Static Linear Panel Data

Bryan S. Graham, UC - Berkeley & NBER

March 8, 2021

Static linear panel data models appear frequently in empirical economics research. They are especially common in the context of program/policy evaluation, where they sometimes appear under the guise of the method of "difference-in-differences". This note presents some basic results, largely drawn from Arellano & Bover (1995). This paper synthesized and extended a variety of results developed in the decade after the seminal *Handbook of Econometrics* chapter by Chamberlain (1984). The textbook presentations in Arellano (2003) and Wooldridge (2010) are also recommended.

Strict exogeneity and panel data

We begin with the model of agricultural production featured in Chamberlain (1984). Let O_t be the agricultural output of a randomly sampled farm in period t; L_t the level of a variable input (e.g., labor or fertilizer), A a *time invariant* production factor, observed by the farmer, but not the econometrician (e.g., land quality, managerial talent etc.). Finally U_t is a stochastic input which is outside the farmers control (e.g., weather). Production takes the Cobb-Douglas form.

$$O_t = L_t^{\beta} \exp\left(A + U_t\right).$$

The farmer choose the period t variable input level, L_t, to maximize profits

$$L_t = \arg\max_{l_t} \mathbb{E}\left[P_t l_t^{\beta} \exp\left(A + U_t\right) - W_t l_t \middle| \mathcal{I}_t \right],$$

with P_t and W_t denoting the output and input price. The beginning-of-period t information set $\mathcal{I}_t$ includes A, $W_1^t = (W_1, \ldots, W_t)'$, $P_1^t = (P_1, \ldots, P_t)'$ and, $U_1^{t-1} = (U_1, \ldots, U_{t-1})'$.

The profit-maximizing first order condition is

$$P_t L_t^{\beta-1} \exp(A) \mathbb{E}\left[\exp(U_t) | \mathcal{I}_t\right] - W_t = 0,$$

which after taking logs and re-arranging yields

$$X_t = \frac{\ln(\beta)}{1-\beta} + \frac{1}{1-\beta} \ln\left(\frac{P_t}{W_t}\right) + \frac{1}{1-\beta} A + \frac{1}{1-\beta} \ln\left(\mathbb{E}\left[\exp(U_t) | \mathcal{I}_t\right]\right)$$

for $X_t = \ln L_t$. The log-output equation is

$$Y_t = X_t'\beta + A + U_t,$$

with $Y_t = \ln O_t$.

We assume that $Y_{i1}, \ldots, Y_{iT}, X_{i1}, \ldots, X_{iT}, A_i \overset{iid}{\sim} F$. Note that we make no assumptions about the joint distribution of $X_{i1}, \ldots, X_{iT}$ and A_i. Building models which allow for dependence across these two random variables is one of the main attractions of panel data.

For, simplicity assume that $F_t \overset{def}{\equiv} \ln\left(\mathbb{E}\left[\exp(U_t) | \mathcal{I}_t\right]\right)$ and $\ln\left(\frac{P_t}{W_t}\right)$vary independently of each other as well as of A *and* U_t. In this case we have

$$\mathbb{C}(X_t, Y_t) = \mathbb{V}(X_t)\beta + \frac{1}{1-\beta}\mathbb{V}(A)$$

and $\mathbb{V}(X_t) = \left[\frac{1}{1-\beta}\right]^2 \mathbb{V}\left(\ln\left(\frac{P_t}{W_t}\right)\right) + \left[\frac{1}{1-\beta}\right]^2 \mathbb{V}(A) + \left[\frac{1}{1-\beta}\right]^2 \mathbb{V}(F_t)$, such that the coefficient on X_t in the linear regression of Y_t onto X_t is

$$b = \beta + (1-\beta)\frac{\mathbb{V}(A)}{\mathbb{V}\left(\ln\left(\frac{P_t}{W_t}\right)\right) + \mathbb{V}(A) + \mathbb{V}(F_t)} > \beta.$$

Because unobserved land quality, A, and the variable input, L_t, are complements, the short regression coefficient is biased upwards for the output elasticity. If land quality varies substantially across farms, the bias can be large. If the main source of input variation across farms is instead variation in input costs and prices (i.e., in $\ln\left(\frac{P_t}{W_t}\right)$ which is assumed independent of A), then the bias can be low. This bias is called *correlated heterogeneity bias.*

Following Chamberlain (1984) decompose A into its linear projection onto the variable input choices in all periods and a projection error:

$$\begin{aligned} A &= \mathbb{E}^* \left[A | X_1, \ldots, X_T \right] + A^* \\ &= \alpha + X_1' \pi_1 + \cdots + X_T' \pi_T + A^* \end{aligned} \tag{1}$$

Note that A^* is uncorrelated with X_t for $t = 1, \ldots, T$ by construction. I wish to emphasize that (1) is nothing more than a decomposition; no substantive assumption is being made beyond time invariance of A. In addition to time invariance we will make a strict exogeneity assumption of

$$\mathbb{E} \left[U_t^* | X_1, \ldots, X_t \right] = 0. \tag{2}$$

Restriction (2) *is* substantive. In particular, as explained further below, it rules out feedback from current values of U_t to future values of X_{t+s}. To see how this assumption might be violated imagine the stochastic input U_t follows an autoregressive process of order 1:

$$U_t = \lambda_t + \rho U_{t-1} + V_t$$

with V_t Gaussian. In this case F_t will equal

$$\mathbb{E} \left[\exp \left(U_t \right) | \mathcal{I}_t \right] = \exp \left(\lambda_t + \rho U_{t-1} + \frac{\sigma^2}{2} \right),$$

yielding a variable input demand equation of

$$X_t = \frac{\ln (\beta)}{1 - \beta} + \lambda_t + \frac{\sigma^2}{2} + \frac{1}{1 - \beta} \ln \left(\frac{P_t}{W_t} \right) + \frac{1}{1 - \beta} A + \rho U_{t-1}.$$

Hence (2) will not hold: past weather shocks – since they help the farmer predict future weather – influence future input choices. Violations of strict exogeneity are often associated with issues of forecastability and information sets in empirical applications. Additional examples will be discussed below.

For now we will maintain restriction (2). Plugging (1) expression into the log-output equation yields

$$Y_t = \alpha + \eta_t + X_t \beta + X_1 \pi_1 + \cdots + X_T \pi_T + A^* + U_t^*,$$

where η_t is the period specific mean of U_t and $U_t^* = U_t - \eta_t$.

Now consider the pooled least squares regression fit of Y_t onto a vector of time dummies, X_t and $X_1, \ldots, X_T$. Under time invariance of A and strict exogeneity (2), the coefficient on

X_t in this regression fit will provide a consistent estimate of β. It might appear the rank condition for this OLS fit to be well-defined is violated, since X_t appears to enter twice as a regressor; this is not the case. Consider the structure of the underlying dataset used to compute the OLS fit. The first $2T$ rows of this dataset, corresponding to the information for the first two farms, is displayed below:

Y_t	D_1	D_2	$\cdots$	D_T	X_t	X_1	X_2	$\cdots$	X_T	i
Y_{11}	1	0	$\cdots$	0	X_{11}	X_{11}	X_{12}	$\cdots$	X_{1T}	1
Y_{12}	0	1	$\cdots$	0	X_{12}	X_{11}	X_{12}	$\cdots$	X_{1T}	1
$\vdots$	$\vdots$	$\vdots$	$\ddots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\ddots$	$\vdots$	$\vdots$
Y_{1T}	0	0	$\cdots$	1	X_{1T}	X_{11}	X_{12}	$\cdots$	X_{1T}	2
Y_{21}	1	0	$\cdots$	0	X_{21}	X_{21}	X_{22}	$\cdots$	X_{2T}	2
Y_{22}	0	1	$\cdots$	0	X_{22}	X_{21}	X_{22}	$\cdots$	X_{2T}	2
$\vdots$	$\vdots$	$\vdots$	$\ddots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\ddots$	$\vdots$	$\vdots$
Y_{2T}	0	0	$\cdots$	1	X_{2T}	X_{21}	X_{22}	$\cdots$	X_{2T}	2

As is apparent from inspection of the table X_t is not a linear combination of $X_1 \cdots X_T$ because it varies across rows corresponding to the same unit (the "within" variation). For accurate inference we will need to "cluster" our standard errors on i or "farms". Similarly either the constant regressor or one of the time dummies will need to be excluded from our fitted model.

One feature of this estimator that is potentially attractive is that it provides an estimate of $\pi = (\pi_1', \ldots, \pi_T')'$. This can be quite useful. For example, if X_{it} is a public health program availability indicator, and A_i measure of latent health in region i, then π provides information on the implicit/de facto targeting rule used for program placement. See Pitt et al. (1993) for an analysis along these lines using Indonesian data.

The above estimation procedure, while consistent for β under strict exogeneity, is generally inefficient. More efficient estimators can be constructed by using a different instrument matrix (a more formal discussion of the relevant semiparametric efficiency theory is given below). An approach, which may be viewed as a GMM formulation of the minimum distance procedure introduced by Chamberlain (1984), is as follows. Let $X_i = (X_{i1}', \ldots, X_{iT}')'$ and append the following T^2K columns to the dataset described above (here $K = \dim(X_{it})$).

$$\begin{array}{cccc}
Z_1 & Z_2 & \cdots & Z_T \\
- & - & - & - \\
X_1' & 0 & \cdots & 0 \\
0 & X_1' & & 0 \\
\vdots & & \ddots & \vdots \\
0 & 0 & & X_1' \\
X_2' & 0 & \cdots & 0 \\
0 & X_2' & & 0 \\
\vdots & & \ddots & \vdots \\
0 & 0 & & X_2'
\end{array}$$

(note $X_1 = (X_{11}', \ldots, X_{1T}')'$ in the above table, not the $t = 1$ regressor for a generic random draw). With one's dataset augmented as above, a simple linear instrumental variables regression of Y_{it} onto $D_{i1}, \ldots, D_{iT}$, X_{it}, and $X_{i1}, \ldots, X_{iT}$ using $Z_{i1}, \ldots, Z_{iT}$ as excluded instruments for X_{it}, and $X_{i1}, \ldots, X_{iT}$ will provide a consistent, and generally more efficient, estimate of β. Note a constant is not included in this fit as all T time dummies enter. Again, standard errors should be clustered on the i subscript.

This model has $T^2K - TK - K$ over-identifying restrictions. Due to concerns about the finite sample properties of Hansen's (1982) optimal two-step GMM setting in models with many overidentifying restrictions, it is common to just report one step estimates (e.g., those based on a vanilla linear instrumental variables command like `ivreg` in Stata). A Generalized Empirical Likelihood (GEL) estimator would be another alternative. The Sargan-Hansen test of overidentifying restrictions for this model is asymptotically equivalent to the omnibus specification test introduced by Chamberlain (1984).

Arellano & Bover (1995) formulation

Chamberlain (1984) inspired a great deal of work on panel data in the decade after its publication. One contribution of the paper by Arellano & Bover (1995) was to synthesize that work, making comparisons between different procedures more straightforward. Some of the results in this paper may now appear somewhat dated (e.g., those which specialize to homoscedasticity assumptions), but it is worth studying the paper in detail. It efficiently summarizes a large literature, presents new results and hints at possible areas for future research (even today). They begin with the following model

$$Y_t = X_t'\beta + W'\gamma + A + U_t,$$

with

$$\mathbb{E}\left[U_t \right| X_1, \ldots, X_T, W, A] = 0 \tag{3}$$

and

$$\mathbb{E}\left[A \right| X_{11}, \ldots, X_{1T}, W_1] = 0 \tag{4}$$

with $X_t = (X_{1t}', X_{2t}')'$ and $W = (W_1', W_2')'$. Relative to the model discussed above, we now include a vector of observed time-invariant regressors, W (with $J = \dim(W)$). We also now allow for the possibility that only some components of X_t and W covary with A (as might occur when some regressors are agent-chosen, while others are randomly assigned (or otherwise exogenously determined).

Let $\mathbf{X} = (X_1, \ldots, X_T)'$, $\mathbf{R} = (\mathbf{X}, \iota W')$ and $\delta = (\beta', \gamma')'$. We can write the T outcome equations for unit i compactly in matrix form as

$$\mathbf{Y}_i = \mathbf{R}_i\delta + \mathbf{V}_i$$

with $\mathbf{V}_i = \iota A_i + (U_{i1}, \ldots, U_{iT})'$. Let $\mathbf{K}$ be any $(T-1) \times T$ matrix of rank $T-1$ such that $\mathbf{K}\iota = 0$ and define

$$\mathbf{H} = \begin{bmatrix} \mathbf{K} \\ T^{-1}\iota' \end{bmatrix}.$$

Two common choices for $\mathbf{K}$ would be the first $T-1$ rows of the within-group operator

$$\mathbf{K} = \left[I_T - \frac{\iota_T \iota_T{}'}{T} \right]_{(1:T-1, 1:T)}$$

or the first difference operator

$$\mathbf{K} = \begin{pmatrix} -1 & 1 & & & 0 \\ & -1 & 1 & & \\ & & & \ddots & \\ 0 & & & -1 & 1 \end{pmatrix}.$$

The $\mathbf{H}$ matrix is useful for implementing within- and between-group transformations. Define:

$$\mathbf{V}_i^+ = \mathbf{H}\mathbf{V}_i = \begin{bmatrix} \mathbf{K}\mathbf{V}_i \\ \bar{V}_i \end{bmatrix}$$

with $\bar{V}_i = \frac{1}{T}\sum_{t=1}^{T} V_{it}$. Next let $R = (X_1', \ldots, X_T', W')'$ and define the following instrument matrix

$$\mathbf{Z}_i = \begin{bmatrix} R_i' & & & 0 \\ & \ddots & & \\ & & R_i' & \\ 0 & & & Q_i' \end{bmatrix} \tag{5}$$

where Q_i is a vector of functions of $X_{1i1}, \ldots, X_{1iT}$ and W_{1i}.

Putting things together, the two conditional moment restrictions (3) and (4) together yield an unconditional moment condition vector of

$$\mathbb{E}\left[\mathbf{Z}_i'\mathbf{H}\mathbf{V}_i\right] = 0.$$

The GMM estimator based on the above moment restriction is closely related to the one described earlier. As in that case it is advisable to avoid two-step GMM estimates; either use a one-step estimator with a non-optimal weight matrix, or a GEL estimator.

Semiparametric efficiency bound

Let $R_{1i} = (X_{1i1}', \ldots, X_{1iT}', W_{1i}')'$ and similarly define R_{2i}. Chamberlain (1992b) showed that bound for β based on the strict exogeneity assumption (3) coincides with the bound for β based on the following conditional moment restriction

$$\mathbb{E}\left[\left.\mathbf{K}\left(\mathbf{Y} - \mathbf{X}\beta\right)\right| R_1, R_2\right] = 0. \tag{6}$$

This is not entirely obvious, But it makes some sense if you think about it a bit. The strict exogeneity assumption implies that $\mathbf{U}$ is mean independent of R_1 and R_2. Restriction (6) asserts the same for linear combinations of $\mathbf{U}$; see Chamberlain (1992b) for details. We can write condition (4) as

$$\mathbb{E}\left[\left.\mathbf{Y} - \mathbf{R}\delta\right| R_1\right] = 0. \tag{7}$$

Rewriting we get

$$\mathbb{E}\left[\left.\mathbf{V}\right| R_1\right] = 0 \tag{8}$$

$$\mathbb{E}\left[\left.\mathbf{K}\mathbf{V}\right| R_1, R_2\right] = 0. \tag{9}$$

Chamberlain (1992a) showed how to calculate efficiency bounds for sequential moment restrictions like (8) and (9). The general idea is not unrelated to our earlier discussion of GMM with auxiliary information. The key "trick" is to appropriately orthogonalize (8) and (9) so that the information on δ in each of the two (orthogonalized) moments can simply be added up. Following Arellano & Bover (1995) define

$$\begin{aligned}\rho_{1i} &= \mathbf{Y}_i - \mathbf{R}_i\delta = \mathbf{V}_i \\ \rho_{2i} &= K\left(\mathbf{Y}_i - \mathbf{X}_i\beta\right) = K\mathbf{V}_i.\end{aligned}$$

We then compute – the conditional on R_i – linear projection of ρ_{1i} onto ρ_{2i}. The appropriate matrix of conditional linear predictor coefficients is

$$\begin{aligned}\Pi\left(R_i\right) &= \mathbb{E}\left[\rho_{1i}\rho_{2i}'\middle| R_i\right] \times \mathbb{E}\left[\rho_{2i}\rho_{2i}'\middle| R_i\right]^{-1} \\ &= \mathbb{E}\left[\mathbf{V}_i\mathbf{V}_i'\mathbf{K}'\middle| R_i\right] \times \mathbb{E}\left[\mathbf{K}\mathbf{V}_i\mathbf{V}_i'K'\middle| R_i\right]^{-1} \\ &= \Omega\left(R_i\right)\mathbf{K}' \times \left(\mathbf{K}\Omega\left(R_i\right)\mathbf{K}'\right)^{-1}\end{aligned}$$

with $\Omega\left(R_i\right) = \mathbb{E}\left[\mathbf{V}_i\mathbf{V}_i'\middle| R_i\right]$. We then compute

$$\rho_{1i}^* = \rho_{1i} - \Pi\left(R_i\right)\rho_{2i},$$

with $\mathbb{E}\left[\rho_{1i}^*\rho_{2i}\middle| R_i\right] = 0$ by the properties of projections. Since the transformed moments are conditionally uncorrelated the bound is just the sum of the two bounds associated with each set of moments separately. Arellano & Bover (1995) show that

$$\begin{aligned}\rho_{1i}^* &= \mathbf{V}_i - \Omega\left(R_i\right)K' \times \left(K\Omega\left(R_i\right)K'\right)^{-1}K\mathbf{V}_i \\ &= \left[I_T - \Omega\left(R_i\right)K' \times \left(K\Omega\left(R_i\right)K'\right)^{-1}K\right]\mathbf{V}_i \\ &= \left(\iota_T'\Omega\left(R_i\right)^{-1}\iota_T\right)^{-1}\iota_T'\Omega\left(R_i\right)^{-1}\mathbf{V}_i.\end{aligned}$$

Our orthogonalized system of conditional moment restrictions is therefore:

$$\begin{aligned}\mathbb{E}\left[K\mathbf{V}\middle| R_1, R_2\right] =& 0 \\ \mathbb{E}\left[\left(\iota_T'\Omega_i^{-1}\iota_T'\right)^{-1}\iota_T'\Omega_i^{-1}\mathbf{V}_i\middle| R_1\right] =& 0.\end{aligned}$$

with a semiparametric efficiency bound for δ_0 of

$$\begin{aligned}\mathcal{I}(\delta_0) =& \mathbb{E}\left[\mathbf{R}_i' K' \left(K\Omega_i K'\right)^{-1} K\mathbf{R}_i\right. \\ &\left.+\frac{\mathbb{E}\left[\mathbf{R}_i\Omega_i^{-1}\iota_T\left(\iota_T'\Omega_i^{-1}\iota_T\right)^{-1}\Big| R_{1i}\right]\mathbb{E}\left[\mathbf{R}_i\Omega_i^{-1}\iota_T\left(\iota_T'\Omega_i^{-1}\iota_T\right)^{-1}\Big| R_{1i}\right]'}{\mathbb{E}\left[\left(\iota_T'\Omega_i^{-1}\iota_T\right)^{-1}\Big| R_{1i}\right]}\right].\end{aligned}$$

None of the estimators discussed above will attain this bound.

Feedback and weak exogeneity

Strict exogeneity is a strong assumption is some contexts. Consider the case where $X_{it}$is a indicator for whether a certain policy is in place in region i during period t. We might worry about *feedback* from current values of the outcome to future values of the policy. For example a negative public health shock in region i might induce a government to improve health infrastructure in that region in subsequent $s > t$ periods. In such settings U_t will covary with X_s for $s > t$. However, we might still be willing to assume:

$$\mathbb{E}\left[U_t\right| X_1, \ldots, X_t, W, A] = 0 \tag{10}$$

$$\mathbb{E}\left[A\right| X_{11}, \ldots, X_{1T}, W_1] = 0 \tag{11}$$

The first of these restrictions is a weakening of strict exogeneity to something called weak exogeneity. This assumption rules out current values of the policy from covarying with future values of the time varying shock, but does allow for past values of this shock to covary with future values of the policy. This is often a reasonable assumption in policy evaluation settings (although most extant empirical work is based – whether implicitly or explicitly – on the stronger strict exogeneity assumption).

In order to construct a feasible estimator based upon (10) and (11) we will use Helmert's transformation

$$V_{it} = C_t\left[V_{it} - \frac{1}{T-t}\left(V_{it+1} + \cdots + V_{iT}\right)\right],\ t = 1, \ldots, T-1$$

where $C_t = (T-t)/(T-t+1)$. In words: to each of the first $(T-1)$ observations of V_t we subtract the mean of the remaining future observations in the sample; C_t equalizes the variances of these transformations. You can think of this as a specialization of the within-group transform specifically tailored for the peculiarities of estimation based upon sequential moment restrictions like (10). The appropriate form for the $\mathbf{K}$ is

$$\begin{aligned}\mathbf{K} =& \operatorname{diag}\left\{\frac{T-1}{T},\ldots,\frac{1}{2}\right\}^{1/2}\\ &\times\begin{bmatrix} 1 & -\frac{1}{T-1} & -\frac{1}{T-1} & \cdots & -\frac{1}{T-1} & -\frac{1}{T-1} & -\frac{1}{T-1}\\ 0 & 1 & -\frac{1}{T-2} & \cdots & -\frac{1}{T-2} & -\frac{1}{T-2} & -\frac{1}{T-2}\\ \vdots & \vdots & \vdots & & \vdots & \vdots & \vdots\\ 0 & 0 & 0 & \cdots & 1 & -\frac{1}{2} & -\frac{1}{2}\\ 0 & 0 & 0 & \cdots & 0 & 1 & -1 \end{bmatrix}.\end{aligned}$$

If we define $\bar{Y}_t = \frac{1}{T-t}\sum_{s=t+1}^{T} Y_s$, this transformation gives

$$\mathbf{KV} = \begin{bmatrix} \frac{T-1}{T}\left\{Y_1 - \bar{Y}_1 - \left(X_1 - \bar{X}_1\right)'\beta\right\}\\ \frac{T-2}{T-1}\left\{Y_2 - \bar{Y}_2 - \left(X_2 - \bar{X}_2\right)'\beta\right\}\\ \vdots\\ \frac{2}{3}\left\{Y_{T-2} - \bar{Y}_{T-2} - \left(X_{T-2} - \bar{X}_{T-2}\right)'\beta\right\}\\ \frac{1}{2}\left\{Y_{T-1} - Y_T - \left(X_{T-1} - X_T\right)'\beta\right\} \end{bmatrix}$$

so that as with the within-group and first difference operators, premultiplication by $\mathbf{K}$ eliminates the unit-specific intercepts A_i. But there is also a crucial difference. The first row of $\mathbf{KV}$ is a linear combination of $V_1,\ldots,V_T$, the second of just $V_2,\ldots,V_T$, the third of $V_3,\ldots,V_T$, while the last row depends only on V_{T-1} and V_T.

This feature of the transformation implies that if we choose our instrument matrix to be

$$\mathbf{Z} = \begin{bmatrix} X_1' & & & 0\\ & \begin{pmatrix} X_1' & X_2' \end{pmatrix} & & \\ & & \begin{pmatrix} X_1' & \cdots & X_{T-1}' \end{pmatrix} & \\ 0 & & & Q' \end{bmatrix}, \tag{12}$$

then $\mathbb{E}\left[\mathbf{Z}_i'\mathbf{H}\mathbf{V}_i\right] = 0$ will be a valid moment condition under (10) and (11). A test for "strict exogeneity" verses weak exogeneity" can be constructed by comparing the Sargan-Hansen test statistic associated with the efficient two-step GMM estimator which uses instrument matrix (5) given in the previous section, with the Sargan-Hansen test statistic associated with the efficient two-step GMM estimator based upon instrument matrix (12) immediately above. To alleviate finite-sample bias concerns, point estimates should be based upon a one-step estimator. A high profile policy evaluation paper which explicitly invokes weak exogeneity instead of strict exogeneity is the one by Chay & Greenstone (2003).

References

Arellano, M. (2003). *Panel Data Econometrics*. Oxford: Oxford University Press.

Arellano, M. & Bover, O. (1995). Another look at the instrumental variables estimation of error-component models. *Journal of Econometrics*, 68(1), 29 – 51.

Chamberlain, G. (1984). *Handbook of Econometrics*, volume 2, chapter Panel Data, (pp. 1247 – 1318). North-Holland: Amsterdam.

Chamberlain, G. (1992a). Comment: sequential moment restrictions in panel data. *Journal of Business and Economic Statistics*, 10(2), 20 – 26.

Chamberlain, G. (1992b). Efficiency bounds for semiparametric regression. *Econometrica*, 60(3), 567 – 596.

Chay, K. Y. & Greenstone, M. (2003). The impact of air pollution on infant mortality: evidence from geographic variation in pollution shocks induced by a recession. *Quarterly Journal of Economics*, 118(3), 1121 – 1167.

Hansen, L. P. (1982). Large sample properties of generalized method of moments estimators. *Econometrica*, 50(4), 1029 – 1054.

Pitt, M. M., Rosenzweig, M. R., , & Gibbons, D. M. (1993). The determinants and consequences of the placement of government programs in indonesia. *World Bank Economic Review*, 7(1), 319 – 348.

Wooldridge, J. M. (2010). *Econometric Analysis of Cross Section and Panel Data*. Cambridge, MA: MIT Press, 2nd edition.

