Technical Appendix for "Real Exchange Rate Fluctuations and the Dynamics of Retail Trade Industries on the U.S.-Canada Border"*

Jeffrey R. Campbell† Beverly Lapham‡

August 2004

The purpose of this appendix is twofold. First, we present our procedure for replacing missing or erroneous observations in the County Business Patterns Dataset. Second, we present the GMM estimator we use, based on Blundell and Bond (1998).

I Data Correction Procedure

The County Business Patterns data that we employ present two difficulties that must be surmounted prior to estimation. First, the Census withholds the employment observation

*The opinions expressed herein are those of the authors and are not necessarily those of the Federal Reserve Bank of Chicago or the Federal Reserve System. The National Science Foundation supported Campbell's research through grant SBR-9730442, "Business Cycles and Industry Dynamics." The Social Science and Humanities Research Council of Canada supported Lapham's research. The latest versions of this technical appendix are available on the World Wide Web at http://www.nber.org/~jrc. Please direct correspondence to Campbell at Economic Research, Federal Reserve Bank of Chicago, 230 South LaSalle Street, Chicago, Illinois 60604.

†Economic Research, Federal Reserve Bank of Chicago and NBER

‡Department of Economics, Queen's University

for any county-industry-year combination when that datum may disclose information about a single producer. Thus, our panel of average employment observations is not balanced. Second, the County Business Patterns' industry-level reports of employment for specific counties in specific years appear to be erroneously high. This section describes our approach to overcoming these two limitations of the data.

A Data Imputation

The basic idea underlying our procedure for replacing withheld data is to use the information that we do have on establishment counts by employment size class at the county level and total employment at the state level to estimate the relationship between the number of establishments and total employment among those counties where the data has been withheld. Fitted values from this estimated regression then serve as estimates of the withheld employment data.

To begin with, focus on a particular retail trade industry during a particular year. Let E_c^s denote the total employment in that industry in county c of state s, and let E^s denote the statewide employment in that industry for state s. If $\mathfrak{C}(s)$ is the set of all counties in state s, then

$$E^s = \sum_{c \in \mathfrak{C}(s)} E_c^s.$$

We assume that observations of E^s are available for every state. Because the number of retail establishments in a given state is usually large, data suppression is typically not a problem at the state level in this data set. On the other hand, suppression of observations of E_c^s for individual counties is common. What is always reported for each county is the number of establishments belonging to several predetermined size classes (based on mid-March employment). Let J denote the set of such size classes and $N_c^s(j)$ denote the number of establishments in class j in county c of state s. The data replacement procedure is based on a regression model of E_c^s on $N_c^s(j)$ restricted to those counties where the census has withheld publication of E_c^s. Let $\mathfrak{W}(s)$denote the set of all counties in state s for which the

Census has withheld publication of E_c^s. Then the basic regression model is

$$E_c^s = \sum_{j \in J} \beta_j N_c^s(j) + u_c^s \tag{1}$$

$$\mathbf{E}[u_c^s] = 0$$

for all $c \in \mathfrak{W}(s)$. The coefficients β_j are constant across both counties and states. That is, the regression equation specifies that the total employment in a county equals a linear function of the number of establishments in each size class plus a mean zero error term.

The obvious impediment to estimating the equation is that the dependent variable is withheld for all of the observations of interest. To overcome this, we can aggregate the equation to the state level, where the aggregated dependent variable is observable. To do so, define $\widetilde{E}^s$ as the employment in all counties in state s for which employment data is withheld. This can be constructed as statewide employment minus employment at all counties at which employment was reported. That is

$$\widetilde{E}^s = E^s - \sum_{c \in \overline{\mathfrak{W}}(s)} E_c^s,$$

where $\overline{\mathfrak{W}}(s)$ is the complement of $\mathfrak{W}(s)$. If we then define

$$\widetilde{N}^s(j) = \sum_{c \in \mathfrak{W}(s)} N_c^s(j)$$

then aggregating (1) for state s yields

$$\widetilde{E}^s = \sum_{j \in J} \beta_j \widetilde{N}^s(j) + \widetilde{u}^s, \tag{2}$$

where

$$\widetilde{u}^s = \sum_{c \in \mathfrak{W}(s)} u_c^s.$$

If we calculate the dependent variables and regressors for (2) for each state, then the coefficients β_j can be estimated by applying the regression to the state-level data. The fitted values of this estimated model can then be used to construct estimates of the withheld county-level employment data. When implementing this procedure, we construct separate estimates of β_j for each year and industry in our sample.

B Data Correction

Next, consider the problem of identifying and correcting erroneous observations of employment in the County Business Patterns. As we noted in the text, the data file and printed versions of the *CBP* contain identical observations of employment for some county-industry pairs in some years that are extremely high. For example, the *CBP* editions for 1994, 1995, and 1996 report employment for Food Stores in Whatcom County, Washington (a border county) to equal $1,937$; $1,875$; and $4,002$ employees. This last observation lies far outside of the historically observed values for this variable. The *CBP* for that year also reports the existence of an establishment in Food Stores with over $1,000$ employees. This is the only establishment with more than 500 employees reported in these three years in the entire state of Washington. When we spoke with representatives of the local chamber of commerce, they indicated that there was never a grocery store in Whatcom county with so many employees. Publically available observations of this industry's employment derived from unemployment insurance records (ES-202) data do not reflect such a large jump in employment. They indicate that March employment in 1996 equals $2,160$.

The historically unprecidented employment measure for Whatcom County in 1996, the unusually large establishment reported to exist there during that year, the failure of local observers to verify this event, and the discrepancy between the *CBP* and the ES-202 data all lead us to suspect that clerical error in producing the *CBP* led to this very high observation. Accordingly, we have replaced the *CBP*'s observation with that from the ES-202 data. We searched for other similar clerical errors throughout our data set using the following procedure: For each county-industry pair, we calculated the (time-series) interquartile range of employment. If the employment observation for a particular year differed from the median employment by more than $3/1.35$ times the estimated interquartile range, the observation was examined further to determine whether or not it also reported an unusually large establishment.[1] The definition of "unusually large" is more than 250 employees

[1] With this multiple of the interquartile range, the resulting quantity equals three standard deviations if

for Food Stores and Eating Places, and it is more than 100 employees for Gasoline Service Stations and Drinking Places. If the observation also had an unusually large establishment, we flagged it as "problematic." In Food Stores, we compared the *CBP*'s employment observations with that county's March observation in the Bureau of Labor Statistics' Covered Wages and Employment (ES-202) data set. If that employment observation differed from the median by less than 3/1.35 times the interquartile range, we replaced the *CBP*'s employment report with the BLS's. Otherwise, we left the *CBP*'s data report intact. The Covered Wages and Employment does not contain observations of our other industries, so we cannot use it to replace any of their problematic observations. For those industries, we drop any county-industry pair with a problematic observation in any year.

II GMM Estimation

In this section, we consider the GMM estimation procedure described in the text in more detail. Several aspects of the paper's empirical model generalize quite readily. Here we consider the appropriately generalized version with m dependent variables, an autoregressive order of p, and k current and lagged values of the real exchange rate. The resulting estimating equation is

$$\underset{(m\times 1)}{y_{it}} = \underset{(m\times 1)}{\alpha_i} + \underset{(m\times 1)}{\mu_t} + \sum_{l=1}^{p} \underset{(m\times m)}{\Lambda_l} \underset{(m\times 1)}{y_{it-l}} + \underset{(m\times k)}{\beta'} \left(s_i \times \underset{(k\times 1)}{e_t} \right) + \underset{(m\times 1)}{\varepsilon_{it}}. \tag{3}$$

The dimensions of all vectors and matrices appear below them. The vector y_{it} contains the period t values of the m variables describing a particular retail trade industry in county i. In the baseline model, m equals 2 and these variables are the logarithms of total establishments and their average employment. The vector α_i is the county specific intercept term, and the vector μ_t is an aggregate disturbance that impacts all counties' industries in period t. The matrices Λ_l contain the model's autoregressive coefficients. The scalar s_i is the index of county i's sensitivity to real exchange rates described in the text, and the vector e_t contains

the data are normally distributed.

the k current and lagged realizations of the real exchange rate. For most counties in our sample, $s_i = 0$. Finally, the matrix β contains the elasticities of y_{it} with respect to a change in e_t for a county with s_i equal to one, and ε_{it} is a disturbance vector.

We observe a balanced panel of the variables in y_{it} for T periods and N counties. Equation (3) only describes the evolution of y_{it} for t between p and T. The initial p realizations of y_{it} will play a key role in estimation of the unknown parameters in (3). We make the following assumptions on the model's error terms and parameters.

1. The roots of $|I - \sum_{l=1}^{p} \Lambda_l L^p|$ all lie strictly outside of the unit circle.

2. $\Pr[s_i = 0] > 0$.

3. The individual specific intercept α_i and the error terms ε_{it}, $t = p+1, \ldots, T$ are independently distributed across individuals and

   (a) $\mathbf{E}[\alpha_i | s_i = 0] = 0$,

   (b) $\mathbf{E}[\varepsilon_{it}] = 0$, $t = p+1, \ldots, T$,

   (c) $\mathbf{E}[\varepsilon_{it}\varepsilon'_{i\tau} | s_i = 0] = 0$, if $t \neq \tau$,

   (d) $\mathbf{E}[\alpha_i \varepsilon'_{it}] = 0$, $t = p+1, \ldots, T$,

   (e) $\mathbf{E}[s_i \varepsilon_{it}] = 0$, $t = p+1, \ldots, T$,

   (f) $\mathbf{E}[\alpha_i \alpha'_i] < \infty$

   (g) $\mathbf{E}[\varepsilon_{it}\varepsilon'_{it}] < \infty$, $t = p+1, \ldots, T$.

4. If $s_i = 0$, then the first p values of y_{it} satisfy

   (4) $$y_{it} = \mu_t + \left(I - \sum_{l=1}^{p} \Lambda_l\right)^{-1} \alpha_i + u_{it}, \;\; t = 1, \ldots, p,$$

   where

   (a) $\mathbf{E}[u_{it}] = 0$, $t = 1, \ldots, p$,

(b) $\mathbf{E}[\alpha_i u'_{it}] = 0$, $t = 1, \ldots, p$,

(c) $\mathbf{E}[s_i u_{it}] = 0$, $t = 1, \ldots, p$,

(d) $\mathbf{E}[u_{i\tau} \varepsilon'_{it}] = 0$, for all $\tau = 1, \ldots, p$ and $t = p+1, \ldots, T$,

(e) $\mathbf{E}[u_{it} u'_{i\tau}] < \infty$ for all $\tau = 1, \ldots, p$ and $t = 1, \ldots, p$.

5. The regressors $e_{p+1}, e_{p+2}, \ldots, e_T$ are known constants.

Assumption 1 implies that the autoregressive system in (3) is stable, and Assumption 2 asserts that cross-border shopping does not impact a positive fraction of our sample counties. This is clearly the case in our sample. Given the presence of the time effects in (3), 3(a) and 3(b) are normalizations. Assumption 3(c) restricts the error term in (3) to be uncorrelated through time. Assumptions 3(d) and 3(e) assert that ε_{it} cannot be forecasted using a linear function of α_i and s_i. In the case where $s_i = 0$, Assumptions 4(a) and 4(b) assert that the deviations of y_{i1} through y_{ip} from their unconditional means are uncorrelated with those means. Assumption 4(d) asserts that ε_t cannot be forecasted using linear functions of $u_{i\tau}$. Notice that assumptions 3(d) and 4(b) do not restrict the higher moments of ε_{it} or u_{it} from being dependent on α_i, so the model allows for general forms of heteroskedasticity. Also note that we do not constrain the covariance between α_i and s_i to equal zero. The remaining assumptions are regularity conditions that guarantee existence of second moments for y_{it}.

A Moment Conditions

To estimate the unknown parameters in (3), we derive moment conditions which are functions of the observed data that are satisfied only at the true parameter values. We then use these moment conditions in a GMM estimation procedure to produce consistent parameter estimates. Our derivation of the moment conditions closely follows Blundell and Bond (1998). The differences between our derivation and theirs is minor, and only allow for the inclusion of the independent variables $s_i e_t$ and a vector (as opposed to univariate) autoregression. Our distributional theory for the estimator is the same as Blundell and Bond's, letting N go to infinity while T is held fixed.

If $s_i = 0$, we can use (3) and (4) to write that

$$y_{ip+1} = \widetilde{\alpha}_i + \mu_{p+1} + \sum_{l=1}^{p} \Lambda_l \left(\widetilde{\alpha}_i + u_{ip+1-l}\right) + \varepsilon_{ip+1-l}.$$

In general, for $t \geq p+1$, we get

$$y_{it} = \widetilde{\alpha}_i + \sum_{j=0}^{t-p-1} \psi_j \varepsilon_{it-j} + \sum_{j=t-p}^{t-1} \psi_j u_{it-j} + \sum_{j=0}^{t-1} \psi_j \mu_{t-j}, \tag{5}$$

where ψ_j is defined recursively with

$$\psi_0 = I_m$$

$$\psi_j = 0, \ \forall j < 0$$

$$\psi_j = \sum_{l=1}^{p} \Lambda_l \psi_{j-l}, \ \forall j > 0.$$

Equation (5) and assumptions 3(c) and 4(d) imply that

$$\mathbf{E}\left[I\left\{s_i = 0\right\} \Delta\varepsilon_{it} \cdot y'_{it-\tau}\right] = 0, \ \forall t \geq p+2, \ 2 \leq \tau \leq t-1, \tag{6}$$

where $I\{s_i = 0\}$ is an indicator function that equals one if and only if $s_i = 0$. Furthermore, (5) and assumptions 3(a), 3(b), 3(c), 3(d), 4(a), 4(b), and 4(d) imply that

$$\mathbf{E}\left[I\left\{s_i = 0\right\} \times \left(\alpha_i + \varepsilon_{it}\right) \cdot \Delta y'_{it-\tau}\right] = 0, \ \forall t \geq p+1, \ \tau \geq 1 \tag{7}$$

Note that many of the additional moment conditions implied by (7) are redundant. If we define $\widetilde{t}(t) = \max\{t+1, p+1\}$ and we choose a $t^\star > \widetilde{t}(t)$, then we can write that

$$\begin{aligned}
\mathbf{E}\left[I\left\{s_i = 0\right\} \times \left(\alpha_i + \varepsilon_{it^\star}\right) \cdot \Delta y'_{it}\right] &= \mathbf{E}\left[I\left\{s_i = 0\right\}\left(\alpha_i + \varepsilon_{i\widetilde{t}(t)} + \sum_{\tau=\widetilde{t}(t)+1}^{t^\star} \Delta\varepsilon_{i\tau}\right) \cdot \Delta y'_{it}\right] \\
&= \mathbf{E}\left[I\left\{s_i = 0\right\}\left(\alpha_i + \varepsilon_{i\widetilde{t}(t)}\right) \cdot \Delta y'_{it}\right] \\
&\quad + \sum_{\tau=\widetilde{t}(t)+1}^{t^\star} \mathbf{E}\left[I\left\{s_i = 0\right\} \Delta\varepsilon_{i\tau} \cdot y'_{it}\right] - \sum_{\tau=\widetilde{t}(t)+1}^{t^\star} \mathbf{E}\left[I\left\{s_i = 0\right\} \Delta\varepsilon_{i\tau} \cdot y'_{it-1}\right].
\end{aligned}$$

Therefore, imposing

$$\mathbf{E}\left[I\left\{s_i=0\right\}\left(\alpha_i+\varepsilon_{i\widetilde{t}(t)}\right)\cdot\Delta y_{it}'\right]=0,\ t=2,\ldots,T-1 \tag{8}$$

and (6) suffices to impose the entire set of moment conditions implied by (6) and (7).

Finally, 3(a) and 3(b) imply that

$$\mathbf{E}\left[I\left\{s_i=0\right\}\left(\alpha_i+\varepsilon_{it}\right)\right]=0, \tag{9}$$

and 3(e) implies that

$$\mathbf{E}\left[\Delta\varepsilon_{it}\cdot s_i\right]=0,\ \forall t\geq p+2. \tag{10}$$

Our full set of moment conditions used for parameter estimation is given by (6), (8), (9), and (10).

B Parameter Estimation

Let

$$\gamma=\left(\operatorname{vec}\left(\Lambda_1'\right)',\operatorname{vec}\left(\Lambda_2'\right)',\ldots,\operatorname{vec}\left(\Lambda_p'\right)',\operatorname{vec}\left(\beta'\right)',\mu_{p+1}',\mu_{p+2}',\ldots,\mu_T'\right)'$$

denote the vector of parameters of interest, and define

$$u_{it}\left(\gamma\right)=y_{it}-\mu_t+\sum_{l=1}^{p}\Lambda_l y_{it-l}+\beta'\left(s_i\times e_t\right),$$

for $t\geq p+1$. Let γ_0 denote the true parameter values. Then the moment conditions (6), (8), (9), and (10) can be rewritten as

$$\mathbf{E}\left[\left\{s_i=0\right\}\Delta u_{it}\left(\gamma_0\right)\cdot y_{it-\tau}'\right]=0,\ t=p+2,\ldots,T,\ 2\leq\tau\leq t-1$$

$$\mathbf{E}\left[I\left\{s_i=0\right\}\times u_{i\widetilde{t}(t)}\left(\gamma_0\right)\cdot\Delta y_{it}'\right]=0,\ t=2,\ldots,T-1.$$

$$\mathbf{E}\left[I\left\{s_i=0\right\}u_{it}\left(\gamma_0\right)\right]=0,\ \ t=p+1,\ldots,T$$

$$\mathbf{E}\left[\Delta u_{it}\cdot s_i\right]=0,\ t=p+2,\ldots,T.$$

To express these moment conditions in matrix form for a given individual, define the error vector $u_i(\gamma)$ to be

$$u_i(\gamma) = \begin{bmatrix} \Delta u_{ip+2}(\gamma) \\ \Delta u_{ip+3}(\gamma) \\ \vdots \\ \Delta u_{iT}(\gamma) \\ u_{ip+1}(\gamma) \\ u_{ip+2}(\gamma) \\ \vdots \\ u_{iT}(\gamma) \end{bmatrix},$$

and define the instrument vector z_i to be

$$z_i = \begin{bmatrix} I\{s_i = 0\} \\ s_i \\ I\{s_i = 0\} \cdot y_{i1} \\ I\{s_i = 0\} \cdot y_{i2} \\ \vdots \\ I\{s_i = 0\} \cdot y_{iT-2} \\ I\{s_i = 0\} \cdot \Delta y_{i2} \\ I\{s_i = 0\} \cdot \Delta y_{i3} \\ \vdots \\ I\{s_i = 0\} \cdot \Delta y_{iT-1} \end{bmatrix}.$$

Finally, define the moment selector matrix C to be a sparse matrix with row dimension equal to the number of valid moment conditions and a single element in the j'th column equal to one if

$$\mathbf{E}\left[(z_i \otimes u_i(\gamma_0))_j\right] = 0$$

where the subscript j refers to the j'th element of that vector.

Let A_N be a square, positive definite matrix that has dimensionality equal to the row dimension of C. This matrix may be data dependent. Define the sample moment function $g_N(\gamma)$ as

$$g_N(\gamma) = C \cdot \frac{1}{N} \sum_{i=1}^{N} z_i \otimes u_i(\gamma).$$

Then the GMM estimator is the value of γ that minimizes

$$J_N(\gamma) = g_N(\gamma)' \cdot A_N \cdot g_N(\gamma).$$

To characterize the solution to this minimization problem, we can apply the rule for differentiating a quadratic form to get the first-order necessary condition which the GMM estimator, $\widehat{\gamma}_N$, must satisfy

$$\frac{\partial J_N(\widehat{\gamma}_N)}{\partial \gamma'} = 2 g_N(\widehat{\gamma}_N)' A_N \frac{\partial g_N(\widehat{\gamma}_N)}{\partial \gamma'} = 0.$$

To find a closed-form solution for $\widehat{\gamma}_N$, it is helpful to define

$$Y_i = \begin{bmatrix} \Delta y_{ip+2} \\ \Delta y_{ip+3} \\ \vdots \\ \Delta y_{iT} \\ y_{ip+1} \\ y_{ip+2} \\ \vdots \\ y_{iT} \end{bmatrix},$$

and

$$
X_i = \begin{bmatrix}
I_m \otimes \Delta y'_{ip+1} & I_m \otimes \Delta y'_{ip} & \cdots & I_m \otimes \Delta y'_{i2} & I_m \otimes s_i \Delta e'_{p+2} \\
I_m \otimes \Delta y'_{ip+2} & I_m \otimes \Delta y'_{ip+1} & \cdots & I_m \otimes \Delta y'_{i3} & I_m \otimes s_i \Delta e'_{p+3} \\
\vdots & \vdots & \ddots & \vdots & \vdots \\
I_m \otimes \Delta y'_{iT-1} & I_m \otimes \Delta y'_{iT-2} & \cdots & I_m \otimes \Delta y'_{iT-p} & I_m \otimes s_i \Delta e'_{T} \\
I_m \otimes y'_{ip} & I_m \otimes y'_{ip-1} & \cdots & I_m \otimes y'_{i1} & I_m \otimes s_i e'_{p+1} \\
I_m \otimes y'_{ip+1} & I_m \otimes y'_{ip} & \cdots & I_m \otimes y'_{i2} & I_m \otimes s_i e'_{p+2} \\
\vdots & \vdots & \ddots & \vdots & \vdots \\
I_m \otimes y'_{iT-1} & I_m \otimes y'_{iT-2} & \cdots & I_m \otimes y'_{iT-p} & I_m \otimes s_i e'_{T}
\end{bmatrix.}
$$

$$
\begin{matrix}
-I_m & I_m & 0 & \cdots & 0 & 0 \\
0 & -I_m & I_m & \cdots & 0 & 0 \\
\vdots & \vdots & \vdots & \ddots & \vdots & \vdots \\
0 & 0 & 0 & \cdots & -I_m & I_m \\
I_m & 0 & 0 & \cdots & 0 & 0 \\
0 & I_m & 0 & \cdots & 0 & 0 \\
\vdots & \vdots & \vdots & \ddots & \vdots & \vdots \\
0 & 0 & 0 & \cdots & 0 & I_m
\end{matrix} \Bigg] .
$$

Then we can write that

$$
u_i(\gamma) = Y_i - X_i \gamma,
$$

and that

$$
\begin{aligned}
C(z_i \otimes u_i(\gamma)) &= C(z_i \otimes (Y_i - X_i\gamma)) \\
&= C(z_i \otimes Y_i) - C(z_i \otimes X_i\gamma) \\
&= C(z_i \otimes Y_i) - C(z_i \otimes X_i)\gamma.
\end{aligned}
$$

This final equality follows from the fact that z_i is a single column vector.

Using this expression, we can rewrite $g_N(\gamma)$ and its derivative as

$$g_N(\gamma) = C \cdot \frac{1}{N}\left(\sum_{i=1}^{N} z_i \otimes Y_i - \sum_{i=1}^{N} z_i \otimes X_i \gamma\right)$$
$$\frac{\partial g_N(\gamma)}{\partial \gamma'} = -C \cdot \frac{1}{N}\sum_{i=1}^{N} z_i \otimes X_i.$$

Using these expressions, we can write the first-order condition for minimization of the GMM criterion function as

$$2\left(C \cdot \sum_{i=1}^{N} z_i \otimes Y_i - \left(C \cdot \sum_{i=1}^{N} z_i \otimes X_i\right)\widehat{\gamma}_N\right)' A_N \left(C \cdot \sum_{i=1}^{N} z_i \otimes X_i\right).$$

Rearrainging this yields

$$\widehat{\gamma}_N' \left(C \cdot \sum_{i=1}^{N} z_i \otimes X_i\right)' A_N \left(C \cdot \sum_{i=1}^{N} z_i \otimes X_i\right) = \left(C \cdot \sum_{i=1}^{N} z_i \otimes Y_i\right)' A_N \left(C \cdot \sum_{i=1}^{N} z_i \otimes X_i\right).$$

So the final expression for the GMM estimator is

$$\widehat{\gamma}_N = \left[\left(C \cdot \sum_{i=1}^{N} z_i \otimes X_i\right)' A_N \left(C \cdot \sum_{i=1}^{N} z_i \otimes X_i\right)\right]^{-1} \left(C \cdot \sum_{i=1}^{N} z_i \otimes X_i\right)' A_N \left(C \cdot \sum_{i=1}^{N} z_i \otimes Y_i\right)$$

B.1 The Large-Sample Distribution of $\widehat{\gamma}_N$

To characterize the distribution of the $\widehat{\gamma}_N$, we apply standard asymptotic distributional arguments, letting N go to infinity while holding T fixed. Towards this end, it is straightforward to show that

$$\widehat{\gamma}_N - \gamma_0 = \left[\left(C \cdot \sum_{i=1}^{N} z_i \otimes X_i\right)' A_N \left(C \cdot \sum_{i=1}^{N} z_i \otimes X_i\right)\right]^{-1} \left(C \cdot \sum_{i=1}^{N} z_i \otimes X_i\right)' A_N \left(C \cdot \sum_{i=1}^{N} z_i \otimes u_i(\gamma_0)\right).$$

Using this, we can apply standard cross-sectional asymptotic theory to show that

$$\operatorname*{plim}_{N\to\infty} \widehat{\gamma}_N = \gamma_0,$$

and that

$$\sqrt{N}\left(\widehat{\gamma}_N - \gamma_0\right) \xrightarrow{d} N(0, V),$$

where

$$V = D'^{-1} S D^{-1},$$

$$D = \operatorname*{plim}_{N\to\infty} \frac{1}{N^2} \left(C \cdot \sum_{i=1}^{N} z_i \otimes X_i \right)' A_N \left(C \cdot \sum_{i=1}^{N} z_i \otimes X_i \right),$$

$$S = \operatorname*{plim}_{N\to\infty} \frac{1}{N^3} \left(C \cdot \sum_{i=1}^{N} z_i \otimes X_i \right)' A_N \cdot \left(\sum_{i=1}^{N} C \left(z_i \otimes u_i \left(\gamma_0 \right) \right) \left(z_i \otimes u_i \left(\gamma_0 \right) \right)' C' \right) \cdot A_N \left(C \cdot \sum_{i=1}^{N} z_i \otimes X_i \right).$$

For a given sequence of weighing matrices, D and S can be consistently estimated using their sample analogues.

B.2 The Weighing Matrix

We rely on one-step GMM estimators, calculated using the weighing matrix.

$$A_N = \left(\frac{1}{N} \sum_{i=1}^{N} C \left(z_i z_i' \otimes \Sigma \right) C' \right),$$

where

$$\underset{m\cdot(2\cdot(T-p)-1)\times m\cdot(2\cdot(T-p)-1)}{\Sigma} = \begin{bmatrix} \underset{m\cdot(T-p-1)\times m\cdot(T-p-1)}{A} & \underset{m\cdot(T-p-1)\times m\cdot(T-p)}{0} \\ \underset{m\cdot(T-p)\times m\cdot(T-p-1)}{0} & \underset{m\cdot(T-p)\times m\cdot(T-p)}{I} \end{bmatrix}$$

where

$$A = \begin{bmatrix} 2I_m & -I_m & 0 & 0 & \cdots & 0 \\ -I_m & 2I_m & -I_m & 0 & \cdots & 0 \\ 0 & -I_m & 2I_m & -I_m & \cdots & 0 \\ 0 & 0 & -I_m & 2I_m & \cdots & 0 \\ \vdots & \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & 0 & 0 & \cdots & 2I_m \end{bmatrix}.$$

This weighing matrix is the multivariate extension of the initial weighing matrix used by Blundell and Bond (1998).

References

Blundell, Richard and Stephen Bond. (1998) Initial Conditions and Moment Restrictions in Dynamic Panel Data Models. *Journal of Econometrics,* **87**(1), 115–143.