Australian National University

Crawford School of Public Policy

CAMA

Centre for Applied Macroeconomic Analysis

Getting the ROC into Sync

CAMA Working Paper 1/2022
January 2022

Liu Yang
School of Economics, Nanjing University

Kajal Lahiri
Department of Economics, University at Albany, SUNY

Adrian Pagan
School of Economics, University of Sydney
Centre for Applied Macroeconomic Analysis, ANU

Abstract

Judging the conformity of binary events in macroeconomics and finance has often been done with indices that measure synchronization. In recent years, the use of Receiver Operating Characteristic (ROC) curve has become popular for this task. This paper shows that the ROC and synchronization approaches are closely related, and each can be represented as a weighted average of correlation coefficients between a set of binary indicators and the target event. An advantage of such a representation is that inferences on the degree of conformity can be made robust to serial dependence in the underlying series in the standard framework of a linear regression model. Such serial correlation is common in macroeconomic and financial data.

Keywords

Receiver operating characteristic curve, Synchronization, Correlation, Economic recession, Serial dependence.

JEL Classification

C14, C52, C53, E37

Address for correspondence:

(E) cama.admin@anu.edu.au

ISSN 2206-0332

Getting the ROC into Sync

Liu Yang

School of Economics, Nanjing University
Kajal Lahiri
Department of Economics, University at Albany, SUNY
Adrian Pagan
School of Economics, University of Sydney

January 7, 2022

Abstract

Judging the conformity of binary events in macroeconomics and finance has often been done with indices that measure synchronization. In recent years, the use of Receiver Operating Characteristic (ROC) curve has become popular for this task. This paper shows that the ROC and synchronization approaches are closely related, and each can be represented as a weighted average of correlation coefficients between a set of binary indicators and the target event. An advantage of such a representation is that inferences on the degree of conformity can be made robust to serial dependence in the underlying series in the standard framework of a linear regression model. Such serial correlation is common in macroeconomic and financial data.

JEL Classifications: C14, C52, C53, E37

Keywords: Receiver operating characteristic curve, Synchronization, Correlation, Economic recession, Serial dependence.

1 Introduction

A large literature has emerged on evaluating how well some indicator matches a binary outcome. Judging this conformity has long been an issue with diagnostic testing in medicine, where an individual either receives a drug or not and the outcome that is observed is the effect on the recipient. One wants to know whether the use of the drug produces a good outcome in order to decide if the drug should be used or not. There have been applications of this setup in economics as well, e.g. using credit scores to predict bankruptcy or asking if certain readily available indicators can signal a financial crisis or a recession. The latter was the concern of Schularick and Taylor (2012) and Berge and Jordà (2011). Consequently the focus in these studies is on trying to assess the utility of certain series for matching a binary outcome, mainly with the idea of real time prediction of the latter.

More generally, consider a Bernoulli random variable Z which captures some outcome, say $Z = 1$ is a recession. Suppose we have information on some continuous variable Y that can be used to construct a binary indicator $I(\tau) = \mathbb{I}(Y \leq \tau)$, where τ is some threshold and $\mathbb{I}(\cdot)$ is the indicator function, which equals one if the condition in the parenthesis holds, and zero otherwise. The question of primary interest is then how well $I(\tau)$ conforms to Z.

One popular method for evaluating the utility of the indicator for this task involves calculating the *Receiver Operating Characteristic* (ROC) curve—both Schularick and Taylor (2012) and Berge and Jordà (2011) did this. Basically, it involves asking if there is a τ that would produce a successful match of $I(\tau)$ with Z. The ROC curve looks at the relation between the hit rate $H(\tau)$ (true positives) and the false alarm rate $F(\tau)$ (false positives) as the threshold τ varies. These constituents are

$$\begin{aligned} H(\tau) \equiv& Pr(Y \leq \tau | Z = 1) \\ F(\tau) \equiv& Pr(Y \leq \tau | Z = 0). \end{aligned} \tag{1}$$

The question is how to utilize the information contained in these components in a useful way. The ROC curve is just a graph of $H(\tau)$ against $F(\tau)$. Rather than a graph, one seeks some summary measures. One of these that has been used is what is referred to as the *area under the ROC curve* ($AUROC$). When that is 0.5, it is said that $I(\tau)$ is uninformative about Z, and that points to tests for whether the empirical values of $AUROC$ depart from 0.5.

A different literature has asked whether $I(\tau)$ and Z are synchronized. One way to proceed is to use a concordance correlation coefficient (Lin, 1989). Harding and Pagan (2006) looked at concordance between the two series to measure the conformity. A drawback associated with the concordance index is that it could take a high value even when $I(\tau)$ is independent of Z. However, one can easily re-express the concordance index in terms of the Pearson correlation coefficient between $I(\tau)$ and Z. Unlike concordance, the magnitude of the correlation is always zero when $I(\tau)$ and Z are independent and does not depend upon the occurrence probability of the target event. Due to these nice properties, the correlation coefficient is perceived as a very useful description of how well $I(\tau)$ is associated with Z.

An interesting question would seem to be what the relationship between the ROC and synchronization approaches is? That is the subject of the current paper. In particular, we will demonstrate that both the ROC constituents, i.e. $H(\tau)$ and $F(\tau)$, and the concordance index are closely connected with the correlation coefficient between $I(\tau)$ and Z at a given threshold value τ. When one's interest lies in the global conformity of Y and Z over all threshold values, there is a need to work with some "average" measure (over τ), rather than just a single value of τ. Accordingly, it is shown that a number of global evaluation metrics can be converted into a weighted average of the correlation coefficients. Specifically, we consider four global measures, namely, (i) the area under the correlation curve ($AUCOR$), (i) the area under the concordance curve ($AUCON$), (iii) the area under the Kuipers-score curve ($AUKSC$) and (iv) the area under the ROC curve ($AUROC$). For each measure, we determine what the corresponding weighting scheme is and discuss its implication for the choice of accuracy measures in practice.

Is there any advantage to converting the ROC approach into one that focuses on a correlation coefficient? This paper argues there is. Firstly, statistical inference on a weighted average of correlations just involves method of moments. In particular, the parameter of interest is represented as the slope coefficient of a linear regression model and thereby can be estimated by ordinary least square (OLS) method. Secondly, most macroeconomic and financial data exhibit serial dependence, and this has to be allowed for when making inferences. How to do this is well known in method of moments and regression frameworks. In looking at synchronization of cycles, Harding and Pagan (2006) showed that conclusions about the degree of synchronization could be very different once dependence in data was allowed for. Although some adjustments for serial correlation have been proposed in the ROC literature, they are not yet fully developed. By and large it assumes that the series are independently distributed ($i.d$). One exception is Lahiri and Yang (2018) who constructed various autocorrelation-robust confidence bands for ROC curves in a parametric bi-normal model. This study takes a further step by accommodating serial dependence in a non-parametric manner since our estimators are free of any distributional assumption.

The rest of the paper is organized as follows. Section 2 demonstrates the connection between the methodologies and we convert each index into a Pearson correlation coefficient at a given threshold value. To measure global accuracy, each index can be aggregated over all threshold values. We show that such measure can be interpreted as a weighted average of correlation coefficients. The weighting schemes for the four global accuracy measures mentioned above are given in Section 3. Section 4 provides a unified framework to conduct statistical inference on them in the presence of serial dependence and the methodology is applied in Section 5 to assess the performance of the yield spread in signalling economic recessions. Section 6 concludes.

2 The Methodologies of ROC and Sync

2.1 The ROC Approach

As mentioned before, the ROC curve depends on the two elements $H(\tau)$ and $F(\tau)$ defined in (1). $H(\tau)$ is the probability that an observation is correctly classified when $Z = 1$ and $F(\tau)$ is the probability that an observation is misclassified when $Z = 0$. A ROC curve is generated by plotting the pair $(F(\tau), H(\tau))$ in a unit square for every $\tau \in [a, b]$, where $[a, b]$ is the support of Y and we assume $\mathscr{R} \equiv b - a \in (0, \infty)$. Figure 1 presents three ROC curves summarizing the effectiveness of the yield spread in signalling upcoming economic recessions as the forecast horizon grows. In terms of the classification accuracy, the yield spread performs best when predicting recessions 9 months in the future (M9) compared with M0 or M3.

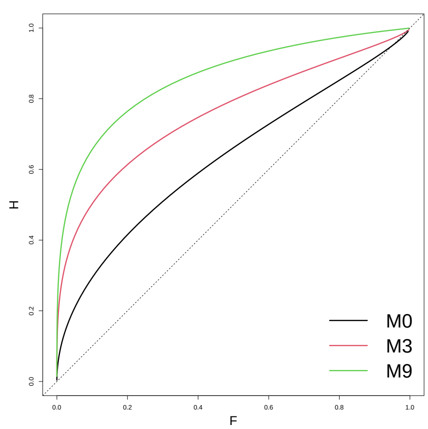


Figure 1: Estimated ROC curves at three forecast horizons

Sometimes, we may only want a single index to summarize all the information contained in a ROC curve. Amongst all existing measures, the area under the ROC curve is probably the most commonly used global index of diagnostic accuracy. As Lieli and Hsu (2019) said, "$AUROC$ is ubiquitous as a statistic to characterize the overall predictive power of binary forecasting models much like an R^2 statistic is used to gauge the fit of a linear regression".

If the ROC curve is continuous, $AUROC$ can be represented by the following integral

$$AUROC = \int_0^1 ROC(u)du, \tag{2}$$

where $ROC(\cdot)$ is the functional form of the ROC curve. For simplicity, we assume both $H(\tau)$ and $F(\tau)$ are strictly increasing with respect to τ so that $\tau = F^{-1}(u)$ and $ROC(u) = H(\tau) = H(F^{-1}(u))$ for $u \in [0, 1]$. By the change-of-variable formula, (2) can be rewritten as

$$AUROC = \int_a^b H(\tau)F'(\tau)d\tau \tag{3}$$

when $F(\tau)$ is continuously differentiable in τ. In particular, a comparison of $AUROC$ to 0.5 is commonly done, and we return to the rationale for this later.

2.2 The Methodology of Synchronization

Now it is useful to look at what the synchronization literature would produce. For a given τ, this works with the *concordance index* between $I(\tau)$ and Z. This is the probability that $I(\tau)$ and Z are jointly zero or unity, namely, $C(\tau) \equiv Pr(I(\tau) = Z)$. Lin (1989) defined concordance as the value $E[(I(\tau) - Z)^2]$, which can be re-formulated as

$$\begin{aligned}
&E[(I(\tau) - Z)^2] = E[I^2(\tau) + Z^2 - 2ZI(\tau)] = E[I(\tau) + Z - 2ZI(\tau)] \\
=&Pr(I(\tau) = 1) + Pr(Z = 1) - 2Pr(Z = 1, I(\tau) = 1) \\
=&Pr(I(\tau) = 1) - Pr(Z = 1, I(\tau) = 1) + Pr(Z = 1) - Pr(Z = 1, I(\tau) = 1) \\
=&Pr(Z = 0, I(\tau) = 1) + Pr(Z = 1, I(\tau) = 0) \\
=&Pr(I(\tau) \neq Z) = 1 - C(\tau).
\end{aligned} \tag{4}$$

So the two are linearly related. In the subsequent analysis, we always use μ_X (σ_X) to denote the population mean (standard error) of a random variable X. It follows from (4) that

$$
\begin{aligned}
C(\tau) = 1 - E[(I(\tau) - Z)^2] = 1 - \mu_{I(\tau)} - \mu_Z + 2E[ZI(\tau)] \\
=& 1 - \mu_{I(\tau)} - \mu_Z + 2\left[Cov(I(\tau), Z) + \mu_{I(\tau)}\mu_Z\right] \\
=& 1 - \mu_{I(\tau)} - \mu_Z + 2\mu_{I(\tau)}\mu_Z + 2\rho(\tau)\sigma_{I(\tau)}\sigma_Z,
\end{aligned} \tag{5}
$$

where $\rho(\tau)$ is the Pearson correlation coefficient between $I(\tau)$ and Z. It is clearly the case that the concordance index is positively related to $\rho(\tau)$. Harding and Pagan (2006) tested if $I(\tau)$ and Z are independent by testing if $\rho(\tau)$ is zero. One notices that even when the correlation is zero, $C(\tau) = 1 - \mu_{I(\tau)} - \mu_Z + 2\mu_{I(\tau)}\mu_Z$, which clearly depends on the marginal distributions of $I(\tau)$ and Z and is not zero in general.

Another popular measure of association is the so-called *Kuipers score*, which is defined as

$$
\begin{aligned}
KS(\tau) = H(\tau) - F(\tau) \\
=& \frac{Pr(I(\tau) = 1, Z = 1)}{Pr(Z = 1)} - \frac{Pr(I(\tau) = 1, Z = 0)}{Pr(Z = 0)} \\
=& \frac{(1 - \mu_Z)Pr(I(\tau) = 1, Z = 1) - \mu_Z Pr(I(\tau) = 1, Z = 0)}{\mu_Z(1 - \mu_Z)} \\
=& \frac{(1 - \mu_Z)E[I(\tau)Z] - \mu_Z E[I(\tau)(1 - Z)]}{\mu_Z(1 - \mu_Z)} \\
=& \frac{E[I(\tau)Z] - \mu_{I(\tau)}\mu_Z}{\mu_Z(1 - \mu_Z)} = \frac{Cov(I(\tau), Z)}{\sigma_Z^2} \\
=& \frac{\rho(\tau)\sigma_{I(\tau)}\sigma_Z}{\sigma_Z^2} = \frac{\sigma_{I(\tau)}}{\sigma_Z}\rho(\tau),
\end{aligned} \tag{6}
$$

which is also positively related to the correlation.

For later reference we observe that there is a moment condition defining $\rho(\tau)$, namely, $\rho(\tau)\sigma_{I(\tau)}\sigma_Z = E[I(\tau)Z] - \mu_{I(\tau)}\mu_Z$. Because $\sigma_{I(\tau)} = \sqrt{(1 - \mu_{I(\tau)})\mu_{I(\tau)}}$ and $\sigma_Z = \sqrt{(1 - \mu_Z)\mu_Z}$, so we need two more moment conditions, namely, $E(I(\tau)) = \mu_{I(\tau)}$ and $E(Z) = \mu_Z$, to complete the set of moments. Method of moments can be used to produce estimates of $\hat{\rho}(\tau), \hat{\mu}_{I(\tau)}$ and $\hat{\mu}_Z$. The distribution of $\sqrt{T}(\hat{\rho}(\tau) - \rho(\tau))$ then follows from standard theory and the inferences can be made robust to both serial correlation and heteroskedasticity in $I(\tau)$ and Z.[1] An alternative way of estimating and testing values of $\rho(\tau)$ is to define $I^*(\tau) = (I(\tau) - \mu_{I(\tau)})/\sigma_{I(\tau)}, Z^* = (Z - \mu_Z)/\sigma_Z$ and then regress the sample values Z_t^* on $I_t^*(\tau)$ (without a constant). The resulting coefficient estimate is $\hat{\rho}(\tau)$ and so inferences can be made robust to serial dependence and heteroskedasticity in the data. This was the way it was motivated in Harding and Pagan (2006).

2.3 Comparing ROC and Sync

In order to compare the two approaches, we need to express $C(\tau)$ in terms of the ROC components in (1). Note that

$$C(\tau) = Pr(I(\tau) = Z) = Pr(I(\tau) = 1, Z = 1) + Pr(I(\tau) = 0, Z = 0)$$
$$= Pr(I(\tau) = 1|Z = 1)Pr(Z = 1) + Pr(I(\tau) = 0|Z = 0)Pr(Z = 0)$$
$$= H(\tau)\mu_Z + (1 - F(\tau))(1 - \mu_Z)$$
$$= \mu_Z H(\tau) - (1 - \mu_Z)F(\tau) - \mu_Z + 1,$$

showing that they are linearly related for a given τ. So concordance combines the two types of outcomes, i.e. correct hits and false alarms, in a particular way. Now we would ultimately like to compare concordances to some benchmark value as τ changes. Within the ROC literature, a suggestion has been to compare the ROC curve to a 45 degree line. On this line, $H(\tau) = F(\tau)$ so that $KS(\tau) = \rho(\tau) = 0$, which further implies $I(\tau)$ and Z are independent by (6). Therefore, the correlation between $I(\tau)$ and Z essentially measures the departure of the ROC curve from the 45 degree line.

Rather than choose a benchmark that varies with τ, Berge and Jordà (2011) used a utility function where the benefits of hits equal the costs of misses in magnitude, and chose the τ that maximizes such a function. So τ optimizes $2\mu_Z H(\tau) - 2(1 - \mu_Z)F(\tau) - 1$ (see their equation on page 268), which is just a linear function of concordance. So concordance

[1] Newey (1984) is a good reference.

involves treating the hit and false alarm rates as having equal utility in absolute value. The τ that maximizes such a utility function will be the τ that maximizes concordance $C(\tau)$ and this is identical to the value which Berge and Jordà used in their work.

3 Global Accuracy Measures Based on Correlations

So far, we have re-expressed various evaluation metrics at a given threshold τ in terms of the correlation between $I(\tau)$ and Z. Rather than choose a single value for τ, one might follow the ROC literature and seek to work with some average over τ. That is, one might look at the area under the concordance curve ($AUCON$) defined as

$$\begin{aligned} AUCON \equiv & \int_a^b C(\tau) d\tau \\ = & \int_a^b \left[1 - \mu_{I(\tau)} - \mu_Z + 2\mu_{I(\tau)}\mu_Z + 2\rho(\tau)\sigma_{I(\tau)}\sigma_Z\right] d\tau \\ = & \mathscr{R}(1-\mu_Z) + (2\mu_Z - 1)\int_a^b \mu_{I(\tau)} d\tau + 2\sigma_Z \int_a^b \rho(\tau)\sigma_{I(\tau)} d\tau \\ = & \mathscr{R}(1-\mu_Z) + (2\mu_Z - 1)\int_a^b [1 - Pr(Y > \tau)] d\tau + 2\sigma_Z \int_a^b \rho(\tau)\sigma_{I(\tau)} d\tau \\ = & \mathscr{R}\mu_Z - (2\mu_Z - 1)\int_a^b Pr(Y > \tau) d\tau + 2\sigma_Z \int_a^b \rho(\tau)\sigma_{I(\tau)} d\tau \\ = & \mathscr{R}\mu_Z - (2\mu_Z - 1)(\mu_Y - a) + 2\sigma_Z A \int_a^b \rho(\tau)\frac{\sigma_{I(\tau)}}{A} d\tau, \end{aligned} \tag{7}$$

where $A \equiv \int_a^b \sigma_{I(\tau)} d\tau$ is the normalizing constant. The area under the Kuipers-score curve ($AUKSC$) is defined as

$$AUKSC \equiv \int_a^b KS(\tau) d\tau = \frac{1}{\sigma_Z}\int_a^b \rho(\tau)\sigma_{I(\tau)} d\tau = \frac{A}{\sigma_Z}\int_a^b \rho(\tau)\frac{\sigma_{I(\tau)}}{A} d\tau. \tag{8}$$

It is worth stressing that even though $AUCON$ and $AUKSC$ are linear functions of the same weighted average of correlations, they have different intercepts and slopes and hence will assume different numerical values. For $AUCON$, the intercept is $\mathscr{R}\mu_Z - (2\mu_Z - 1)(\mu_Y - a)$ while the slope is $2\sigma_Z A$. For $AUKSC$, the intercept is zero while the slope is A/σ_Z. Note that both the intercept and slope for each accuracy measure only capture the features of the marginal distributions of Y and Z, and they do not contain any information on how well Y is associated with Z. For the purpose of measuring conformity of Y and Z, it is only the weighted average of correlations that matters.

The area under the correlation curve ($AUCOR$) is defined as

$$AUCOR \equiv \int_a^b \rho(\tau)d\tau = \mathscr{R}\int_a^b \rho(\tau)\frac{1}{\mathscr{R}}d\tau. \tag{9}$$

Finally, the area under the ROC curve is

$$\begin{aligned} AUROC &= \int_a^b H(\tau)F'(\tau)d\tau \\ &= \int_a^b F(\tau)F'(\tau)d\tau + \int_a^b KS(\tau)F'(\tau)d\tau \\ &= \frac{1}{2} + \frac{1}{\sigma_Z}\int_a^b \rho(\tau)\sigma_{I(\tau)}F'(\tau)d\tau \\ &= \frac{1}{2} + \frac{B}{\sigma_Z}\int_a^b \rho(\tau)\frac{\sigma_{I(\tau)}F'(\tau)}{B}d\tau, \end{aligned} \tag{10}$$

where $B \equiv \int_a^b \sigma_{I(\tau)}F'(\tau)d\tau$.

To sum up, all global conformity measures captured by the areas in (7)-(10) can be represented as $E + G\int_a^b \rho(\tau)\omega(\tau)d\tau$, where $\omega(\tau) \geq 0$ and $\int_a^b \omega(\tau)d\tau = 1$.[2] They differ only in the use of different weighting functions. For $AUCOR$, $\omega(\tau) = 1/\mathscr{R}$ is a constant, meaning that the conformity of $I(\tau)$ with Z, as measured by the correlation $\rho(\tau)$, is equally important for all τ. For $AUCON$ and $AUKSC$,

$$\omega(\tau) = \frac{\sigma_{I(\tau)}}{A} = \frac{\sqrt{\mu_{I(\tau)} - \mu^2_{I(\tau)}}}{A} = \frac{\sqrt{Pr(Y \leq \tau) - Pr^2(Y \leq \tau)}}{A}, \tag{11}$$

which clearly depends on τ. When τ lies in either tail of the distribution of Y, $\omega(\tau)$ is close to zero. Conversely, the maximum $\omega(\tau)$ is achieved when $Pr(Y \leq \tau) = 0.5$, i.e. τ is the median of Y. So the conformity of $I(\tau)$ with Z in the center of the distribution is more important than those in both tails. Finally, the weighting scheme of $AUROC$ is

$$\omega(\tau) = \frac{\sigma_{I(\tau)}F'(\tau)}{B} = \frac{\sqrt{Pr(Y \leq \tau) - Pr^2(Y \leq \tau)}f(\tau)}{B}, \tag{12}$$

[2]For evaluating conformity, both the intercept E and the slope G can be ignored since they only reflect the marginal information on Y and Z.

where $f(\cdot) \equiv F'(\cdot)$. Since $F(\tau) = Pr(Y \leq \tau | Z = 0)$ is the conditional distribution function of Y given $Z = 0$, $f(\tau)$ must be the corresponding conditional density (if it exists). Again, $\omega(\tau)$ is small in either tail. However, the median of Y may not be the maximizer of $\omega(\tau)$ although it is still the maximizer of $\sqrt{Pr(Y \leq \tau) - Pr^2(Y \leq \tau)}$. In the case of $AUROC$, the value of τ at which the conformity of $I(\tau)$ with Z receives the largest weight not only depends on the median of Y but also depends on the conditional mode of Y given $Z = 0$, i.e. the value of τ maximizing $f(\tau)$. When these two are different, the optimal τ is likely to lie between them.

The merit of representing various global measures in the unified form above is three-fold. First, it maps the range of the original measure into the closed interval $[-1, 1]$ to make different measures comparable in magnitude. Second, it suggests a natural way of selecting an appropriate evaluation metric in a given empirical setting. Before assessing the utility of a predictor, a forecast user has to determine which threshold value(s) might be important in the current decision-making context. A global measure that assigns the highest weights to these relevant values should be preferred to any alternative that just treats all threshold values as equally important. Finally, all statistical inferences can be carried out in a unified framework instead of dealing with each measure separately, as shown in the next section.

4 Non-parametric Estimators of Global Measures

In this section, we attempt to develop a non-parametric estimator using a partitioning approach for the averaged correlation of the form $\int_a^b \rho(\tau)\omega(\tau)d\tau$. We will show the estimator is asymptotically normally distributed under a set of quite general assumptions. Depending on whether the weighting scheme $\omega(\tau)$ is prescribed or depends on data, the asymptotic variance takes two distinct forms. The reason being that, when $\omega(\tau)$ has to be estimated, extra moment conditions need to be added to those used to define $\rho(\tau)$. Section 4.1 deals with the case when $\omega(\tau)$ is prescribed and Section 4.2 discusses how to estimate $\omega(\tau)$ when it is unknown.

4.1 The Weighting Scheme is Prescribed

Given a sample $\{(Y_t, Z_t) : t = 1, 2, \cdots, T\}$, we first approximate the integral $\int_a^b \rho(\tau)\omega(\tau)d\tau$ by a finite sum. Specifically, let $a = \tau_0 < \tau_1 < \cdots < \tau_M < \tau_{M+1} = b$ be a partition of the support of Y satisfying $\tau_j - \tau_{j-1} = \mathscr{R}/(M+1)$ for a positive integer M. Furthermore, define $\omega^*(\tau_j) = \omega(\tau_j)/\sum_{j=1}^M \omega(\tau_j)$. The finite sum $\sum_{j=1}^M \omega^*(\tau_j)\rho(\tau_j)$ provides a good approximation to $\int_a^b \rho(\tau)\omega(\tau)d\tau$ provided the number of partitions M is sufficiently large.[3]

Now let $I \equiv \sum_{j=1}^M \omega^*(\tau_j)I(\tau_j)/\sigma_{I(\tau_j)}$ and let $Z^* \equiv Z/\sigma_Z$, so that

$$\begin{aligned}
\rho \equiv \frac{Cov(I, Z^*)}{Var(Z^*)} &= Cov\left(\sum_{j=1}^M \frac{\omega^*(\tau_j)I(\tau_j)}{\sigma_{I(\tau_j)}}, \frac{Z}{\sigma_Z}\right) \\
=&\sum_{j=1}^M Cov\left(\frac{\omega^*(\tau_j)I(\tau_j)}{\sigma_{I(\tau_j)}}, \frac{Z}{\sigma_Z}\right) = \sum_{j=1}^M \omega^*(\tau_j)\frac{Cov(Z, I(\tau_j))}{\sigma_Z\sigma_{I(\tau_j)}} \\
=&\sum_{j=1}^M \omega^*(\tau_j)Corr(Z, I(\tau_j)) = \sum_{j=1}^M \omega^*(\tau_j)\rho(\tau_j).
\end{aligned} \tag{13}$$

Consequently, the quantity we want to estimate is the slope coefficient in the linear projection of I on Z^*. One then supplements the moment condition in (13) with moments defining the means of $I(\tau_j)$ and Z. Afterwards, GMM can be applied to produce inferences about the global measure ρ that is robust to heteroskedasticity and serial correlation.

Sometimes, one is interested in comparing two predictors, say Y_1 and Y_2, in terms of how close their global conformities are. Following Diebold and Mariano (1995), this can be done by testing if the difference between the global measures is significantly different from zero. Define $I_k(\tau_j^k)$ as the binary indicator $\mathbb{I}(Y_k \leq \tau_j^k)$ for $k = 1, 2$ and let $I_k \equiv \sum_{j=1}^{M_k} \omega^*(\tau_j^k)I_k(\tau_j^k)/\sigma_{I_k(\tau_j^k)}$, where M_k and τ_j^k depend on k to allow for the possibility that the support of Y_1 may not overlap with that of Y_2. The global accuracy measure for Y_k

[3]By construction, both $\rho(\tau_0)$ and $\rho(\tau_{M+1})$ are zero so they are omitted from the finite sum.

is given by $\rho_k \equiv \sum_{j=1}^{M_k} \omega^*(\tau_j^k)\rho_k(\tau_j^k)$, with $\rho_k(\tau_j^k)$ being the Pearson correlation coefficient between $I_k(\tau_j^k)$ and Z. The question of interest is how to test whether the difference $\rho_1 - \rho_2$ is zero or not. To do this, define

$$\begin{aligned}\varphi \equiv \rho_1 - \rho_2 = &\sum_{j=1}^{M_1} \omega^*(\tau_j^1)\rho_1(\tau_j^1) - \sum_{j=1}^{M_2} \omega^*(\tau_j^2)\rho_2(\tau_j^2) \\ =&Cov\left(\sum_{j=1}^{M_1} \frac{\omega^*(\tau_j^1)I_1(\tau_j^1)}{\sigma_{I_1(\tau_j^1)}}, \frac{Z}{\sigma_Z}\right) - Cov\left(\sum_{j=1}^{M_2} \frac{\omega^*(\tau_j^2)I_2(\tau_j^2)}{\sigma_{I_2(\tau_j^2)}}, \frac{Z}{\sigma_Z}\right) \\ =&Cov\left(\sum_{j=1}^{M_1} \frac{\omega^*(\tau_j^1)I_1(\tau_j^1)}{\sigma_{I_1(\tau_j^1)}} - \sum_{j=1}^{M_2} \frac{\omega^*(\tau_j^2)I_2(\tau_j^2)}{\sigma_{I_2(\tau_j^2)}}, \frac{Z}{\sigma_Z}\right) \\ =&Cov\left(I_1 - I_2, \frac{Z}{\sigma_Z}\right) = \frac{Cov(I_1 - I_2, Z^*)}{Var(Z^*)}.\end{aligned}$$

Once again, this is a moment condition that needs to be augmented by moments capturing the means of $I_k(\tau_j^k)$ and Z. Using the moment conditions to get $\hat{\varphi}$, $\sqrt{T}(\hat{\varphi} - \varphi)$ will be expected to be asymptotically normal, so that $\varphi = 0$ can be tested with HAC standard errors. Because the asymptotic properties of $\hat{\rho}$ and $\hat{\varphi}$ are much the same, we only focus on $\hat{\rho}$ in the rest of the paper.[4]

To implement the above procedure, all unknown quantities in Z^* and I must be replaced with their sample estimates. It is worth emphasizing here that the estimator proposed here does not rely on any distributional assumption about (Y, Z). For this reason, it is a non-parametric estimator, which is expected to be fairly robust compared with its parametric counterpart (Lahiri and Yang, 2018).

As we also said in Section 2.2, one can also think of estimating ρ by running a regression of I on Z^*. The asymptotic theory for the regression-based estimator of ρ relies on the following standard assumptions.

Assumption 1

(i) $\{(Y_t, Z_t) : t = 1, 2, \cdots, T\}$ *is strictly stationary and mixing with a strong mixing coefficient of size* $2r'/(r' - 1)$ *for some* $r' > 1$.

[4]The asymptotics of $\hat{\varphi}$ is available upon request.

(ii) $\mu_Z \in (0,1)$ *and* $Pr(Y \in [a,b]) = 1$ *with* $\mathscr{R} = b - a \in (0, \infty)$.

(iii) The distribution function $\mathbb{F}(y) \equiv Pr(Y \leq y)$ *is strictly increasing for* $y \in [a,b]$.

(iv) For $j = 1, 2, \cdots, M$, $\omega^*(\tau_j) \in [0, \infty)$ *and* $\sum_{j=1}^{M} \omega^*(\tau_j) = 1$.

Assumption 1*(i)* implies that the weak law of large numbers and the central limit theorem are applicable in the current context. Assumption 1*(ii)* rules out the degenerate case of $\mu_Z = 0/1$, which is not interesting from the perspective of forecasting. To facilitate the asymptotic analysis, we restrict the support of Y to be a finite closed interval $[a,b]$. Assumptions 1*(iii)* ensures $\sigma_{I(\tau_j)} > 0$ for $j = 1, 2, \cdots, M$. Under Assumptions 1*(iv)*, ρ is a weighted average of the correlation coefficients.

Theorem 1 establishes the asymptotic normality of the OLS estimator of ρ, which is applicable in the case of $AUCOR$ by noting that $\omega^*(\tau_j) = 1/M$ for each $j = 1, 2, \cdots, M$.

Theorem 1 *Under Assumption 1(i)-(iv),*

$$\sqrt{T}(\hat{\rho} - \rho) \xrightarrow{d} N(0, \Omega),$$

where

$$\begin{aligned}
\Omega &\equiv \frac{1}{\sigma_Z^2} Var\left(\frac{1}{\sqrt{T}} \sum_{t=1}^{T} S_t\right) \\
S_t &\equiv Z_t I_t + C_0 Z_t + \sum_{j=1}^{M} C_j \mathbb{I}(Y_t \leq \tau_j) \\
C_0 &\equiv -\left(\mu_I + \frac{\rho(1 - 2\mu_Z)}{2\sigma_Z}\right) \\
C_j &\equiv \frac{(2\mu_{I(\tau_j)} - 1)\omega^*(\tau_j)}{2\sigma_{I(\tau_j)}^3} E(Z\mathbb{I}(Y \leq \tau_j)) - \frac{\mu_Z \omega^*(\tau_j)}{2\sigma_{I(\tau_j)}(1 - \mu_{I(\tau_j)})}.
\end{aligned}$$

The proof is provided in the accompanying mathematical appendix.

4.2 The Weighting Scheme is Unknown

In general, $\omega^*(\tau_j)$ is unknown and needs to be estimated. For example, in the case of $AUCON$ and $AUKSC$, $\omega^*(\tau_j) = \sigma_{I(\tau_j)} / \sum_{j=1}^{M} \sigma_{I(\tau_j)}$. For these two global measures, it is only necessary to estimate $\sigma_{I(\tau_j)}$ and ρ simultaneously. This can be done by adjoining the moment condition for ρ to those for the means and variances of $I(\tau_j)$ and Z, since $\omega^*(\tau_j)$ can be constructed from those quantities. However, for $AUROC$,

$$\omega^*(\tau_j) = \frac{\omega(\tau_j)}{\sum_{j=1}^{M} \omega(\tau_j)} = \frac{\sigma_{I(\tau_j)} F'(\tau_j)}{\sum_{j=1}^{M} \sigma_{I(\tau_j)} F'(\tau_j)},$$

so that one needs to estimate the conditional distribution function $F(\cdot)$ as well.

More generally, suppose the weights depend on some unknown parameter $\theta^* \in \Theta \subset \mathbb{R}^k$ and denote $\hat{\theta}$ as an estimator of θ^* satisfying the conditions in Assumption 2. This $\hat{\theta}$ is then used to first compute $\omega^*(\tau_j; \hat{\theta})$ and then $\tilde{\rho}$ is found conditional on $\hat{\theta}$. Newey (1984) provides general results for GMM when a sequential estimation strategy is employed. To analyze this estimation strategy in our case, we state Assumption 2.

Assumption 2

(i) Θ *is a non-empty compact set in* $\mathbb{R}^k$ *and* θ^* *is an interior point of* Θ.

(ii) For $j = 1, 2, \cdots, M$, $\omega^*(\tau_j; \theta)$ *is continuously differentiable in a neighbourhood of* θ^*.

(iii) For any $\theta \in \Theta$, $0 \leq \omega^*(\tau_j; \theta) < \infty$ *and* $\sum_{j=1}^{M} \omega^*(\tau_j; \theta) = 1$.

(iv) $\sqrt{T}(\hat{\theta} - \theta^*) = \sum_{t=1}^{T} \eta(Y_t, Z_t, \theta^*)/\sqrt{T} + o_p(1)$, *where* $\eta(y, z, \theta)$ *is a* $\mathbb{R}^k$*-valued function such that* $E[\eta(Y, Z, \theta^*)] = 0$ *and* $E\|\eta(Y, Z, \theta^*)\|^{2+u} < \infty$ *for some* $u > 0$.

Assumption 2*(i)* is quite standard in the literature of M-estimators (Newey and McFadden, 1994). Assumption 2*(ii)* imposes a smoothness restriction on the weighting function $\omega^*(\tau; \theta)$ in a neighbourhood of θ^*. Assumption 2*(iii)* is a natural extension of Assumption 1*(iv)* when the weighting scheme is unknown. Assumption 2*(iv)* ensures $\hat{\theta}$ has an asymptotically linear representation with $\eta(y, z, \theta^*)$ being its influence function.

Theorem 2 below shows that the OLS estimator $\tilde{\rho}$ constructed using the estimated weights $\{\omega^*(\tau_j; \hat{\theta}) : j = 1, 2, \cdots, M\}$ is still asymptotically normally distributed.

Theorem 2 *Under Assumptions 1(i)-(iii) and 2(i)-(iv),*

$$\sqrt{T}(\tilde{\rho} - \rho) \xrightarrow{d} N(0, V),$$

where

$$V \equiv \frac{1}{\sigma_Z^2} Var\left(\frac{1}{\sqrt{T}}\sum_{t=1}^{T}\left(S_t + C_\theta' \eta(Y_t, Z_t, \theta^*)\right)\right)$$

and

$$C_\theta \equiv \sigma_Z \sum_{j=1}^{M} \frac{\partial \omega^*(\tau_j;\theta^*)}{\partial \theta}\rho(\tau_j).$$

The proof is provided in the appendix. Comparing V in Theorem 2 with Ω in Theorem 1, it is clear that using the estimated weights, rather than the true weights, introduces additional uncertainty that is captured by the influence function of the first-step estimator $\hat{\theta}$. The two asymptotic variances are identical when $C_\theta = 0$. In such an instance, one can safely ignore the first-step estimation uncertainty and conduct statistical inferences by conditioning on $\omega^*(\tau_j;\hat{\theta})$. For instance, when Y and Z are independent, we have $\rho(\tau_j) = 0$ for $j = 1, 2, \cdots, M$. As a result, $C_\theta = 0$. A useful implication is that one can construct a t-test for the null hypothesis $\rho = 0$ in the linear regression model whether the weighting scheme is known or not. However, this conclusion might not hold in a general case, where accounting for the first-step estimation error is necessary for a correct inference.

For $AUCON$ and $AUKSC$,

$$\frac{\omega(\tau_j)}{\sum_{j=1}^{M}\omega(\tau_j)} = \frac{\sigma_{I(\tau_j)}}{\sum_{j=1}^{M}\sigma_{I(\tau_j)}} = \frac{\sqrt{Pr(Y \le \tau_j) - Pr^2(Y \le \tau_j)}}{\sum_{j=1}^{M}\sqrt{Pr(Y \le \tau_j) - Pr^2(Y \le \tau_j)}} = \omega^*(\tau_j)$$

and the unknown parameter vector $\theta^* = (\mathbb{F}(\tau_1), \cdots, \mathbb{F}(\tau_M))'$.[5] θ^* can be estimated by $\hat{\theta} = (\hat{\mathbb{F}}(\tau_1), \cdots, \hat{\mathbb{F}}(\tau_M))'$, where

$$\hat{\mathbb{F}}(\tau_j) = \frac{\sum_{t=1}^{T}\mathbb{I}(Y_t \le \tau_j)}{T} \tag{14}$$

for $j = 1, 2, \cdots, M$. It is straightforward to verify that all assumptions in Theorem 2 hold and the OLS estimator $\tilde{\rho}$ is asymptotically normally distributed.

Similarly,

$$\omega^*(\tau_j) = \frac{\omega(\tau_j)}{\sum_{j=1}^{M}\omega(\tau_j)} = \frac{\sigma_{I(\tau_j)}F'(\tau_j)}{\sum_{j=1}^{M}\sigma_{I(\tau_j)}F'(\tau_j)} \approx \frac{\sigma_{I(\tau_j)}(F(\tau_j) - F(\tau_{j-1}))}{\sum_{j=1}^{M}\sigma_{I(\tau_j)}(F(\tau_j) - F(\tau_{j-1}))}$$

[5] Note that both $\mathbb{F}(\tau_0) = Pr(Y \le \tau_0) = \mathbb{F}(a) = 0$ and $\mathbb{F}(\tau_{M+1}) = Pr(Y \le \tau_{M+1}) = \mathbb{F}(b) = 1$ are known.

for $AUROC$, where the last equality uses the fact that $F'(\tau_j)$ can be approximated well by

$$\frac{F(\tau_j) - F(\tau_{j-1})}{\tau_j - \tau_{j-1}} = \frac{F(\tau_j) - F(\tau_{j-1})}{\mathscr{R}/(M+1)}$$

provided M is large enough. The unknown parameter vector is

$$\theta^* = (\mathbb{F}(\tau_1), \cdots, \mathbb{F}(\tau_M), F(\tau_1), \cdots, F(\tau_M))',$$

where $\mathbb{F}(\tau_j)$ is estimated by $\hat{\mathbb{F}}(\tau_j)$ in (14) and

$$\hat{F}(\tau_j) = \frac{\sum_{t=1}^{T}(1 - Z_t)\mathbb{I}(Y_t \leq \tau_j)}{\sum_{t=1}^{T}(1 - Z_t)}$$

is a natural estimator of $F(\tau_j)$ for $j = 1, 2, \cdots, M$. Again, Theorem 2 can be applied to find the asymptotic distribution of $\tilde{\rho}$.

4.3 Estimating Asymptotic Variances

The asymptotic normality results in Theorems 1 and 2 are essential for constructing confidence intervals and statistics appropriate for testing hypotheses. To this end, we need consistent estimators of the unknown asymptotic variances Ω and V. It is straightforward to estimate each population moment by the corresponding sample average. In particular, μ_Z is estimated by $\hat{\mu}_Z = \sum_{t=1}^{T} Z_t/T$ while $\mu_{I(\tau_j)}$ is estimated by $\hat{\mathbb{F}}(\tau_j)$ in (14).[6] So in the situation where the weighting scheme is prescribed, $\hat{\mu}_I = \sum_{j=1}^{M} \omega^*(\tau_j)\hat{\mathbb{F}}(\tau_j)/\hat{\sigma}_{I(\tau_j)}$ is an estimate of μ_I. In cases where $\omega^*(\tau_j)$ is unknown, we can modify the estimators by replacing $\omega^*(\tau_j)$ with $\omega^*(\tau_j; \hat{\theta})$, that is, $\tilde{\mu}_I = \sum_{j=1}^{M} \omega^*(\tau_j; \hat{\theta})\hat{\mathbb{F}}(\tau_j)/\hat{\sigma}_{I(\tau_j)}$.

Having obtained estimators for all population moments, the long-run variance Ω (V) can be estimated by the class of *Heteroskedasticity and Autocorrelation Consistent* (HAC) estimators. One leading example is the standard Newey-West type estimator defined as

$$\hat{\Omega} \equiv \sum_{l=-L}^{L} \left(1 - \frac{|l|}{L+1}\right) \hat{\Gamma}_T(l),$$

[6] Since Z and $\mathbb{I}(Y_t \leq \tau_j)$ are Bernoulli random variables, the corresponding variance estimates are $\hat{\sigma}_Z^2 = \hat{\mu}_Z(1 - \hat{\mu}_Z)$ and $\hat{\sigma}^2_{I(\tau_j)} = \hat{\mathbb{F}}(\tau_j)(1 - \hat{\mathbb{F}}(\tau_j))$, respectively.

where L is sometimes called the lag truncation number (or bandwidth parameter), and $\hat{\Gamma}_T(l)$ is the sample auto-covariance at lag l, that is,

$$\hat{\Gamma}_T(l) \equiv \frac{1}{T-l} \sum_{t=1}^{T-l} (S_t - \bar{S})(S_{t+l} - \bar{S})$$

and $\hat{\Gamma}_T(-l) = \hat{\Gamma}_T(l)$. So $\hat{\Omega}$ can be interpreted as a weighted sum of the sample auto-covariances, with the weight being $1 - |l|/(L+1)$—the so-called *Bartlett* kernel. Of course, other weight functions can be used as well. Andrews (1991) gave a comprehensive list of popular kernel functions.

Newey and West (1987) presented a set of sufficient conditions for $\hat{\Omega}$ to be non-negative and consistent. Basically, the key assumption requires the lag truncation number grow no faster than the sample size. Alternatively, the ratio of the truncation lag to the sample size, often denoted as b, must shrink towards zero at a certain rate. As pointed out by Kiefer and Vogelsang (2005), a major difficulty with this procedure lies in how to select the kernel function and the smoothing parameter b, as the asymptotic distribution of the resulting t statistic does not depend on the choice of kernel and b. They proposed a new asymptotic theory by holding b fixed at a given value. However, their asymptotic distribution is non-standard, and the critical values have to be found by simulation. In order to keep our procedure simple, we are going to rely on the asymptotic F test with a data dependent optimal bandwidth developed by Sun (2013, 2014) to conduct inferences. This has an advantage because of its computational ease.

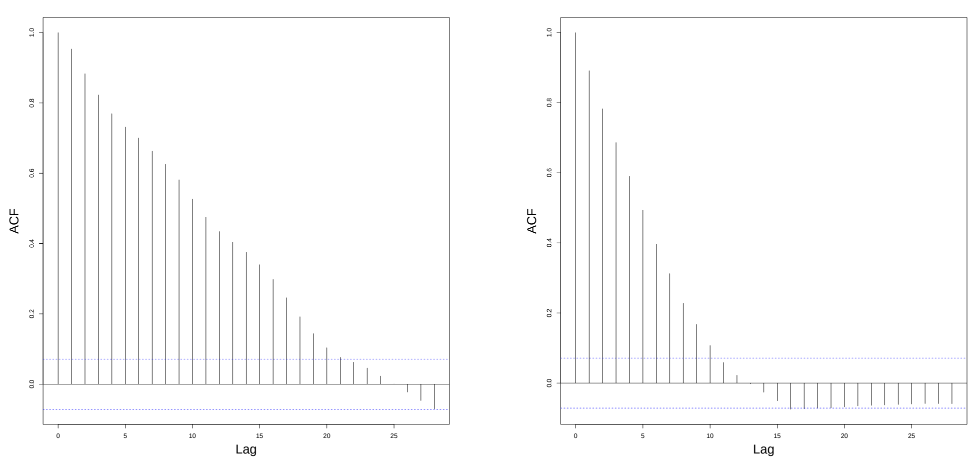
5 An Empirical Illustration

In this section, we will demonstrate the usefulness of the proposed methodology for gauging the effectiveness of the yield spread in signalling future economic recessions. A large body of literature has documented the enduring power of the slope of the yield curve in forecasting real GDP growth and recessions (Ang et al., 2006; Estrella and Mishkin, 1998; Lahiri and

Wang, 1996; Rudebusch and Williams, 2009). In the current illustration, the yield spread Y_t is defined as the difference between the yield on a ten-year U.S. Treasure bond and the federal funds rate. The economic recession is determined by the National Bureau of Economic Research (NBER) business cycle dating committee. That is, we define $Z_t = 1$ if month t is between a business cycle peak and trough identified by the NBER. Our monthly data covers the period from 1959:01 to 2021:11. In Figure 1 (section 2.1), we have presented the ROC curves at three horizons—0, 3 and 9 months.

Figure 2 depicts the sample auto-correlation functions of the yield spread and recession indicator. Clearly, both variables, especially the former, exhibit strong serial correlation. All coefficients up to 21 months lags for the yield spread and 10 months lags for the recession indicator are significantly different from zero. For this reason, we will rely on HAC estimators to conduct statistical inferences and we will then compare the results with those obtained by assuming data is independent.

Figure 2: Auto-correlation functions of the yield spread and recession indicator



(a) Yield spread (b) Recession indicator

Notes: Auto-correlation values outside the dotted band are significantly different from zero at the usual 5% level.

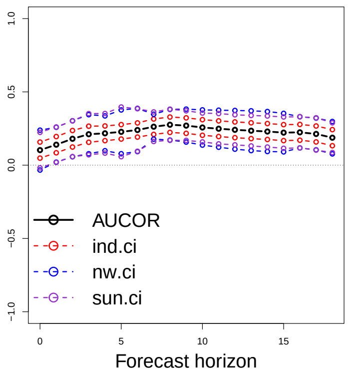
As suggested by existing studies, the yield spread (Y) is negatively related to the occurrence probability of an economic recession. So, we take $\mathbb{I}(Y \leq \tau)$ as the forecasting rule, i.e. a recession alarm is issued when the observed yield spread is below a threshold value τ. An inversion of the yield curve (τ=0) is often used by financial analysts as the threshold. Researchers have found that the predictive accuracy of the yield spread differs depending on the horizons. To allow for this heterogeneity in correlations across models, we will assess the performance of the yield spread during the current month (M0) and each of the next 18 months (M1-M18). Figure 3 summarizes the estimates of the global accuracy measures developed earlier for each forecast horizon. To implement the weighting scheme, we set $M = 100$ - results for $M > 100$ are virtually the same as those in Figure 3. As expected, the predictive power of the yield spread gets stronger as the horizon increases for all of the three global measures. Consistent with previous literature, the maximum predictive accuracy is achieved at around 9-12 month horizon (depending on which global measure is used). Beyond that, the quality of the prediction slightly deteriorates but remains relatively high.

Figure 3 also presents three types of confidence intervals for measures that involve a weighted average of correlations. Among them, the intervals generated by assuming serial independence ("ind.ci") is the narrowest. In contrast, the two HAC-based intervals are considerably wider at each forecast horizon, which is not surprising given the strong evidence of serial dependence in Figure 2. This indicates that ignoring the dependence structure implicit in the data leads to a downward bias in estimating the sampling variability of the proposed estimators. Thus, the 95% intervals when assuming independence in the data exclude zero. This points to a high conformity of the yield spread to the recession indicator at all considered horizons. However, the degree of conformity becomes much less significant when looking at HAC-robust intervals. For $AUCON$ and $AUROC$, the robust intervals up to M6 include zero correlation, so one cannot reject the null hypotheses of zero correlation at short horizons. For longer horizons, the three types of intervals always stay above zero when predicting recessions. Observe that Sun's intervals ("sun.ci") are somewhat wider than those based on the traditional Newey-West type HAC estimators of the long-run variances ("nw.ci"), especially at relatively short horizons. This pattern was also noted by Lahiri and Yang (2018), whose simulation experiment lends some evidence for the superior finite-sample performance of Sun's HAC estimator when data exhibits strong serial dependence. Hence, we recommend the use of Sun's intervals in the current context as well.

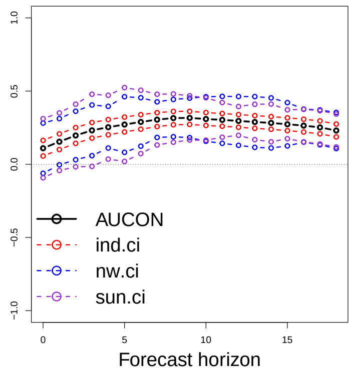
Figure 4 summarizes all of the three global accuracy measures, where we observe $AUCOR < AUCON \simeq AUROC$ uniformly at each horizon, $AUCON > AUROC$ before M10 and $AUCON < AUROC$ after M10. These three measures are weighted averages of a common set of correlation coefficients—they differ only in the use of different weighting schemes. For this reason, one cannot claim $AUCOR$ is worse or less accurate than $AUROC$ simply due to its lower value at each horizon. Whether one measure is superior to the other crucially depends on the forecaster's subjective judgement as to which value of the yield spread is more important, and which is less important. In the absence of such information, it is meaningless to rank various global accuracy measures simply by their magnitudes but we have indicated a way to put them on an equal footing.

Figure 5 illustrates how the weighting schemes of the three global measures differ for four selected forecast horizons. For each horizon, the black solid line is the estimated correlation coefficient $\hat{\rho}(\tau)$ at each value of the threshold. The weighting scheme of $AUCOR$, represented by the red dashed line, is a constant, meaning all correlations are viewed as equally important. The weighting scheme of $AUCON$ (broken blue line) has a single peak at the median of the yield spread, as we have discussed following equation (11). Furthermore, the weighting scheme of $AUCON$ is invariant to the forecast horizon because it is completely determined by the marginal distribution of the yield spread, which remains

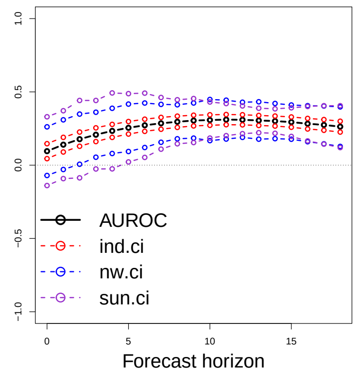
Figure 3: Estimates of global measures and the corresponding 95% confidence intervals



(a) *AUCOR*



(b) *AUCON/AUKSC*



(c) *AUROC*

Notes: "ind.ci" refers to the 95% confidence interval generated by assuming serial independence, while "nw.ci" and "sun.ci" represent the 95% confidence intervals using Newey-West and Sun's HAC estimators of the long-run variances respectively.

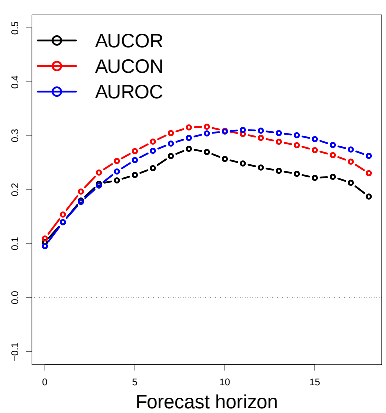
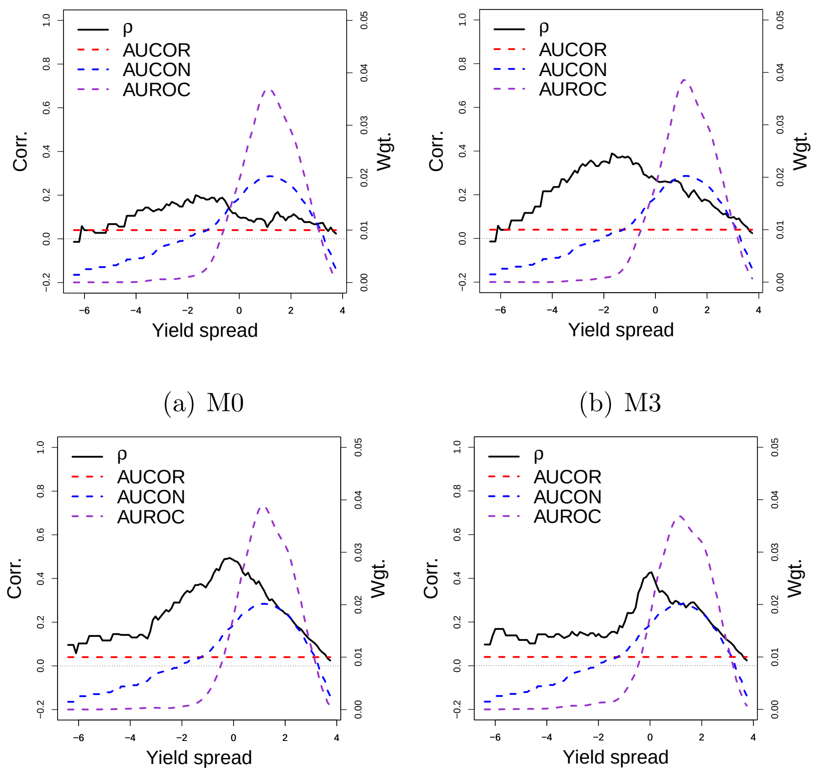


Figure 4: Estimates for three global accuracy measures

unchanged as the horizon increases.

Figure 5: Estimated correlation coefficients and weighting schemes for 4 horizons



(a) M0 (b) M3

(c) M12 (d) M18

Interestingly, the weighting scheme for $AUROC$ (Figure 5) looks very similar to that of $AUCON$ and varies little across forecast horizons, see Figure 5. Theoretically, we know from (12), that $\omega(\tau)$ not only depends on the marginal distribution but also on the conditional distribution of the yield spread given expansion and the latter is horizon specific. However, in our empirical context, since the number of months spent in expansions is much larger than the number of months in recessions (87.4% vs. 12.6%), the conditional density during expansions almost overlaps with the marginal density and stays the same across horizons. In other words, the marginal density of the yield spread is mainly determined by the conditional density during expansions and thus the latter is also insensitive to the change in horizon. In other empirical cases where the two outcomes are more balanced, the weighting scheme of $AUROC$ can be different from that of $AUCON$ and be more responsive to changes in horizon. Also, note that the weighting schemes for $AUCON$ and $AUROC$ do not follow the correlation curve because, as elaborated before, they effectively weigh the hit rate and false alarm rate differently.

In a decision-making scenario, a forecast user often translates the value of the yield spread into a 0/1 binary action by choosing a threshold value τ. As pointed out by Elliott and Lieli (2013), the optimal threshold is completely determined by the user's utility function. As each forecast user likely has a different utility function, the optimal threshold value is user dependent. Once the distribution over the utility functions is specified as in Lieli and Nieto-Barthaburu (2010), one is able to deduce the corresponding distribution of the optimal threshold that serves as a benchmark for comparison. Among a collection of global accuracy measures, the one whose weighting scheme best approximates the benchmark distribution should be preferred.

6 Conclusion

The paper has looked at how the approaches of ROC and synchronization to measuring the conformity of binary events in macroeconomics and finance relate. We have shown that both can be re-expressed in terms of the correlation coefficient between a binary outcome and an indicator. We argue that an important advantage to thinking about conformity in terms of correlation lies in the technical convenience to perform statistical inference that is robust to the serial correlation in the data. Specifically, the standard inferential procedure in a linear regression model is shown to be valid once a HAC estimator of the long-run variance is properly chosen.

To illustrate the empirical relevance of the proposed method, we re-examine the quality of the yield spread in signalling economic recessions across a wide range of horizons. As expected, the conventional confidence intervals based on the independence assumption are shown too narrow to capture the sampling uncertainty. After accounting for the autocorrelation, the yield spread is found to be significantly informative only at horizons beyond six months, while it is always superior to an independent benchmark at all horizons if the serial correlation is ignored.

References

Andrews, D. W. K. (1991), 'Heteroskedasticity and Autocorrelation Consistent Covariance Matrix Estimation', *Econometrica* **59**, 817–858.

Ang, A., Plazzesi, M. and Wei, M. (2006), 'What does the Yield Curve Tell us about GDP Growth?', *Journal of Econometrics* **131**, 359–403.

Berge, T. J. and Jordà, O. (2011), 'Evaluating the Classification of Economic Activity into Recessions and Expansions', *American Economic Journal: Macroeconomics* **3**, 246–277.

Diebold, F. X. and Mariano, R. S. (1995), 'Comparing Predictive Accuracy', *Journal of Business & Economic Statistics* **13**, 253–263.

Doukhan, P. (1994), *Mixing: Properties and Examples*, Springer.

Elliott, G. and Lieli, R. P. (2013), 'Predicting Binary Outcomes', *Journal of Econometrics* **174**, 15–26.

Estrella, A. and Mishkin, F. S. (1998), 'Predicting U.S. Recessions: Financial Variables as Leading Indicators', *The Review of Economics and Statistics* **80**, 45–61.

Harding, D. and Pagan, A. R. (2006), 'Synchronization of Cycles', *Journal of Econometrics* **132**, 59–79.

Kiefer, N. M. and Vogelsang, T. J. (2005), 'A New Asymptotic Theory for Heteroskedasticity-Autocorrelation Robust Tests', *Econometric Theory* **21**, 1130–1164.

Lahiri, K. and Wang, J. G. (1996), Interest Rate Spreads as Predictors of Business Cycles, *in* G. S. Maddala and C. R. Rao, eds, 'Handbook of Statistics 14', North-Holland Amsterdam, 297–315.

Lahiri, K. and Yang, L. (2018), 'Confidence Bands for ROC Curves with Serially Dependent Data', *Journal of Business & Economic Statistics* **36**, 115–130.

Lieli, P. R. and Hsu, Y. (2019), 'Using the Area under an Estimated ROC Curve to Test the Adequacy of Binary Predictors', *Journal of Nonparametric Statistics* **31**, 100–130.

Lieli, R. P. and Nieto-Barthaburu, A. (2010), 'Optimal Binary Prediction for Group Decision Making', *Journal of Business* & *Economic Statistics* **28**, 308–319.

Lin, L. I. (1989), 'A Concordance Correlation Coefficient to Evaluate Reproducibility', *Biometrics* **45**, 255–268.

Newey, W. K. (1984), 'A Method of Moments Interpretation of Sequential Estimators', *Economics Letters* **14**, 201–206.

Newey, W. K. and McFadden, D. L. (1994), Large Sample Estimation and Hypothesis Testing, *in* R. F. Engle and D. L. McFadden, eds, 'Handbook of Econometrics Volume 4', North-Holland Amsterdam.

Newey, W. K. and West, K. D. (1987), 'A Simple, Positive Semi-definite, Heteroskedasticity and Autocorrelation Consistent Covariance Matrix', *Econometrica* **55**, 703–708.

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.

Schularick, M. and Taylor, A. M. (2012), 'Credit Booms Gone Bust: Monetary Policy, Leverage Cycles, and Financial Crises, 18702008', *American Economic Review* **102**, 1029–1061.

Sun, Y. (2013), 'A Heteroskedasticity and Autocorrelation Robust F Test using an Orthonormal Series Variance Estimator', *Econometrics Journal* **16**, 1–26.

Sun, Y. (2014), 'Let's Fix it: Fixed-b Asymptotics versus Small-b Asymptotics in Heteroscedasticity and Autocorrelation Robust Inference', *Journal of Econometrics* **178**, 659–677.

Mathematical Appendix

This appendix accompanies the paper "Getting the ROC into Sync". It contains proofs of all theorems in the text. We start with some additional notations. Let $\hat{\mu}_{I(\tau_j)} = \sum_{t=1}^{T} \mathbb{I}(Y_t \leq \tau_j)/T$ and $\hat{\mu}_Z = \sum_{t=1}^{T} Z_t/T$ be sample means of $I(\tau_j)$ and Z, respectively, with $\hat{\sigma}^2_{I(\tau_j)} = \hat{\mu}_{I(\tau_j)}(1 - \hat{\mu}_{I(\tau_j)})$ and $\hat{\sigma}^2_Z = \hat{\mu}_Z(1 - \hat{\mu}_Z)$ being the corresponding sample variances. Our linear regression models are based on the sample $\{(\hat{Z}_t, \hat{I}_t, \tilde{I}_t) : t = 1, 2, \cdots, T\}$, where $\hat{Z}_t = Z_t/\hat{\sigma}_Z$ is an estimate of $Z_t^* = Z_t/\sigma_Z$, $\hat{I}_t = \sum_{j=1}^{M} \omega^*(\tau_j)\mathbb{I}(Y_t \leq \tau_j)/\hat{\sigma}_{I(\tau_j)}$ and $\tilde{I}_t = \sum_{j=1}^{M} \omega^*(\tau_j; \hat{\theta})\mathbb{I}(Y_t \leq \tau_j)/\hat{\sigma}_{I(\tau_j)}$ are estimates of $I_t = \sum_{j=1}^{M} \omega^*(\tau_j)\mathbb{I}(Y_t \leq \tau_j)/\sigma_{I(\tau_j)}$.

Proof (Theorem 1): The OLS estimator of ρ can be written as

$$\begin{aligned}
\hat{\rho} = &\frac{\sum_{t=1}^{T}(\hat{Z}_t - \hat{\mu}_Z/\hat{\sigma}_Z)\hat{I}_t}{\sum_{t=1}^{T}(\hat{Z}_t - \hat{\mu}_Z/\hat{\sigma}_Z)^2} = \hat{\sigma}_Z \frac{\sum_{t=1}^{T}(Z_t - \hat{\mu}_Z)\hat{I}_t}{\sum_{t=1}^{T}(Z_t - \hat{\mu}_Z)^2} = \hat{\sigma}_Z \frac{\sum_{t=1}^{T} Z_t\hat{I}_t - \hat{\mu}_Z\sum_{t=1}^{T}\hat{I}_t}{\sum_{t=1}^{T}(Z_t^2 - 2Z_t\hat{\mu}_Z + \hat{\mu}_Z^2)} \\
=&\hat{\sigma}_Z \frac{\sum_{t=1}^{T} Z_t\hat{I}_t - \hat{\mu}_Z\sum_{t=1}^{T}\hat{I}_t}{T\hat{\mu}_Z - T\hat{\mu}_Z^2} = \frac{1}{\hat{\sigma}_Z}\frac{\sum_{t=1}^{T} Z_t\hat{I}_t - \hat{\mu}_Z\sum_{t=1}^{T}\hat{I}_t}{T} \\
=&\frac{1}{\hat{\sigma}_Z}\left(\frac{1}{T}\sum_{t=1}^{T} Z_t\hat{I}_t - \hat{\mu}_Z\sum_{j=1}^{M}\frac{\omega^*(\tau_j)\hat{\mu}_{I(\tau_j)}}{\hat{\sigma}_{I(\tau_j)}}\right).
\end{aligned}$$

To show the asymptotic distribution of $\hat{\rho}$, consider the rescaled difference

$$\begin{aligned}
&\sqrt{T}(\hat{\rho} - \rho) = \sqrt{T}\left[\frac{1}{\hat{\sigma}_Z}\left(\frac{1}{T}\sum_{t=1}^{T} Z_t\hat{I}_t - \hat{\mu}_Z\sum_{j=1}^{M}\frac{\omega^*(\tau_j)\hat{\mu}_{I(\tau_j)}}{\hat{\sigma}_{I(\tau_j)}}\right) - \rho\right] \\
=&\frac{\sqrt{T}}{\sigma_Z}\left(\frac{1}{T}\sum_{t=1}^{T} Z_t\hat{I}_t - \hat{\mu}_Z\sum_{j=1}^{M}\frac{\omega^*(\tau_j)\hat{\mu}_{I(\tau_j)}}{\hat{\sigma}_{I(\tau_j)}} - \hat{\sigma}_Z\rho\right)\left(1 + \frac{\sigma_Z - \hat{\sigma}_Z}{\hat{\sigma}_Z}\right) \\
=&\frac{\sqrt{T}}{\sigma_Z}\left(\frac{1}{T}\sum_{t=1}^{T} Z_t\hat{I}_t - \hat{\mu}_Z\sum_{j=1}^{M}\frac{\omega^*(\tau_j)\hat{\mu}_{I(\tau_j)}}{\hat{\sigma}_{I(\tau_j)}} - \hat{\sigma}_Z\rho\right) + h.o \\
=&\frac{\sqrt{T}}{\sigma_Z}\left(\frac{1}{T}\sum_{t=1}^{T} Z_t\hat{I}_t - \hat{\mu}_Z\sum_{j=1}^{M}\sqrt{\frac{\hat{\mu}_{I(\tau_j)}}{1 - \hat{\mu}_{I(\tau_j)}}}\omega^*(\tau_j) - \sqrt{\hat{\mu}_Z(1 - \hat{\mu}_Z)}\rho\right) + h.o,
\end{aligned}$$

where "$h.o$" denotes the higher order term. By Taylor's expansion, we have

$$\begin{aligned}
&\frac{1}{T}\sum_{t=1}^{T} Z_t\hat{I}_t - \hat{\mu}_Z \sum_{j=1}^{M}\sqrt{\frac{\hat{\mu}_{I(\tau_j)}}{1-\hat{\mu}_{I(\tau_j)}}}\omega^*(\tau_j) - \sqrt{\hat{\mu}_Z(1-\hat{\mu}_Z)}\rho \\
=&\frac{1}{T}\sum_{t=1}^{T} Z_t I_t - \mu_Z\mu_I - \sigma_Z\rho - \left(\mu_I + \frac{\rho(1-2\mu_Z)}{2\sigma_Z}\right)(\hat{\mu}_Z - \mu_Z) \\
&+\sum_{j=1}^{M}\left(\frac{(2\mu_{I(\tau_j)}-1)\omega^*(\tau_j)}{2\sigma^3_{I(\tau_j)}}\frac{1}{T}\sum_{t=1}^{T} Z_t\mathbb{I}(Y_t\le\tau_j) - \frac{\mu_Z\omega^*(\tau_j)}{2\sigma_{I(\tau_j)}(1-\mu_{I(\tau_j)})}\right)(\hat{\mu}_{I(\tau_j)} - \mu_{I(\tau_j)}) + h.o \\
=&\frac{1}{T}\sum_{t=1}^{T} Z_t I_t - E(ZI) - \left(\mu_I + \frac{\rho(1-2\mu_Z)}{2\sigma_Z}\right)(\hat{\mu}_Z - \mu_Z) \\
&+\sum_{j=1}^{M}\left(\frac{(2\mu_{I(\tau_j)}-1)\omega^*(\tau_j)}{2\sigma^3_{I(\tau_j)}}E(Z\mathbb{I}(Y\le\tau_j)) - \frac{\mu_Z\omega^*(\tau_j)}{2\sigma_{I(\tau_j)}(1-\mu_{I(\tau_j)})}\right)(\hat{\mu}_{I(\tau_j)} - \mu_{I(\tau_j)}) \\
&+\sum_{j=1}^{M} o_p(1)(\hat{\mu}_{I(\tau_j)} - \mu_{I(\tau_j)}) + h.o \\
=&\frac{1}{T}\sum_{t=1}^{T} Z_t I_t - E(ZI) + C_0(\hat{\mu}_Z - \mu_Z) + \sum_{j=1}^{M} C_j(\hat{\mu}_{I(\tau_j)} - \mu_{I(\tau_j)}) + \sum_{j=1}^{M} o_p(1)(\hat{\mu}_{I(\tau_j)} - \mu_{I(\tau_j)}) + h.o.
\end{aligned}$$

Hence,

$$\begin{aligned}
&\sqrt{T}(\hat{\rho}-\rho) = \frac{\sqrt{T}}{\sigma_Z}\left(\frac{1}{T}\sum_{t=1}^{T} Z_t\hat{I}_t - \hat{\mu}_Z \sum_{j=1}^{M}\sqrt{\frac{\hat{\mu}_{I(\tau_j)}}{1-\hat{\mu}_{I(\tau_j)}}}\omega^*(\tau_j) - \sqrt{\hat{\mu}_Z(1-\hat{\mu}_Z)}\rho\right) + h.o \\
=&\frac{\sqrt{T}}{\sigma_Z}\left(\frac{1}{T}\sum_{t=1}^{T} Z_t I_t - E(ZI) + C_0(\hat{\mu}_Z - \mu_Z) + \sum_{j=1}^{M} C_j(\hat{\mu}_{I(\tau_j)} - \mu_{I(\tau_j)}) + \sum_{j=1}^{M} o_p(1)(\hat{\mu}_{I(\tau_j)} - \mu_{I(\tau_j)})\right) + h.o \\
=&\frac{\sqrt{T}}{\sigma_Z}\left(\frac{1}{T}\sum_{t=1}^{T} Z_t I_t - E(ZI) + \frac{C_0}{T}\sum_{t=1}^{T}(Z_t - \mu_Z) + \sum_{j=1}^{M}\frac{C_j}{T}\sum_{t=1}^{T}(\mathbb{I}(Y_t\le\tau_j) - \mu_{I(\tau_j)})\right) + o_p(1) + h.o \\
=&\frac{\sqrt{T}}{\sigma_Z}\frac{1}{T}\sum_{t=1}^{T}\left(Z_t I_t - E(ZI) + C_0(Z_t - \mu_Z) + \sum_{j=1}^{M} C_j(\mathbb{I}(Y_t\le\tau_j) - \mu_{I(\tau_j)})\right) + o_p(1) + h.o \\
=&\frac{1}{\sigma_Z}\frac{1}{\sqrt{T}}\sum_{t=1}^{T}(S_t - E(S_t)) + h.o.
\end{aligned}$$

The central limit theorem for a mixing sequence (Doukhan, 1994) implies that $\sqrt{T}(\hat{\rho}-\rho) \xrightarrow{d} N(0,\Omega)$. □

Proof (Theorem 2): By definition, the OLS estimator $\tilde{\rho}$ can be written as

$$\tilde{\rho} = \frac{\sum_{t=1}^{T}(\hat{Z}_t - \hat{\mu}_Z/\hat{\sigma}_Z)\tilde{I}_t}{\sum_{t=1}^{T}(\hat{Z}_t - \hat{\mu}_Z/\hat{\sigma}_Z)^2} = \frac{1}{\hat{\sigma}_Z}\left(\frac{1}{T}\sum_{t=1}^{T} Z_t\tilde{I}_t - \hat{\mu}_Z\sum_{j=1}^{M}\frac{\omega^*(\tau_j;\hat{\theta})\hat{\mu}_{I(\tau_j)}}{\hat{\sigma}_{I(\tau_j)}}\right).$$

As in the proof of Theorem 1, we can express the rescaled difference as

$$\sqrt{T}(\tilde{\rho}-\rho) = \frac{\sqrt{T}}{\sigma_Z}\left(\frac{1}{T}\sum_{t=1}^{T} Z_t\tilde{I}_t - \hat{\mu}_Z\sum_{j=1}^{M}\sqrt{\frac{\hat{\mu}_{I(\tau_j)}}{1-\hat{\mu}_{I(\tau_j)}}}\omega^*(\tau_j;\hat{\theta}) - \sqrt{\hat{\mu}_Z(1-\hat{\mu}_Z)}\rho\right) + h.o.$$

By Taylor's expansion, we have

$$
\begin{aligned}
&\frac{1}{T}\sum_{t=1}^{T} Z_t\tilde{I}_t - \hat{\mu}_Z \sum_{j=1}^{M}\sqrt{\frac{\hat{\mu}_{I(\tau_j)}}{1-\hat{\mu}_{I(\tau_j)}}}\omega^*(\tau_j;\hat{\theta}) - \sqrt{\hat{\mu}_Z(1-\hat{\mu}_Z)}\rho \\
=&\frac{1}{T}\sum_{t=1}^{T} Z_t I_t - \mu_Z\mu_I - \sigma_Z\rho - \left(\mu_I + \frac{\rho(1-2\mu_Z)}{2\sigma_Z}\right)(\hat{\mu}_Z - \mu_Z) \\
&+\sum_{j=1}^{M}\left(\frac{(2\mu_{I(\tau_j)}-1)\omega^*(\tau_j)}{2\sigma^3_{I(\tau_j)}}\frac{1}{T}\sum_{t=1}^{T} Z_t\mathbb{I}(Y_t\le\tau_j) - \frac{\mu_Z\omega^*(\tau_j)}{2\sigma_{I(\tau_j)}(1-\mu_{I(\tau_j)})}\right)(\hat{\mu}_{I(\tau_j)} - \mu_{I(\tau_j)}) \\
&+\sum_{j=1}^{M}\left(\frac{\partial\omega^*(\tau_j;\theta^*)}{\partial\theta}\right)'\frac{1}{\sigma_{I(\tau_j)}}\left(\frac{1}{T}\sum_{t=1}^{T} Z_t\mathbb{I}(Y_t\le\tau_j) - \mu_Z\mu_{I(\tau_j)}\right)(\hat{\theta}-\theta^*) + h.o \\
=&\frac{1}{T}\sum_{t=1}^{T} Z_t I_t - E(ZI) - \left(\mu_I + \frac{\rho(1-2\mu_Z)}{2\sigma_Z}\right)(\hat{\mu}_Z - \mu_Z) \\
&+\sum_{j=1}^{M}\left(\frac{(2\mu_{I(\tau_j)}-1)\omega^*(\tau_j)}{2\sigma^3_{I(\tau_j)}}E(Z\mathbb{I}(Y\le\tau_j)) - \frac{\mu_Z\omega^*(\tau_j)}{2\sigma_{I(\tau_j)}(1-\mu_{I(\tau_j)})}\right)(\hat{\mu}_{I(\tau_j)} - \mu_{I(\tau_j)}) \\
&+\sum_{j=1}^{M}\left(\frac{\partial\omega^*(\tau_j;\theta^*)}{\partial\theta}\right)'\frac{E(Z\mathbb{I}(Y\le\tau_j)) - \mu_Z\mu_{I(\tau_j)}}{\sigma_{I(\tau_j)}}(\hat{\theta}-\theta^*) \\
&+\sum_{j=1}^{M} o_p(1)(\hat{\mu}_{I(\tau_j)} - \mu_{I(\tau_j)}) + \sum_{j=1}^{M} o_p(1)(\hat{\theta}-\theta^*) + h.o \\
=&\frac{1}{T}\sum_{t=1}^{T} Z_t I_t - E(ZI) + C_0(\hat{\mu}_Z - \mu_Z) + \sum_{j=1}^{M} C_j(\hat{\mu}_{I(\tau_j)} - \mu_{I(\tau_j)}) + C'_\theta(\hat{\theta}-\theta^*) \\
&+\sum_{j=1}^{M} o_p(1)(\hat{\mu}_{I(\tau_j)} - \mu_{I(\tau_j)}) + \sum_{j=1}^{M} o_p(1)(\hat{\theta}-\theta^*) + h.o.
\end{aligned}
$$

Hence,

$$
\begin{aligned}
&\sqrt{T}(\tilde{\rho}-\rho) = \frac{\sqrt{T}}{\sigma_Z}\left(\frac{1}{T}\sum_{t=1}^{T} Z_t\tilde{I}_t - \hat{\mu}_Z \sum_{j=1}^{M}\sqrt{\frac{\hat{\mu}_{I(\tau_j)}}{1-\hat{\mu}_{I(\tau_j)}}}\omega^*(\tau_j;\hat{\theta}) - \sqrt{\hat{\mu}_Z(1-\hat{\mu}_Z)}\rho\right) + h.o \\
=&\frac{\sqrt{T}}{\sigma_Z}\left(\frac{1}{T}\sum_{t=1}^{T} Z_t I_t - E(ZI) + C_0(\hat{\mu}_Z - \mu_Z) + \sum_{j=1}^{M} C_j(\hat{\mu}_{I(\tau_j)} - \mu_{I(\tau_j)}) + C'_\theta(\hat{\theta}-\theta^*)\right) \\
&+\frac{1}{\sigma_Z}\sum_{j=1}^{M} o_p(1)\sqrt{T}(\hat{\mu}_{I(\tau_j)} - \mu_{I(\tau_j)}) + \frac{1}{\sigma_Z}\sum_{j=1}^{M} o_p(1)\sqrt{T}(\hat{\theta}-\theta^*) + h.o \\
=&\frac{\sqrt{T}}{\sigma_Z}\left(\frac{1}{T}\sum_{t=1}^{T} Z_t I_t - E(ZI) + \frac{C_0}{T}\sum_{t=1}^{T}(Z_t-\mu_Z) + \sum_{j=1}^{M}\frac{C_j}{T}\sum_{t=1}^{T}(\mathbb{I}(Y_t\le\tau_j) - \mu_{I(\tau_j)}) + \frac{C'_\theta}{T}\sum_{t=1}^{T}\eta(Y_t,Z_t,\theta^*)\right) \\
&+o_p(1) + h.o \\
=&\frac{\sqrt{T}}{\sigma_Z}\frac{1}{T}\sum_{t=1}^{T}\left(Z_t I_t - E(ZI) + C_0(Z_t-\mu_Z) + \sum_{j=1}^{M} C_j(\mathbb{I}(Y_t\le\tau_j) - \mu_{I(\tau_j)}) + C'_\theta\eta(Y_t,Z_t,\theta^*)\right) + o_p(1) + h.o \\
=&\frac{1}{\sigma_Z}\frac{1}{\sqrt{T}}\sum_{t=1}^{T}(S_t - E(S_t) + C'_\theta\eta(Y_t,Z_t,\theta^*)) + h.o,
\end{aligned}
$$

which converges in distribution to $N(0,V)$ by the central limit theorem. □