 10.

D'AGOSTINO, A., K. MCQUINN and K. WHELAN (2012): "Are Some Forecasters Really Better Than Others?", *Journal of Money, Credit and Banking*, 48(4) pp. 715-843.

DAVIES, A. and K. LAHIRI (1995): "A New Framework for Analyzing Survey Forecasts Using Three-dimensional Panel Data", *Journal of Econometrics*, 68(1) pp. 205-27.

EFRON, B. (2010): *Large-Scale Inference*, Cambridge University Press.

ELLIOT, G., I. KOMUNJER and A. TIMMERMANN (2008): "Biased in Macroeconomic Forecasts: Irrationality or Asymmetric Loss?", *Journal of the European Economic Association*, 6(1) pp. 122-157.

GRANGER, C.W.J. (1969): "Prediction with a Generalized Cost of Error Function", *OR*, 20(2) pp. 199-207.

GRANGER, C. W. J. and P. NEWBOLD (1977): *Forecasting Economic Time Series*, Academic Press.

HOLM, S. (1979): "A Simple Sequentially Rejective Multiple Test Procedure", *Scandinavian Journal of Statistics*, 6(2) pp. 65-70.

KEANE, M. and D. RUNKLE (1990): "Testing the Rationality of Price Forecasters: New Evidence from Panel Data", *American Economic Review*, 80(4) pp. 714-35

KOMUNJER, I. and M. OWYANG (2012): "Multivariate Forecast Evaluation and Rationality Testing", *The Review of Economics and Statistics*, 94(4) pp. 1066-80.

LEHMANN, E. and J. ROMANO (2005): *Testing Statistical Hypothesis*, Springer

LIM, T. (2001): "Rationality and Analysts' Forecast Bias", *Journal of Finance*, 56(1) pp. 369-85.

MINCER, J. and V. ZARNOWITZ (1969): "The Evaluation of Economic Forecasts", In *Economic Forecasts and Expectations*, J. Mincer (ed.), National Bureau of Economic Research.

PESARAN, M.H. and M. WEALE (2006): "Survey Expectations", in Elliot, Granger and Timmermann (eds.), *Handbook of Economic Forecasting*, Vol. 1, Elsevier.

ROMANO, J., A. SHAIKH and M. WOLF (2010): "Hypothesis Testing in Econometrics", *Annual Review of Economics*, 2(1) pp. 75-104.

SARKAR, S. (2002): "Some Results on False Discovery Rate in Stepwise Multiple Testing Procedures", *Annals of Statistics*, 30(1) pp. 239-57.

SCHUH, S. (2001): "An Evaluation of Recent Macroeconomic Forecast Errors", *New England*

Economic Review, Issue 1, pp. 35-56.

STEKLER, H. (1987): "Who Forecasts Better?", *Journal of Business and Economic Statistics*, 5(1) pp. 155-158.

STOCK, J. and M. WATSON (2007): "Why Has U.S. Inflation Become Harder to Forecast?", *Journal of Money, Credit and Banking*, 39(1) pp 3-33.

THEIL, H. (1958): *Economic Forecasts and Policy*, North-Holland.

ZARNOWITZ, V. (1985): "Rational Expectations and Macroeconomic Forecasts", *Journal of Business and Economic Statistics*, 3(4) pp. 293-311.

Panel B. Mixture

Panel A. Mixture Components

Figure 1. False Discoveries

This figure illustrates the presence of false discoveries in a mixture model where the components are F-densities. The dashed line in both panels correspond to the threshold point defining the rejection region associated to the uncorrected procedure.

Table I. *Monte Carlo Exercises - Average Discoveries and False Discovery Rates*

				Average Discoveries					False Discovery Rate				
γ	n	T	Bias	*BC*	*HC*	*BH*	*BY*	*UC*	*BC*	*HC*	*BH*	*BY*	*UC*
(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)	(13)	(14)
1%	10	50	3	1.6	1.7	1.8	1.6	2.0	.20%	.20%	.63%	.21%	1.9%
1%	10	100	3	2.0	2.0	2.1	2.0	2.2	.17%	.24%	.72%	.20%	1.8%
1%	10	150	3	2.1	2.1	2.2	2.1	2.4	.21%	.29%	.55%	.25%	1.4%
1%	50	50	15	7.0	7.1	8.8	7.4	10.3	.02%	.03%	.45%	.08%	2.4%
1%	50	100	15	10.0	10.0	10.6	10.2	11.4	.03%	.06%	.63%	.12%	2.2%
1%	50	150	15	10.3	10.4	11.2	10.5	12.2	.02%	.03%	.55%	.08%	2.1%
1%	100	50	30	12.7	13.0	17.5	14.4	20.7	.14%	.15%	.74%	.20%	2.7%
1%	100	100	30	19.8	19.9	21.2	20.3	22.9	.05%	.07%	.62%	.17%	2.4%
1%	100	150	30	20.5	20.6	22.5	21.2	24.6	0%	.03%	.66%	.15%	2.4%
5%	10	50	3	1.9	1.9	2.1	1.9	2.6	1.1%	1.2%	2.9%	1.0%	9.5%
5%	10	100	3	2.1	2.2	2.4	2.2	2.7	.84%	1.1%	3.2%	.99%	8.0%
5%	10	150	3	2.3	2.3	2.5	2.3	2.9	.75%	.89%	2.8%	.85%	7.4%
5%	50	50	15	8.4	8.6	10.4	8.9	12.7	.17%	.20%	2.9%	.52%	10.0%
5%	50	100	15	10.3	10.4	11.7	10.6	13.8	.25%	.33%	3.0%	.72%	9.5%
5%	50	150	15	10.7	10.8	12.6	11.2	14.8	.15%	.21%	3.1%	.60%	9.4%
5%	100	50	30	15.7	16.0	21.1	17.7	25.7	.48%	.55%	3.3%	.73%	10.5%
5%	100	100	30	20.3	20.4	23.5	21.1	28.0	.19%	.21%	3.1%	.59%	10.2%
5%	100	150	30	21.3	21.4	25.4	22.5	29.9	.15%	.19%	3.3%	.65%	9.9%
10%	10	50	3	2.0	2.1	2.4	2.0	3.0	1.9%	2.3%	6.8%	2.1%	16.4%
10%	10	100	3	2.2	2.3	2.6	2.3	3.3	1.8%	2.3%	6.4%	2.2%	15.1%
10%	10	150	3	2.4	2.4	2.8	2.4	3.3	1.4%	2.0%	6.1%	1.8%	14.0%
10%	50	50	15	9.0	9.1	11.5	9.6	15.0	.58%	.63%	5.7%	1.2%	17.8%
10%	50	100	15	10.5	10.6	12.8	11.0	16.1	.54%	.65%	6.0%	1.3%	16.4%
10%	50	150	15	11.0	11.2	13.9	11.7	17.1	.36%	.54%	6.4%	1.2%	16.5%
10%	100	50	30	17.0	17.2	23.4	19.0	30.5	.67%	.72%	6.3%	1.2%	18.3%
10%	100	100	30	20.6	20.8	25.7	21.9	32.8	.29%	.34%	6.4%	1.2%	17.8%
10%	100	150	30	21.7	22.2	27.8	23.4	34.7	.2 %	.3%	6.6 %	1.3%	17.5%

This table presents estimates, based on simulated data, of the number of forecasters for whom the null hypothesis of unbiasedness could be rejected (discoveries) and the ratio between the number of those forecasters for whom the null hypothesis of unbiasedness could be erroneously rejected and the number of discoveries (false discovery rate) according to the Bonferroni procedure (BC), Holm procedure (HC), Benjamini-Hochberg procedure (BH), Benjamini-Yekutieli procedure (BY) and the uncorrected procedure (UC) described in the previous section. The number of Monte Carlo replications is 1000.

Table II. *Estimated Number of Biased Forecasters: Real Output (Real Time)*

		Homoske'y-only F-stat						Heteroske'y-robust F-stat					
γ	min T_i	n	*BC*	*HC*	*BH*	*BY*	*UC*	n	*BC*	*HC*	*BH*	*BY*	*UC*
Panel A: 1968:4 - 1979:1													
.01	12	79	6	6	19	7	30	57	10	10	13	11	13
.01	20	45	5	5	8	5	14	36	7	7	9	8	9
.01	30	17	0	0	3	0	4	13	2	2	3	1	3
.05	12	79	9	9	34	19	38	57	12	12	13	13	13
.05	20	45	5	7	14	8	15	36	8	9	9	9	9
.05	30	17	3	3	4	3	4	13	3	3	3	3	3
.1	12	79	15	16	34	23	37	57	13	13	13	13	14
.1	20	45	8	9	15	10	16	36	9	9	9	9	10
.1	30	17	4	4	4	4	4	13	3	3	3	3	3
Panel B: 1981:3 - 2006:3													
.01	12	88	29	31	44	34	51	81	1	1	1	1	9
.01	20	54	17	17	22	18	29	50	0	0	0	0	7
.01	30	31	5	5	5	5	14	28	0	0	0	0	2
.05	12	88	34	35	59	43	60	81	1	1	8	1	20
.05	20	54	18	18	32	22	33	50	1	1	6	0	15
.05	30	31	5	5	17	5	18	28	0	0	0	0	10
.1	12	82	36	37	62	52	64	81	2	2	9	1	28
.1	20	51	18	20	34	30	36	50	2	2	8	0	23
.1	30	31	5	6	18	15	19	28	0	0	0	0	15
Panel C: 1981:3 - 2012:4													
.01	12	109	59	63	73	66	74	85	21	21	24	21	26
.01	20	70	39	41	44	42	45	54	18	18	21	18	23
.01	30	45	25	26	27	26	28	32	11	11	11	11	11
.05	12	109	65	66	81	72	87	85	21	21	26	24	30
.05	20	68	42	42	51	45	54	54	18	19	23	21	25
.05	30	45	26	27	32	28	34	32	11	11	11	11	13
.1	12	109	65	66	87	74	87	85	21	22	27	24	40
.1	20	68	42	43	54	46	54	54	19	21	25	22	33
.1	30	45	26	27	34	29	34	32	11	11	13	11	20

This table presents estimates, based on data from the Survey of Professional Forecasters, of the number of forecasters for whom the null hypothesis of unbiasedness could be rejected according to the Bonferroni procedure (BC), Holm procedure (HC), Benjamini-Hochberg procedure (BH), Benjamini-Yekutieli procedure (BY) and the uncorrected procedure (UC) described in the previous section. The target variable is the real-time (second revision) growth rate of real output (GNP for the subsample 1968:4 - 1979:1 and GDP for the subsamples 1981:3 - 2006:3 and 1981:3 - 2012:4).

Table III. *Estimated Number of Biased Forecasters: Real Output (Fully Revised)*

		Homoske'y-only F-stat						Heteroske'y-robust F-stat					
γ	min T_i	n	BC	HC	BH	BY	UC	n	BC	HC	BH	BY	UC
1968:4 - 1979:1													
.01	12	79	5	5	11	6	14	50	10	10	10	10	13
.01	20	45	4	4	7	4	9	31	8	8	8	8	10
.01	30	17	1	1	1	1	3	15	3	3	3	3	4
.05	12	79	9	10	14	11	21	50	10	10	13	10	18
.05	20	45	7	7	9	8	13	31	8	8	12	8	14
.05	30	17	2	2	3	2	6	15	3	3	5	3	7
.10	12	79	10	11	16	12	28	50	10	10	18	10	20
.10	20	45	8	8	12	8	17	31	8	8	14	10	16
.10	30	17	3	3	5	2	8	15	4	4	7	4	8
1981:3 - 2006:3													
.01	12	88	59	64	73	67	74	39	8	8	8	8	9
.01	20	54	39	40	45	41	45	24	7	7	7	7	7
.01	30	31	24	24	26	24	26	13	5	5	5	5	5
.05	12	88	65	67	79	73	79	39	8	8	9	8	10
.05	20	54	40	42	48	45	48	24	7	7	7	7	8
.05	30	31	24	26	27	26	27	13	5	5	5	5	6
.10	12	88	67	70	80	76	80	39	8	9	9	9	10
.10	20	54	41	45	48	47	48	24	7	7	7	7	8
.10	30	31	24	27	27	27	27	13	5	5	6	5	6
1981:3 - 2012:4													
.01	12	109	40	43	56	47	60	72	6	6	7	6	8
.01	20	70	24	25	33	27	35	49	5	5	7	6	7
.01	30	45	12	13	18	13	19	33	5	5	6	5	6
.05	12	109	45	47	68	56	72	72	7	7	8	7	10
.05	20	70	27	30	40	33	42	49	6	6	7	7	8
.05	30	45	15	16	22	18	24	33	6	6	7	6	7
.10	12	109	52	56	79	67	82	72	8	8	9	8	15
.10	20	70	31	32	44	37	48	49	7	7	8	7	13
.10	30	45	16	18	24	19	28	33	6	6	7	6	11

This table presents estimates, based on data from the Survey of Professional Forecasters, of the number of forecasters for whom the null hypothesis of unbiasedness could be rejected according to the Bonferroni procedure (BC), Holm procedure (HC), Benjamini-Hochberg procedure (BH), Benjamini-Yekutieli procedure (BY) and the uncorrected procedure (UC) described in the previous section. The target variable is the fully revised (latest revision) growth rate of real output (GNP for the subsample 1968:4 - 1979:1 and GDP for the subsamples 1981:3 - 2006:3 and 1981:3 - 2012:4).