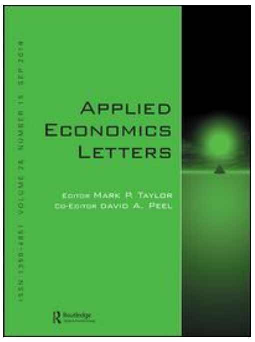

Applied Economics Letters

ISSN: (Print) (Online) Journal homepage: https://www.tandfonline.com/loi/rael20

Testing for bias in forecasts for independent binary outcomes

Philip Hans Franses

To cite this article: Philip Hans Franses (2020): Testing for bias in forecasts for independent binary outcomes, Applied Economics Letters, DOI: 10.1080/13504851.2020.1838429

To link to this article: https://doi.org/10.1080/13504851.2020.1838429

Published online: 27 Oct 2020.

APPLIED ECONOMICS LETTERS
https://doi.org/10.1080/13504851.2020.1838429

Routledge
Taylor & Francis Group

ARTICLE

OPEN ACCESS Check for updates

Testing for bias in forecasts for independent binary outcomes

Philip Hans Franses

Erasmus School of Economics, Econometric Institute, Rotterdam, The Netherlands

ABSTRACT
This letter deals with a test on forecast bias in predicting independent binary outcomes, where the outcomes are either 1 or 0, and the predictions are probabilities. The test concerns two parameter restrictions in a simple logit model. Size-corrected power experiments show remarkable power.

KEYWORDS
Binary outcomes; Probability forecasts; Forecast bias; Logit model

JEL CODES
C25; C53

I. Introduction and motivation

This letter deals with a test on forecast bias in predicting independent binary outcomes, where the outcomes are either 1 or 0, and the predictions are probabilities. There is no need to know how the predictions were created, that is, the predictions can be based on a logit model (Cramer 1991) or a probit model or a linear probability model, or by expert judgement.

In a standard regression model for continuous outcomes, one can consider the auxiliary regression which links the predictions with the realizations. If realizations y_i and forecasts $\hat{y}_i$ would be continuous variables, and these can be cross sectional data or time series data, where the forecast sample is $i = 1, 2, .., N$, then one can examine bias using the auxiliary regression

$$y_i = \alpha + \beta \hat{y}_i + \varepsilon_i$$

The parameters are estimated using Ordinary Least Squares. The Wald type test of interest concerns the hypothesis that saranyaMand $\beta = 1$, jointly. Under the null hypothesis, there is no forecast bias. This regression is called the Mincer Zarnowitz regression, see Mincer and Zarnowitz (1969).

In this letter, I propose a similar test but now for independent binary outcomes, that is, there are realizations that can be either 1 or 0, and the predictions are estimated probabilities that the outcome is 1. The question is if these probabilities are unbiased or not. Note that if the predictions are also 1 or 0, one can resort to variants of tests on hit rates, see for example Franses and Paap (2001, page 65), but a test for the hit rate is not the focus here. The new test turns out to be similarly easy as based on the Mincer Zarnowitz regression. Power simulations show that the test works quite well. The test for forecast bias is illustrated using the 2018 Goldman Sachs predictions for the football teams that supposedly would make it to the second round of the 2018 World Championship football in Russia.

II. The main idea

Consider N forecasts $\hat{p}_i$ and N observations y_i, where y_i can take values 1 or 0, and where the forecasts are numbers in between 0 and 1. An example is the dataset in Table 1, which refers to the Goldman Sachs forecasts. The interest is to see if there is forecast bias.

The key identity to design the test is

$$Prob(y_i) = \hat{p}_i$$

where simple algebra gives

$$Prob(y_i) = \frac{\exp\left(\log\left(\frac{\hat{p}_i}{1-\hat{p}_i}\right)\right)}{1 + \exp\left(\log\left(\frac{\hat{p}_i}{1-\hat{p}_i}\right)\right)} = \hat{p}_i$$

The middle term can be recognized as the expression for the logit model (Franses and Paap 2001, page, 54), that for a single variable x_i is given by

CONTACT Philip Hans Franses franses@ese.eur.nl Econometric Institute, Erasmus School of Economics, 3062PA Rotterdam, The Netherlands

Table 1. Realizations and forecasts concerning surviving the first round of the 2018 World Cup in Russia. Data source is Exhibit 2 of the 11 June 2018 Global Macro Research report of the Goldman Sachs Group, Inc.

Brazil	1	0.875
France	1	0.814
Germany	0	0.805
Portugal	1	0.752
Belgium	1	0.785
England	1	0.723
Argentina	1	0.731
Spain	1	0.797
Colombia	1	0.749
Uruguay	1	0.744
Poland	0	0.685
Mexico	1	0.478
Denmark	1	0.520
Sweden	1	0.459
Iran	0	0.354
Peru	0	0.373
Russia	1	0.335
Switzerland	1	0.479
Australia	0	0.498
Croatia	1	0.428
Iceland	0	0.452
Costa Rica	0	0.368
Tunisia	0	0.329
Saudi Arabia	0	0.365
Serbia	0	0.434
Japan	1	0.352
Egypt	0	0.344
Morocco	0	0.216
Nigeria	0	0.171
Senegal	0	0.252
Korea Republic	0	0.201
Panama	0	0.132

$$Prob(y_i) = \frac{\exp(\alpha + \beta x_i)}{1 + \exp(\alpha + \beta x_i)}$$

Hence, a Mincer Zarnowitz type test for the null hypothesis of no forecast bias can be based on the logit model

$$Prob(y_i) = \Lambda\left(\alpha + \beta \log\left(\frac{\hat{p}_i}{1 - \hat{p}_i}\right)\right) \tag{1}$$

with Λ the logistic function, and on the Wald test for $\alpha = 0$and $\beta = 1$, jointly.

III. Simulations

To see how the test works in practice, I consider various simulation experiments. As there is no such test around,[1] I focus only on the proposed test. For sample size N, I generate 2 N observations, where the first half will be used to estimate the model parameters, and the second half will be used to create and evaluate the forecasts. The Data Generating Process (DGP) is

$$Prob(y_i) = \frac{\exp(0 + 2x_i)}{1 + \exp(0 + 2x_i)}$$

where $x_i\tilde{N}(0, 1)$ The binary data on y_i are created as follows:

$$y_i = 1\text{when}Prob(y_i) > 0.5$$

$$y_i = 0\text{when}Prob(y_i) \leq 0.5$$

Next, the parameters in the logit model are estimated using Maximum Likelihood, see Franses and Paap (2001, section 4.2). The estimated parameters are used for the second set of N observations to create $\hat{p}_i$. Finally, the logit model in (1) is considered and the Wald test is computed. I use 1000 simulation runs.

First, I examine if the test has proper size. This turns out not to be the case, as even in case $N = 10000$, the rejection rate is 16.2%. To obtain a new 5% critical value, the 95$^{\text{th}}$ Wald test value is taken, and this is equal to 10.71. With this new critical value, size-corrected power experiments can be run.

To create data under the alternative hypothesis, I replace observations on y_i in the second set of N observations. Each time 5%, 10%, 15%, until 90% of the observations with $y_i = 1$ is replaced by $y_i = 0$. The size-corrected power for $N = 50, 100, 500\text{and}1000$ is displayed in Table 1.

Clearly, the size-corrected power is quite high, even for small samples.

IV. Illustration

To illustrate the new test, consider the data in Table 2. There are 32 countries of which 16 attained the knockout stage of the 2018 World Cup football tournament. Attaining this stage is labelled as 1, having to leave the tournament after the first round is 0. The third column of Table 2 presents the probabilities assigned by Goldman Sachs of attaining the second round.

The Maximum Likelihood based parameter estimates (using Eviews version 8.0) are 0.033 (0.440) and 1.596 (0.569) for α and β, respectively, with

[1]There are tests on the so-called hit rate, that is the fraction of correctly predicted 1 and 0 observations, but that concerns another feature of the forecasts.

Table 2. Size-corrected power. The 5% critical value is set at 10.71.

	N			
Replacement	50	100	500	1000
5%	0.241	0.240	0.558	0.847
10%	0.373	0.508	0.973	1.000
15%	0.543	0.733	1.000	1.000
20%	0.701	0.896	1.000	1.000
25%	0.803	0.970	1.000	1.000
30%	0.894	0.995	1.000	1.000
35%	0.944	0.999	1.000	1.000
40%	0.966	1.000	1.000	1.000
45%	0.981	1.000	1.000	1.000
50%	0.996	1.000	1.000	1.000
55%	1.000	1.000	1.000	1.000
60%	1.000	1.000	1.000	1.000
65%	1.000	1.000	1.000	1.000
70%	1.000	1.000	1.000	1.000
75%	1.000	1.000	1.000	1.000
80%	1.000	1.000	1.000	1.000
85%	1.000	1.000	1.000	1.000
90%	1.000	1.000	1.000	1.000

estimated standard errors in parentheses. The McFadden R-squared (Franses and Paap 2001, page, 64) is 0.282, so the logit model fits the data quite well. Finally, the Wald test for the joint hypothesis that $\alpha = 0$and $\beta = 1$ appears to equal 1.100, which is substantially smaller than 10.71. This suggests that the Goldman Sachs forecast were unbiased.

Acknowledgments

Thanks to Richard Paap for helpful comments and Max Welz for programming.

Disclosure statement

No potential conflict of interest was reported by the author.

ORCID

Philip Hans Franses http://orcid.org/0000-0002-2364-7777

References

Cramer, J. S. 1991. *The Logit Model: An Introduction for Economists*. New York: Routlegde.

Franses, P. H., and R. Paap. 2001. *Quantitative Models in Marketing Research*. Cambridge UK: Cambridge University Press.

Mincer, J., and V. Zarnowitz. 1969. "The Evaluation of Economic Forecasts." In *Economic Forecasts and Expectations*, edited by J. Mincer. New York: National Bureau of Economic Research.