

Lahiri, Kajal; Yang, Liu

**Working Paper**

## Asymptotic Variance of Brier (Skill) Score in the Presence of Serial Correlation

CESifo Working Paper, No. 5290

**Provided in Cooperation with:**

Ifo Institute – Leibniz Institute for Economic Research at the University of Munich

*Suggested Citation:* Lahiri, Kajal; Yang, Liu (2015) : Asymptotic Variance of Brier (Skill) Score in the Presence of Serial Correlation, CESifo Working Paper, No. 5290, Center for Economic Studies and ifo Institute (CESifo), Munich

This Version is available at:

<https://hdl.handle.net/10419/110796>

**Standard-Nutzungsbedingungen:**

Die Dokumente auf EconStor dürfen zu eigenen wissenschaftlichen Zwecken und zum Privatgebrauch gespeichert und kopiert werden.

Sie dürfen die Dokumente nicht für öffentliche oder kommerzielle Zwecke vervielfältigen, öffentlich ausstellen, öffentlich zugänglich machen, vertreiben oder anderweitig nutzen.

Sofern die Verfasser die Dokumente unter Open-Content-Lizenzen (insbesondere CC-Lizenzen) zur Verfügung gestellt haben sollten, gelten abweichend von diesen Nutzungsbedingungen die in der dort genannten Lizenz gewährten Nutzungsrechte.

**Terms of use:**

*Documents in EconStor may be saved and copied for your personal and scholarly purposes.*

*You are not to copy documents for public or commercial purposes, to exhibit the documents publicly, to make them publicly available on the internet, or to distribute or otherwise use the documents in public.*

*If the documents have been made available under an Open Content Licence (especially Creative Commons Licences), you may exercise further usage rights as specified in the indicated licence.*



## Asymptotic Variance of Brier (Skill) Score in the Presence of Serial Correlation

Kajal Lahiri  
Liu Yang

CESIFO WORKING PAPER NO. 5290  
CATEGORY 12: EMPIRICAL AND THEORETICAL METHODS  
MARCH 2015

*An electronic version of the paper may be downloaded*

- from the SSRN website: [www.SSRN.com](http://www.SSRN.com)
- from the RePEc website: [www.RePEc.org](http://www.RePEc.org)
- from the CESifo website: [www.CESifo-group.org/wp](http://www.CESifo-group.org/wp)

ISSN 2364-1428

# Asymptotic Variance of Brier (Skill) Score in the Presence of Serial Correlation

## Abstract

We propose autocorrelation-robust asymptotic variances of the Brier score and Brier skill score, which are generally applicable in circumstances with weak serial correlation. An empirical application in macroeconomics underscores the importance of taking care of serial correlation. We find that the conventional variances are too conservative to account for the sampling variability in estimating the Brier (skill) score.

JEL-Code: C010, C120, C250, C520, C530.

Keywords: probability forecasts, serial correlation, Brier score, Brier skill score, survey of professional forecasters.

*Kajal Lahiri*  
*Department of Economics*  
*University at Albany, SUNY*  
*USA – NY 12222*  
*klahiri@albany.edu*

*Liu Yang\**  
*School of Economics*  
*Nanjing University*  
*P.R. China – Jiangsu 210093*  
*lyang2@albany.edu*

\*corresponding author

# 1 Introduction

Probability forecasts of a future event are routinely used in diverse fields of application. For example, macroeconomists issue probability forecasts of an economic recession at varying horizons in each quarter. Banks are required by regulators to evaluate the probabilities of default associated with loans they have made. Assessing the efficacy of the probability forecasts is of utmost importance to guide the decision-makings in these contexts. Typically, a number of scores, which are functions of the forecasts and actuals, are employed as evaluation metrics. Lahiri and Yang (2013) provided a survey on these scores from the perspective of economic forecasting. Among them, the Brier score (or the quadratic probability score) is probably the most commonly used and it is the probabilistic analogue of the mean squared error. A large body of literature on economic forecasting takes the Brier score as the primary statistic to summarize the predictive performance of probability forecasts. Recent examples include Lahiri *et al.* (2013), Levanon *et al.* (2014) and Rudebusch and Williams (2009), only to name a few.

The Brier score is computed on the basis of a given sample on binary events and probability forecasts. Thus, the sampling uncertainty in estimating the score has to be properly accounted for in order to make a statistically meaningful inference. When observations are independently and identically distributed (i.i.d.) across the sample, Bradley *et al.* (2008) derived approximations to the variances of the Brier score and related Brier skill score. In macroeconomic forecasting, it is widely accepted that the sample displays positive serial correlation as a consequence of the persistence in economic series, such as real GDP or inflation. One implication of serial correlation is that the usual sampling variance of the score obtained by assuming independence is no longer valid. This phenomenon has been recognized by Lahiri and Yang (2015) and Pesaran and Timmermann (2009) in various scenarios. Wilks (2010) has shown that the failure to accommodate positive serial correlation will significantly underestimate the standard error of the Brier (skill) score and the magnitude of underestimation depends on the event probability and the quality of the forecast. To correct the effects of serial correlation, Wilks proposed an adjustment factor, whose legitimacy is jus-

tified by a simulation experiment. Though useful by itself under his specific data generating process, Wilks' adjusted variance might lose its ground when the underlying process deviates a lot from the assumed process. The main contribution of our paper is to extend the variances given in Wilks (2010) and to derive asymptotic variances of the Brier score and Brier skill score in general settings with weak serial correlation.

The remainder of the paper is organized as follows. Section 2 develops the asymptotic variance of the Brier (skill) score when the sample used to produce it is serially dependent. In Section 3, we apply the proposed methodology to examine the quality of the *Survey of Professional Forecasters* (SPF), with respect to its capacity of predicting real GDP declines in the United States. Some concluding remarks are given in Section 4.

## 2 Asymptotic variances of the Brier score and Brier skill score under weak serial correlation

Throughout the paper,  $Z$  is the binary variable to be predicted. When the target event occurs,  $Z = 1$ , and  $Z = 0$  otherwise.  $P$  is the probabilistic forecast of the target event. A macroeconomic example of  $Z$  and  $P$  is given in Section 3. To measure how well the forecast is related to the actual, Brier (1950) developed a score function based on a sequence  $\{(Z_t, P_t) : t = 1, \dots, T\}$  as  $BS \equiv \frac{1}{T} \sum_{t=1}^T (Z_t - P_t)^2$ . By construction, the Brier score  $BS$  lies strictly between 0 and 1, and it has a negative orientation in that lower  $BS$  indicates higher accuracy of  $P$ . When  $P$  exactly coincides with  $Z$ , that is,  $P$  is a perfect forecast,  $BS = 0$ . Another forecast, which is often taken as the benchmark, is  $P = \pi \equiv P(Z = 1)$ . In practice, the population probability  $P(Z = 1)$  is rarely known, and thus it is usually replaced by its sample analogue  $\bar{Z} \equiv \frac{1}{T} \sum_{t=1}^T Z_t$ . The Brier score of this naive forecast is denoted as  $BS_0 \equiv \frac{1}{T} \sum_{t=1}^T (Z_t - \bar{Z})^2$ . A real-life forecast with some skill is not perfect, yet it can beat the naive benchmark. Consequently, the Brier score of a real-life forecast is often higher than 0 but lower than  $BS_0$ .

Sometimes, it is likely to yield a misleading conclusion regarding the performance of  $P$  if we merely look at the Brier score. For example, suppose  $P(Z = 1)$  is very close to 0, that

is, the target event  $Z = 1$  is rare. In this case,  $BS_0$  could be very close to 0 as well. To see this, let  $BS_0 \equiv \frac{1}{T} \sum_{t=1}^T Z_t^2 - 2 \frac{1}{T} \sum_{t=1}^T Z_t \bar{Z} + \bar{Z}^2 = \bar{Z}(1 - \bar{Z})$ . With a rare event,  $\bar{Z}$  is very small, which makes  $BS_0$  quite close to 0. Although the naive forecast seemingly performs fairly well by its Brier score, it is clearly of no skill at all because this benchmark cannot distinguish between the occasions when  $Z = 1$  occurs and those when  $Z = 1$  does not occur. To circumvent this pitfall of the Brier score in the case of rare events, the Brier skill score can be used. Given any forecast  $P$ , the Brier skill score is defined as  $BSS \equiv \frac{BS_0 - BS}{BS_0} = 1 - \frac{BS}{BS_0}$  and it is the improvement of the forecast  $P$  relative to the naive baseline. The Brier skill score of the benchmark is 0. When  $P$  outperforms the benchmark,  $BSS > 0$ . Otherwise,  $BSS < 0$ . Other details of  $BSS$  can be found in Stephenson (2000).

Define  $\Omega_T$  to be the covariance matrix of  $\frac{1}{\sqrt{T}} (\sum_{t=1}^T (Z_t - P_t)^2, \sum_{t=1}^T (Z_t - \pi)^2)'$ . The  $ij$ th component of a matrix  $A$  is denoted by  $A_{ij}$ . Let  $BS^* \equiv E(Z_t - P_t)^2$ ,  $BS_0^* \equiv E(Z_t - \pi)^2 = \pi(1 - \pi)$  and  $BSS^* \equiv 1 - \frac{BS^*}{BS_0^*}$ . The goal of this section is to identify the asymptotic variances of  $BS$  and  $BSS$  when the sample used to generate them is serially correlated. The following assumptions are sufficient for this purpose.

**Assumption 1** For each  $t$ ,  $P_t \in [0, 1]$  and  $Z_t \sim \text{Bernoulli}(\pi)$ , where  $\pi \in (0, 1)$ .

**Assumption 2** For some  $r' > 1$ , the process  $\{(Z_t, P_t) : t = 1, \dots\}$  is a mixing sequence with either uniform mixing coefficient  $\phi_m$  or strong mixing coefficient  $\alpha_m$  of size  $2r'/(r' - 1)$ .

**Assumption 3**  $(Z_t, P_t)$  is identically distributed across  $t$ .

**Assumption 4**  $\Omega_T$  is positive definite for each  $T \in N$  and there exists  $\varepsilon > 0$  and a natural number  $N(\varepsilon)$  such that  $|\Omega_T| > \varepsilon$  for all  $T > N(\varepsilon)$ .

Assumption 1 rules out the case of non-stochastic  $Z$  when  $\pi = 0$  or  $\pi = 1$ , which is of no interest. Assumption 2 allows for a certain degree of serial correlation in the sample, as long as its dependence shrinks towards zero at the stated rate. The population Brier (skill) score is not well defined unless Assumption 3 holds. Assumption 4 is required to ensure the existence of a positive definite long run covariance matrix  $\Omega$  in Lemma 1.

**Lemma 1** Under Assumptions 1-4, there exists a symmetric positive definite matrix  $\Omega$  such that  $\Omega_T \rightarrow \Omega$  as  $T \rightarrow \infty$ .

**Theorem 1** Under Assumptions 1-4,  $\sqrt{T} \begin{pmatrix} BS - BS^* \\ BS_0 - BS_0^* \end{pmatrix} \xrightarrow{d} N(0, \Omega)$ .

**Collorary 1** Under Assumptions 1-4,  $\sqrt{T}(BS - BS^*) \xrightarrow{d} N(0, \Omega_{11})$ .

**Collorary 2** Under Assumptions 1-4,  $\sqrt{T}(BSS - BSS^*) \xrightarrow{d} N(0, \frac{1}{BS_0^2}(\Omega_{11} + \frac{BS^*}{BS_0^2}\Omega_{22} - 2\frac{BS^*}{BS_0}\Omega_{12}))$ .

Colloraries 1 and 2 present the asymptotic distributions of  $BS$  and  $BSS$  respectively. If the data is independently identically distributed,  $\Omega_T = \Omega$ , which is the covariance matrix of  $((Z_t - P_t)^2, (Z_t - \pi)^2)'$ . However, in the presence of serial correlation,  $\Omega_T \neq \Omega$  and  $\Omega$  is the sum of the covariance matrix of  $((Z_t - P_t)^2, (Z_t - \pi)^2)'$  and its autocovariance matrices of various orders. For example,  $\Omega_{11} = Var((Z_t - P_t)^2) + 2\sum_{m=1}^{\infty} Cov((Z_t - P_t)^2, (Z_{t+m} - P_{t+m})^2)$  in Collorary 1. If the data exhibits positive autocorrelation,  $Cov((Z_t - P_t)^2, (Z_{t+m} - P_{t+m})^2) > 0$  for any  $m$ , and thus  $\Omega_{11} > Var((Z_t - P_t)^2)$ . This implies that the conventional variance based on independence assumption could underestimate the true uncertainty in estimating the Brier score, and the degree of underestimation depends on the strength of autocorrelation.

The asymptotic variance of the Brier skill score in Collorary 2 is more complex. Motivated by Diebold and Mariano (1995) and Lopez (2001), an alternative expression of this variance is given by

$$\frac{\Omega_{11} + \Omega_{22}(1 - BSS^*)^2 - 2(1 - BSS^*)\Omega_{12}}{\pi^2(1 - \pi)^2},$$

as shown in the appendix. As above, we can write  $\Omega_{ij}$  as the sum of two parts, that is,  $\Omega_{ij} = \Omega_{ij}^1 + \Omega_{ij}^2$ , where  $\Omega_{ij}^1$  is the variance part and  $\Omega_{ij}^2$  is the autocovariance part.<sup>1</sup> Accordingly, the asymptotic variance of  $BSS$  is

$$\frac{\Omega_{11}^1 + \Omega_{22}^1(1 - BSS^*)^2 - 2(1 - BSS^*)\Omega_{12}^1}{\pi^2(1 - \pi)^2} + \frac{\Omega_{11}^2 + \Omega_{22}^2(1 - BSS^*)^2 - 2(1 - BSS^*)\Omega_{12}^2}{\pi^2(1 - \pi)^2}. \quad (1)$$

In view of (1), the inflation in variance due to serial correlation is characterized by  $\frac{\Omega_{11}^2 + \Omega_{22}^2(1 - BSS^*)^2 - 2(1 - BSS^*)\Omega_{12}^2}{\pi^2(1 - \pi)^2}$ , whose magnitude is determined by the event probability  $\pi$ , the quality of the forecast  $BSS^*$  and the strength of serial correlation in  $\Omega_{ij}^2$ . Consistent with the simulation evidence in Wilks (2010), a lower  $\pi$ , other things being equal, induces a larger inflation in variance.

To make use of Colloraries 1 and 2 to conduct statistical inference, the asymptotic variances must be estimated. For example,  $BS_0 = \bar{Z}(1 - \bar{Z})$  can be used to estimate  $BS_0^*$ . Similarly,

<sup>1</sup>For instance,  $\Omega_{11}^2 = 2\sum_{m=1}^{\infty} Cov((Z_t - P_t)^2, (Z_{t+m} - P_{t+m})^2)$ .

$BS^*$  is estimated by  $BS = \frac{1}{T} \sum_{t=1}^T (Z_t - P_t)^2$ . By Theorem 1, both  $BS_0$  and  $BS$  are consistent. The long run covariance matrix  $\Omega$  can be estimated in different ways. All estimators are based on an appropriately chosen weighting scheme (kernel). The basic idea is to use a finite sum of sample autocovariance matrices to approximate the population infinite sum, allowing for the truncation lag to increase to infinity at an appropriate rate as the sample size grows. Under some weak regularity conditions, these estimators can be shown to be consistent. See Andrews (1991) and Newey and West (1987, 1994) for details.

### 3 An empirical illustration

In this section, we will use the Brier score and Brier skill score to assess the performance of the probability forecasts of real GDP declines in the *Survey of Professional Forecasters*. The main purpose here is to compare the asymptotic confidence intervals of BS and BSS using our autocorrelation-robust variances with those constructed by assuming independence. In each quarter since 1968:Q4, the respondents of this survey are asked to indicate the probability they would attach to a decline in the level of real GDP in the current and the next four quarters. The target variable is the same for all of these five horizons. Our sample spans from 1968:Q4 to 2015:Q1. During this period, the fraction of real GDP declines is about 12.9%, meaning that it is a relatively uncommon event. Lahiri and Wang (2013) carried out a comprehensive evaluation on the accuracy of these subjective forecasts.

Our robust analysis is motivated by Figure 1, which presents the autocorrelation functions of SPF forecasts and actuals. Clearly, all series display positive autocorrelation. All of the autocorrelation coefficients up to three-quarter lags are significantly different from zero. Table 1 shows BS and BSS for SPF forecasts. As horizon rises, the performance of professional forecasters deteriorates, as reflected by rising BS and declining BSS. In this example, the Brier score of the naive benchmark is about 0.116. As a result, the four-quarter-ahead forecasts are even worse than this benchmark, as is obvious from its negative BSS. Two types of 95% confidence intervals of BS and BSS are also given in Table 1.<sup>2</sup> By ignoring the posi-

---

<sup>2</sup>To construct independent intervals, we use the sample covariance matrix of  $((Z_t - P_t)^2, (Z_t - \pi)^2)'$  to es-



tive serial correlation in Figure 1, the independent intervals are uniformly narrower than the autocorrelation-robust intervals.

Table 1: Brier (skill) scores and their 95% confidence intervals

Statistic	Q0	Q1	Q2	Q3	Q4
BS	0.062	0.083	0.098	0.115	0.119
Ind.	(0.042,0.082)	(0.062,0.105)	(0.073,0.124)	(0.084,0.146)	(0.086,0.152)
Corr.	(0.035,0.090)	(0.049,0.117)	(0.053,0.143)	(0.064,0.166)	(0.070,0.168)
BSS	0.446	0.259	0.126	0.013	-0.040
Ind.	(0.258,0.635)	(0.094,0.425)	(0.006,0.245)	(-0.056,0.082)	(-0.109,0.030)
Corr.	(0.196,0.696)	(0.042,0.477)	(-0.006,0.258)	(-0.078,0.104)	(-0.139,0.059)

**Notes:** “Ind.” is the 95% confidence interval based on independence assumption. “Corr.” is the 95% confidence interval with  $\Omega$  being estimated by Andrew’s approach.

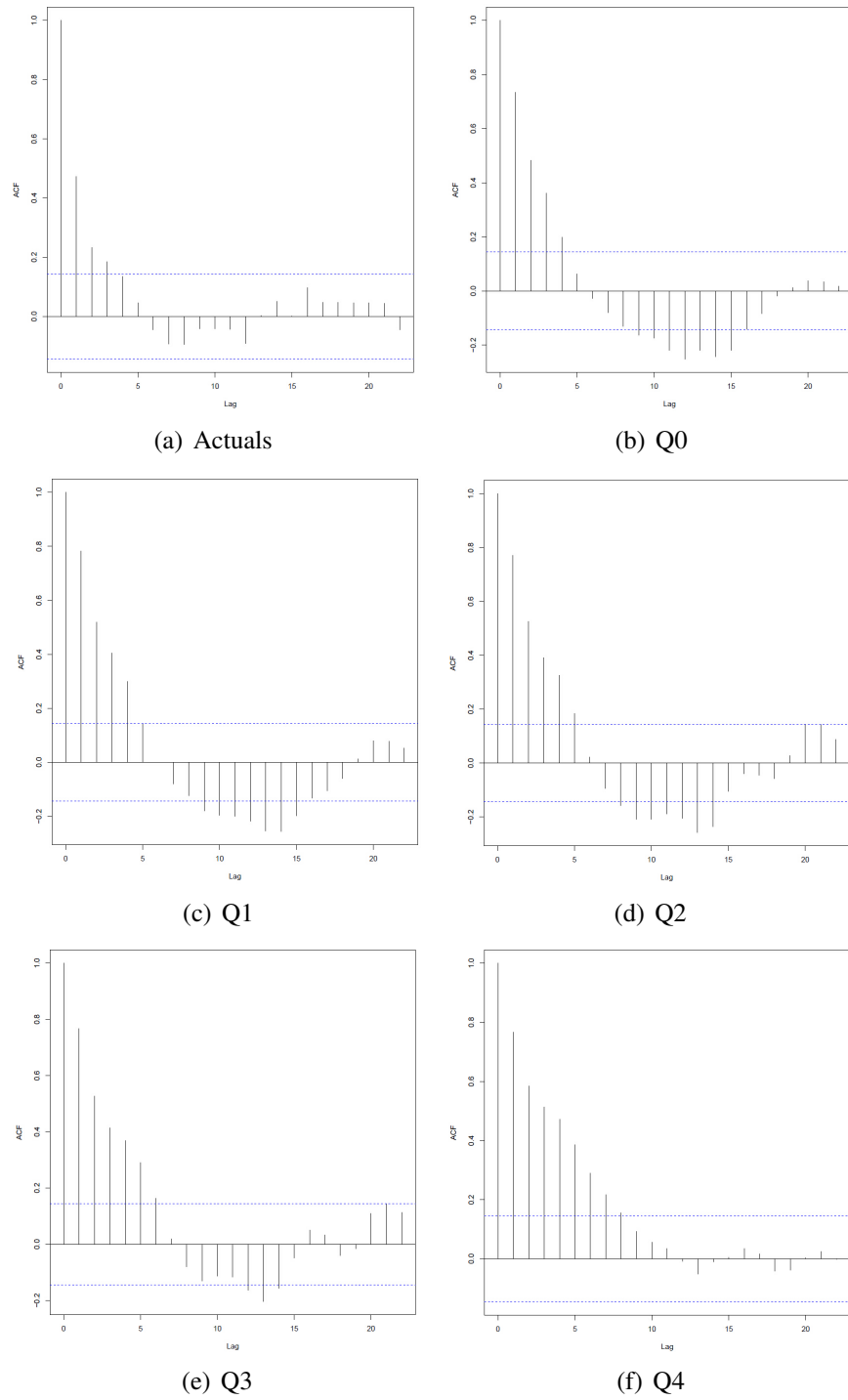
## 4 Conclusions

This paper addresses the problem of correcting the effects of serial correlation on the sampling properties of the Brier (skill) score, initially investigated by Wilks (2010). The proposed asymptotic variance is more general and thus applicable in circumstances with weak serial correlation. Using an empirical example with SPF probability forecasts, we confirm that by ignoring the positive serial correlation, the conventional variance is too conservative to account for the sampling uncertainty in estimating the Brier (skill) score.

---

timinate  $\Omega$ . Prior to calculating the robust intervals, the data is filtered by AR(1) prewhitening procedure as advocated by Andrew and Monahan (1992). The quadratic spectral kernel is used in generating the robust intervals

Figure 1: Autocorrelation functions of SPF and Actuals



**Notes:** The dotted lines are 95% confidence band about the zero line. “Q0” is the current-quarter SPF forecast. “Q1” is the one-quarter-ahead SPF forecast. All other notations are self-explained.

## Mathematical Appendix

**Proof** (Lemma 1): This lemma can be shown following the same reasoning in Lemma 1 of Lahiri and Yang (2015).  $\square$

**Proof** (Theorem 1): According to the central limit theorem for mixing sequences (cf. White (2000)),  $\frac{1}{\sqrt{T}} \left( \frac{\sum_{t=1}^T ((Z_t - P_t)^2 - E(Z_t - P_t)^2)}{\sum_{t=1}^T ((Z_t - \pi)^2 - E(Z_t - \pi)^2)} \right)$  converges in distribution to  $N(0, \Omega)$ . Since  $\sqrt{T}(BS_0 - BS_0^*) = \frac{1}{\sqrt{T}} \sum_{t=1}^T ((Z_t - \bar{Z})^2 - E(Z_t - \pi)^2)$ , we have

$$\begin{aligned}
 & \sqrt{T}(BS_0 - BS_0^*) - \frac{1}{\sqrt{T}} \sum_{t=1}^T ((Z_t - \pi)^2 - E(Z_t - \pi)^2) \\
 &= \frac{1}{\sqrt{T}} \sum_{t=1}^T ((Z_t - \bar{Z})^2 - E(Z_t - \pi)^2) - \frac{1}{\sqrt{T}} \sum_{t=1}^T ((Z_t - \pi)^2 - E(Z_t - \pi)^2) \\
 &= \frac{1}{\sqrt{T}} \sum_{t=1}^T ((Z_t - \bar{Z})^2 - (Z_t - \pi)^2) \\
 &= \frac{1}{\sqrt{T}} \sum_{t=1}^T (2Z_t(\pi - \bar{Z}) + \bar{Z}^2 - \pi^2) \\
 &= (\bar{Z} - \pi) \left( \frac{1}{\sqrt{T}} \sum_{t=1}^T (-2Z_t + \bar{Z} + \pi) \right) \\
 &= \frac{1}{\sqrt{T}} \sum_{t=1}^T (Z_t - \pi)(\pi - \bar{Z}).
 \end{aligned}$$

Since  $\pi - \bar{Z} = o_p(1)$  by law of large number and  $\frac{1}{\sqrt{T}} \sum_{t=1}^T (Z_t - \pi) = O_p(1)$  by central limit theorem,  $\sqrt{T}(BS_0 - BS_0^*) - \frac{1}{\sqrt{T}} \sum_{t=1}^T ((Z_t - \pi)^2 - E(Z_t - \pi)^2) = o_p(1)$ . Consequently,  $\sqrt{T} \begin{pmatrix} BS - BS^* \\ BS_0 - BS_0^* \end{pmatrix} \xrightarrow{d} N(0, \Omega)$ .  $\square$

**Proof** (Collorary 1): It is a direct consequence of Theorem 1.  $\square$

**Proof** (Collorary 2): Note that  $BSS = 1 - \frac{BS}{BS_0}$  and  $BSS^* = 1 - \frac{BS^*}{BS_0^*}$ . The result follows from Theorem 1 and the Delta method.  $\square$

**An alternative proof** (Collorary 2): Note that the two competing forecasts to be compared are  $P_t$  and  $\bar{Z}$ . By Diebold and Mariano (1995),  $\sqrt{T} \left( \frac{1}{T} \sum_{t=1}^T ((Z_t - \bar{Z})^2 - (Z_t - P_t)^2) - E(Z_t - \pi)^2 + E(Z_t - P_t)^2 \right) \xrightarrow{d} N(0, \Omega_{11} + \Omega_{22} - 2\Omega_{12})$ . Following the same line of Theorem 1, we

have

$$\sqrt{T} \begin{pmatrix} \frac{1}{T} \sum_{t=1}^T ((Z_t - \bar{Z})^2 - (Z_t - P_t)^2) - E(Z_t - \pi)^2 + E(Z_t - P_t)^2 \\ \frac{1}{T} \sum_{t=1}^T (Z_t - \bar{Z})^2 - E(Z_t - \pi)^2 \end{pmatrix} \xrightarrow{d} N(0, \tilde{\Omega}),$$

where  $\tilde{\Omega} = \begin{pmatrix} \Omega_{11} + \Omega_{22} - 2\Omega_{12} & \Omega_{22} - \Omega_{12} \\ \Omega_{22} - \Omega_{12} & \Omega_{22} \end{pmatrix}$ . By the Delta method,

$$\sqrt{T}(BSS - BSS^*) \xrightarrow{d} N(0, \frac{1}{BS_0^{*2}}(1, -BSS^*)\tilde{\Omega}(1, -BSS^*)').$$

The result follows by noting that

$$\begin{aligned} & \frac{1}{BS_0^{*2}}(1, -BSS^*)\tilde{\Omega}(1, -BSS^*)' = \frac{\Omega_{11} + \Omega_{22}(1 - BSS^*)^2 - 2(1 - BSS^*)\Omega_{12}}{\pi^2(1 - \pi)^2} \\ &= \frac{\Omega_{11} + \Omega_{22} - 2\Omega_{12} + BSS^{*2}\Omega_{22} - 2BSS^*\Omega_{22} + 2BSS^*\Omega_{12}}{\pi^2(1 - \pi)^2} \\ &= \frac{1}{BS_0^{*2}}(\Omega_{11} + \Omega_{22} - 2\Omega_{12} + \Omega_{22} - 2\frac{BS^*}{BS_0^*}\Omega_{22} + \frac{BS^{*2}}{BS_0^{*2}}\Omega_{22} - 2\Omega_{22} + 2\Omega_{12} + 2\frac{BS^*}{BS_0^*}\Omega_{22} - 2\frac{BS^*}{BS_0^*}\Omega_{12}) \\ &= \frac{1}{BS_0^{*2}}(\Omega_{11} + \frac{BS^{*2}}{BS_0^{*2}}\Omega_{22} - 2\frac{BS^*}{BS_0^*}\Omega_{12}). \end{aligned}$$

□

## References

- Andrews, D. W. K. (1991), ‘Heteroskedasticity and Autocorrelation Consistent Covariance Matrix Estimation’, *Econometrica* **59**, 817–858.
- Andrews, D. W. K. and Monahan, J. C. (1992), ‘An Improved Heteroskedasticity and Autocorrelation Consistent Covariance Matrix Estimator’, *Econometrica* **60**, 953–966.
- Bradley, A. A., Schwartz, S. S. and Hashino, T. (2008), ‘Sampling Uncertainty and Confidence Intervals for the Brier Score and Brier Skill Score’, *Weather and Forecasting* **23**, 992–1006.
- Brier, G. W. (1950), ‘Verification of Forecasts Expressed in Terms of Probability’, *Monthly Weather Review* **78**, 1–3.
- Diebold, F. X. and Mariano, R. S. (1995), ‘Comparing Predictive Accuracy’, *Journal of Business & Economic Statistics* **13**, 253–263.
- Lahiri, K., Monokroussos, G. and Zhao, Y. (2013), ‘The Yield Spread Puzzle and the Information Content of SPF Forecasts’, *Economics Letters* **118**, 219–221.
- Lahiri, K. and Wang, J. G. (2013), ‘Evaluating Probability Forecasts for GDP Declines using Alternative Methodologies’, *International Journal of Forecasting* **29**, 175–190.
- Lahiri, K. and Yang, L. (2013), Forecasting Binary Outcomes, in A. Timmermann and G. Elliott, eds, ‘Handbook of Economic Forecasting Volume 2B’, North-Holland Amsterdam, pp. 1025–1106.
- Lahiri, K. and Yang, L. (2015), ‘Confidence Bands for ROC Curves with Serially Dependent Data’, Working paper, Department of Economics, University at Albany, SUNY.
- Levanon, G., Manini, J., Ozyildirim, A., Schaitkin, B. and Tanchua, J. (2014), ‘Using Financial Indicators to Predict Turning Points in the Business Cycle: The Case of the Leading Economic Index for the United States’, *International Journal of Forecasting* forthcoming.

- Lopez, J. A. (2001), ‘Evaluating the Predictive Accuracy of Volatility Models’, *Journal of Forecasting* **20**, 87–109.
- Newey, W. K. and West, K. D. (1987), ‘A Simple, Positive Semi-definite, Heteroskedasticity and Autocorrelation Consistent Covariance Matrix’, *Econometrica* **55**, 703–708.
- Newey, W. K. and West, K. D. (1994), ‘Automatic Lag Selection in Covariance Matrix Estimation’, *Review of Economic Studies* **61**, 631–653.
- Pesaran, M. H. and Timmermann, A. (2009), ‘Testing Dependence among Serially Correlated Multi-Category Variables’, *Journal of the American Statistical Association* **104**, 325–337.
- Rudebusch, G. D. and Williams, J. C. (2009), ‘Forecasting Recessions: The Puzzle of the Enduring Power of the Yield Curve’, *Journal of Business & Economic Statistics* **27**, 492–503.
- Stephenson, D. B. (2000), ‘Use of the ‘Odds Ratio’ for Diagnosing Forecast Skill’, *Weather Forecasting* **15**, 221–232.
- White, H. (2000), *Asymptotic Theory for Econometricians*, Academic Press.
- Wilks, D. S. (2010), ‘Sampling Distributions of the Brier Score and Brier Skill Score under Serial Dependence’, *Quarterly Journal of the Royal Meteorological Society* **136**, 2109–2118.

## R Code

```
library(RODBC)
library(sandwich)
channel <- odbcConnectExcel("Datapath")
spf <- sqlFetch(channel,"sheetname")
P <- -spf[,1]
Z <- -spf[,2]
obs <- nrow(spf)
alpha <- 0.05
BS <- -mean((Z - P)^2)
BS0 <- -mean(Z) * (1 - mean(Z))
BSS <- -1 - BS/BS0
Gradient <- matrix(nrow = obs, ncol = 2)
for(i in 1 : obs) Gradient[i,] = c((Z[i] - P[i])^2, (Z[i] - mean(Z))^2)
AsyCovI = cov(Gradient)/obs
AsyCov = lrvar(Gradient, type = "Andrews", prewhite = TRUE, adjust = TRUE)
c(BS - qnorm(1 - alpha/2) * sqrt(AsyCovI[1,1]), BS + qnorm(1 - alpha/2) *
sqrt(AsyCovI[1,1]))
c(BS - qnorm(1 - alpha/2) * sqrt(AsyCov[1,1]), BS + qnorm(1 - alpha/2) *
sqrt(AsyCov[1,1]))
c(BSS - qnorm(1 - alpha/2) * sqrt(1/BS0^2 * (AsyCovI[1,1] + (BS/BS0)^2 * AsyCovI[2,2] -
2 * BS/BS0 * AsyCovI[1,2])), BSS + qnorm(1 - alpha/2) * sqrt(1/BS0^2 * (AsyCovI[1,1] +
(BS/BS0)^2 * AsyCovI[2,2] - 2 * BS/BS0 * AsyCovI[1,2])))
c(BSS - qnorm(1 - alpha/2) * sqrt(1/BS0^2 * (AsyCov[1,1] + (BS/BS0)^2 * AsyCov[2,2] -
2 * BS/BS0 * AsyCov[1,2])), BSS + qnorm(1 - alpha/2) * sqrt(1/BS0^2 * (AsyCov[1,1] +
(BS/BS0)^2 * AsyCov[2,2] - 2 * BS/BS0 * AsyCov[1,2])))
```