Oliver B. Linton

Nonparametric inference for unbalanced time series data

Originally published in Econometric theory, 21 (1). pp. 143-157 © 2005 Cambridge University Press.

Econometric Theory, **21**, 2005, 143–157. Printed in the United States of America.
DOI: 10.1017/S0266466605050097

NONPARAMETRIC INFERENCE FOR UNBALANCED TIME SERIES DATA

Oliver Linton
London School of Economics

This paper is concerned with the practical problem of conducting inference in a vector time series setting when the data are unbalanced or incomplete. In this case, one can work with only the common sample, to which a standard HAC/bootstrap theory applies, but at the expense of throwing away data and perhaps losing efficiency. An alternative is to use some sort of imputation method, but this requires additional modeling assumptions, which we would rather avoid. We show how the sampling theory changes and how to modify the resampling algorithms to accommodate the problem of missing data. We also discuss efficiency and power. Unbalanced data of the type we consider are quite common in financial panel data; see, for example, Connor and Korajczyk (1993, *Journal of Finance* 48, 1263–1291). These data also occur in cross-country studies.

1. INTRODUCTION

Estimation of heteroskedasticity and autocorrelation consistent covariance matrices (HACs) is a well established problem in time series. Results have been established under a variety of weak conditions on temporal dependence and heterogeneity that allow one to conduct inference on a variety of statistics; see Jowett (1955), Hannan (1957), Newey and West (1987), Andrews (1991), Hansen (1992), de Jong and Davidson (2000), and Robinson (2005). Alternative methods for conducting inference include the bootstrap, for which there is also now a very active research program in time series especially; see Lahiri (2003) for an overview. One convenient method for time series is the subsampling approach of Politis, Romano, and Wolf (1999). This method was used by Linton, Maasoumi, and Whang (2003) in the context of testing for stochastic dominance.

This paper is concerned with the practical problem of conducting inference in a vector time series setting when the data are unbalanced or incomplete. In this case, one can work with only the common sample, to which a standard HAC/bootstrap theory applies, but at the expense of throwing away data and perhaps losing efficiency. An alternative is to use some sort of imputation method to complete the data set, but this requires additional modeling assumptions, which

I thank Greg Connor, Esfandiar Maasoumi, Peter Phillips, Peter Robinson, Yoon Whang, and two referees for helpful comments. Address correspondence to Oliver Linton, Department of Economics, London School of Economics, Houghton Street, London WC2A 2AE, United Kingdom; e-mail: O.Linton@lse.ac.uk.

 0266-4666/05 $12.00

we would rather avoid.[1] We show how the sampling theory changes and how to modify the resampling algorithms to accommodate the problem of missing data. We also discuss efficiency and power. Unbalanced data of the type we consider are quite common in financial panel data; see, for example, Connor and Korajczyk (1993). These data also occur in cross-country studies.

2. MODEL AND SETUP

The most general setting would be a multivariate data series $\{X_{it}, i \in N_t, t \in I_i\}$ where the set N_t contains the number of cross-sectional units at time t and I_i contains the number of time series observations for unit i. We are primarily concerned with the case where $I_i \cap I_j \neq \emptyset$ (although in a multivariate setting you could have $I_i \cap I_j = \emptyset$ for some pairs but not for all pairs and still obtain the main results we find subsequently). The set $I_i \backslash I_i \cap I_j$ could in general contain just contiguous observations, or it could contain several isolated sets before and after $I_i \cap I_j$. The setup costs in establishing notation for the most general case are quite high, and so we have chosen to concentrate on the bivariate special case with staggered samples so that $I_i \backslash I_i \cap I_j$, say, is a series of contiguous observations prior to $I_i \cap I_j$, whereas $I_j \backslash I_i \cap I_j$ is a series of contiguous observations subsequent to $I_i \cap I_j$. This captures the main issues we wish to concentrate on.

Suppose we have two samples denoted I_X and I_Y on X and Y, respectively, with cardinalities T_X and T_Y, where $I_X = \{X_1, \ldots, X_{T_X}\}$ and $I_Y = \{Y_{T^X+1}, \ldots, Y_{T^X+T_Y}\}$. These observations can be partitioned into T^{XY} common observations, denoted $I^{XY} = \{(X_{T^X+1}, Y_{T^X+1}), \ldots, (X_{T^X+T^{XY}}, Y_{T^X+T^{XY}})\}$, T^X separate observations on X, denoted $I^X = \{X_1, \ldots, X_{T^X}\}$, and T^Y separate observations on Y, denoted $I^Y = \{Y_{T^X+T^{XY}+1}, \ldots, Y_{T^X+T_Y}\}$, so that $T_X = T^X + T^{XY}$ and $T_Y = T^Y + T^{XY}$. There are a number of cases of interest with regard to the relative magnitudes of T^X, T^Y, and T^{XY}. We shall suppose that these quantities are all large.

ASSUMPTION A. $T^X = T^X(N) \to \infty, T^Y = T^Y(N) \to \infty$, and $T^{XY} = T^{XY}(N) \to \infty$ as the magnitude parameter $N \to \infty$. In the sequel all limits are taken as $N \to \infty$.

The main case of interest theoretically is where T^X, T^Y, and T^{XY} are all of approximately the same size, but we shall allow other cases. The case where T^{XY} is large relative to T^X, T^Y is trivial, whereas the case where T^X, T^Y are large relative to T^{XY} can be viewed as a limiting version of the main case. Denote by

$$T = T_X T_Y / (T_X + T_Y)$$

the dominant (i.e., smaller) magnitude; thus $T = T(N) \to \infty$ under Assumption A.

We suppose that the data are temporally and cross-sectionally dependent but are stationary and mixing. We assume that the missing at random (MAR) condition of Little and Rubin (1987) holds; that is, the process by which the observations are missing is unrelated to the underlying data distribution.

We are concerned with testing hypotheses about the marginal distributions of X_t and Y_t. There are two general types of hypotheses of interest.

Example 1

We want to test the hypothesis that

$$\mathbf{H}_0 : \mu_X = E(X_t) = E(Y_t) = \mu_Y \tag{1}$$

with alternative either one sided or two sided. This is a special case of the problem of testing whether $f(m_X) = f(m_Y)$, where m_X, m_Y are vectors of moments (including quantiles) from the distributions X, Y, respectively, and f is a smooth function. A more general version of this would involve regression on a benchmark variable Z_t. Thus suppose that $Y_t = \beta_Y^\top Z_t + u_{Yt}$ and $X_t = \beta_X^\top Z_t + u_{Xt}$, where $E(u_t|Z_t) = 0$ with $u_t = (u_{Yt}, u_{Xt})^\top$, and we observe Y_t, X_t as stated previously but Z_t is observed throughout $t = 1, \ldots, T^X + T_Y$. We want to test whether $f(\beta_Y) = f(\beta_X)$ for some smooth function f. A leading example here would concern comparison of the alphas of two different funds (where these are computed relative to a benchmark fund Z_t).

Example 2

We want to test the hypothesis that the distribution of X_t first-order dominates the distribution of Y_t. Let F_X, F_Y denote the cumulative distribution functions (c.d.f.s) of X and Y, respectively. The hypothesis can be stated as

$$\mathbf{H}_0 : \sup_z \{F_X(z) - F_Y(z)\} \leq 0 \tag{2}$$

with the alternative hypothesis that $\sup_z\{F_X(z) - F_Y(z)\} > 0$. More generally we can consider tests of higher order dominance.

In cases like Example 1, we can expect a normal distribution theory to apply under moment and mixing conditions, with the possibility of obtaining asymptotically pivotal test statistics, whereas in cases like Example 2 we expect a more complicated nonnormal distribution theory, with complicated dependence on nuisance parameters precluding asymptotically pivotal statistics.

In Example 1 a natural test statistic to use is

$$\tau = \sqrt{T}(\overline{X} - \overline{Y}), \tag{3}$$

where $\overline{X} = T_X^{-1} \sum_{t=1}^{T_X} X_t$ and $\overline{Y} = T_Y^{-1} \sum_{t=T^X}^{T^X+T_Y} Y_t$. Under a variety of additional conditions $\tau/\hat{\sigma} \Rightarrow N(0,1)$ under the null hypothesis, where $\sigma^2 = \text{avar}(\sqrt{T}(\overline{X} - \overline{Y}))$ and $\hat{\sigma}^2$ is a consistent estimate thereof; we discuss the computation of $\hat{\sigma}^2$ subsequently. The test is based on comparing the studentized τ

with standard normal critical values. An alternative test statistic would be based on only the common sample I^{XY}, $\tau^{XY} = \sqrt{T^{XY}}(\overline{X}^{XY} - \overline{Y}^{XY})$, where $\overline{X}^{XY} = (T^{XY})^{-1}\sum_{t\in I^{XY}} X_t$ and $\overline{Y}^{XY} = (T^{XY})^{-1}\sum_{t\in I^{XY}} Y_t$. In this case also $\tau^{XY}/\hat{\omega} \Rightarrow N(0,1)$ under the null, where $\omega^2 = \mathrm{avar}(\sqrt{T^{XY}}(\overline{X}^{XY} - \overline{Y}^{XY}))$ and $\hat{\omega}^2$ is a consistent estimate thereof. In some cases this test may be an attractive option, but when T^X and/or T^Y is large, this approach, although convenient, may lose power.

In Example 2 a natural test statistic is

$$\delta = \sqrt{T}\sup_z\{\hat{F}_X(z) - \hat{F}_Y(z)\}, \tag{4}$$

where $\hat{F}_X(z) = T_X^{-1}\sum_{t=1}^{T_X} 1(X_t \le z)$ and $\hat{F}_Y(z) = T_Y^{-1}\sum_{t=T^X}^{T^X+T_Y} 1(Y_t \le z)$ are the empirical distribution functions. In practice, the supremum in (2) is approximated by the maximum over a large grid. In this case, the limiting null distribution is $\Delta_F = \sup_z W_F(z)$, where W_F is a Gaussian process with covariance function depending on the joint distribution of X, Y and on the joint autodependence of these processes. The only feasible way of conducting inference here is to use some sort of bootstrap procedure. Linton et al. (2003) have proposed a subsampling algorithm for the statistic $\delta^{XY} = \sqrt{T^{XY}}\sup_z\{\hat{F}_X^{XY}(z) - \hat{F}_Y^{XY}(z)\}$, where $\hat{F}_X^{XY}(z) = (T^{XY})^{-1}\sum_{t\in I^{XY}} 1(X_t \le z)$ and $\hat{F}_Y^{XY}(z) = (T^{XY})^{-1}\sum_{t\in I^{XY}} 1(Y_t \le z)$ are the empirical distributions based on the common sample. Because δ^{XY} uses less data it might be expected to be less powerful than δ. We show subsequently how to modify the Linton et al. (2003) subsampling algorithm to obtain a consistent test based on δ.

3. INFERENCE

3.1. Estimation of Long-Run Variance

Here we show how to estimate σ^2 and conduct the test based on a studentized version of τ. Let $\gamma_X(j)$ and $\gamma_Y(j)$ be the marginal covariance functions of the processes X, Y, respectively, and let $\gamma_{XY}(j) = \mathrm{cov}(X_t, Y_{t-j})$. We use the symbol $\simeq$ to denote asymptotic equivalence as $N \to \infty$, that is, $X_N \simeq Y_N$ if $X_N / Y_N \to^p 1$; in the matrix case this is interpreted element by element.

THEOREM 1. *Suppose that (X_t, Y_t) is jointly stationary with 1-summable covariance function, that is, $\sum_{j=1}^{\infty} j|\gamma_{XY}(j)| < \infty$. Suppose that Assumption A holds. Then*

$$\mathrm{var}\begin{bmatrix}\overline{X}\\ \overline{Y}\end{bmatrix} \simeq \begin{bmatrix} \dfrac{1}{T_X}\displaystyle\sum_{j=-\infty}^{\infty}\gamma_X(j) & \dfrac{T^{XY}}{T_X T_Y}\displaystyle\sum_{j=-\infty}^{\infty}\gamma_{XY}(j) \\ \dfrac{T^{XY}}{T_X T_Y}\displaystyle\sum_{j=-\infty}^{\infty}\gamma_{XY}(j) & \dfrac{1}{T_Y}\displaystyle\sum_{j=-\infty}^{\infty}\gamma_Y(j) \end{bmatrix}.$$

This shows that the marginal variances are the usual terms proportional to the full marginal sample sizes, whereas the covariance is proportional to the common sample size T^{XY}. The reason is basically because terms like $\sum_{t=T^X+1}^{T^X+T^{XY}} X_t$ and $\sum_{t=T^X+T^{XY}+1}^{T^X+T_Y} Y_t$ are asymptotically independent. The restriction $\sum_{j=1}^{\infty} j|\gamma_{XY}(j)| < \infty$ is only needed for the covariance term; if this condition does not hold, the asymptotic covariance term may be different.

A consequence of Theorem 1 is that

$$\sigma^2 \simeq \frac{T_Y}{T_X + T_Y} \sum_{j=-\infty}^{\infty} \gamma_X(j) + \frac{T_X}{T_X + T_Y} \times \sum_{j=-\infty}^{\infty} \gamma_Y(j) - 2\frac{T^{XY}}{T_X + T_Y} \sum_{j=-\infty}^{\infty} \gamma_{XY}(j), \tag{5}$$

whereas $\omega^2 \simeq \sum_{j=-\infty}^{\infty} \gamma_X(j) + \sum_{j=-\infty}^{\infty} \gamma_Y(j) - 2\sum_{j=-\infty}^{\infty} \gamma_{XY}(j)$. When T^X, T^Y, T^{XY} are all of the same magnitude, all three terms in (5) remain in the limit; otherwise one or more of them may vanish asymptotically. To estimate these quantities we now apply the HAC theory. Specifically, we can estimate the long-run variances $\mathrm{lrv}(X) = \sum_{j=-\infty}^{\infty} \gamma_X(j)$, $\mathrm{lrv}(Y) = \sum_{j=-\infty}^{\infty} \gamma_Y(j)$, and $\mathrm{lrcov}(X,Y) = \sum_{j=-\infty}^{\infty} \gamma_{XY}(j)$ by corresponding HAC estimators based, respectively, on the full sample of $X's$, the full sample of $Y's$, and the common sample I^{XY}. For example, let $\hat{\gamma}_X(j) = (T_X - j)^{-1} \sum_{s=1}^{T_X - j} (X_s - \overline{X})(X_{s+j} - \overline{X})$ for $j = 1, \ldots, J(T_X)$ and let

$$\widehat{\mathrm{lrv}}(X) = \sum_{j=-J}^{J} k\left(\frac{j}{J}\right) \hat{\gamma}_X(j), \tag{6}$$

where $k(.)$ is a weight function with support $[-1,1]$ and J is a bandwidth parameter satisfying $J \to \infty$ and $J/T_X \to 0$. See Andrews (1991) for methods and results on how to choose J and Xiao and Linton (2002) and Phillips (2005) for alternative strategies.

We now turn to the properties of the studentized tests $\tau/\hat{\sigma}$ and $\tau^{XY}/\hat{\omega}$, where $\hat{\omega}, \hat{\sigma}$ are consistent estimates of ω, σ. Under local alternatives of the form $\mu_X = \mu_Y + \lambda\big/\sqrt{T}$, we have

$$\frac{\tau^{XY}}{\hat{\omega}} \Rightarrow N(\pi^{XY}, 1) \quad \text{and} \quad \frac{\tau}{\hat{\sigma}} \Rightarrow N(\pi, 1),$$

where

$$\pi = \frac{\lambda}{\sigma} \quad \text{and} \quad \pi^{XY} = \frac{\lambda}{\omega} \lim_{N\to\infty} \sqrt{\frac{T^{XY}}{T}}.$$

Clearly, when $T^{XY}/T \to 0$ the common sample test has no power against these alternatives and τ is preferable. However, the ranking could go the other way.

Suppose that $T^X = T^Y = T^{XY}$, in which case $T = T_X/2 = T_Y/2 = T^{XY}$, so that $\pi^{XY} = \lambda/\omega$. We then have $\sigma^2 \simeq (\frac{1}{2})\sum_{j=-\infty}^{\infty}\gamma_X(j) + (\frac{1}{2})\sum_{j=-\infty}^{\infty}\gamma_Y(j) - (\frac{1}{2})\sum_{j=-\infty}^{\infty}\gamma_{XY}(j)$, and it is possible that $\omega^2 \leq \sigma^2$, at least when $\sum_{j=-\infty}^{\infty}\gamma_{XY}(j) > 0$. For example, suppose that $\sum_{j=-\infty}^{\infty}\gamma_X(j) = \sum_{j=-\infty}^{\infty}\gamma_Y(j) = \vartheta$ and $\sum_{j=-\infty}^{\infty}\gamma_{XY}(j) = \rho\vartheta$; then $\omega^2 - \sigma^2 = \vartheta(2-3\rho)/2$, which can be negative for $\rho > \frac{2}{3}$.[2]

In conclusion, we have found that although $\bar{X}$ is always more efficient than $\bar{X}^{XY}$, the ranking of τ^{XY}, τ as test statistics could go either way—it depends on the relative sample sizes and on their mutual dependence. Subsequently we discuss further the issue of efficiency and local power.

3.2. Subsampling

In the second class of testing problems it is not possible to obtain a pivotal statistic by studentizing, and inference is usually based on some sort of resampling scheme. We concentrate on the subsampling method because it has certain advantages in Example 2; see Linton et al. (2003) for more discussion. The problem here is that just subsampling through the data as usual gives missing data or confines the researcher only to I^{XY}, which would not adequately reflect the sampling error of τ or δ.

We propose a simple modification of the subsampling procedure suitable for the full data set and show that it works in our Example 2. Rewrite $\delta = g(I^X, I^{XY}, I^Y)$ for some function g. Define subsample sizes b^X, b^{XY}, and b^Y with $b^j \to \infty$ and $b^j/T^j \to 0$ for $j = X, Y, XY$. Then define subsamples I^{X,i,b^X} from I^X with

$$I^{X,i,b^X} = \{X_i, \ldots, X_{i+b^X-1}\} \quad \text{for } i = 1, \ldots, T^X - b^X + 1,$$

likewise define subsamples I^{Y,i,b^Y} from I^Y

$$I^{Y,i,b^Y} = \{Y_{T^X+T^{XY}+i}, \ldots, Y_{T^X+T^{XY}+i+b^Y-1}\} \quad \text{for } i = 1, \ldots, T^Y - b^Y + 1,$$

and define subsamples $I^{XY,i,b^{XY}}$ from I^{XY}

$$I^{XY,i,b^{XY}} = \{(X_{T^X+i}, Y_{T^X+i}) \ldots, (X_{T^X+i+b^{XY}-1}, Y_{T^X+i+b^{XY}-1})\} \qquad \text{for } i = 1, \ldots, T^{XY} - b^{XY} + 1.$$

Then define the subsample statistic $\delta_{T,b,i} = g(I^{X,i,b^X}, I^{XY,i,b^{XY}}, I^{Y,i,b^Y})$ and likewise $\tau_{T,b,i}$, specifically

$$\delta_{T,b,i} = \sup_z \sqrt{b}\left(\frac{1}{b^X + b^{XY}}\left[\sum_{s=i}^{i+b^X-1} 1(X_s \leq z) + \sum_{s=T^X+i}^{T^X+i+b^{XY}-1} 1(X_s \leq z)\right] - \frac{1}{b^Y + b^{XY}}\left[\sum_{s=T^X+i}^{T^X+i+b^{XY}-1} 1(Y_s \leq z) + \sum_{s=T^X+T^{XY}+i}^{T^X+T^{XY}+i+b^Y-1} 1(Y_s \leq z)\right]\right).$$

Here, $b(T)$ is chosen to satisfy (asymptotically)

$$\frac{T_Y}{T_X + T_Y} = \frac{b}{b^X + b^{XY}}, \qquad \frac{T_X}{T_X + T_Y} = \frac{b}{b^Y + b^{XY}},$$
$$\frac{T^{XY}}{T_X + T_Y} = \frac{bb^{XY}}{(b^X + b^{XY})(b^Y + b^{XY})}. \tag{7}$$

For example, when $T_X = T_Y = 2T^{XY}$ and $b^X = b^Y = b^{XY}$ we can take $b = b^X$.

We approximate the sampling distribution of δ (or τ) using the distribution of the values of $\delta_{T,b,i}$ (or $\tau_{T,b,i}$) computed over the different subsamples. That is, we approximate the sampling distribution G_T of δ by

$$\hat{G}_{T,b}(w) = \frac{1}{B}\sum_{i=1}^{B} 1(\delta_{T,b,i} \le w), \tag{8}$$

where $B(T) = \min\{T^X - b^X + 1, T^Y - b^Y + 1, T^{XY} - b^{XY} + 1\}$ is the number of different feasible subsamples.[3] Let $g_{T,b}(1-\alpha)$ denote the $(1-\alpha)$th sample quantile of $\hat{G}_{T,b}(\cdot)$, that is,

$$g_{T,b}(1-\alpha) = \inf\{w : \hat{G}_{T,b}(w) \ge 1-\alpha\}.$$

We call it the *subsample critical value* of significance level α. Thus, we reject the null hypothesis at the significance level α if $\tau > g_{T,b}(1-\alpha)$.

Although this algorithm does not seem to replicate precisely the temporal ordering (for example, the sample I^{X,i,b^X} is separated temporally from $I^{XY,i,b^{XY}}$) this does not matter for the first-order asymptotics because of the asymptotic independence argument.

THEOREM 2. *Suppose that* (X_t, Y_t) *is jointly stationary and alpha mixing random sequence with 1-summable mixing coefficients and suppose that under the null hypothesis (2)* δ *converges in distribution to the random variable* Δ_F *whose* $(1-\alpha)$*th quantile is denoted by* $g(1-\alpha)$. *Suppose that Assumption A holds. Then, under the null hypothesis (2),*

$$g_{T,b}(1-\alpha) \xrightarrow{p} \begin{cases} g(1-\alpha) & \text{if } \sup_z\{F_X(z) - F_Y(z)\} = 0 \\ -\infty & \text{if } \sup_z\{F_X(z) - F_Y(z)\} < 0. \end{cases}$$

4. EFFICIENT ESTIMATION AND TESTING

It is well known that the sample mean is an efficient estimate of a population mean when the data are independent and identically distributed (i.i.d.) (Bickel, Klaassen, Ritov, and Wellner, 1993, pp. 67–68) and in some time series cases (Grenander, 1954). Indeed, this is a case where "OLS = GLS"; see Amemiya

(1985, pp. 182–183). We show that this does not hold in the unbalanced case and one can obtain a more efficient estimator than the sample mean. The more efficient estimator translates into a more powerful test. This result carries over to estimation of other quantities such as distribution functions. Bickel, Ritov, and Wellner (1991) treat a related problem of estimating $E[a(X_t, Y_t)]$ for known function a when the marginal distributions of X and of Y are known, which corresponds to the case where T^X and T^Y are very large relative to T^{XY}.

Define the vector of sample moments

$$m = \left[\frac{1}{T^X} \sum_{t \in I^X} X_t, \frac{1}{T^{XY}} \sum_{t \in I^{XY}} X_t, \frac{1}{T^{XY}} \sum_{t \in I^{XY}} Y_t, \frac{1}{T^Y} \sum_{t \in I^Y} Y_t \right]^\top = [m_1, m_2, m_3, m_4]^\top.$$

The vector m contains unbiased estimators of the parameter vector $\theta = (\mu_X, \mu_Y)^\top$. We consider estimators that minimize the minimum distance criterion $(m - A\theta)^\top \Psi (m - A\theta)$, where A is the 4×2 matrix of zeros and ones that takes $(\mu_X, \mu_Y)^\top$ into $(\mu_X, \mu_X, \mu_Y, \mu_Y)^\top$, whereas Ψ is a symmetric positive definite weighting matrix. The resulting estimator has closed form $\hat{\theta} = (A^\top \Psi A)^{-1} A^\top \Psi m$, that is, it is a linear combination of the elements of m.[4] Therefore, $\mathrm{var}(\hat{\theta}) \simeq (A^\top \Psi A)^{-1} A^\top \Psi V \Psi A (A^\top \Psi A)^{-1}$, where V is the asymptotic variance of m:

$$V = \begin{bmatrix} \frac{1}{T^X} \sum_{j=-\infty}^{\infty} \gamma_X(j) & 0 & 0 & 0 \\ 0 & \frac{1}{T^{XY}} \sum_{j=-\infty}^{\infty} \gamma_X(j) & \frac{1}{T^{XY}} \sum_{j=-\infty}^{\infty} \gamma_{XY}(j) & 0 \\ 0 & \frac{1}{T^{XY}} \sum_{j=-\infty}^{\infty} \gamma_{XY}(j) & \frac{1}{T^{XY}} \sum_{j=-\infty}^{\infty} \gamma_Y(j) & 0 \\ 0 & 0 & 0 & \frac{1}{T^Y} \sum_{j=-\infty}^{\infty} \gamma_Y(j) \end{bmatrix}.$$

The optimal choice of Ψ is proportional to V^{-1}, in which case $\hat{\theta}$ has asymptotic variance proportional to $(A^\top V^{-1} A)^{-1}$.[5] The full sample mean $\bar{\theta} = (\bar{X}, \bar{Y})^\top$ is also a linear combination of m, $\bar{\theta} = Sm$, where S is the 2×4 matrix with first row $S_1 = T_X^{-1}(T^X, T^{XY}, 0, 0)$ and second row $S_2 = T_Y^{-1}(0, 0, T^{XY}, T^Y)$. Likewise the subsample mean $\bar{\theta}^{XY} = (\bar{X}^{XY}, \bar{Y}^{XY})^\top = S^{XY} m$, where S^{XY} is the 2×4 matrix with first row $S_1^{XY} = (0,1,0,0)$ and second row $S_2^{XY} = (0,0,1,0)$. It is easy to show that $SVS^\top \geq (A^\top V^{-1} A)^{-1}$ and $S^{XY} V (S^{XY})^\top \geq (A^\top V^{-1} A)^{-1}$ in the matrix partial order so that $\hat{\theta}$ is more efficient than both $\bar{\theta}$ and $\bar{\theta}^{XY}$. Whether the weak inequality is strict depends on the relative magnitudes of T^X, T^Y, T^{XY}. We conjecture that $\hat{\theta}$ is semiparametrically efficient for estimation of θ under some conditions. To support this conjecture we can show that $\hat{\theta}$ achieves the asymptotic Cramèr–Rao lower bound when the data are i.i.d. over time and Gaussian. A feasible version of $\hat{\theta}$, which shares its limiting distribution, can be obtained from estimates of V, which can be obtained from the estimates of $\mathrm{lrv}(X)$, $\mathrm{lrv}(Y)$, and $\mathrm{lrcov}(X,Y)$ defined as in (6).

We illustrate these general results with an example. Suppose that $T^X = T^Y = T^{XY}$ and that $\sum_{j=-\infty}^{\infty}\gamma_X(j) = \sum_{j=-\infty}^{\infty}\gamma_Y(j) = \vartheta$ and $\sum_{j=-\infty}^{\infty}\gamma_{XY}(j) = \rho\vartheta$. Then

$$\mathrm{var}(\hat{\theta}) \simeq \frac{\vartheta}{T}\begin{bmatrix} \dfrac{4-2\rho^2}{4-\rho^2} & \dfrac{2\rho}{4-\rho^2} \\ \dfrac{2\rho}{4-\rho^2} & \dfrac{4-2\rho^2}{4-\rho^2} \end{bmatrix}; \qquad \mathrm{var}(\bar{\theta}) \simeq \frac{\vartheta}{T}\begin{bmatrix} 1 & \dfrac{\rho}{2} \\ \dfrac{\rho}{2} & 1 \end{bmatrix};$$

$$\mathrm{var}(\bar{\theta}^{XY}) \simeq \frac{\vartheta}{T}\begin{bmatrix} 2 & 2\rho \\ 2\rho & 2 \end{bmatrix}.$$

For all ρ, $\mathrm{var}(\bar{\theta}) - \mathrm{var}(\hat{\theta})$ is positive definite, strictly so for $\rho \neq 0$. For all ρ, $\mathrm{var}(\bar{\theta}^{XY}) - \mathrm{var}(\hat{\theta})$ is positive definite, strictly so for $\rho \neq 1$.

We now turn to the testing problem. Define $\tau_E = \sqrt{T}(1,-1)\hat{\theta}$ and let $\widehat{\sigma_E}$ be a consistent estimate of σ_E, which can be obtained from the estimates of V as already discussed. It follows that under local alternatives $\mu_X = \mu_Y + \lambda\big/\sqrt{T}$,

$$\frac{\tau_E}{\widehat{\sigma_E}} \Rightarrow N(\pi_E,1),$$

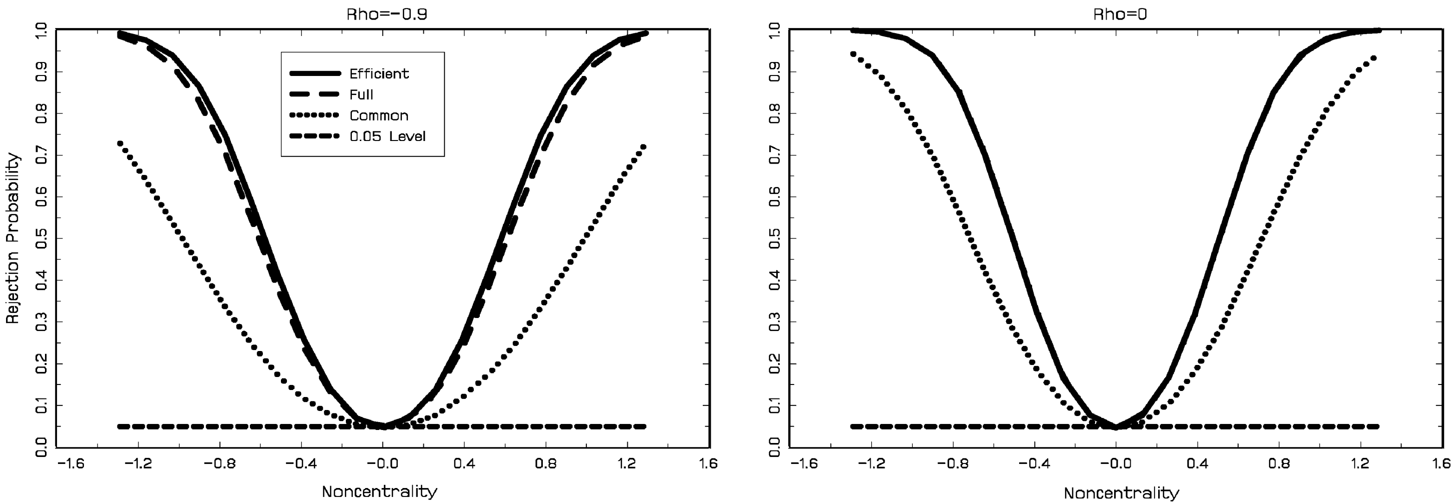
where $\pi_E = \lambda/\sigma_E$. Furthermore, $|\pi_E| \geq \max\{|\pi|,|\pi^{XY}|\}$ so that $\tau_E/\widehat{\sigma_E}$ is the most powerful test in this class. Consider the special case that $T^X = T^Y = T^{XY}$, $\sum_{j=-\infty}^{\infty}\gamma_X(j) = \sum_{j=-\infty}^{\infty}\gamma_Y(j) = \vartheta$, and $\sum_{j=-\infty}^{\infty}\gamma_{XY}(j) = \rho\vartheta$. We have

$$\pi_E^2 = \frac{\lambda^2}{\vartheta}\,\frac{2-\rho}{2-2\rho} \geq \max\{(\pi^{XY})^2,\pi^2\} = \frac{\lambda^2}{\vartheta}\max\left\{\frac{1}{2-2\rho},\frac{2}{2-\rho}\right\}.$$

For the range $\rho \in [-1,0.5]$, π_E^2/π^2 is quite modest; it lies in $[1,1.12]$, but as $\rho \to 1$, $\pi_E^2/\pi^2 \to \infty$. On the other hand $\pi_E^2/(\pi^{XY})^2 = 2-\rho \in [1,3]$.[6]

We briefly report the results of a simulation study that investigates τ_E, τ, τ^{XY} in the case where $X_t = X_t^* + \lambda\big/\sqrt{T}$ with $X_t^* = \phi X_{t-1}^* + \varepsilon_t, Y_t = \phi Y_{t-1} + \eta_t$, where (ε_t,η_t) are jointly standard normal with correlation ρ. In this case, $\sum_{j=-\infty}^{\infty}\gamma_X(j) = \sum_{j=-\infty}^{\infty}\gamma_Y(j) = (1-\phi)^{-2}$ and $\sum_{j=-\infty}^{\infty}\gamma_{XY}(j) = (1-\phi)^{-2}\rho$. We take $T^X = T^Y = T^{XY} = 60$ corresponding to 5 years of monthly data and $\phi = 0.5$ throughout, while varying $\rho \in \{-0.9, 0, 0.5, 0.9\}$. The power curves for the 0.05 level two-sided tests are shown in Figure 1 calculated from 100,000 replications.

Throughout, the test based on τ_E has the higher power curve, but which test comes second changes according to the design: the common sample test does very poorly when $\rho = -0.9$, whereas the full sample test does very poorly when $\rho = 0.9$, as predicted by the theory. We acknowledge that the feasible


Rho=-0.9
Rejection Probability
Noncentrality
Efficient
Full
Common
0.05 Level
0.0
0.1
0.2
0.3
0.4
0.5
0.6
0.7
0.8
0.9
1.0
-1.6
-1.2
-0.8
-0.4
-0.0
0.4
0.8
1.2
1.6
Rho=0
Noncentrality

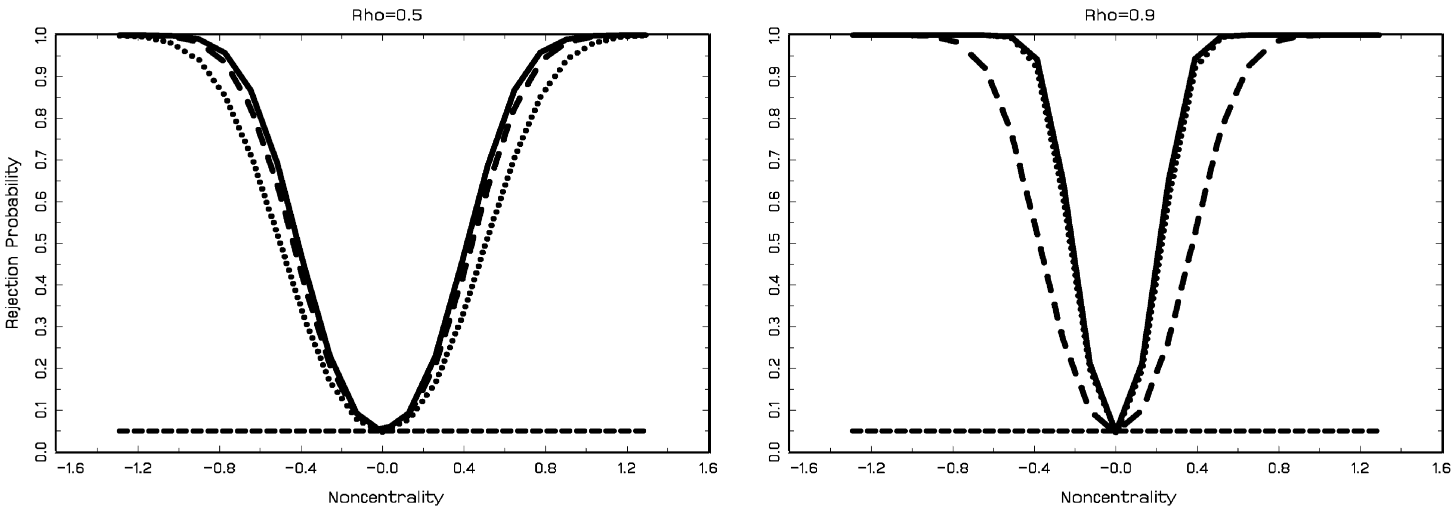


FIGURE 1. Power curves for the 0.05 level two-sided tests.

version of τ_E can suffer from small sample effects that might diminish its edge, and we intend to investigate this in future work.

Finally, this estimation/testing strategy can also be applied to the c.d.f.s in Example 2. Specifically, define for each z the vector of sample moments

$$m_z = \left[\frac{1}{T^X} \sum_{t \in I^X} 1(X_t \le z), \frac{1}{T^{XY}} \sum_{t \in I^{XY}} 1(X_t \le z), \right.$$
$$\left. \frac{1}{T^{XY}} \sum_{t \in I^{XY}} 1(Y_t \le z), \frac{1}{T^Y} \sum_{t \in I^Y} 1(Y_t \le z) \right]^\top$$

and define estimates $\hat{F}_X^E(z)$ and $\hat{F}_Y^E(z)$ by the preceding minimum distance strategy. Then define $\delta^E = \sqrt{T}\sup_z\{\hat{F}_X^E(z) - \hat{F}_Y^E(z)\}$, which converges weakly under the null hypothesis to $\Delta_{F^E} = \sup_z W_{F^E}(z)$, where $W_{F^E}(\cdot)$ is some Gaussian process. By construction $\hat{F}_X^E(z)$ and $\hat{F}_Y^E(z)$ are more efficient than $\hat{F}_X(z)$ and $\hat{F}_Y(z)$, and it may be possible to show that tests based on δ^E are more powerful than those based on δ.[7] The same subsampling algorithm described in Section 3.2 could be used to set critical values.

5. CONCLUDING REMARKS

We have shown how to modify inference procedures in some special unbalanced data cases. In particular, we showed how to conduct valid inference for the "natural" full sample test statistics τ, δ in our two examples. We also showed that these may not be the most powerful tests, and indeed there are situations where using only the common sample may be superior. We proposed more efficient tests that use all the data and require estimates of long-run variances to do the optimal weighting.

Our results can be generalized in a number of ways. First, in the multivariate case there is possibility for further efficiency/power improvements. Second, we can consider more general nonseparable hypotheses. For example, consider the hypothesis that $E[a(X_t, Y_t, \ldots, X_{t-p}, Y_{t-p})] = 0$ for some known function a and lag length p. This involves the joint distribution of $(X_t, Y_t, \ldots, X_{t-p}, Y_{t-p})$, not just its marginals. Nevertheless, one can generally improve on the test statistic $(T^{XY})^{-1} \sum_{t:t,\ldots,t-p \in I^{XY}} a(X_t, Y_t, \ldots, X_{t-p}, Y_{t-p})$ using the minimum distance strategy discussed in Section 4 by taking appropriate choice of moments.

NOTES

1. But if we did go down that path we would advocate a general to specific approach.
2. The extreme case of i.i.d. data with perfect mutual correlation makes the intuition clear—in that case τ^{XY} is constant, whereas τ will have randomness due to the unmatched samples.
3. A more general approach can be based on $\delta_{T,b,i,i',i''} = f(I^{X,i,b^X}, I^{XY,i',b^{XY}}, I^{Y,i'',b^Y})$ and then taking the empirical distribution across all consistent $\{i, i', i''\}$.

4. Suppose that $T^X = T^Y = T^{XY}$ and that $\sum_{j=-\infty}^{\infty} \gamma_X(j) = \sum_{j=-\infty}^{\infty} \gamma_Y(j) = \vartheta$ and $\sum_{j=-\infty}^{\infty} \gamma_{XY}(j) = \rho\vartheta$. The estimator has the natural form

$$\hat{\theta} = \frac{1}{4-\rho^2}\begin{bmatrix} (2-\rho^2)m_1 + 2m_2 + \rho(m_4 - m_3) \\ \rho(m_1 - m_2) + 2m_3 + (2-\rho^2)m_4 \end{bmatrix}.$$

5. More formally, under additional conditions $D_N(\hat{\theta} - \theta) \Rightarrow N(0, (A_0^\top V_0^{-1} A_0)^{-1})$, where

$$D_N = \begin{bmatrix} \sqrt{T_X} & 0 \\ 0 & \sqrt{T_Y} \end{bmatrix}; \qquad A_0 = \lim_{N\to\infty} \begin{bmatrix} \sqrt{T^X/T_X} & 0 \\ \sqrt{T^{XY}/T_X} & 0 \\ 0 & \sqrt{T^{XY}/T_Y} \\ 0 & \sqrt{T^Y/T_Y} \end{bmatrix}$$

$$V_0 = \begin{bmatrix} \sum_{j=-\infty}^{\infty} \gamma_X(j) & 0 & 0 & 0 \\ 0 & \sum_{j=-\infty}^{\infty} \gamma_X(j) & \sum_{j=-\infty}^{\infty} \gamma_{XY}(j) & 0 \\ 0 & \sum_{j=-\infty}^{\infty} \gamma_{XY}(j) & \sum_{j=-\infty}^{\infty} \gamma_Y(j) & 0 \\ 0 & 0 & 0 & \sum_{j=-\infty}^{\infty} \gamma_Y(j) \end{bmatrix}.$$

6. In this case we can write

$$\tau_E = \sqrt{T}\,\frac{1}{2-\rho}\,[(1-\rho)(m_1 - m_4) + (m_2 - m_3)],$$

which gives a nice interpretation—as ρ increases more weight is put on the common sample difference.

7. Suppose it can be shown that $W_F(\cdot) = W_{F^E}(\cdot) + U(\cdot)$, where $U(\cdot)$ is independent of $W_{F^E}(\cdot)$. If $\hat{F}_X^E(\cdot) - \hat{F}_Y^E(\cdot)$ were semiparametrically efficient, this structure would be expected by the Hajek–Le Cam convolution theorem (Bickel et al., 1993, p. 182). It follows that the critical value of the one-sided test based on $\sup_z \sqrt{T}\{\hat{F}_X^E(z) - \hat{F}_Y^E(z)\}$ is smaller than that based on $\sup_z \sqrt{T}\{\hat{F}_X(z) - \hat{F}_Y(z)\}$ (see Bickel et al., 1993, p. 194), and so the former test should be more powerful under local alternatives that shift the location of the limiting processes equally.

REFERENCES

Amemiya, T. (1985) *Advanced Econometrics*. Harvard University Press.

Andrews, D.W.K. (1991) Heteroskedasticity and autocorrelation consistent covariance matrix estimation. *Econometrica* 59, 817–858.

Bickel, P., C.A.J. Klaassen, Y. Ritov, & J. Wellner (1993) *Efficient and Adaptive Estimation for Semiparametric Models*. Springer-Verlag.

Bickel, P., Y. Ritov, & J. Wellner (1991) Efficient estimation of linear functionals. *Annals of Statistics* 19, 1316–1346.

Connor, G. & R.A. Korajczyk (1993) A test for the number of factors in an approximate factor model. *Journal of Finance* 48, 1263–1291.

Grenander, U. (1954) On the estimation of regression coefficients in the case of an autocorrelated disturbance. *Annals of Mathematical Statistics* 25, 252–272.

Hannan, E.J. (1957) The variance of the mean of a stationary process. *Journal of the Royal Statistical Society, Series B* 19, 282–285.

Hansen, B.E. (1992) Consistent covariance matrix estimation for dependent heterogenous processes. *Econometrica* 60, 967–972.

de Jong, R.M. & J. Davidson (2000) Consistency of kernel estimators of heteroskedastic and autocorrelated covariance matrices. *Econometrica* 68, 407–423.

Jowett, G.H. (1955) The comparison of means of sets of observations from sections of independent stochastic series. *Journal of the Royal Statistical Society, Series B* 17, 208–227.

Lahiri, S.N. (2003) *Resampling Methods for Dependent Data*. Springer-Verlag.

Linton, O., E. Maasoumi, & Y.J. Whang (2003) Consistent Testing for Stochastic Dominance under General Sampling Schemes. Cowles Foundation Discussion paper, Yale University.

Little, R.J.A. & D.B. Rubin (1987) *Statistical Analysis with Missing Data*. Wiley.

Newey, W.K. & K. West (1987) A simple positive semi-definite heteroscedasticity and autocorrelation consistent covariance matrix. *Econometrica* 55, 703–708.

Phillips, P.C.B. (2005) HAC estimation by automated regression. *Econometric Theory* (this issue).

Politis, D.N., J.P. Romano, & M. Wolf (1999) *Subsampling*. Springer-Verlag.

Robinson, P.M. (2005) Robust covariance matrix estimation: "HAC" estimates with long memory/antipersistence correction. *Econometric Theory* (this issue).

Xiao, Z. & O. Linton (2002) A nonparametric prewhitened covariance estimator. *Journal of Time Series Analysis* 23, 215–250.

APPENDIX

Proof of Theorem 1. *By standard arguments*

$$\operatorname{var}(\bar{X}) \simeq \frac{1}{T_X} \sum_{j=-\infty}^{\infty} \gamma_X(j) \quad \text{and} \quad \operatorname{var}(\bar{Y}) \simeq \frac{1}{T_Y} \sum_{j=-\infty}^{\infty} \gamma_Y(j).$$

It remains to calculate $\operatorname{cov}(\bar{X}, \bar{Y})$. For notational brevity write $x_t = X_t - E(X_t)$ and $y_t = Y_t - E(Y_t)$. Then

$$\begin{aligned}
\operatorname{cov}(\bar{X}, \bar{Y}) &= \frac{1}{T_X T_Y} E\left[\left(\sum_{t=1}^{T^X} x_t + \sum_{t=T^X+1}^{T^X+T^{XY}} x_t\right)\left(\sum_{t=T^X+T^{XY}+1}^{T^X+T_Y} y_t + \sum_{t=T^X+1}^{T^X+T^{XY}} y_t\right)\right] \\
&= \frac{1}{T_X T_Y} E\left[\sum_{t=T^X+1}^{T^X+T^{XY}} x_t \sum_{t=T^X+T^{XY}+1}^{T^X+T_Y} y_t\right] + \frac{1}{T_X T_Y} E\left[\sum_{t=1}^{T^X} x_t \sum_{t=T^X+T^{XY}+1}^{T^X+T_Y} y_t\right] \\
&\quad + \frac{1}{T_X T_Y} E\left[\sum_{t=T^X+1}^{T^X+T^{XY}} x_t \sum_{t=T^X+1}^{T^X+T^{XY}} y_t\right] + \frac{1}{T_X T_Y} E\left[\sum_{t=1}^{T^X} x_t \sum_{t=T^X+1}^{T^X+T^{XY}} y_t\right] \\
&= I + II + III + IV.
\end{aligned}$$

We have

$$III = \frac{T^{XY}}{T_X T_Y} \sum_{|j| \le T^{XY}} \left(1 - \frac{|j|}{T^{XY}}\right) \gamma_{XY}(j) \simeq \frac{T^{XY}}{T_X T_Y} \sum_{j=-\infty}^{\infty} \gamma_{XY}(j) = O(T^{-1})$$

by dominated convergence using the fact that $T^{XY} \le \min\{T_X, T_Y\}$. Define the integer sets

$$I_u = \{t : s - t = u; s = T^X + T^{XY} + 1, \ldots, T^X + T_Y; t = 1, \ldots, T^X\},$$

$$I'_u = \{t : s - t = u; s = T^X + T^{XY} + 1, \ldots, T^X + T_Y; t = T^X + 1, \ldots T^X + T^{XY}\}, \qquad u \ge 1$$

and let n_u (n'_u) denote the cardinality of I_u (I'_u), noting that $n_u, n'_u \le u$ for all u. Then

$$II = \frac{1}{T_X T_Y} \sum_{t=1}^{T^X} \sum_{s=T^X+T^{XY}+1}^{T^X+T_Y} \gamma_{XY}(s-t) = \frac{1}{T_X T_Y} \sum_{u=T^{XY}+1}^{T^X+T_Y-1} n_u \gamma_{XY}(u)$$

$$\le \frac{1}{T_X T_Y} \sum_{u=T^{XY}+1}^{\infty} u|\gamma_{XY}(u)| = o(T^{-2})$$

because $\sum_{u=1}^{\infty} u|\gamma_{XY}(u)| < \infty$ and $T^2/T_X T_Y = T_X T_Y/(T_X + T_Y)^2$ is bounded. Also,

$$I = \frac{1}{T_X T_Y} \sum_{t=T^X+1}^{T^X+T^{XY}} \sum_{s=T^X+T^{XY}+1}^{T^X+T_Y} \gamma_{XY}(s-t) = \frac{1}{T_X T_Y} \sum_{u=1}^{T_Y-1} n'_u \gamma_{XY}(u) = O(T^{-2}),$$

by the same reasoning. Likewise $IV = O(T^{-2})$. ■

Proof of Theorem 2. *The proof is based on showing that*

$$U(\cdot) = \sqrt{b}\left(\frac{1}{b^X + b^{XY}}\left[\sum_{s=i}^{i+b^X-1} 1(X_s \le \cdot) + \sum_{s=T^X+i}^{T^X+i+b^{XY}-1} 1(X_s \le \cdot)\right]\right.$$
$$\left. - \frac{1}{b^Y + b^{XY}}\left[\sum_{s=T^X+i}^{T^X+i+b^{XY}-1} 1(Y_s \le \cdot) + \sum_{s=T^X+T^{XY}+i}^{T^X+T^{XY}+i+b^Y-1} 1(Y_s \le \cdot)\right]\right)$$

satisfies a functional central limit theorem with limit $W_F(\cdot)$. The main step is to show that $U(z)$ has asymptotically the same variance as $\sqrt{T}\{\hat{F}_X(z) - \hat{F}_Y(z)\}$, and this follows using the proof of Theorem 1, that is,

$$\mathrm{var}(U(z)) \simeq \frac{b}{b^X + b^{XY}} \sum_{j=-\infty}^{\infty} \gamma_{F_X(z)}(j) + \frac{b}{b^Y + b^{XY}} \sum_{j=-\infty}^{\infty} \gamma_{F_Y(z)}(j)$$
$$- 2\frac{bb^{XY}}{(b^X + b^{XY})(b^Y + b^{XY})} \sum_{j=-\infty}^{\infty} \gamma_{F_{XY}(z,z)}(j),$$

where $\gamma_{F_X(z)}(j) = \mathrm{cov}(1(X_t \le z), 1(X_{t-j} \le z))$, $\gamma_{F_Y(z)}(j) = \mathrm{cov}(1(Y_t \le z), 1(Y_{t-j} \le z))$, and $\gamma_{F_{XY}(z,z)}(j) = \mathrm{cov}(1(X_t \le z)1(Y_t \le z), 1(X_{t-j} \le z)1(Y_{t-j} \le z))$. The two variances coincide when (7) holds. ■