Nonparametric Inference for Unbalanced Time Series Data

by

Oliver Linton*
London School of Economics and Political Science

Contents:

Discussion Paper
No.EM/04/474
April 2004

The Suntory Centre
Suntory and Toyota International Centres for
Economics and Related Disciplines
London School of Economics and Political Science
Houghton Street
London WC2A 2AE
Tel.: 020 7955 6698

---

* I would like to thank Greg Connor, Esfandiar Maasoumi, Peter Robinson, and Yoon Whang for helpful comments.

Abstract

This paper is concerned with the practical problem of conducting inference in a vector time series setting when the data is unbalanced or incomplete. In this case, one can work only with 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 modelling 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). These data also occur in cross-country studies.

Keywords: Bootstrap, efficient, HAC estimation, missing data, subsampling.

JEL Nos.: C14, C22.

Contact addresses: Professor Oliver Linton, Department of Economics, London School of Economics and Political Science, Houghton Street, London WC2A 2AE, UK. Email: o.linton@lse.ac.uk/

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 Newey and West (1987), Hansen (1992), de Jong and Davidson (2000), and Robinson (2004). Indeed there is an extensive literature on automating these procedures starting with Andrews (1991). 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) (henceforth LMW) 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 is unbalanced or incomplete. In this case, one can work only with 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 modelling assumptions, which 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 Set-up

Suppose we have two samples denoted I_X and I_Y on X and Y respectively with cardinalities T_X and T_Y. We will suppose that the samples are staggered and in particular $I_X = \{X_1, \ldots, X_{T_X}\}$ 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}. The main case of interest theoretically is where they are all of approximately the same size. The case where T^{XY} is large relative to T^X, T^Y is trivial, while the case where T^X, T^Y are large relative to T^{XY} can be viewed as a limiting version of the main case. In any case we assume throughout that $T_X, T_Y \to \infty$, and denote by $T = T_X T_Y/(T_X + T_Y)$

[1] But if we did go down that path we would advocate a general to specific approach.

the dominant magnitude. We suppose that the data are temporally and cross-sectionally dependent, but are stationary and mixing. We assume that the 'missing data' arises exogenously, i.e., the MAR assumption applies, Little and Rubin (1987).

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

$$H_0 : \mu_X = E(X_t) = E(Y_t) = \mu_Y \tag{1}$$

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 above but that Z_t is observed throughout $t = 1, \ldots, T^X + T_Y$. Want to test whether $f(\beta_Y) = f(\beta_X)$ for some smooth function f. A leading example here would concern comparison of two funds $\alpha's$ (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 c.d.f. of X and Y respectively, the hypothesis can be stated as

$$H_0 : \sup_z \{F_X(z) - F_Y(z)\} \leq 0 \tag{2}$$

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

In the former case, we can expect a normal distribution theory to apply under moment and mixing conditions, with the possibility of obtaining asymptotically pivotal test statistics, while in the latter type we expect a more complicated non-normal distribution theory, with complicated dependence on nuisance parameters precluding asymptotic pivotality.

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 standard conditions $\tau/\widehat{\sigma} \Longrightarrow N(0,1)$ under the null hypothesis, where $\sigma^2 = \mathrm{avar}(\sqrt{T}(\overline{X} - \overline{Y}))$ and $\widehat{\sigma}^2$ is a consistent estimate thereof. 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 $\tau^{XY}/\widehat{\omega} \Longrightarrow N(0,1)$

under the null, where $\omega^2 = \text{avar}(\sqrt{T^{XY}}(\overline{X}^{XY} - \overline{Y}^{XY}))$ and $\widehat{\omega}^2$ 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 while convenient, may lose power.

In example 2 a natural test statistic is

$$\delta = \sqrt{T} \max_{1 \leq z_\ell \leq L(T)} \{\widehat{F}_X(z_\ell) - \widehat{F}_Y(z_\ell)\}, \tag{4}$$

where $\widehat{F}_X(z) = T_X^{-1} \sum_{t=1}^{T_X} 1(X_t \leq z)$ and $\widehat{F}_Y(z) = T_Y^{-1} \sum_{t=T^X}^{T^X+T_Y} 1(Y_t \leq z)$ are the empirical distribution functions. Here, z_ℓ are grid points whose cardinality $L(T)$ increases with sample size. 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. LMW have proposed a subsampling algorithm for the statistic $\delta^{XY} = \sqrt{T^{XY}} \max_{1 \leq z_\ell \leq L(T)} \{\widehat{F}_X^{XY}(z_\ell) - \widehat{F}_Y^{XY}(z_\ell)\}$, where $\widehat{F}_X^{XY}(z) = (T^{XY})^{-1} \sum_{t \in I^{XY}} 1(X_t \leq z)$ and $\widehat{F}_Y^{XY}(z) = (T^{XY})^{-1} \sum_{t \in I^{XY}} 1(Y_t \leq z)$ are the empirical distributions based on the common sample. Because δ^{XY} is using less data it can also be less powerful than δ. We show below how to modify the LMW 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 $\gamma_{XY}(j) = \text{cov}(X_t, Y_{t-j})$.

Theorem 1. *Suppose that (X_t, Y_t) is jointly stationary with absolutely summable covariance function such that $\sum_{j=1}^{\infty} j|\gamma_{XY}(j)| < \infty$. Suppose that $T_X, T_Y \to \infty$, and let $T = T_X T_Y/(T_X + T_Y) \to \infty$. Then*

$$\text{var} \begin{bmatrix} \overline{X} \\ \overline{Y} \end{bmatrix} = \begin{bmatrix} \frac{1}{T_X} \sum_{j=-\infty}^{\infty} \gamma_X(j) & \frac{T^{XY}}{T_X T_Y} \sum_{j=-\infty}^{\infty} \gamma_{XY}(j) \\ \frac{T^{XY}}{T_X T_Y} \sum_{j=-\infty}^{\infty} \gamma_{XY}(j) & \frac{1}{T_Y} \sum_{j=-\infty}^{\infty} \gamma_Y(j) \end{bmatrix} + o(T^{-1}).$$

This shows that the marginal variances are the usual terms proportional to the full marginal sample sizes, while 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, but in its absence this term may change.

A consequence of Theorem 1 is that

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

while $\omega^2 \simeq \sum_{j=-\infty}^{\infty}\gamma_X(j)+\sum_{j=-\infty}^{\infty}\gamma_Y(j)-2\sum_{j=-\infty}^{\infty}\gamma_{XY}(j)$. 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 on the common sample I^{XY}. For example, let $\widehat{\gamma}_X(j)=(T_X-j)^{-1}\sum_{s=1}^{T_X-j}(X_s-\overline{X})(X_{s+j}-\overline{X})$ for $j=1,\ldots,J(T_X)$ and let

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

where $k(.)$ is a weight function with support $[-1,1]$ and $J(T_X)$ is a bandwidth parameter satisfying $J(T_X)\to\infty$ and $J(T_X)/T_X\to 0$. See Andrews (1991) for methods and results on how to choose $J(T_X)$, and Xiao and Linton (2002) and Phillips (2004) for alternative strategies.

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

$$\frac{\tau^{XY}}{\widehat{\omega}}\Longrightarrow N\left(\pi^{XY},1\right)\text{ and }\frac{\tau}{\widehat{\sigma}}\Longrightarrow N\left(\pi,1\right),$$

where

$$\pi=\frac{\lambda}{\sigma}\text{ and }\pi^{XY}=\frac{\lambda}{\omega}\lim_{T_X,T_Y\to\infty}\sqrt{\frac{T^{XY}(T_X+T_Y)}{T_XT_Y}}.$$

Clearly, when $T^{XY}/\min\{T_X,T_Y\}\to 0$ 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 $\sigma^2\simeq(1/2)\sum_{j=-\infty}^{\infty}\gamma_X(j)+(1/2)\sum_{j=-\infty}^{\infty}\gamma_Y(j)-(1/2)\sum_{j=-\infty}^{\infty}\gamma_{XY}(j)$, 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 $\rho>2/3$.[2]

In conclusion, we have found that although $\overline{X}$ is always more efficient than $\overline{X}^{XY}$, 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. We discuss further below the issue of efficiency and local power.

[2]The extreme case of i.i.d. data with perfect mutual correlation makes the intuition clear - in that case τ^{XY} is constant, while τ will have randomness due to the unmatched samples.

3.2 Subsampling

In the second class of testing problems it is not possible to obtain pivotality by studentizing the statistic, 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 LMW for more discussion. The problem here is that just subsampling through the data as usual gives you missing data or confines you 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 dataset 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

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

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

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

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

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

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

$$\begin{aligned}\delta_{T,b,i} &= \max_{1\le z_\ell \le L(T)} \sqrt{b}\left(\frac{1}{b^X+b^{XY}}\left[\sum_{s=i}^{i+b^X-1} 1(X_s \le z_\ell) + \sum_{s=T^X+i}^{T^X+i+b^{XY}-1} 1(X_s \le z_\ell)\right]\right. \\ &\quad \left. -\frac{1}{b^Y+b^{XY}}\left[\sum_{s=T^X+i}^{T^X+i+b^{XY}-1} 1(Y_s \le z_\ell) + \sum_{s=T^X+T^{XY}+i}^{T^X+T^{XY}+i+b^Y-1} 1(Y_s \le z_\ell)\right]\right).\end{aligned}$$

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

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

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

$$\widehat{G}_{T,b}(w) = \frac{1}{N}\sum_{i=1}^{N} 1\left(\delta_{T,b,i} \le w\right), \tag{8}$$

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

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

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, 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 $g(1-\alpha)$. 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 in both the i.i.d. case, 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". We show that this no longer holds in the unbalanced case and one can obtain a more efficient estimator than the sample mean. This result carries over to estimation of other quantities like distribution functions. See Bickel, Ritov, and Wellner (1991) for a more general problem of this type in the i.i.d. case. The more efficient estimator translates into a more powerful test. In this section we assume that T^X, T^Y, and T^{XY} are of similar magnitude to avoid degeneracy.

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 $(\mu_X, \mu_X, \mu_Y, \mu_Y)^\top$, while Ψ is a symmetric positive definite weighting matrix. The resulting estimator has closed form $\widehat{\theta} = \left(A^\top\Psi A\right)^{-1} A^\top\Psi m$, i.e., it is

[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 $\{i, i', i''\}$.

a linear combination of the elements of m.[4] This estimator has asymptotic variance proportional to $\left(A^{\top}\Psi A\right)^{-1} A^{\top}\Psi V \Psi A \left(A^{\top}\Psi A\right)^{-1}$, 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 $\widehat{\theta}$ has asymptotic variance proportional to $\left(A^{\top}V^{-1}A\right)^{-1}$. The full sample mean $\overline{\theta} = (\overline{X}, \overline{Y})^{\top}$ is also a linear combination of m, $\overline{\theta} = Sm$, 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 $\overline{\theta}^{XY} = (\overline{X}^{XY}, \overline{Y}^{XY})^{\top} = S^{XY}m$, 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 $SVS^{\top} \geq \left(A^{\top}V^{-1}A\right)^{-1}$ and $S^{XY}V(S^{XY})^{\top} \geq \left(A^{\top}V^{-1}A\right)^{-1}$ in the matrix partial order so that $\widehat{\theta}$ is more efficient than both $\overline{\theta}$ and $\overline{\theta}^{XY}$. 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:

$$\mathrm{var}(\widehat{\theta}) \simeq \frac{\vartheta}{T}\begin{bmatrix} \frac{4-2\rho^2}{4-\rho^2} & \frac{2\rho}{4-\rho^2} \\ \frac{2\rho}{4-\rho^2} & \frac{4-2\rho^2}{4-\rho^2} \end{bmatrix};\ \mathrm{var}(\overline{\theta}) \simeq \frac{\vartheta}{T}\begin{bmatrix} 1 & \frac{\rho}{2} \\ \frac{\rho}{2} & 1 \end{bmatrix};\ \mathrm{var}(\overline{\theta}^{XY}) \simeq \frac{\vartheta}{T}\begin{bmatrix} 2 & 2\rho \\ 2\rho & 2 \end{bmatrix}.$$

For all ρ, $\mathrm{var}(\overline{\theta}) - \mathrm{var}(\widehat{\theta})$ is positive definite, strictly so for $\rho \neq 0$. For all ρ, $\mathrm{var}(\overline{\theta}^{XY}) - \mathrm{var}(\widehat{\theta})$ is positive definite, strictly so for $\rho \neq 1$. We conjecture that $\widehat{\theta}$ is semiparametrically efficient for estimation of θ. A feasible version of $\widehat{\theta}$, 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 like in (6).

We now turn to the testing problem. Define $\tau_E = \sqrt{T}(1, -1)\widehat{\theta}$ 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 $\mu_X = \mu_Y + \lambda/\sqrt{T}$,

$$\frac{\tau_E}{\widehat{\sigma_E}} \Longrightarrow N\left(\pi_E, 1\right),$$

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

[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:

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

$\sum_{j=-\infty}^{\infty} \gamma_{XY}(j) = \rho\vartheta$. We have

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

For the range $\rho \in [-1, 0.5]$, π_E^2/π^2 is quite modest, it lies in $[1, 1 \cdot 12]$, but as $\rho \to 1$, $\pi_E^2/\pi^2 \to \infty$. On the other hand $\pi_E^2/(\pi^{XY})^2 = 2 - \rho \in [1, 3]$.[5]

We briefly report the results of a simulation study that investigates τ_E, τ, τ^{XY} in the case where $X_t = X_t^* + \lambda/\sqrt{T}$ 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.

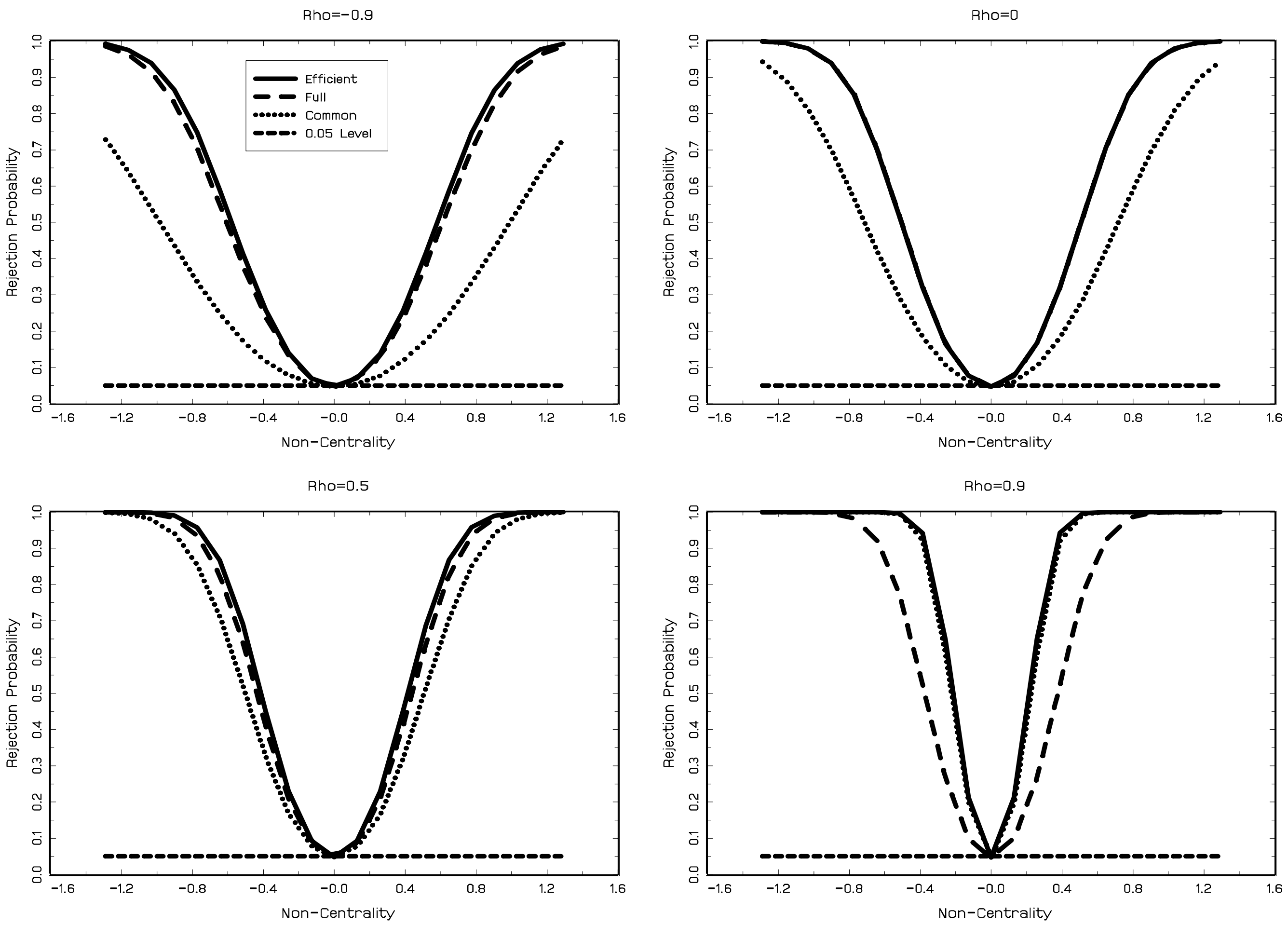

Figure 1.

[5]In this case we can write

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

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

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

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

and define estimates $\widehat{F}_X^E(z)$ and $\widehat{F}_Y^E(z)$ by the above minimum distance strategy. Then define $\delta^E = \sqrt{T} \max_{1 \le z_\ell \le L(T)} \{ \widehat{F}_X^E(z_\ell) - \widehat{F}_Y^E(z_\ell) \}$. By construction $\widehat{F}_X^E(z)$ and $\widehat{F}_Y^E(z)$ are more efficient than $\widehat{F}_X(z)$ and $\widehat{F}_Y(z)$, and it may be possible to show that tests based on δ^E are more powerful than those based on δ. 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 the case of unbalanced data. 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.

6 Appendix

Proof of Theorem 1. By standard arguments

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

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

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

We have

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

by dominated convergence. Define the integer sets

$$\begin{aligned}
I_u &= \{t : s - t = u; s = T^X + T^{XY} + 1, \ldots, T^X + T_Y; t = 1, \ldots, T^X\}, \\
I'_u &= \{t : s - t = u; s = T^X + T^{XY} + 1, \ldots, T^X + T_Y; t = T^X + 1, \ldots T^X + T^{XY}\}, \ u \geq 1,
\end{aligned}$$

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

$$\begin{aligned}
II &= \frac{1}{T_X T_Y} \sum_{t=1}^{T^X} \sum_{s=T^X+T^{XY}+1}^{T^X+T_Y} \gamma_{XY}(s-t) = \frac{1}{T_X T_Y} \sum_{u=T^{XY}+1}^{T^X+T_Y-1} n_u \gamma_{XY}(u) \\
&\leq \frac{1}{T_X T_Y} \sum_{u=T^{XY}+1}^{\infty} u |\gamma_{XY}(u)| = o(T^{-2}),
\end{aligned}$$

because $\sum_{u=1}^{\infty} u|\gamma_{XY}(u)| < \infty$. 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

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

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 $\sqrt{T}\{\widehat{F}_X(z) - \widehat{F}_Y(z)\}$, and this follows using the proof of Theorem 1, i.e.,

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

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

■

REFERENCES

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

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

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

Connor, G., and R.A. Korajczyk (1993). A test for the number of factors in an approximate factor model. *The Journal of Finance* XLVIII, 1263-1291.

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

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

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

Lahiri, S.N. (2003). *Resampling methods for Dependent data.* Springer Verlag, Berlin.

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

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

Phillips, P.C.B. (2004). HAC Estimation by Automated Regression. Mimeo.

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

Robinson, P.M., (2004). Robust Covariance Matrix Estimation: "HAC" Estimates with Long Memory/Antipersistence Correction. STICERD discussion paper

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