NBER TECHNICAL PAPER SERIES

MULTIVARIATE REGRESSION MODELS FOR PANEL DATA

Gary Chamberlain

Technical Paper No. 8

NATIONAL BUREAU OF ECONOMIC RESEARCH
1050 Massachusetts Avenue
Cambridge MA 02138

December 1980

I am grateful to Arthur Goldberger, Zvi Griliches, and Ariel Pakes for helpful discussions. Financial support was provided by the National Science Foundation (Grant No. SOC 7925959). The research reported here is part of the NBER's research program in Labor Economics and in Productivity. Any opinions expressed are those of the author and not those of the National Bureau of Economic Research.

NBER Technical Working Paper #8
December 1980

Multivariate Regression Models For Panel Data

ABSTRACT

Under stationarity, the heterogeneous stochastic processes are the non-ergodic ones. We show that if a distributed lag is of finite order, then its coefficients are unconditional means of the underlying random coefficients. This result is applied to linear transformations of the process. The estimation framework is a multivariate wide-sense regression function. The identification analysis requires certain restrictions on the coefficients. The actual regression function is nonlinear, and so we provide a theory of inference for linear approximations. It rests on obtaining the asymptotic distribution of functions of sample moments. Restrictions are imposed by using a minimum distance estimator; it is generally more efficient than the conventional estimators.

Professor Gary Chamberlain
Department of Economics
Social Science Building
1180 Observatory Drive
University of Wisconsin
Madison, WI 53706

(608) 262-7789

MULTIVARIATE REGRESSION MODELS FOR PANEL DATA

by

Gary Chamberlain
University of Wisconsin-Madison

1. INTRODUCTION

We shall consider a stochastic process $\{(x_t, y_t), t\in T\}$, where T is the set of all integers. We are primarily interested in the distribution of $\{y_t, t\in T\}$ conditional on $\{x_t, t\in T\}$. The paper has two main sections: one on identification and one on estimation. The identification problem is that the parameters of interest depend on the entire distribution of $\{(x_t, y_t), t\in T\}$, but we are only given the distribution of $\{(x_t, y_t), t=1, \ldots, T\}$. The estimation problem is to use a sample, $(x_{i1}, y_{i1}, \ldots, x_{iT}, y_{iT})$, $i=1, \ldots, N$, to make inferences about the distribution of $\{(x_t, y_t), t=1, \ldots, T\}$.

A typical application will be based on following people or firms through time. We do not assume that observing a long time series on a single individual allows us to infer the distribution of the stochastic process. In the case of a stationary process, the time average $\sum_{t=1}^{T} g(x_t, y_t)/T$ converges to a well defined random variable $(\hat{g})$ as $T \to \infty$ if $E|g(x_t, y_t)| < \infty$. We do not assume that $\hat{g} = E[g(x_t, y_t)]$; this is true if the variance of $\hat{g}$ is zero, but in general the variance will not be zero. If there is some $\hat{g}$ with non-zero variance, then we say that the $\{x_t, y_t\}$ process is heterogeneous. Then the heterogeneous stationary processes are precisely the non-ergodic ones.

The associated identification problem is that a parameter of interest may be $E(\hat{b}) = E[\lim_{T\to\infty} \sum_{t=1}^{T} (x_t - \bar{x}_T)(y_t - \bar{y}_T) / \sum_{t=1}^{T} (x_t - \bar{x}_T)^2]$. If the $\{x_t, y_t\}$ process is ergodic, then $E(\hat{b}) = Cov(x_1, y_1)/V(x_1)$, and we can identify $E(\hat{b})$ with $T = 1$. With a non-ergodic process, however, we can only say that $E(\lim_{T\to\infty} \bar{x}_T) = E(x_1)$, $E(\lim_{T\to\infty} \bar{y}_T) = E(y_1)$, $E(\lim_{T\to\infty} \sum_{t=1}^{T} x_t^2/T) = E(x_1^2)$, and $E(\lim_{T\to\infty} \sum_{t=1}^{T} x_t y_t/T) = E(x_1 y_1)$. Since in general $E(m_1 m_2) \neq E(m_1)\,E(m_2)$ and $E(m_1/m_2) \neq E(m_1)/E(m_2)$, we need more restrictions and may need $T > 1$. The population parameter $E(\hat{b})$ corresponds to the following estimator: accumulate a long time series on each of a large number of randomly sampled individuals; compute a linear regression coefficient for each individual; then average these coefficients across the individuals. This does not, in general, correspond to the regression coefficient obtained from a large cross-section sample at a single point in time.

Our solution to the identification problem takes the following form: we assume that $E(y_t|\ldots, x_{-1}, x_o, x_1, \ldots, c, \ldots, b_{-1}, b_o, b_1, \ldots) = c + \sum_{j=-\infty}^{\infty} b_j x_{t-j}$, where c and the b_j are constant over time but may vary across individuals; then if the $\{x_t\}$ process satisfies a regularity condition, our theorem asserts that if $E(y_t|\ldots, x_{-1}, x_0, x_1, \ldots) = E(y_t|x_t)$, then $E(y_t|x_t) = \gamma + \beta\, x_t$, where $\beta = E(b_o) = E(\hat{b})$ and $E(b_j) = 0$, $j \neq 0$. There are corresponding results for the case in which $E(y_t|\ldots, x_{-1}, x_o, x_1, \ldots)$ is not contemporaneous but displays lags and leads of finite order. We show how the theorem can be applied to linear transformations of the process; for example, if $\tilde{y}_t = y_t - y_{t-1}$, $\tilde{x}_t = x_t - x_{t-1}$, then $E(\tilde{y}_t|\ldots, \tilde{x}_{-1}, \tilde{x}_o, \tilde{x}_1, \ldots) = E(\tilde{y}_t|\tilde{x}_t)$ implies that $E(\tilde{y}_t|\tilde{x}_t) = \gamma + \beta\, \tilde{x}_t$, where $\beta = E(b_o)$ and $E(b_j) = 0$, $j \neq 0$.

Alternative linear transformations allow us to treat general rational distributed lag schemes.

Our estimation results focus on linear approximations. We assume that $(y_{i1}, \ldots, y_{iT}, x_{i1}, \ldots, x_{iT}) = (y_i', x_i')$, i=1, ..., N, are independently and identically distributed (i.i.d). We consider the following minimum mean square error linear predictors: $E^*(y_{it}|x_i) = \eta_t + \pi_t' x_i$, which we write as $E^*(y_i|x_i) = \eta + \Pi x_i$. The identification results suggest examining certain restrictions on Π. It is not appropriate, however, to assume that $E^*(y_i|x_i) = E(y_i|x_i)$; so we require a theory of inference for linear approximations.

This is provided by noting that Π is a function of population moments. The least squares estimator is the corresponding function of the sample moments. Hence the problem is to derive the limiting distribution of a differentiable function of sample moments, under i.i.d. sampling. This is straightforward and gives $\sqrt{N}(\text{vec } \hat{\Pi}' - \text{vec } \Pi') \xrightarrow{D} N(0, \Omega)$ as $N \to \infty$. We simplify the formula for Ω and provide a consistent estimator $\hat{\Omega}$.

We can impose restrictions within this framework by using a minimum distance estimator. We find the Π matrix satisfying the restrictions that is closest to $\hat{\Pi}$ in the norm provided by $\hat{\Omega}^{-1}$. This leads to some surprising results. For example, consider a univariate linear predictor (T=1): $E^*(y_i|x_{i1}, x_{i2}) = \eta + \pi_1 x_{i1} + \pi_2 x_{i2}$. We can impose the restriction that $\pi_2 = 0$ by using the least squares regression of y_i on x_{i1} to provide an estimator of π_1; however, that estimator is less efficient, in general, than our minimum distance estimator of π_1.

2. IDENTIFICATION

Consider a bivariate stochastic process $\{(x_t, y_t), t\varepsilon T\}$, where T is the set of all integers. We are interested in the distribution of $\{y_t, t\varepsilon T\}$ conditional on $\{x_t, t\varepsilon T\}$. A realization of the process is a point $\omega\varepsilon\Omega$ consisting of a doubly infinite sequence of vectors: $\omega = \{\ldots, \omega_{-1}, \omega_o, \omega_1, \ldots\} = \{\omega_t, t\varepsilon T\}$, where $\omega_t = (\omega_{1t}, \omega_{2t})$; $(x_t(\omega), y_t(\omega)) = (\omega_{1t}, \omega_{2t})$ is the $t^{\underline{th}}$ coordinate function.

A distinctive feature is that we shall consider the case in which more than one (partial) realization of the process is observed. Let $\underset{\sim}{z}_i = (x_{i1}, y_{i1}, \ldots, x_{iT}, y_{iT})$. We have a sample $\underset{\sim}{z}_i$, i=1, ..., N, where we regard the $\underset{\sim}{z}_i$ as independent and identically distributed (i.i.d) according to the probability law of $\{(x_t, y_t), t=1, \ldots T\}$. Typically i will index people or firms.

We shall consider in some detail the case of a stationary process. This will allow us to state clearly the key definitions and to obtain some sharp results. These results should provide guidance in attempts to deal with the additional identification problems created by non-stationarity. Our estimation results will not rely on any stationarity assumptions.

Consider the sample mean for a single individual (point):

$$\bar{x}_T(\omega) = \frac{1}{T}\sum_{t=1}^{T} x_t(\omega).$$

We shall not assume that observing a very long time series for a single individual allows us to infer the distribution of the stochastic process. In particular, we shall not assume that $\lim \bar{x}_T(\omega) = E(x_t)$ as $T \to \infty$. The

assumption that $\{(x_t, y_t), t\varepsilon T\}$ is a stationary stochastic process implies that $\lim \bar{x}_T(\omega)$ exists except for ω in a set with probability measure zero, which we shall abbreviate by saying that it exists a.e. We can set $\hat{x}(\omega) = \lim \bar{x}_T(\omega)$ and $\hat{x}$ is a well defined random variable. If it is degenerate, then $\hat{x} = E(x_t)$ a.e.; but the variance of $\hat{x}$, which we shall denote denote by $V(\hat{x})$, need not be zero.

More generally, consider time averages of the following form:

$$\lim_{T\to\infty} \frac{1}{T} \sum_{t=1}^{T} g(x_t(\omega), y_t(\omega)) = \lim_{T\to\infty} \bar{g}_T(\omega).$$

If $E|g(x_t, y_t)| < \infty$, then the ergodic theorem assets that $\hat{g}(\omega) = \lim \bar{g}_T(\omega)$ exists a.e. and $E(\hat{g}) = E[g(x_t, y_t)]$; in addition, we can interpret $\hat{g}$ as a conditional expectation: $\hat{g} = E[g(x_t, y_t)|J]$, where J is the σ-algebra of shift invariant sets. (See Billingsley (1965), theorem 1.3, p. 13 or Rozanov (1967), theorem 5.1, p. 157). If $V(\hat{g}) = 0$, then $\hat{g} = E[g(x_t, y_t)]$ a.e. In general, however, $V(\hat{g}) \neq 0$. We shall say that there is heterogeneity if $V(\hat{g}) \neq 0$ for some g. Then there is heterogeneity if and only if the process is not ergodic.

What is heterogeneity bias? Consider the following time average:

$$\hat{C}(x_t(\omega), y_t(\omega)) = \lim_{T\to\infty} \frac{1}{T} \sum_{t=1}^{T} x_t(\omega)y_t(\omega) - \left(\lim_{T\to\infty} \frac{1}{T} \sum_{t=1}^{T} x_t(\omega)\right)\left(\lim_{T\to\infty} \frac{1}{T} \sum_{t=1}^{T} y_t(\omega)\right).$$

If $E(x_t^2) < \infty$ and $E(y_t^2) < \infty$, then $\hat{C}(x_t, y_t)$ is a well defined, integrable random variable. Say that the population parameter of interest is $E[\hat{C}(x_t, y_t)]$. Consider measuring $E[\hat{C}(x_t, y_t)]$ by using $\text{Cov}(x_t, y_t) = C(x_t, y_t) = E(x_t y_t) - E(x_t)\,E(y_t)$. We shall say that there is a heterogeneity bias if $E[\hat{C}(x_t, y_t)] \neq C(x_t, y_t)$. Since

$$E[\hat{C}(x_t, y_t)] = E(x_t\, y_t) - E(\hat{x}\,\hat{y}), \quad \text{where}$$

$$\hat{x}(\omega) = \lim_{T\to\infty} \frac{1}{T} \sum_{t=1}^{T} x_t(\omega),\ \hat{y}(\omega) = \lim_{T\to\infty} \frac{1}{T} \sum_{t=1}^{T} y_t(\omega)\ ,$$

and since $E(\hat{x}) = E(x_t)$ and $E(\hat{y}) = E(y_t)$, we see that there is no heterogeneity bias if and only if $\mathrm{Cov}(\hat{x}, \hat{y}) = 0$. The absence of heterogeneity in x or in y is a sufficient condition for $\mathrm{Cov}(\hat{x}, \hat{y}) = 0$, for then $V(\hat{x}) = 0$ or $V(\hat{y}) = 0$.

The associated identification problem is somewhat more general. Clearly $E[\hat{C}(x_t, y_t)]$ depends on the entire distribution of $\{(x_t, y_t), t\varepsilon T\}$. We would like to know whether $E[\hat{C}(x_t, y_t)]$ can be recovered from a knowledge of the marginal distribution of $\{(x_t, y_t), t = 1, \ldots, T\}$, where T is small. This problem is relevant if we have an i.i.d. sample of $(x_{i1}, y_{i1}, \ldots, x_{iT}, y_{iT})$, $i = 1, \ldots, N$, where N is large. The use of $C(x_t, y_t)$ is an attempt to recover $E[\hat{C}(x_t, y_t)]$ from the marginal distribution of (x_1, y_1) when $T = 1$.

For a regression example, consider the following:

$$b_o(\omega) = \lim_{T\to\infty} \frac{\frac{1}{T}\sum_{t=1}^{T} x_t(\omega)y_t(\omega) - \left(\frac{1}{T}\sum_{t=1}^{T} x_t(\omega)\right)\left(\frac{1}{T}\sum_{t=1}^{T} y_t(\omega)\right)}{\frac{1}{T}\sum_{t=1}^{T} x_t^2(\omega) - \left(\frac{1}{T}\sum_{t=1}^{T} x_t(\omega)\right)^2} = \frac{\hat{C}(x_t, y_t)}{\hat{V}(x_t)}\ .$$

This is a well defined random variable if $\hat{V}(x_t) \neq 0$ a.e. This restriction on $\hat{V}(x_t)$ is an example of a regularity condition on the $\{x_t\}$ process; it is violated if there is a set of individuals with non-zero probability for whom x_t is constant over time. Assume that $E|b_o| < \infty$ and say that $E(b_o)$ is the population parameter of interest. Consider measuring $E(b_o)$ by using

$\beta_o = C(x_t, y_t)/V(x_t)$. We shall say that there is a heterogeneity bias if $E(b_o) \neq \beta_o$.

In order to analyze the bias, assume that

$$E^*(y_t|x_t, c, b_o) = c + b_o x_t$$

where

$$b_o = \frac{\hat{C}(x_t, y_t)}{\hat{V}(x_t)}, \qquad c = \hat{y} - b_o \hat{x},$$

and E^* is the minimum mean square error linear predictor -- the linear or wide sense regression function. (We would obtain this directly if we used $E^*(y_t|x_t, J)$ since $b_o = \hat{C}(x_t, y_t)/\hat{V}(x_t) = C(x_t, y_t|J)/V(x_t|J)$.) Since $\hat{C}(x_t, y_t) = b_o \hat{V}(x_t)$, we have

$$\frac{E[\hat{C}(x_t, y_t)]}{E[\hat{V}(x_t)]} = E(b_o)$$

if $C(b_o, \hat{V}(x_t)) = 0$, so that there is no correlation between the conditional variance of x and the slope of the linear predictor. In that case we can use

$$C(x_t, y_t) = E[\hat{C}(x_t, y_t)] + C(\hat{x}, \hat{y})$$

$$V(x_t) = E[\hat{V}(x_t)] + V(\hat{x})$$

to obtain

$$\beta = \frac{C(x_t, y_t)}{V(x_t)} = \frac{E[\hat{C}(x_t, y_t)]}{E[\hat{V}(x_t)]} \frac{E[\hat{V}(x_t)]}{V(x_t)} + \frac{C(\hat{x}, \hat{y})}{V(\hat{x})} \frac{V(\hat{x})}{V(x_t)}$$

$$= \lambda E(b_o) + (1-\lambda) \frac{C(\hat{x}, \hat{y})}{V(\hat{x})} ,$$

where $\lambda = \frac{E[\hat{V}(x_t)]}{V(x_t)}$, $0 < \lambda \leq 1$.

Hence there is no heterogeneity bias if and only if $V(\hat{x}) = 0$ or

$$\frac{C(x_t, y_t)}{V(x_t)} = \frac{C(\hat{x}, \hat{y})}{V(\hat{x})}$$

-- the cross sectional linear regression must coincide with the liner regression of the long run mean of y on the long run mean of x.

This suggests that the linear regression of $\bar{y}_T = \sum_{t=1}^{T} y_t/T$ on $\bar{x}_T = \sum_{t=1}^{T} x_t/T$ should have the same slope as the linear regression of y_t on x_t. If T = 2, this is equivalent to the linear regression of $y_2 - y_1$ on $x_2 - x_1$ having the same slope as the linear regression of y_1 on x_1 or y_2 on x_2. A sufficient condition is that

$$E^*(y_t|x_1, x_2) = E^*(y_t|x_t) = \gamma + \beta_o x_t, \qquad t=1,2,$$

so that

$$E^*(y_2-y_1|x_1, x_2) = \beta_o(x_2 - x_1).$$

2.1. THE THEOREM

A general version of a related argument is presented in appendix A. We assume that the following regression function is in fact linear:

$$E(y_t|\ldots x_{-1}, x_o, x_1, \ldots, c, \ldots, b_{-1}, b_o, b_1, \ldots)$$

$$= c + \sum_{j=-\infty}^{\infty} b_j x_{t-j}, \qquad t\epsilon T,$$

where the series is assumed to be absolutely convergent a.e. and integrable. The coefficients c, b_j are integrable, invariant random variables. The $\{x_t\}$ process is assumed to satisfy a regularity condition (C), which will be discussed below. Consider the regression function $E(y_t|\ldots, x_{-1}, x_o, x_1, \ldots)$, which we shall denote by $E(y_t|x)$; it does not condition on the random coefficients. We show that if this regression function displays finite lags and leads, then it must be linear and its coefficients are unconditional means of c and the b_j:

$$E(y_t|x) = E(y_t|x_{t-J}, \ldots, x_{t+M}) \qquad (0 \leq J, M < \infty)$$

implies that

$$E(y_t|x) = \gamma + \sum_{j=-M}^{J} \beta_j x_{t-j}, \text{ where}$$

$$\gamma = E(c) = E(c|x) \quad \text{a.e.}$$

$$\beta_j = E(b_j) = E(b_j|x) \text{ a.e. for } j\epsilon[-M, J]$$

$$E(b_j) = E(b_j|x) = 0 \text{ a.e. for } j \notin [-M, J].$$

So in this case there is no heterogeneity bias. If invariant random coefficients are correlated with x, then they must be correlated with the values of x in the distant future and distant past.

Next we shall indicate how the theorem can be applied to linear transformations of the model. Then we shall provide some interpretation of the regularity condition on $\{x_t\}$.

2.2 LINEAR TRANSFORMATIONS

Our assumption that $E(y_t|x, c, \ldots, b_{-1}, b_o, b_1, \ldots)$ is a linear regression function implies that

$$E(\tilde{y}_t|x, c, \ldots, b_{-1}, b_o, b_1, \ldots)$$

$$= \tilde{c} + \sum_{j=-\infty}^{\infty} b_j \tilde{x}_{t-j} = \tilde{c} + \sum_{j=-\infty}^{\infty} \tilde{b}_j x_{t-j} ,$$

where

$$\tilde{y}_t = y_t - \delta_1 y_{t-1} - \ldots - \delta_L y_{t-L}$$

$$\tilde{x}_t = x_t - \delta_1 x_{t-1} - \ldots - \delta_L x_{t-L}$$

$$\tilde{b}_j = b_j - \delta_1 b_{j-1} - \ldots - \delta_L b_{j-L}, \quad \tilde{c} = c(1-\delta_1 - \ldots - \delta_L).$$

Hence there are two ways in which we can apply the theorem, depending on which representation we use for the convolution.

We shall first consider the important case $\tilde{y}_t = y_t - y_{t-1}$, $\tilde{x}_t = x_t - x_{t-1}$; this is the basis for the analysis of covariance estimator.[1] It will be more fruitful to apply the theorem to the first representation of the convolution. Assume that the $\{\tilde{x}_t\}$ process satisfies the regularity condition (C). Then the theorem asserts that if

$$E(\tilde{y}_t|\tilde{x}) = E(\tilde{y}_t|\tilde{x}_{t-J}, \ldots, \tilde{x}_{t+M}) \quad (0 \leq J,M < \infty),$$

then

$$E(\tilde{y}_t|\tilde{x}) = \sum_{j=-M}^{J} \beta_j \tilde{x}_{t-j} \qquad \text{where}$$

$$\beta_j = E(b_j) = E(b_j|x) \text{ a.e.} \quad \text{for } j\in[-M, J]$$

$$E(b_j) = E(b_j|x) = 0 \quad \text{a.e.} \quad \text{for } j\notin[-M, J].$$

Say that we do not find finite lags and leads with $E(y_t|x)$ but we do with $E(\tilde{y}_t|\tilde{x})$. An interpretation is that $E(c|x) \neq E(c)$, which causes a heterogeneity bias in $E(y_t|x)$ that is eliminated by differencing in $E(\tilde{y}_t|\tilde{x})$. Another possibility is that $E(b_j|x) \neq E(b_j)$ but $E(b_j|\tilde{x}) = E(b_j)$. This is quite plausible since it is always true that $E^*(b_j|\tilde{x}) = E(b_j)$: by stationarity $\text{Cov}(b_j, \tilde{x}_t) = E(b_j x_t) - E(b_j x_{t-1}) = 0$ since b_j is an invariant random variable. So $E(\tilde{y}_t|x)$ will not necessarily display finite lags and leads, but we do not require that for the identification of $E(b_j)$. This is the reason that we chose the first form for the convolution.

Consider next the case $\tilde{y}_t = y_t - \rho\, y_{t-1}$, where $|\rho| < 1$. Here it is more fruitful to use the second form for the convolution:

$$E(\tilde{y}_t|x, c, \ldots, b_{-1}, b_o, b_1, \ldots) = \tilde{c} + \sum_{j=-\infty}^{\infty} \tilde{b}_j x_{t-j} .$$

If $\{x_t\}$ satisfies the regularity condition (C), then

$$E(\tilde{y}_t|x) = E(\tilde{y}_t|x_{t-J}, \ldots, x_{t+M}) \qquad (0 \leq J, M < \infty)$$

implies that

$$E(\tilde{y}_t|x) = \tilde{\gamma} + \sum_{j=-M}^{J} \tilde{\beta}_j x_{t-j}, \qquad \text{where}$$

$$\tilde{\gamma} = E(\tilde{c}) = E(\tilde{c}|x) \quad \text{a.e.}$$

$$\tilde{\beta}_j = E(\tilde{b}_j) = E(\tilde{b}_j|x) \quad \text{a.e.} \quad \text{for } j \in [-M, J]$$

$$E(\tilde{b}_j) = E(\tilde{b}_j|x) = 0 \quad \text{a.e.} \quad \text{for } j \notin [-M, J].$$

For example, say that $E(\tilde{y}_t|x) = E(\tilde{y}_t|x_t)$. Then $E(\tilde{y}_t|x) = \tilde{\gamma} + \tilde{\beta}_o x_t$, $E(\tilde{b}_o) = \tilde{\beta}_o$, $E(\tilde{b}_j) = 0$ for $j \neq 0$. The solution is

$$E(b_j) = \rho^j \tilde{\beta}_o, \quad j \geq 0$$

$$E(b_j) = 0, \quad j < 0.$$

So the $E(b_j)$ follow a geometric distributed lag.

More generally, let $\tilde{y}_t = y_t - \delta_1 y_{t-1} - \ldots - \delta_L y_{t-L}$, where the roots of the lag polynomial lie outside the unit circle. Then $E(y_t|x) = E(y_t|x_t, \ldots x_{t-J})$ implies that the $E(b_j)$ follow a general rational distributed lag scheme. For example, if $\tilde{y}_t = y_t - (\rho_1 + \rho_2) y_{t-1} + \rho_1\rho_2 y_{t-2}$, $|\rho_1| < 1$, $|\rho_2| < 1$, and $E(\tilde{y}_t|x) = E(\tilde{y}_t|x_t, x_{t-1})$, then we have

$$E(\tilde{y}_t|x) = \tilde{\gamma} + \tilde{\beta}_o x_t + \tilde{\beta}_1 x_{t-1}$$

$$E(b_j) = \tilde{\beta}_o \sum_{m=o}^{j} \rho_1^m \rho_2^{j-m} + \tilde{\beta}_1 \sum_{m=o}^{j-1} \rho_1^m \rho_2^{j-1-m}, \quad j \geq 1,$$

and $E(b_o) = \tilde{\beta}_o$, $E(b_j) = 0$ for $j < 0$.

2.3 REGULARITY CONDITIONS

The regularity condition (C) has two parts. The first part (C.1) rules out mover-stayer type processes for which there is non-zero probability that $x_t = x_{t-1}$ for all t. We can see the need for this condition by considering $\tilde{y}_t = y_t - y_{t-1}$, $\tilde{x}_t = x_t - x_{t-1}$, $E(\tilde{y}_t|\tilde{x}) = \beta_o \tilde{x}_t$; say we also assume that $b_j = 0$ for $j \neq 0$ so that $E(\tilde{y}_t|\tilde{x}, b_o) = b_o \tilde{x}_t$. Does this imply that $\beta_o = E(b_o)$? We can say $E(\tilde{y}_t|\tilde{x}) = E(b_o|\tilde{x})\, \tilde{x}_t$, which implies that $E(b_o|\tilde{x}) = \beta_o$ if $x_t \neq x_{t-1}$; but $E(b_o|\tilde{x})$ can take on any value if $x_t = x_{t-1}$ for all t, and so $E(b_o)$ can be arbitrarily far from β_o if there is non-zero probability that $x_t = x_{t-1}$ for all t.

If the probability that $x_t = x_{t-1}$ for all t is zero, then for almost all ω we can find a t, which may depend on ω, such that $x_t(\omega) \neq x_{t-1}(\omega)$; then for that t, $E(b_o|x)_\omega \tilde{x}_t(\omega) = \beta_o \tilde{x}_t(\omega)$ implies that $E(b_o|x)_\omega = \beta_o$. Hence $E(b_o|x) = \beta_o$ a.e. and so $E(b_o) = \beta_o$.

Condition (C.2) applied to $\tilde{x}_t = x_t - x_{t-1}$ rules out a process in which $\tilde{x}_t(\omega) = 0$ or $u(\omega)$ where $V(u) \neq 0$; in this process, x changes by the same amount for a given individual whenever it changes, and this amount is a non-degenerate random variable. To see the need for this condition, say that $E(\tilde{y}_t|\tilde{x}) = E(\tilde{y}_t|\tilde{x}_t) = E(b_o|\tilde{x})\tilde{x}_t$. This does not imply that $E(b_o|\tilde{x}) = E(b_o)$ a.e. since we can set $E(\tilde{y}_t|\tilde{x}_t) = \tilde{x}_t^2$ and $E(b_o|\tilde{x}) = u$. This is a valid solution since $\tilde{x}_t^2 = u\, \tilde{x}_t$ for all t; but $E(b_o|\tilde{x}) = u \neq E(b_o)$ if $V(u) > 0$.

2.4 IDENTIFICATION UNDER ALTERNATIVE ASSUMPTIONS

An alternative approach is to begin with the assumption that $E(y_t|x, c, \ldots, b_{-1}, b_o, b_1, \ldots)$ displays finite lags and leads in addition to

being linear:

$$E(y_t|x, c, \ldots, b_{-1}, b_o, b_1, \ldots) = c + \sum_{j=-M}^{J} b_j x_{t-j} .$$

We shall modify this formula by limiting ourselves to probability statements that involve only the finite dimensional distribution of $\{(x_t, y_t), t = 1, \ldots, T\}$. We shall illustrate by considering the case $T = 2$ and a purely contemporaneous relationship ($M = 0$, $J = 0$):

$$E(y_t|x_1, x_2, c, b_o) = c + b_o x_t, \qquad t = 1,2.$$

Then if $E(y_2 - y_1|x_2 - x_1)$ is linear, we have

$$E(y_2-y_1|x_2-x_1) = E(b_o|x_2-x_1)(x_2-x_1) = \beta_o(x_2-x_1).$$

As noted in our discussion of regularity condition (C.1), this does not by itself imply that $E(b_o) = \beta_o$; but if $P(x_2 = x_1) = 0$, then we do obtain $E(b_o|x_2-x_1) = \beta_o$ a.e. and so $E(b_o) = \beta_o$. The restriction $P(x_2 = x_1) = 0$ excludes the case in which x is discrete; for example, if $x_t = 0$ or 1, then there will be a non-zero probability that $x_1 = x_2 = 0$ or $x_1 = x_2 = 1$.

If we assume that $P(x_2 = x_1) = 0$, then there is a more powerful approach that does not rely on $E(b_o|x_2 - x_1) = E(b_o)$. We simply say

$$E(y_2-y_1|x_2-x_1) = E(b_o|x_2-x_1)\ (x_2-x_1) \quad \text{a.e.}$$

$$E\left(\frac{y_2-y_1}{x_2-x_1}\ |x_2-x_1\right) = E(b_o|x_2 - x_1) \quad \text{a.e.}$$

Hence if $E|(y_2-y_1)/(x_2-x_1)| < \infty$, we have

$$E\left(\frac{y_2-y_1}{x_2-x_1}\right) = E(b_o).$$

So if we have an i.i.d. sample of $z_i = (x_{i1}, y_{i1}, x_{i2}, y_{i2})$, $i=1, \ldots, N$, we can obtain a consistent estimator of $E(b_o)$ as $N \to \infty$ from the sample mean of $(y_{i2} - y_{i1})/(x_{i2} - x_{i1})$.

For $T > 2$ we compute a linear least squares regression of y_{it} on x_{it}, $t=1, \ldots, T$, for each individual i to obtain a slope coefficient $\hat{b}_{oi}$; then $\sum_{i=1}^{N} \hat{b}_{oi}/N$ converges almost surely to $E(b_o)$ as $N \to \infty$ if $E|\hat{b}_{oi}| < \infty$. This works since $E(\hat{b}_{oi}|x_{i1}, \ldots, x_{iT}, b_o) = b_o$; hence $E(\hat{b}_{oi}) = E(b_o)$ if $E|\hat{b}_{oi}| < \infty$. So we are averaging unbiased estimators. It would fail miserably if $x_t = y_{t-1}$ — it is essential that there be strict exogeneity: $E(y_t|x_1, \ldots, x_T, c, b_o) = c + b_o x_t$ and not just $E(y_t|x_t, c, b_o) = c + b_o x_t$.

Note that strict exogeneity is usually taken to mean[2]

$$E^*(y_t|x) = E^*(y_t|x_t, x_{t-1}, \ldots).$$

In addition to using regression functions instead of linear predictors, we require finite lags, not just one-sided lags. If we condition on all past

values of x and then find that the first few future values of x are irrelevant, this is not evidence against heterogeneity bias. For we show in appendix A that

$$E(c|x) = E(c|x_t, x_{t-1}, \ldots) = E(c|x_{t+1}, x_{t+2}, \ldots)$$

and the same is true for $E(b_j|x)$.

3. ESTIMATION

Consider a sample $\underset{\sim}{z}_i' = (\underset{\sim}{x}_i', \underset{\sim}{y}_i')$, i=1, ..., N, where $\underset{\sim}{x}_i' = (x_{i1}, \ldots, x_{iT})$ and $\underset{\sim}{y}_i' = (y_{i1}, \ldots, y_{iT})$. We shall assume that the $\underset{\sim}{z}_i$, i=1, ..., N, are independent and identically distributed (i.i.d.) according to some multivariate distribution with $E(x_{it}^4) < \infty$, $E(y_{it}^4) < \infty$, t=1, ..., T, and $V(\underset{\sim}{x}_i)$ non-singular. The results on estimation will not require stationarity assumptions. We shall focus on the estimation of the linear predictor of y_{it} conditional on $x_{i1}, \ldots, x_{iT}$. The key simplification is linearity in the parameters; the variables x_{it} could be non-linear functions of more basic variables. Also x_{it} could be a vector of variables, but we simplify notation by discussing the scaler case.

Consider the following minimum mean square error linear predictors:

$$E^*(y_{it}|\underset{\sim}{x}_i) = \eta_t + \underset{\sim}{\pi}_t' \underset{\sim}{x}_i , \qquad t=1, \ldots, T,$$

which we write as

$$E^*(\underset{\sim}{y}_i|\underset{\sim}{x}_i) = \underset{\sim}{\eta} + \underset{\sim}{\Pi} \underset{\sim}{x}_i , \qquad i=1, \ldots, N,$$

where

$$\underset{\sim}{\Pi}' = V^{-1}(\underset{\sim}{x}_i) \operatorname{Cov}(\underset{\sim}{x}_i, \underset{\sim}{y}_i').$$

Our theorem suggests that we look for finite lags and leads. This may be most promising for y_t near the end of the sample period, since insufficient lags may result in a full set of leads. (The point is, of course, symmetric;

if $b_{-j} \neq 0$, then insufficient leads may result in a full set of lags.)

In general it will not be true that $E^*(y_{-i}|x_{-i}) = E(y_{-i}|x_{-i})$. For small t there may be omitted lags and $E(x_{-j}|x_1, \ldots, x_T)$ is not linear, in general; throughout the sample period there may be non-linear terms arising from $E(c|x_1, \ldots, x_T) \neq E^*(c|x_1, \ldots, x_T)$ and $E(b_j|x_1, \ldots, x_T) \neq E(b_j)$. So in making inferences about $\underset{\sim}{\Pi}$, we shall be careful to treat $\underset{\sim}{\Pi}\, x_{-i}$ as a linear predictor and not assume that it is the multivariate regression function.

We shall also be interested in estimating linear transformations. Given our focus on linear predictors, there are simple relationships that must hold if a linear transformation results in zero coefficients on certain lags or leads. First consider differencing; say that it produces a contemporaneous relationship that is stationary apart from additive period effects:

$$E^*(y_t - y_{t-1}|x_2 - x_1, \ldots, x_T - x_{T-1}) = \psi_t + \beta\,(x_t - x_{t-1}), \quad t=2, \ldots, T.$$

Then the original $\underset{\sim}{\Pi}$ matrix must have been of the form

$$\underset{\sim}{\Pi} = \beta\, \underset{\sim}{I}_T + \underset{\sim}{\ell}\, \underset{\sim}{\delta}' ,$$

where $\underset{\sim}{\ell}$ is a T×1 vector of ones and $\underset{\sim}{\delta}$ is an unrestricted T×1 vector. So our criterion of looking for transformations that produce finite lags or leads can be stated in terms of restrictions on $\underset{\sim}{\Pi}$.

It is clear that a contemporaneous relationship holds a somewhat special position. It is easier to justify than a claim that two lagged values of x are necessary but not three lags. For an example where a contemporaneous

relationship might be expected to hold for the differences, say that $y_{it} = \ell n$ (hours worked), $x_{it} = \ell n$ (wage rate) for individual i in period t. We ignore the problems created by non-participation. Ghez and Becker (1975), MaCurdy (1980), and Heckman and MaCurdy (1980) have argued that $E^*(y_t|x_1, \ldots, x_T)$ will depend on all the x_t, since wages in all periods affect the individual's life time budget constraint. In a model in which individuals have perfect foresight, however, there are restrictions on utility functions that will justify a purely contemporaneous relationship between $y_t - y_{t-1}$ and $x_t - x_{t-1}$. Then it is of interest to test whether past and future values of $x_t - x_{t-1}$ really have zero coefficients.

Consider next the transformation $\tilde{y}_t = y_t - \rho\, y_{t-1}$, $|\rho| < 1$. In the contemporaneous case, we have

$$E^*(y_t - \rho\, y_{t-1}|x_1, \ldots, x_T) = \psi + \beta\, x_t .$$

Then it must be that

$$\underset{\sim}{\Pi} = \beta\, \underset{\sim}{D} + \underset{\sim}{\lambda}\, \underset{\sim}{\delta}',$$

where $\underset{\sim}{D}$ is a T×T lower triangular matrix:

$$\underset{\sim}{D} = \begin{pmatrix} 1 & & & \\ \rho & 1 & & 0 \\ \vdots & \vdots & & \\ \rho^{T-1} & \rho^{T-2} & \ldots \rho & 1 \end{pmatrix},$$

$\underset{\sim}{\delta}$ is an unconstrained T×1 vector, and $\underset{\sim}{\lambda}$ is a T×1 vector with $\lambda_t = \rho^{t-1}$, t=1, ..., T. This interpretation of λ_t was pointed out to me by Zvi Griliches and Ariel Pakes. Combining the two cases gives $\tilde{y}_t = (y_t - y_{t-1}) - \rho(y_{t-1} - y_{t-2})$ and

$$\underset{\sim}{\Pi} = \beta \underset{\sim}{D} + \underset{\sim}{\ell}\, \underset{\sim}{\delta}_1' + \underset{\sim}{\lambda}\, \underset{\sim}{\delta}_2' ,$$

where $\underset{\sim}{\delta}_1$ and $\underset{\sim}{\delta}_2$ are unconstrained T×1 vectors, $\underset{\sim}{\ell}$ is a T×1 vector of ones, and $\underset{\sim}{\lambda}$ is a T×1 vector with $\lambda_t = \rho^{t-1}$, t=1, ..., T.

We should stress once again that this is only a wide sense regression function. It may be that $E^*(\tilde{y}_t | x_2-x_1, \ldots, x_T-x_{T-1})$ is in fact a regression function (indeed, we hope that it is); but in integrating back to $E(y_t | x_1, \ldots, x_T)$ we pick up $E(c|x_1, \ldots, x_T)$ and $E(\sum_{j=o}^{\infty} \rho^j x_{-j} | x_1, \ldots, x_T)$, which are both non-linear in general. So we cannot rely on an asymptotic theory that begins with the assumption that we are estimating a true regression function.

2.1 THE ESTIMATION OF LINEAR APPROXIMATIONS

We are assuming that $\underset{\sim}{z}_i' = (\underset{\sim}{x}_i', \underset{\sim}{y}_i')$ is i.i.d., i=1, ..., N. So $\underset{\sim}{w}_i' = (\underset{\sim}{z}_i', \underset{\sim}{z}_i' \otimes \underset{\sim}{x}_i')$ is i.i.d. and

$$\underset{\sim}{\Pi}' = V^{-1}(\underset{\sim}{x}_i)\, \mathrm{Cov}(\underset{\sim}{x}_i, \underset{\sim}{y}_i')$$

is a function of $E(\underset{\sim}{w}_i)$. The least squares estimator $\hat{\underset{\sim}{\Pi}}'$ is the corresponding function of $\bar{\underset{\sim}{w}} = \sum_{i=1}^{N} \underset{\sim}{w}_i/N$. Hence our problem is to derive the limiting distribution of a differentiable function of a sample mean, under i.i.d. sampling.

This is a straightforward application of the δ-method and gives

$$\sqrt{N}(\text{vec } \hat{\underset{\sim}{\Pi}}' - \text{vec } \underset{\sim}{\Pi}') \xrightarrow{D} N(\underset{\sim}{0}, \underset{\sim}{\Omega}),$$

where

$$\underset{\sim}{\Omega} = E[(\underset{\sim}{y}_i^* - \underset{\sim}{\Pi}\, \underset{\sim}{x}_i^*)(\underset{\sim}{y}_i^* - \underset{\sim}{\Pi}\, \underset{\sim}{x}_i^*)' \otimes V^{-1}(\underset{\sim}{x}_i)(\underset{\sim}{x}_i^*\, \underset{\sim}{x}_i'^*)\, V^{-1}(\underset{\sim}{x}_i)],$$

$\underset{\sim}{x}_i^* = \underset{\sim}{x}_i - E(\underset{\sim}{x}_i)$, $\underset{\sim}{y}_i^* = \underset{\sim}{y}_i - E(\underset{\sim}{y}_i)$. (See appendix B.) A consistent (as $N \to \infty$) estimator of $\underset{\sim}{\Omega}$ is readily available from the corresponding sample moments:

$$\hat{\underset{\sim}{\Omega}} = \frac{1}{N} \sum_{i=1}^{N} \{[(\underset{\sim}{y}_i - \bar{\underset{\sim}{y}}) - \hat{\underset{\sim}{\Pi}}(\underset{\sim}{x}_i - \bar{\underset{\sim}{x}})][(\underset{\sim}{y}_i - \bar{\underset{\sim}{y}}) - \hat{\underset{\sim}{\Pi}}\, (\underset{\sim}{x}_i - \bar{\underset{\sim}{x}})]'$$

$$\otimes\, \underset{\sim}{S}_x^{-1}(\underset{\sim}{x}_i - \bar{\underset{\sim}{x}})(\underset{\sim}{x}_i - \bar{\underset{\sim}{x}})'\, \underset{\sim}{S}_x^{-1}\} ,$$

$$\bar{\underset{\sim}{y}} = \sum_{i=1}^{N} \underset{\sim}{y}_i/N, \quad \bar{\underset{\sim}{x}} = \sum_{i=1}^{N} \underset{\sim}{x}_i/N, \quad \underset{\sim}{S}_x = \sum_{i=1}^{N} (\underset{\sim}{x}_i - \bar{\underset{\sim}{x}})\, (\underset{\sim}{x}_i - \bar{\underset{\sim}{x}})'/N.$$

If $E(\underset{\sim}{y}_i|\underset{\sim}{x}_i) = \underset{\sim}{\eta} + \underset{\sim}{\Pi}\, \underset{\sim}{x}_i$, then we have

$$\underset{\sim}{\Omega} = E[V(\underset{\sim}{y}_i|\underset{\sim}{x}_i) \otimes V^{-1}(\underset{\sim}{x}_i)(\underset{\sim}{x}_i^*\, \underset{\sim}{x}_i'^*)\, V^{-1}(\underset{\sim}{x}_i)].$$

If $V(y_i|x_i)$ is uncorrelated with $x_i^*\, x_i'^*$, then

$$\underset{\sim}{\Omega} = E[V(\underset{\sim}{y}_i|\underset{\sim}{x}_i)] \otimes V^{-1}(\underset{\sim}{x}_i) .$$

If we have homoskedastic variance, so that $V(\underset{\sim}{y}_i|\underset{\sim}{x}_i) = \underset{\sim}{\Sigma}$ does not depend on $\underset{\sim}{x}_i$, then $\underset{\sim}{\Omega} = \underset{\sim}{\Sigma} \otimes V^{-1}(\underset{\sim}{x}_i)$.

This approach was used by Cramèr (1946) to obtain limiting normal distributions for sample correlation and regression coefficients (p. 367); he presents an explicit formula for the asymptotic variance of a simple correlation coefficient (p. 359). Kendall and Stuart (1961, p. 293) and

Goldberger (1974) present the formula for the asymptotic sampling variance of a simple regression coefficient. Using a different approach, White (1980) obtains the asymptotic covariance matrix for univariate regression coefficients.

2.2 IMPOSING RESTRICTIONS

We can impose restrictions on $\underset{\sim}{\Pi}$ by using a minimum distance estimator:

$$\min_{\underset{\sim}{\theta}} [\text{vec } \hat{\underset{\sim}{\Pi}}' - \text{vec } \underset{\sim}{\Pi}'(\theta)] \; \hat{\underset{\sim}{\Omega}}^{-1} [\text{vec } \hat{\underset{\sim}{\Pi}}' - \text{vec } \underset{\sim}{\Pi}'(\underset{\sim}{\theta})],$$

where the constraints on $\underset{\sim}{\Pi}$ are specified by the condition that $\underset{\sim}{\Pi}$ depends only on a p×1 parameter vector $\underset{\sim}{\theta}$. The domain of $\underset{\sim}{\theta}$ is some compact subset of R^p. The relevant asymptotic theory is developed, for example, in Malinvaud (1970, Chap. 9). Since we already have the asymptotic distribution of $\hat{\underset{\sim}{\Pi}}$, we only need to assume that $\underset{\sim}{\Omega}$ is non-singular and that the $\underset{\sim}{\Pi}(\underset{\sim}{\theta})$ functions are continuous and satisfy Malinvaud's assumptions 5 and 6 (p. 349), which impose invertibility and differentiability restrictions in a neighborhood of the true parameter value $\underset{\sim}{\theta}^o$. Then $\sqrt{N}(\hat{\underset{\sim}{\theta}} - \underset{\sim}{\theta}^o) \xrightarrow{D} N(\underset{\sim}{0}, \underset{\sim}{\Lambda})$; a consistent (as $N \to \infty$) estimator of $\underset{\sim}{\Lambda}$ is provided by

$$\hat{\underset{\sim}{\Lambda}} = (\underset{\sim}{G}' \; \hat{\underset{\sim}{\Omega}}^{-1} \; \underset{\sim}{G})^{-1} ,$$

where $\underset{\sim}{G} = \partial \text{ vec } \underset{\sim}{\Pi}'(\underset{\sim}{\theta})/\partial\underset{\sim}{\theta}'$ is evaluated at $\underset{\sim}{\theta} = \hat{\underset{\sim}{\theta}}$. We can test the restrictions by using the limiting distribution of

$$N[\text{vec } \hat{\underset{\sim}{\Pi}}' - \text{vec } \underset{\sim}{\Pi}'(\hat{\underset{\sim}{\theta}})]' \; \hat{\underset{\sim}{\Omega}}^{-1} [\text{vec } \hat{\underset{\sim}{\Pi}}' - \text{vec } \underset{\sim}{\Pi}'(\hat{\underset{\sim}{\theta}})],$$

which is χ^2 with MK−p degrees of freedom if $\underset{\sim}{\Pi}$ is M×K.

This framework leads to some surprising results on efficient estimation. For a simple example, we shall use a univariate linear regression model:

$$E^*(y_i|x_{i1}, x_{i2}) = \eta + \pi_1 x_{i1} + \pi_2 x_{i2} \qquad (i=1, \ldots, N).$$

Consider imposing the restriction $\pi_2 = 0$. If we assume that the distribution of y_i conditional on x_{i1}, x_{i2} is $N(\eta + \pi_1 x_{i1} + \pi_2 x_{i2}, \sigma^2)$, then under random sampling the efficient, maximum likelihood estimator of π_1 is b_{yx_1}, the slope coefficient in the least squares regression of y on x_1. Let $\hat{\pi}_1$, $\hat{\pi}_2$ be the coefficients in the multiple regression of y on x_1, x_2. If $Cov(x_{i1}, x_{i2}) = 0$ and $\pi_2 = 0$, then $\sqrt{N}(b_{yx_1} - \pi_1)$ and $\sqrt{N}(\hat{\pi}_1 - \pi_1)$ have the same limiting distribution -- but $\hat{\pi}_1$ is generally less efficient than the minimum distance estimator if the regression function is non-linear or if there is heteroskedasticity.

To see this, consider the asymptotic covariance between $\hat{\pi}_1$ and $\hat{\pi}_2$:

$$\omega_{12} = \frac{E[(y_i^* - \pi_1 x_{i1}^* - \pi_2 x_{i2}^*)^2 \, x_{i1}^* \, x_{i2}^*]}{V(x_{i1}) \, V(x_{i2})} .$$

This is non-zero, in general, even though $Cov(x_{i1}, x_{i2}) = 0$. So if $\pi_2 = 0$, we gain efficiency in estimating π_1 by considering a linear combination $\hat{\pi}_1 + \tau \hat{\pi}_2$, where τ is chosen to minimize the variance. (We have consistency for any τ since $\hat{\pi}_2$ almost surely converges to $\pi_2 = 0$ as $N \to \infty$.) Hence we want $\tau = -\omega_{12}/\omega_{22}$. It is a general property of minimum distance estimation that substituting a consistent estimator ($\hat{\Omega}$) for Ω does not affect the limiting distribution. So we use

$$\hat{\pi}_1^* = \hat{\pi}_1 - \frac{\hat{\omega}_{12}}{\hat{\omega}_{22}} \hat{\pi}_2 .$$

This has a smaller asymptotic variance than $\hat{\pi}_1$ (or b_{yx_1}) unless

$(y_i - \eta - \pi_1 x_{i1} - \pi_2 x_{i2})^2$ is uncorrelated with x^*_{i1} x^*_{i2}.

If $\text{Cov}(x_{i1}, x_{i2}) \neq 0$, then our minimum distance estimator still has the form $\hat{\pi}^*_1 = \hat{\pi}_1 - (\hat{\omega}_{12}/\hat{\omega}_{22})\,\hat{\pi}_2$, where $\hat{\omega}_{12}$ and $\hat{\omega}_{22}$ are obtained from our general formula for $\hat{\Omega}$. In the special case where $E(y|x_1, x_2)$ is linear and $V(y|x_1, x_2) = \sigma^2$, then $\omega_{12}/\omega_{22} = -\text{Cov}(x_{i1}, x_{i2})/V(x_{i1})$ and $\hat{\pi}^*_1 = b_{yx_1}$; but in general they differ and b_{yx_1} is not efficient since it is correlated with $\hat{\pi}_2$.[3]

2.3 AN EXTENSION

Consider a panel of N farms observed over T seasons. The technology is Cobb-Douglas:

$$y_{it} = \beta\, x_{it} + c_i + u_{it} \qquad (i=1, \ldots, N;\ t=1, \ldots, T),$$

where y = ℓn (output), x = ℓn (input), c represents an unmeasured input (soil quality) that is constant over the period of the sample, and u represents a stochastic input such as rainfall. All farms use the same technology so there is no variation in β.

Assume that farmers choose x to maximize expected profits, conditional on an information set that includes c. Then $E(c_i|x_{i1}, \ldots, x_{iT}) \neq E(c_i)$; under plausible assumptions, farmers on the higher quality soil will use more of the variable factor.[4] The transformation to $\tilde{y}_{it} = y_{it} - y_{i,t-1}$, $\tilde{x}_{it} = x_{it} - x_{i,t-1}$ is appropriate if $\text{Cov}(x_{it}, u_{is}) = 0$, s, t = 1,..., T. This will not hold, however, if the farmers forecast u_t on the basis of past values.

Say that $u_{it} = \rho\, u_{i,t-1} + v_{it}$, where v_{it} is serially independent. Then

$$y_{it} - y_{i,t-1} = \beta\,(x_{it} - x_{i,t-1}) - \beta\,\rho(x_{i,t-1} - x_{i,t-2})$$

$$+ \rho(y_{i,t-1} - y_{i,t-2}) + v_{it} - v_{i,t-1}.$$

Since $\mathrm{Cov}(x_t, v_s) = 0$ for $t \leq s$ and $\mathrm{Cov}(y_t, v_s) = 0$ for $t < s$, we can obtain consistent estimates by using instrumental variables based on $x_{t-1}, x_{t-2}, \cdots, x_1$ and $y_{t-2}, y_{t-3}, \cdots, y_1$.

The asymptotic distribution for the instrumental variable estimator is derived in appendix B. We set

$$\hat{\underset{\sim}{\pi}}_m = \Big[\sum_{i=1}^{N} (\underset{\sim}{q}_{im} - \bar{\underset{\sim}{q}}_m)(\underset{\sim}{x}_{im} - \bar{\underset{\sim}{x}}_m)'\Big]^{-1} \sum_{i=1}^{N} (\underset{\sim}{q}_{im} - \bar{\underset{\sim}{q}}_m)(y_{im} - \bar{y}_m)$$

$$\underset{\sim}{\pi}_m = [\mathrm{Cov}(\underset{\sim}{q}_{im}, \underset{\sim}{x}'_{im})]^{-1} \mathrm{Cov}(\underset{\sim}{q}_{im}, y_{im}) \qquad (m=1, \ldots, M),$$

where $\underset{\sim}{x}_{im}$ is $K_m \times 1$, $\underset{\sim}{q}_{im}$ is $K_m \times 1$, and $\mathrm{Cov}(\underset{\sim}{q}_{im}, \underset{\sim}{x}'_{im})$ is assumed to be non-singular. We assume i.i.d. sampling from a distribution with finite forth moments. Then $\sqrt{N}(\hat{\underset{\sim}{\pi}}_m - \underset{\sim}{\pi}_m)$ and $\sqrt{N}(\hat{\underset{\sim}{\pi}}_n - \underset{\sim}{\pi}_n)$ have a limiting normal distribution with mean $\underset{\sim}{0}$ and covariance matrix

$$\underset{\sim}{\Omega}_{mn} = E[(y^*_{im} - \underset{\sim}{\pi}'_m \underset{\sim}{x}^*_{im})(y^*_{in} - \underset{\sim}{\pi}'_n \underset{\sim}{x}^*_{in})\, \underset{\sim}{\Phi}_m^{-1} (\underset{\sim}{q}^*_{im}\, \underset{\sim}{q}'^*_{in}) \underset{\sim}{\Phi}'^{-1}_n],$$

where $\underset{\sim}{\Phi}_m = \mathrm{Cov}(\underset{\sim}{q}_{im}, \underset{\sim}{x}'_{im})$, $y^*_{im} = y_{im} - E(y_{im})$, $\underset{\sim}{x}^*_{im} = \underset{\sim}{x}_{im} - E(\underset{\sim}{x}_{im})$, $\underset{\sim}{q}^*_{im} = \underset{\sim}{q}_{im} - E(\underset{\sim}{q}_{im})$, $m, n = 1, \ldots, M$.

In our example, y_{im} is $y_{it} - y_{i,t-1}$; $\underset{\sim}{x}_{im}$ contains $(x_{it} - x_{i,t-1})$, $(x_{i,t-1} - x_{i,t-2})$, $(y_{i,t-1} - y_{i,t-2})$; and $\underset{\sim}{q}_{im}$ contains three variables based on $x_{i,t-1}, x_{i,t-2}, \cdots, x_{i1}, y_{i,t-2}, y_{i,t-3}, \cdots, y_{i1}$.

We can impose restrictions on the coefficients, both within and across equations, by using the minimum distance procedure. There is, however, another efficiency issue, since we have more than K_m instrumental variables for the $m^{\underline{th}}$ equation. Given our previous discussion of imposing restrictions, it is perhaps not surprising that the standard two stage least squares estimator is not, in general, an efficient procedure for combining instrumental variables. We shall demonstrate this by means of a simple example.

Say that $(y_i, x_i, q_{i1}, q_{i2})$ is i.i.d. according to some distribution with finite forth moments. Assume that $E(y_i) = E(x_i) = E(q_{i1}) = E(q_{i2}) = 0$, $E(x_i q_{i1}) \neq 0$, $E(x_i q_{i2}) \neq 0$, and

$$\frac{E(y_i q_{i1})}{E(x_i q_{i1})} = \frac{E(y_i q_{i2})}{E(x_i q_{i2})} = \pi.$$

Then we have two instrumental variable estimators that both almost surely converge to π as $N \to \infty$:

$$\hat{\pi}_1 = \sum_{i=1}^{N} y_i q_{i1} / \sum_{i=1}^{N} x_i q_{i1}, \quad \hat{\pi}_2 = \sum_{i=1}^{N} y_i q_{i2} / \sum_{i=1}^{N} x_i q_{i2}.$$

The two stage least squares estimator combines $\hat{\pi}_1$ and $\hat{\pi}_2$ by forming $\hat{q}_3 = \hat{\psi}_1 q_1 + \hat{\psi}_2 q_2$ based on the least squares regression of x on q_1, q_2 (We shall assume that $V(q_{i1}, q_{i2})$ is non-singular):

$$\hat{\pi}_{TSLS} = \sum_{i=1}^{N} y_i \hat{q}_{i3} / \sum_{i=1}^{N} x_i \hat{q}_{i3} = \hat{\delta}\, \hat{\pi}_1 + (1-\hat{\delta})\, \hat{\pi}_2 ,$$

where $$\hat{\delta} = \hat{\psi}_1 \sum_{i=1}^{N} x_i q_{i1} / (\hat{\psi}_1 \sum_{i=1}^{N} x_i q_{i1} + \hat{\psi}_2 \sum_{i=1}^{N} x_i q_{i2}).$$

So we have

$$\sqrt{N}\ (\hat{\pi}_{TSLS} - \pi) = \hat{\delta}\ \sqrt{N}\ (\hat{\pi}_1 - \pi) + (1 - \hat{\delta})\ \sqrt{N}\ (\hat{\pi}_2 - \pi).$$

We can obtain the limiting distribution of $\sqrt{N}(\hat{\pi}_1 - \pi,\ \hat{\pi}_2 - \pi)$ from our general result on instrumental variable estimation: set $M = 2$, $\underset{\sim}{x}_{i1} = x_i$, $\underset{\sim}{x}_{i2} = x_i$, $\underset{\sim}{q}_{i1} = q_{i1}$, $\underset{\sim}{q}_{i2} = q_{i2}$, $y_{i1} = y_i$, $y_{i2} = y_i$. (We never assumed that $V(y_{i1}, y_{i2})$ is non-singular.) So we have

$$\sqrt{N}\ \left[\begin{pmatrix}\hat{\pi}_1\\ \hat{\pi}_2\end{pmatrix} - \begin{pmatrix}\pi\\ \pi\end{pmatrix}\right] \xrightarrow{D} N(\underset{\sim}{0},\ \underset{\sim}{\Omega}),$$

where $\underset{\sim}{\Omega} = (\omega_{jk})$:

$$\omega_{jk} = \frac{E[(y_i - \pi x_i)^2\ q_{ij}\ q_{ik}]}{E(q_{ij}x_i)\ E(q_{ik}x_i)}, \quad j,k = 1,2.$$

Since

$$\hat{\delta} \xrightarrow{a.s.} \frac{\psi_1 E(x_i q_{i1})}{\psi_1 E(x_i q_{i1}) + \psi_2\ E(x_i q_{i2})} = \delta \text{ as } N \to \infty,$$

it follows that

$$\sqrt{N}\ (\hat{\pi}_{TSLS} - \pi) \xrightarrow{D} N(0,\ \underset{\sim}{\delta}'\ \underset{\sim}{\Omega}\ \underset{\sim}{\delta}) \text{ as } N \to \infty$$

where $\underset{\sim}{\delta}' = (\delta,\ 1-\delta)$. (Note that the limiting distribution coincides with what we would obtain using $q_3 = E^*(x|q_1,q_2) = \psi_1 q_1 + \psi_2 q_2$ instead of $\hat{q}_3$.)

We see that the limiting distribution of $\hat{\pi}_{TSLS}$ coincides with that of $\delta\ \hat{\pi}_1 + (1-\delta)\ \hat{\pi}_2$. This suggests finding the τ that minimizes the asymptotic variance of $\tau\ \hat{\pi}_1 + (1-\tau)\ \hat{\pi}_2$; the answer leads to the minimum distance estimator:

$$\min_{\theta} \left\{ \left[\begin{pmatrix} \hat{\pi}_1 \\ \hat{\pi}_2 \end{pmatrix} - \begin{pmatrix} \theta \\ \theta \end{pmatrix} \right]' \underset{\sim}{\Omega}^{-1} \left[\begin{pmatrix} \hat{\pi}_1 \\ \hat{\pi}_2 \end{pmatrix} - \begin{pmatrix} \theta \\ \theta \end{pmatrix} \right] \right\}$$

gives $\hat{\pi}^* = \tau \hat{\pi}_1 + (1-\tau) \hat{\pi}_2$,

where $\tau = (\omega^{11} + \omega^{12})/(\omega^{11} + 2\omega^{12} + \omega^{22})$

and ω^{jk} is the j,k element of $\underset{\sim}{\Omega}^{-1}$. The estimator obtained by using a consistent $\hat{\underset{\sim}{\Omega}}$ instead of $\underset{\sim}{\Omega}$ has the same limiting distribution.

In general $\tau \neq \delta$ since τ is a function of forth moments and δ is not. For example, say that $x_i = q_{i2}$. Then $\delta = 0$ but $\tau \neq 0$ unless

$$E\left[(y_i - \pi x_i)^2 \left(\frac{x_i^2}{E(x_i^2)} - \frac{x_i q_{i1}}{E(x_i q_{i1})} \right) \right] = 0.$$

Hence the asymptotic variance of $\hat{\pi}^*$ is less than that of $\hat{\pi}_{TSLS}$ unless $(y_i - \pi x_i)^2$ is uncorrelated with $[x_i^2/E(x_i^2)] - [(x_i q_{i1})/E(x_i q_{i1})]$.

If we add another equation then we can consider the conventional three stage least squares estimator. Its limiting distribution can be derived in a straightforward fashion since the estimator is a differentiable function of sample moments; however, viewed as a minimum distance estimator, it is using the wrong norm in general.

Instead we should simply recognize that the restrictions can be expressed as restrictions on population moments. We let $\underset{\sim}{z}_i' = (\underset{\sim}{x}_i', y_i')$ and let $\underset{\sim}{w}_i$ be the vector formed from $\underset{\sim}{z}_i$ and the lower triangle of $\underset{\sim}{z}_i \underset{\sim}{z}_i'$. Then the population parameters of interest are functions of $E(\underset{\sim}{w}_i)$. The restrictions can be expressed by the condition that $E(\underset{\sim}{w}_i)$ depends only on a p-dimensional parameter vector $\underset{\sim}{\theta}$: $E(\underset{\sim}{w}_i) = \underset{\sim}{g}(\underset{\sim}{\theta})$, where the domain of $\underset{\sim}{\theta}$ is some compact subset of R^p that

contains the true parameter vector $\underset{\sim}{\theta}^{o}$. Under i.i.d. sampling we have $\sqrt{N}\ [\bar{\underset{\sim}{w}} - E(\underset{\sim}{w}_i)] \xrightarrow{D} N(\underset{\sim}{0}, V(\underset{\sim}{w}_i))$ if $V(\underset{\sim}{w}_i)$ is finite; a consistent estimator of $V(\underset{\sim}{w}_i)$ is provided by

$$\hat{V}(\underset{\sim}{w}_i) = \frac{1}{N} \sum_{i=1}^{N} (\underset{\sim}{w}_i - \bar{\underset{\sim}{w}})(\underset{\sim}{w}_i - \bar{\underset{\sim}{w}})'.$$

If $V(\underset{\sim}{w}_i)$ is non-singular and if $\underset{\sim}{g}$ is continuous and satisfies Malinvaud's assumptions 5 and 6, then

$$\min_{\theta} [\bar{\underset{\sim}{w}} - \underset{\sim}{g}(\underset{\sim}{\theta})]' \hat{V}^{-1}(\underset{\sim}{w}_i) [\bar{\underset{\sim}{w}} - \underset{\sim}{g}(\underset{\sim}{\theta})]$$

gives

$$\sqrt{N}(\hat{\underset{\sim}{\theta}} - \underset{\sim}{\theta}^{o}) \xrightarrow{D} N(\underset{\sim}{0}, \underset{\sim}{\Lambda}) \text{ as } N \to \infty ,$$

where a consistent estimator of $\underset{\sim}{\Lambda}$ is provided by

$$\hat{\underset{\sim}{\Lambda}} = [\underset{\sim}{G}' \hat{\underset{\sim}{V}}^{-1}(\underset{\sim}{w}_i) \underset{\sim}{G}]^{-1}$$

with $\underset{\sim}{G} = \partial \underset{\sim}{g}/\partial \underset{\sim}{\theta}'$ evaluated at $\underset{\sim}{\theta} = \hat{\underset{\sim}{\theta}}$.

4. CONCLUSION

Under stationarity, the heterogeneous stochastic processes are the non-ergodic ones. We showed that if a distributed lag is of finite order, then its coefficients are the unconditional means of the underlying random coefficients. This result can be usefully applied to linear transformations of the process.

The estimation framework is a multivariate wide-sense regression function. The identification analysis suggests that we examine certain restrictions on the coefficients. The actual regression function is non-linear, and so we provide a theory of inference for linear approximations. It rests on obtaining the asymptotic distribution of functions of sample moments. Restrictions are imposed by using a minimum distance estimator; it is generally more efficient than the conventional estimators.

APPENDIX A

Let Ω be a set of points where $\omega \in \Omega$ is a doubly infinite sequence of vectors of real numbers: $\omega = \{\ldots, \omega_{-1}, \omega_0, \omega_1, \ldots\} = \{\omega_t, \ t \in T\}$, where $\omega_t = (\omega_{1t}, \omega_{2t}) \in R^2$ and T is the set of all integers. Let $(x_t(\omega), y_t(\omega)) = (\omega_{1t}, \omega_{2t})$ be the t^{th} coordinate function. Let F be the σ-field consisting of sets of the form

$$A = \{\omega: \ (x_t(\omega), y_t(\omega)) \in B_0, \ldots, (x_{t+k}(\omega), y_{t+k}(\omega)) \in B_k\},$$

where $t,k \in T$ and the B's are two-dimensional Borel sets. Let P be a probability measure defined on F such that $\{(x_t, y_t), \ t \in T\}$ is a stationary stochastic process on the probability space (Ω, F, P).

The shift transformation S is defined by

$$(x_t(S\omega), y_t(S\omega)) = (x_{t+1}(\omega), y_{t+1}(\omega)).$$

It is an invertible, measure preserving transformation. A random variable z is invariant if $z(S\omega) = z(\omega)$ a.e. (We shall use "a.e." as an abbreviation for "except on a set of probability measure zero.") A set $A \in F$ is invariant if its indicator function is an invariant random variable. A σ-field $G \subset F$ is invariant if $SG = G$; that is, for every $G \in G$ we have $SG \in G$ and $SH = G$ for some $H \in G$.

We shall use $E(z|G)_\omega$ to denote the conditional expectation of the random variable z with respect to the σ-field G, evaluated at ω. We shall use $\sigma(x)$ to denote the σ-field generated by $\ldots, x_{-1}, x_0, x_1, \ldots$, and $E(z|x)$ will denote $E(z|\sigma(x))$.

We shall need the following two lemmas.

LEMMA 1. If G is an invariant σ-field, then

$$E(z|G)_{S\omega} = E(v|G)_{\omega} \quad \text{a.e.}$$

where $v(\omega) = z(S\omega)$.

PROOF. A change of variable argument shows that

$$E(z|G)_{S\omega} = E(v|S^{-1}G)_{\omega} .$$

(See Billingsley (1965), example 10.3, p. 109.) The result follows since $S^{-1}G = G$.

Q.E.D.

LEMMA 2. If G is an invariant σ-field and if z is an invariant random variable, then $E(z|G)$ is an invariant random variable.

PROOF. By Lemma 1, $E(z|G)_{S\omega} = E(v|G)_{\omega}$ a.e., where $v(\omega) = z(S\omega)$; since z is invariant, $z(S\omega) = z(\omega)$ a.e.; hence $E(z|G)_{S\omega} = E(z|G)_{\omega}$ a.e.

Q.E.D.

Our theorem will require that the $\{x_t\}$ process satisfy the following regularity condition:

CONDITION (C). 1. Let $\{(z_{1t},z_{2t}), t \in T\}$ be a stationary stochastic process where z_{1t} is measurable $\sigma(x_t,\ldots,x_{t+k})$ and z_{2t} is measurable $\sigma(\ldots,x_{t-1},x_{t+k+1},\ldots)$, $t,k \in T$. Then

$$P(z_{1t} = z_{2t} \text{ for all } t) = 0$$

unless z_{1t} is a constant a.e.

2. Let $A_1 \in \sigma(x_t,\ldots,x_{t+k})$, $A_2 \in \sigma(\ldots,x_{t-1},x_{t+k+1},\ldots)$, $t,k \in T$. Then

$$P(A_1 \cap A_2) = P(A_1)$$

implies $P(A_1) = 0$ or $P(A_2) = 1$.

In addition, we assume that $\mathrm{Var}(x_t) \neq 0$.

THEOREM. If $\{x_t\}$ satisfies the regularity condition (C), if $E|y_t| < \infty$, and if

$$E(y_t|\sigma(G \cup \sigma(x))) = c + \sum_{j=-\infty}^{\infty} b_j x_{t-j} \quad \text{a.e.,} \quad t \in T,$$

where c and the b_j are integrable, invariant random variables and $\sum_{j=-\infty}^{\infty} |b_j x_{t-j}| < \infty$ a.e. and integrable, then

$$E(y_t|x) = E(y_t|x_{t-J},\ldots,x_{t+M}) \text{ a.e.} \quad (t \in T, \quad 0 \leq J,M < \infty)$$

implies that

$$E(y_t|x) = \gamma + \sum_{j=-M}^{J} \beta_j x_{t-j} \quad \text{a.e.,} \quad t \in T,$$

where

$$\gamma = E(c) = E(c|x) \quad \text{a.e.}$$

$$\beta_j = E(b_j) = E(b_j|x) \quad \text{a.e.,} \quad j \in [-M,J]$$

$$E(b_j|x) = 0 \quad \text{a.e.,} \quad j \notin [-M,J].$$

(We can take G to be J, the σ-field of shift invariant sets, but this is not necessary.)

PROOF. $E(y_t|x_{t-J},\ldots,x_{t+M}) = E(y_t|x) = E[E(y_t|\sigma(G \cup \sigma(x)))|x]$

$$= E(c|x) + E\Big(\sum_{j=-\infty}^{\infty} b_j x_{t-j}|x\Big)$$

$$= E(c|x) + \sum_{j=-\infty}^{\infty} E(b_j|x)x_{t-j} \quad \text{a.e.,} \quad t \in T,$$

since

$$\Big|\sum_{j=-L}^{K} b_j x_{t-j}\Big| \le \sum_{j=-L}^{K} |b_j x_{t-j}| \le \sum_{j=-\infty}^{\infty} |b_j x_{t-j}|$$

and $\sum_{j=-\infty}^{\infty} |b_j x_{t-j}|$ is integrable. (See Billingsley (1979), theorem 34.2(v), p. 396.)

Let $T_1 = \bigcap_{t=-\infty}^{0} \sigma(x_t, x_{t-1}, \ldots)$ be the left tail σ-field and $T_2 = \bigcap_{t=0}^{\infty} \sigma(x_t, x_{t+1}, \ldots)$ be the right tail σ-field. Since $\sigma(x)$ is an invariant σ-field, lemma 2 implies that $E(c|x)$ and $E(b_j|x)$ are invariant random variables. Hence they are measurable T_1 and measurable T_2. (See Rozanov, lemma 6.1, p. 162; more precisely, there exist versions of $E(c|x)$ and $E(b_j|x)$ that are measurable T_1 and there exist versions of $E(c|x)$ and $E(b_j|x)$ that are measurable T_2.)

Define

$$z_{t-k,k} = [E(y_t|x_{t-J},\ldots,x_{t+M}) - E(c|x) - \sum_{j\neq k} E(b_j|x)x_{t-j}]/E(b_k|x)$$

for $t \in T$ and $\omega \in A_k$, where

$$A_k = \{\omega: \ E(b_k|x)_\omega \neq 0\};$$

set $z_{t-k,k} = 0$ for $\omega \notin A_k$. Then $\{z_{t-k,k}, t \in T\}$ is a stationary stochastic process since $E(c|x)$, $E(b_j|x)$, $j \in T$, are invariant random variables and A_k is an invariant set.

If $k \notin [-M,J]$, then $E(y_t|x_{t-J},\ldots,x_{t+M})$ is measurable $\sigma(\ldots,x_{t-k-1},x_{t-k+1},\ldots)$. Since $E(c|x)$ and $E(b_j|x)$ are measurable $\mathcal{T}_1$, they are measurable $(\ldots,x_{t-k-2},x_{t-k-1})$; since A_k is an invariant set, its indicator function is measurable $\mathcal{T}_1$ and so measurable $(\ldots,x_{t-k-2},x_{t-k-1})$. Hence $z_{t-k,k}$ is measurable $\sigma(\ldots,x_{t-k-1},x_{t-k+1},\ldots)$, $t \in T$.

Since $x_{t-k} = z_{t-k,k}$ for all $t \in T$ for $\omega \in A_k$, condition (C.1) implies that $\mathrm{Var}(x_t) = 0$ if $P(A_k) > 0$. We conclude that $E(b_k|x) = 0$ a.e. for $k \notin [-M,J]$.

So we have

$$E(c|x) + \sum_{j=-M}^{J} E(b_j|x)x_{t-j} = E(y_t|x_{t-J},\ldots,x_{t+M}) \quad \text{a.e.}$$

for $t \in T$. If we consider this equation for $t = J,J+1,\ldots,2J+M+1$, we can say

$$\underset{\sim}{Q}\,\underset{\sim}{d} = \underset{\sim}{g} \quad \text{a.e.}$$

where $\underset{\sim}{d}' = [E(c|x), E(b_{-M}|x), \ldots, E(b_J|x)]$, $\underset{\sim}{g}' = [g_{2J+M+1}, \ldots, g_J]$ where $g_s = E(y_s|x_{s-J}, \ldots, x_{s+M})$, and $\underset{\sim}{Q}$ is a $(J+M+2)\times(J+M+2)$ matrix:

$$\underset{\sim}{Q} = \begin{bmatrix} 1 & x_{2J+2M+1} & x_{2J+2M} & \cdots & x_{J+M+1} \\ 1 & x_{2J+2M} & x_{2J+2M-1} & \cdots & x_{J+M} \\ \vdots & \vdots & \vdots & & \vdots \\ 1 & x_{J+M} & x_{J+M-1} & \cdots & x_0 \end{bmatrix}.$$

We show below in lemma 3 that $P(D) > 0$ where $D = \{\omega: \underset{\sim}{Q}$ is non-singular$\}$.

Let $\underset{\sim}{h} = \underset{\sim}{Q}^{-1}\underset{\sim}{g}$ for $\omega \in D$; $\underset{\sim}{h} = \underset{\sim}{0}$ otherwise. Then $\underset{\sim}{h}' = (h_1,\ldots,h_{J+M+2}) = \underset{\sim}{d}'$ for $\omega \in D$. We shall show that $E(c|x) = E(c)$ a.e.

If h_1 is not constant a.e. for $\omega \in D$, then we can choose a constant r such that

$$A_1 = \{\omega: \ \omega \in D \text{ and } h_1(\omega) \leq r\}, \quad P(A_1) > 0$$

$$A_1' = D - A_1, \quad P(A_1') > 0 .$$

Then set

$$A_2 = \{\omega: \ E(c|x)_\omega \leq r\}, \quad A_2' = \Omega - A_2 .$$

Since Q, g, and the indicator function for D are measurable $\sigma(x_0,\ldots,x_{2J+2M+1})$, it follows that A_1 and A_1' are in $\sigma(x_0,\ldots,x_{2J+2M+1})$. $E(c|x)$ is measurable T_1 and so A_2 and A_2' are in $\sigma(\ldots,x_{-2},x_{-1})$. $E(c|x) = h_1$ for $\omega \in D$ implies that

$$P(A_1 \cap A_2) = P(A_1), \quad P(A_1' \cap A_2') = P(A_1') .$$

Since $P(A_1) > 0$, $P(A_1') > 0$, condition (C.2) implies that $P(A_2) = 1$ and $P(A_2') = 1$, a contradiction. Hence h_1 is equal to some constant r a.e. for $\omega \in D$.

Now set $A_1 = D$, $A_2 = \{\omega: \ E(c|x)_\omega = r\}$. Then $P(A_1 \cap A_2) = P(A_1)$, and so (C.2) implies $P(A_2) = 1$. Hence $E(c|x) = E(c)$ a.e.

An identical argument establishes that $E(b_j|x) = E(b_j)$ a.e., $j \in [-M,J]$. So we have shown that

$$E(y_t|x) = E(c) + \sum_{j=-M}^{J} E(b_j)x_{t-j} \quad \text{a.e.}$$

Q.E.D.

LEMMA 3. Let

$$\underset{\sim}{Q}_s = \begin{bmatrix} 1 & x_{2s+1} & x_{2s} & \cdots & x_{s+1} \\ 1 & x_{2s} & x_{2s-1} & \cdots & x_s \\ \vdots & \vdots & \vdots & & \vdots \\ 1 & x_s & x_{s-1} & \cdots & x_0 \end{bmatrix}$$

where s is a non-negative integer. Then condition (C) implies that

$$P(\underset{\sim}{Q}_s \text{ non-singular}) > 0 \quad \text{for all } s.$$

PROOF. We shall use induction on s. If $s = 0$, then $\underset{\sim}{Q}_s$ singular implies $x_1 = x_0$. If $P(x_1 = x_0) = 1$, then (C.1) or (C.2) implies $\mathrm{Var}(x_t) = 0$. So $P(\underset{\sim}{Q}_0 \text{ non-singular}) > 0$. Now assume that $P(\underset{\sim}{Q}_{s-1} \text{ non-singular}) > 0$ for some $s \geq 1$ and show that $P(\underset{\sim}{Q}_s \text{ non-singular}) > 0$.

If $\underset{\sim}{Q}_s$ is singular a.e., then expanding $\det(\underset{\sim}{Q}_s)$ in terms of the cofactors of the first row gives

$$x_{2s+1} \det(\underset{\sim}{Q}_{s-1}) = g(x_0,\ldots,x_{2s}) \quad \text{a.e.}$$

where g is a measurable function. By the induction hypothesis, $P(D_{s-1}) > 0$ where $D_{s-1} = \{\omega: \det(\underset{\sim}{Q}_{s-1}) \neq 0\}$. Let

$$h = g(x_0,\ldots,x_{2s})/\det(\underset{\sim}{Q}_{s-1})$$

for $\omega \in D_{s-1}$; $h = 0$ for $\omega \notin D_{s-1}$. Then $x_{2s+1} = h$ for $\omega \in D_{s-1}$. If h is not constant a.e. for $\omega \in D_{s-1}$, we can choose a constant r such that

$$A_1 = \{\omega: \ \omega \in D_{s-1} \ \text{ and } \ h(\omega) \leq r\}, \quad P(A_1) > 0$$

$$A_1' = D_{s-1} - A_1, \quad P(A_1') > 0.$$

Then set

$$A_2 = \{\omega: \ x_{2s+1}(\omega) \leq r\}, \quad A_2' = \Omega - A_2.$$

Since A_1 and A_1' are in $\sigma(x_0,\ldots,x_{2s})$ and

$$P(A_1 \cap A_2) = P(A_1), \quad P(A_1' \cap A_2') = P(A_1'),$$

condition (C.2) implies that $P(A_2) = 1$ and $P(A_2') = 1$, a contradiction. So if Q_s is singular a.e., there must be a constant r such that $h(\omega) = r$ a.e. for $\omega \in D_{s-1}$.

Now set $A_1 = D_{s-1}$, $A_2 = \{\omega: \ x_{2s+1}(\omega) = r\}$. Then $P(A_1 \cap A_2) = P(A_1)$, and so (C.2) implies that $P(A_2) = 1$, which contradicts $\mathrm{Var}(x_t) \neq 0$. We conclude that $P(Q_s \text{ non-singular}) > 0$.

Q.E.D.

APPENDIX B

Let $\underset{\sim}{z}_i' = (\underset{\sim}{x}_i', \underset{\sim}{y}_i')$, $i = 1, \ldots, N$, where $\underset{\sim}{x}_i$ is $K\times 1$, $\underset{\sim}{y}_i$ is $M\times 1$. We shall assume that the $\underset{\sim}{z}_i$ are independent and identically distributed (i.i.d). We shall also assume that $E(x_{ik}^4) < \infty$ $(k = 1, \ldots, K)$, $E(y_{im}^4) < \infty$ $(m = 1, \ldots, M)$, and $V(\underset{\sim}{x}_i)$ non-singular. Let $\underset{\sim}{w}_i' = (\underset{\sim}{w}_{i1}', \underset{\sim}{w}_{i2}') = (\underset{\sim}{z}_i', \underset{\sim}{z}_i' \otimes \underset{\sim}{x}_i')$. Then the $\underset{\sim}{w}_i$ are i.i.d. with $V(\underset{\sim}{w}_i) < \infty$. Hence the central limit theorem for i.i.d. random variables implies that

$$\sqrt{N}\,(\bar{\underset{\sim}{w}} - E(\underset{\sim}{w}_i)) \xrightarrow{D} N(\underset{\sim}{0}, V(\underset{\sim}{w}_i)) \text{ as } N \to \infty ,$$

where $\bar{\underset{\sim}{w}} = \sum_{i=1}^{N} \underset{\sim}{w}_i/N$.

Consider the estimator

$$\hat{\underset{\sim}{\Pi}}' = \left[\sum_{i=1}^{N} (\underset{\sim}{x}_i - \bar{\underset{\sim}{x}})(\underset{\sim}{x}_i - \bar{\underset{\sim}{x}})'\right]^{-1} \sum_{i=1}^{N} (\underset{\sim}{x}_i - \bar{\underset{\sim}{x}})(\underset{\sim}{y}_i - \bar{\underset{\sim}{y}})$$

where $\bar{\underset{\sim}{x}} = \sum_{i=1}^{N} \underset{\sim}{x}_i/N$, $\bar{\underset{\sim}{y}} = \sum_{i=1}^{N} \underset{\sim}{y}_i/N$. $\hat{\underset{\sim}{\Pi}}$ is a function of $\bar{\underset{\sim}{w}}$:

$$\text{vec } \hat{\underset{\sim}{\Pi}}' = \underset{\sim}{g}(\bar{\underset{\sim}{w}}),$$

where vec $\hat{\Pi}'$ is the $MK\times 1$ vector formed by stacking the successive columns of $\hat{\underset{\sim}{\Pi}}'$. Evaluating $\underset{\sim}{g}$ at $E(\underset{\sim}{w}_i)$ gives

$$\text{vec } \underset{\sim}{\Pi}' = \underset{\sim}{g}(E(\underset{\sim}{w}_i)),$$

where

$$\underset{\sim}{\Pi}' = V^{-1}(\underset{\sim}{x}_i)\ \mathrm{Cov}(\underset{\sim}{x}_i, \underset{\sim}{y}_i').$$

Since $\underset{\sim}{g}$ is differentiable at $E(\underset{\sim}{w}_i)$, the δ-method gives

$$\sqrt{N}\ (\mathrm{vec}\ \hat{\underset{\sim}{\Pi}}' - \mathrm{vec}\ \underset{\sim}{\Pi}') \xrightarrow{D} N(\underset{\sim}{0}, \underset{\sim}{\Omega}) \text{ as } N \to \infty,$$

where

$$\underset{\sim}{\Omega} = \frac{\partial \underset{\sim}{g}}{\partial \underset{\sim}{r}'}\ V(\underset{\sim}{w}_i)\ \frac{\partial \underset{\sim}{g}'}{\partial \underset{\sim}{r}}$$

and the partial derivatives are evaluated at $\underset{\sim}{r} = E(\underset{\sim}{w}_i)$. (See, for example, Billingsley [1979, example 29.1, p. 340] or Rao [1973, p. 388].) We shall evaluate the partial derivatives and simplify to show that

$$\underset{\sim}{\Omega} = E[(\underset{\sim}{y}_i^* - \underset{\sim}{\Pi}\ \underset{\sim}{x}_i^*)(\underset{\sim}{y}_i^* - \underset{\sim}{\Pi}\ \underset{\sim}{x}_i^*)' \otimes V^{-1}(\underset{\sim}{x}_i)(\underset{\sim}{x}_i^*\ \underset{\sim}{x}_i'^*)\ V^{-1}(\underset{\sim}{x}_i)],$$

where $\underset{\sim}{x}_i^* = \underset{\sim}{x}_i - E(\underset{\sim}{x}_i)$, $\quad \underset{\sim}{y}_i^* = \underset{\sim}{y}_i - E(\underset{\sim}{y}_i^*)$.

Let $\underset{\sim}{z}_i'^* = (\underset{\sim}{x}_i'^*, \underset{\sim}{y}_i'^*)$, $\underset{\sim}{w}_i'^* = (\underset{\sim}{w}_{i1}'^*, \underset{\sim}{w}_{i2}'^*) = (\underset{\sim}{z}_i'^*, \underset{\sim}{z}_i'^* \otimes \underset{\sim}{x}_i'^*)$, and set $\bar{\underset{\sim}{w}}^* = \sum_{i=1}^{N} \underset{\sim}{w}_i^*/N$. Now note that $\underset{\sim}{g}(\bar{\underset{\sim}{w}}^*) = \underset{\sim}{g}(\bar{\underset{\sim}{w}})$ and $\underset{\sim}{g}(E(\underset{\sim}{w}_i^*)) = \underset{\sim}{g}(E(\underset{\sim}{w}_i))$ since $\underset{\sim}{z}_i^* - \bar{\underset{\sim}{z}}^* = \underset{\sim}{z}_i - \bar{\underset{\sim}{z}}$ and $V(\underset{\sim}{z}_i^*) = V(\underset{\sim}{z}_i)$. Since $\underset{\sim}{w}_i^*$ is i.i.d., we have

$$\underset{\sim}{\Omega} = \frac{\partial \underset{\sim}{g}}{\partial \underset{\sim}{r}'}\ V(\underset{\sim}{w}_i^*)\ \frac{\partial \underset{\sim}{g}'}{\partial \underset{\sim}{r}},$$

where the partial derivatives are evaluated at $E(\underset{\sim}{w}_i^*)$.

We can write

$$\underset{\sim}{g}(\underset{\sim}{r}) = \underset{\sim}{f}[\underset{\sim}{h}(\underset{\sim}{r})], \qquad \underset{\sim}{h}(\underset{\sim}{r}) = \underset{\sim}{r}_2 - \underset{\sim}{r}_1 \otimes \underset{\sim}{r}_1,$$

where the $\underset{\sim}{r}' = (\underset{\sim}{r}_1', \underset{\sim}{r}_2')$ partition conforms to the $\underset{\sim}{w}' = (\underset{\sim}{w}_1', \underset{\sim}{w}_2')$ partition

and $\underset{\sim}{r}_{11}$ is the sub-vector of $\underset{\sim}{r}_1$ containing the first K elements. When $\underset{\sim}{r} = E(\underset{\sim}{w}^*_{i1})$, we have

$$\frac{\partial \underset{\sim}{h}}{\partial \underset{\sim}{r}_1'} = \underset{\sim}{0}, \qquad \frac{\partial \underset{\sim}{h}}{\partial \underset{\sim}{r}_2'} = \underset{\sim}{I}_{K(K+M)} \quad ;$$

hence we can simplify to

$$\underset{\sim}{\Omega} = \frac{\partial \underset{\sim}{f}}{\partial \underset{\sim}{h}'} \; V(\underset{\sim}{w}^*_{i2}) \; \frac{\partial \underset{\sim}{f}'}{\partial \underset{\sim}{h}} \quad ,$$

where the partial derivatives are evaluated at $\underset{\sim}{h} = E(\underset{\sim}{w}^*_{i2})$.

The function $\underset{\sim}{f}$ can be implicitly defined by

$$\underset{\sim}{S} \; \underset{\sim}{f}_m = \underset{\sim}{d}_m \qquad (m = 1, \ldots, M),$$

where $\text{vec}(\underset{\sim}{s}_1, \ldots, \underset{\sim}{s}_K, \underset{\sim}{d}_1, \ldots, \underset{\sim}{d}_M) = \underset{\sim}{h}$, $\underset{\sim}{S} = (\underset{\sim}{s}_1, \ldots, \underset{\sim}{s}_K)$, and $\text{vec}\,(\underset{\sim}{f}_1, \ldots, \underset{\sim}{f}_M) = \underset{\sim}{f}$. So we have

$$\underset{\sim}{S} \; \frac{\partial \underset{\sim}{f}_m}{\partial \underset{\sim}{s}_k'} + f_{mk} \; \underset{\sim}{I}_K = \underset{\sim}{0} \qquad (k = 1, \ldots, K; \; m = 1, \ldots, M),$$

$$\underset{\sim}{S} \; \frac{\partial \underset{\sim}{f}_m}{\partial \underset{\sim}{d}_n'} = \begin{cases} \underset{\sim}{0} & \text{if } n \neq m \\ \underset{\sim}{I}_K & \text{if } n = m \end{cases} \qquad (n, m = 1, \ldots, M).$$

If $\underset{\sim}{h} = E(\underset{\sim}{w}^*_{i2})$, then $\underset{\sim}{S} = V(\underset{\sim}{x}_i)$, $\underset{\sim}{d}_m = \text{Cov}(\underset{\sim}{x}_i, y_{im})$, and $\underset{\sim}{f} = \text{vec}\,\underset{\sim}{\Pi}'$. So the partial derivatives evaluated at $\underset{\sim}{h} = E(\underset{\sim}{w}^*_{i2})$ are

$$\frac{\partial \underset{\sim}{f}}{\partial \underset{\sim}{h}'} = (-\underset{\sim}{\Pi}, \; \underset{\sim}{I}_M) \; \circledast \; V^{-1}(\underset{\sim}{x}_i).$$

Since

$$V(\underset{\sim}{w}^*_{i2}) = V(\underset{\sim}{z}^*_i \circledast \underset{\sim}{x}^*_i)$$

$$= E(\underset{\sim}{z}^*_i \underset{\sim}{z}_i'^* \circledast \underset{\sim}{x}^*_i \underset{\sim}{x}_i'^*) - E(\underset{\sim}{z}^*_i \circledast \underset{\sim}{x}^*_i)\, E(\underset{\sim}{z}_i'^* \circledast \underset{\sim}{x}_i'^*),$$

there are two terms to evaluate:

1. $$\frac{\partial \underset{\sim}{f}}{\partial \underset{\sim}{h}'} [E(\underset{\sim}{z}^*_i \underset{\sim}{z}_i'^* \circledast \underset{\sim}{x}^*_i \underset{\sim}{x}_i'^*)] \frac{\partial \underset{\sim}{f}'}{\partial \underset{\sim}{h}}$$

$$= E\left\{[(-\underset{\sim}{\Pi}, \underset{\sim}{I}_M) \circledast V^{-1}(\underset{\sim}{x}_i)](\underset{\sim}{z}^*_i \underset{\sim}{z}_i'^* \circledast \underset{\sim}{x}^*_i \underset{\sim}{x}_i'^*)[(-\underset{\sim}{\Pi}, \underset{\sim}{I}_M)' \circledast V^{-1}(\underset{\sim}{x}_i)]\right\}$$

$$= E[(\underset{\sim}{y}^*_i - \underset{\sim}{\Pi}\, \underset{\sim}{x}^*_i)(\underset{\sim}{y}^*_i - \underset{\sim}{\Pi}\, \underset{\sim}{x}^*_i)' \circledast V^{-1}(\underset{\sim}{x}_i)(\underset{\sim}{x}^*_i \underset{\sim}{x}_i'^*)\, V^{-1}(\underset{\sim}{x}_i)].$$

2. $$\frac{\partial \underset{\sim}{f}}{\partial \underset{\sim}{h}'} [E(\underset{\sim}{z}^*_i \circledast \underset{\sim}{x}^*_i)\, E(\underset{\sim}{z}_i'^* \circledast \underset{\sim}{x}_i'^*)] \frac{\partial \underset{\sim}{f}'}{\partial \underset{\sim}{h}} = \underset{\sim}{0}$$

since

$$E\left\{[(-\underset{\sim}{\Pi}, \underset{\sim}{I}_M) \circledast V^{-1}(\underset{\sim}{x}_i)]\ (\underset{\sim}{z}^*_i \circledast \underset{\sim}{x}^*_i)\right\}$$

$$= E[(\underset{\sim}{y}^*_i - \underset{\sim}{\Pi}\, \underset{\sim}{x}^*) \circledast V^{-1}(\underset{\sim}{x}_i)\, \underset{\sim}{x}^*_i]$$

and

$$E(y^*_{im} - \underset{\sim}{\pi}_m' \underset{\sim}{x}^*_i)\, \underset{\sim}{x}^*_i = \mathrm{Cov}(y_{im}, \underset{\sim}{x}_i) - V(\underset{\sim}{x}_i)\, \underset{\sim}{\pi}_m = \underset{\sim}{0} \qquad (m = 1, \ldots, M).$$

We conclude that

$$\underset{\sim}{\Omega} = E[(\underset{\sim}{y}^*_i - \underset{\sim}{\Pi}\, \underset{\sim}{x}^*_i)(\underset{\sim}{y}^*_i - \underset{\sim}{\Pi}\, \underset{\sim}{x}^*_i)' \circledast V^{-1}(\underset{\sim}{x}_i)(\underset{\sim}{x}^*_i \underset{\sim}{x}_i'^*)\, V^{-1}(\underset{\sim}{x}_i)].$$

Next we shall apply this procedure to evaluate the limiting distribution of the instrumental variable estimator. Let $\underset{\sim}{q}_i$ be J×1 and assume that $(\underset{\sim}{x}_i, y_i, \underset{\sim}{q}_i)$, i=1, ..., N, is i.i.d. according to some multivariate distribution with finite forth moments. Let $\underset{\sim}{x}_{im}$ be a K_m×1 subvector of $\underset{\sim}{x}_i$; let $\underset{\sim}{q}_{im}$ be a K_m×1 subvector of $\underset{\sim}{q}_i$, and consider the instrumental variable estimators:

$$\hat{\underset{\sim}{\pi}}_m = \left[\sum_{i=1}^{N} (\underset{\sim}{q}_{im} - \bar{\underset{\sim}{q}}_m)(\underset{\sim}{x}_{im} - \bar{\underset{\sim}{x}}_m)'\right]^{-1} \sum_{i=1}^{N} (\underset{\sim}{q}_{im} - \bar{\underset{\sim}{q}}_m)(y_{im} - \bar{y}_m) \quad (m=1, \ldots, M)$$

We shall assume that $\underset{\sim}{\Phi}_m = \mathrm{Cov}(\underset{\sim}{q}_{im}, \underset{\sim}{x}'_{im})$ is non-singular, m=1, ..., M. Then $\hat{\underset{\sim}{\pi}}_m$ almost surely converges to $\underset{\sim}{\pi}_m = [\mathrm{Cov}(\underset{\sim}{q}_{im}, \underset{\sim}{x}'_{im})]^{-1} \mathrm{Cov}(\underset{\sim}{q}_{im}, y_{im})$ as $N \to \infty$. Let $\hat{\underset{\sim}{\pi}}' = (\hat{\underset{\sim}{\pi}}'_1, \ldots, \hat{\underset{\sim}{\pi}}'_M)$ and $\underset{\sim}{\pi}' = (\underset{\sim}{\pi}'_1, \ldots, \underset{\sim}{\pi}'_M)$. Since $\hat{\underset{\sim}{\pi}}$ is a differentiable function of sample moments, we have

$$\sqrt{N}\,(\hat{\underset{\sim}{\pi}} - \underset{\sim}{\pi}) \xrightarrow{D} N(\underset{\sim}{0}, \underset{\sim}{\Omega}) \quad \text{as } N \to \infty .$$

In order to simplify $\underset{\sim}{\Omega}$, we shall repeat the procedure used for the multivariate regression case. Now $\underset{\sim}{f}' = (\underset{\sim}{f}'_1, \ldots, \underset{\sim}{f}'_M)$ is implicitly defined by

$$\underset{\sim}{S}_m \underset{\sim}{f}_m = \underset{\sim}{d}_m \qquad (m=1, \ldots, M)$$

where $\underset{\sim}{S}_m = (\underset{\sim}{s}_{m1}, \ldots, \underset{\sim}{s}_{mK_m})$ is $K_m \times K_m$.

$$\underset{\sim}{S}_m \frac{\partial \underset{\sim}{f}_m}{\partial \underset{\sim}{s}'_{nk}} = \begin{cases} \underset{\sim}{0} & \text{if } n \neq m \\ -f_{mk}\, \underset{\sim}{I}_{K_m} & \text{if } n = m \end{cases}$$

$(k = 1, \ldots K_n;\ n, m = 1, \ldots, M)$.

$$\underset{\sim}{S}_m \quad \frac{\partial \underset{\sim}{f}_m}{\partial \underset{\sim}{d}_n'} = \begin{cases} \underset{\sim}{0} & \text{if } n \neq m \\ \underset{\sim}{I}_{K_m} & \text{if } m = n \end{cases} \qquad (n,m=1, \ldots, M).$$

We have $\underset{\sim}{S}_m = \mathrm{Cov}(\underset{\sim}{q}_{im}, \underset{\sim}{x}_{im}')$, $\underset{\sim}{d}_m = \mathrm{Cov}(\underset{\sim}{q}_{im}, y_m)$, and $\underset{\sim}{f}_m = \underset{\sim}{\pi}_m$ when these functions are evaluated at the population means. Let $\underset{\sim}{h}_m = \mathrm{vec}(\underset{\sim}{s}_{ml}, \ldots, \underset{\sim}{s}_{mK_m}, \underset{\sim}{d}_m)$. Then evaluating the partial derivatives at the population means gives

$$\frac{\partial \underset{\sim}{f}_m}{\partial \underset{\sim}{h}_n'} = \begin{cases} \underset{\sim}{0} & \text{if } n \neq m \\ (-\underset{\sim}{\pi}_m', \ 1) \otimes \underset{\sim}{\Phi}_m^{-1} & \text{if } n = m \end{cases} \qquad (n, m=1, \ldots, M).$$

The m, n submatrix of $\underset{\sim}{\Omega}$ gives the asymptotic covariance between $\hat{\underset{\sim}{\pi}}_m$ and $\hat{\underset{\sim}{\pi}}_n$:

$$\underset{\sim}{\Omega}_{mn} = \frac{\partial \underset{\sim}{f}_m}{\partial \underset{\sim}{h}_m'} \left[E(\underset{\sim}{z}_{im}^* \underset{\sim}{z}_{in}^{*\prime} \otimes \underset{\sim}{q}_{im}^* \underset{\sim}{q}_{in}^{*\prime}) - E(\underset{\sim}{z}_{im}^* \otimes \underset{\sim}{q}_{im}^*) E(\underset{\sim}{z}_{in}^{\prime *} \otimes \underset{\sim}{q}_{in}^{\prime *}) \right] \frac{\partial \underset{\sim}{f}_n'}{\partial \underset{\sim}{h}_n},$$

where $\underset{\sim}{z}_{im}^{\prime *} = (\underset{\sim}{x}_{im}^{\prime *}, y_{im}^*)$, $\underset{\sim}{x}_{im}^* = \underset{\sim}{x}_{im} - E(\underset{\sim}{x}_{im})$, $y_{im}^* = y_{im} - E(y_{im})$, and $\underset{\sim}{q}_{im}^* = \underset{\sim}{q}_{im} - E(\underset{\sim}{q}_{im})$. As before there are two terms to evaluate and the second one is zero since $E(y_{im}^* - \underset{\sim}{\pi}_m' \underset{\sim}{x}_{im}^*) \underset{\sim}{q}_{im}^* = 0$. The first term simplifies to

$$\underset{\sim}{\Omega}_{mn} = E[(y_{im}^* - \underset{\sim}{\pi}_m' \underset{\sim}{x}_{im}^*)(y_{in}^* - \underset{\sim}{\pi}_n' \underset{\sim}{x}_{in}^*) \underset{\sim}{\Phi}_m^{-1} \underset{\sim}{q}_{im}^* \underset{\sim}{q}_{in}^{\prime *} \underset{\sim}{\Phi}_n^{\prime -1}],$$

where $\underset{\sim}{\Phi}_m = \mathrm{Cov}(\underset{\sim}{q}_{im}, \underset{\sim}{x}_{im}')$, $\quad m,n=1, \ldots, M.$

FOOTNOTES

[1]See Mundlak (1961, 1978).

[2]See Sims (1972).

[3]MaCurdy (1979) considers the conventional maximum likelihood estimator in a homoskedastic multivariate regression model. He shows that its asymptotic distribution can be obtained without normality assumptions by treating it as a minimum distance estimator. In our model, however, that maximum likelihood estimator is generally inefficient since it is using the wrong norm.

[4]See Mundlak (1961), Chamberlain (1980).

REFERENCES

Billingsley, P., 1965, Ergodic theory and information (Wiley, New York).

_____ , 1979, Probability and measure (Wiley, New York).

Chamberlain, G., 1980, Analysis of covariance with qualitative data, Review of Economic Studies, 47, 225-238.

Cramèr, H., 1946, Mathematical methods of statistics (Princeton University Press, Princeton).

Ghez, G. R. and G. S. Becker, 1975, The allocation of time and goods over the life cycle (Columbia University Press, New York).

Goldberger, A. S., 1974, Asymptotics of the sample regression slope, unpublished lecture notes, No. 12.

Heckman, J. J. and T. E. MaCurdy, 1980, A life cycle model of female labor supply, Review of Economic Studies, 47, 47-74.

Kendall, M. G. and A. Stuart, 1961, The advanced theory of statistics, Vol. 2 (Griffin, London).

MaCurdy, T. E., 1979, Multiple time series models applied to panel data: specification of a dynamic model of labor supply, unpublished manuscript.

_____, 1980, An empirical model of labor supply in a life cycle setting, National Bureau of Economic Research Working Paper No. 421.

Malinvaud, E., 1970, Statistical methods of econometrics (North-Holland, Amsterdam).

Mundlak, Y., 1961, Empirical production function free of management bias, Journal of Farm Economics, 43, 44-56.

_____, 1978, "On the pooling of time series and cross section data, Econometrica, 46, 69-85.

Rao, C. R., 1973, Linear statistical inference and its applications (Wiley, New York).

Rozanov, Y. A., 1967, Stationary random processes (Holden-Day, San Francisco).

Sims, C. A., 1972, Money, income, and causality, American Economic Review 62, 540-552.

White, H., 1980, Using least squares to approximate unknown regression functions, International Economic Review 21, 149-170.