NEW PRODUCT VARIETIES AND THE MEASUREMENT OF INTERNATIONAL PRICES

Robert C. Feenstra

APPENDIX A

In this appendix an instrumental variables (IV) interpretation is provided for the estimator in section IV. Suppose that import price and quantity data from each country are available over the periods 1,2,...,T. Let $T_i \leq T$ denote the number of time periods for which country $i$ supplies the product in the current and last year, so that the first-differences in (10) can be computed. Stacking (10) over time and then over countries, the total number of observations is $L = \sum_{i=1}^{k} T_i$. Letting $Y$ denote the $L \times 1$ vector with components $Y_{it}$, $X$ the $L \times 2$ matrix with rows $(X_{1it}, X_{2it})$, $u$ the $L \times 1$ vector with components $u_{it}$, and $\theta$ the column vector $(\theta_1, \theta_2)$, equation (10) can be written as:

\[(A1) \quad Y = X\theta + u.\]

To control for the correlation of $u_{it}$ with $X_{1it}$ and $X_{2it}$, use IV that are dummy variables for each good $i \neq k$. Let $l_i$ denote a $T_i \times 1$ vector of 1's, $i = 1,...,N$, $i \neq k$, and define $Z$ as the $L \times (N-1)$ matrix,

\[Z = \begin{pmatrix} l_1 & 0 & \ldots & 0 \\ 0 & l_2 & \ldots & 0 \\ 0 & 0 & \ldots & 0 \\ \vdots & \vdots & \vdots \\ 0 & 0 & \ldots & l_N \end{pmatrix}.\]

Then consider the usual IV estimator:

\[(A2) \quad \hat{\theta} = [X'Z(Z'Z)^{-1}Z'X]^{-1}X'Z(Z'Z)Z'Y = \theta + [X'Z(Z'Z)^{-1}Z'X]^{-1}X'Z(Z'Z)Z'u.\]
Consider taking the probability limit of (A2) as $T \to \infty$, while holding $N$ fixed.

Under condition (12) and the assumptions in Feenstra (1991), it can be shown that: plim$[(X'Z/T)(Z'Z/T)^{-1}(Z'X/T)]$ has full rank of 2, and so is invertible; and plim$(Z'u/T)$ is a $(N-1)x1$ vector with components $\text{plim}[\frac{\sum_t \delta_{it}\xi_{it}}{T(\sigma-1)(1-p)}]$, which equal zero from the assumed independence of $\delta_{it}$ and $\xi_{it}$. Then it follows immediately that $\text{plim}\hat{\theta} = \theta$, so the IV estimator is consistent.

To relate this estimator to equation (10') in the text, let $\bar{Y}_i = \frac{\sum_t Y_{it}}{T_i}$, $\bar{X}_{1i} = \frac{\sum_t X_{1it}}{T_i}$, $\bar{X}_{2i} = \frac{\sum_t X_{2it}}{T_i}$ and $\bar{u}_i = \frac{\sum_t u_{it}}{T_i}$ denote the means of the variables in (A1) over country $i$. Then pre-multiplying (A1) by $Z(Z'Z)^{-1}Z'$, $L$ equations are obtained of the form,

$$\text{(A3)} \quad \bar{Y}_i = \theta_1\bar{X}_{1i} + \theta_2\bar{X}_{2i} + \bar{u}_i,$$

where the equation for country $i$ is repeated $T_i$ times. Thus, the IV estimate $\hat{\theta}$ in (A2) can be equivalently obtained by running weighted least squares (WLS) on the observations $i=1,...,N$, $i \neq k$, in (A3), while using $T_i$ as the weights.

While the IV estimator in (A2) is consistent, it is not the most efficient. To see this, consider that the error term $u_{it}$ in (A1). It has the variance $E(u_{it}^2) = (\sigma^2_{\delta_1} + \sigma^2_{\delta_k})(\sigma^2_{\xi_1} + \sigma^2_{\xi_k})/[((1-p)(\sigma-1))^2]$, which differs across countries $i$ when (12) holds. To correct for this, weight all observations in (A1) for country $i$ by the inverse of $\hat{s}_i^2 = \frac{\sum_t \hat{u}_{it}^2}{T_i}$, where $\hat{u}_{it} = Y_{it} - \hat{\theta}_1X_{1it} - \hat{\theta}_2X_{2it}$ is the computed residual using the initial estimates $\hat{\theta}$. Letting $\hat{S}$ denote the $(L \times L)$ diagonal matrix with $\hat{s}_i^2$ repeated $T_i$ times on the diagonal for $i=1,...,N$, $i \neq k$, and with $\hat{X} = Z(Z'Z)^{-1}Z'X$, the weighted IV estimator is:

$$\text{(A4)} \quad \hat{\theta}^* = [\hat{X}'\hat{S}^{-1}\hat{X}]^{-1}\hat{X}'\hat{S}^{-1}\hat{Y} = \theta + [\hat{X}'\hat{S}^{-1}\hat{X}]^{-1}\hat{X}'\hat{S}^{-1}\hat{u}.$$
where the second line follows since it can be shown that \( \hat{X}'\mathbf{S}^{-1}\hat{X} = \bar{X}'\mathbf{S}^{-1}\bar{X} \).

Halbert White (1982) demonstrates the consistency of \( \hat{e}^* \) for an unbalanced panel, when the errors \( u_{it} \) are independent over \( i \) and \( t \). In this case \( \hat{e}^* \) is the efficient estimator (given the set of instruments), and its covariance matrix is consistently estimated by \( [\bar{X}'\mathbf{S}^{-1}\bar{X}]^{-1} \). This is the formula used for the standard errors reported in Table 2.

Finally, to recognize the measurement error in the unit-values \( U_{it} \) that are used, specify that,

\[
(A5) \quad \Delta \ln U_{it} = \Delta \ln P_{it} + \mu_{it},
\]

where \( P_{it} \) are the true but unobserved prices, and \( \mu_{it} \) is the measurement error. It is assumed that \( \mu_{it} \) is stationary with equal variance across supplying countries \( i \), and that \( \mu_{it} \) is independent of \( \varepsilon_{js} \) and \( \delta_{js} \). Then using (A5), (10) can be rewritten as,

\[
(A6) \quad (\Delta \ln U_{it} - \Delta \ln U_{kt})^2 = 2\sigma_{\mu}^2 + \theta_1(\Delta \ln s_{it} - \Delta \ln s_{kt})^2 \\
+ \theta_2(\Delta \ln U_{it} - \Delta \ln U_{kt})(\Delta \ln s_{it} - \Delta \ln s_{kt}) + v_{it},
\]

where,

\[
v_{it} = u_{it} + [(\mu_{it} - \mu_{kt})^2 - 2\sigma_{\mu}^2] + 2(\Delta \ln P_{it} - \Delta \ln P_{kt})(\mu_{it} - \mu_{kt}) \\
- \theta_2(\Delta \ln s_{it} - \Delta \ln s_{kt})(\mu_{it} - \mu_{kt}).
\]

Because of the independence of the measurement error \( \mu_{it} \) from \( \varepsilon_{js} \) and \( \delta_{js} \), the error \( v_{it} \) will have expected value of zero. Then it can be shown that the instruments \( Z \) are orthogonal to \( v_{it} \), but correlated with the right-hand side variables in (A6) provided that (12) holds. It follows that the IV estimator is consistent, where the term \( 2\sigma_{\mu}^2 \) in (A6) is replaced by a constant \( \theta_0 \). The efficient estimator and standard errors are constructed in the same manner as discussed above.
APPENDIX B

The variances of $\hat{\sigma}$ and $\hat{\rho}$ are computed by using the first-order approximation to (11e) around the true parameters $\sigma$ and $\rho$. Letting $V(\hat{\sigma}, \hat{\rho})$ and $V(\hat{\sigma}_1, \hat{\sigma}_2)$ denote the variance-covariance matrices, this procedure gives:

$$V(\hat{\sigma}, \hat{\rho}) = M' V(\hat{\sigma}_1, \hat{\sigma}_2) M$$

where

$$M = \begin{bmatrix}
(\hat{\sigma}-1)^3(1-\hat{\rho}) & -(\hat{\sigma}-1)^2(1-\hat{\rho})^2(1-2\hat{\rho}) \\
(\hat{\sigma}-1)^2(1-\hat{\rho}) & 2(\hat{\sigma}-1)\hat{\rho}(1-\hat{\rho})^2
\end{bmatrix}.$$

In addition, the supply elasticity is

$$(1/\hat{\omega}) = [(\hat{\sigma}-1)/\hat{\rho}] - \hat{\sigma},$$

and it follows that:

$$\text{var}(1/\hat{\omega}) = \frac{1}{\hat{\rho}^4} \left( \hat{\rho}^2(1-\hat{\rho})^2 \text{var} \hat{\sigma} + (\hat{\sigma}-1)^2 \text{var} \hat{\rho} - 2(\hat{\sigma}-1)\hat{\rho}(1-\hat{\rho}) \text{cov}(\hat{\sigma}, \hat{\rho}) \right).$$

These formulas are used to compute the standard errors reported in the second row of Table 2 for each product. The confidence intervals in the third row are obtained by considering all values of $\hat{\theta} = (\hat{\theta}_1, \hat{\theta}_2)$ in the 95% confidence ellipse, defined using the $F$-distribution as

$$(\hat{\theta}-\theta)' V(\hat{\theta})^{-1} (\hat{\theta}-\theta) \leq 2F_{0.95}(2, L-3).$$

For each value of $\hat{\theta}$ in this region, the formulas in Proposition 2 were applied, and the minimum and maximum values obtained for $\sigma$ define its 95% confidence interval.

REFERENCES


LIBNAME LIB 'D:DATA\M6';
options pagesize=65;

TITLE 'AS';

data COUNTRY1;
   set LIB.ASDATA;
   if COUNTRY = "JAP";
   DLSHARE1 = DLSHARE;
   DLPRICE1 = DLPRICE;
   keep YEAR DLSHARE1 DLPRICE1;
run;

data SORT;
   set LIB.ASDATA;
   proc sort; by YEAR;
run;

data RDL1;
   merge SORT COUNTRY1; by YEAR;
run;

proc sort; by COUNTRY YEAR;
run;

data D1;
   set RDL1;
   RDLSHARE = DLSHARE - DLSHARE1;
   RDPRICE = DLPRICE - DLPRICE1;
   drop DLSHARE1 DLPRICE1;
run;

data G;
   set D1;
   RDLP2 = RDPRICE**2;
   RDLS2 = RDLSHARE**2;
   RDLP = RDPRICE*RDLSHARE;
   keep RDLP2 RDLS2 RDLP S COUNTRY TYPE;
run;

proc glm
   data=G; class COUNTRY;
   model RDLS2 RDLP S = COUNTRY/noint p nouni;
   output out=NEW p=S2HAT PS3HAT;
run;

proc reg
   data=NEW;
   model RDLP2 = S2HAT PS3HAT/covb;
run;

data DAT2;
   set D1;
   RDLP2 = RDPRICE**2;
   RDLS2 = RDLSHARE**2;
   RDLP = RDPRICE*RDLSHARE;
   RES = RDLP2 - .051559*RDLS2 -.005433*RDLP S - 0.047146;
RES2 = RES**2;
keep RDLPS2 RDLPS RES RES2
TYPE YEAR COUNTRY;
run;

proc sort
data=DAT2; by COUNTRY;
run;

proc glm
data=DAT2; class COUNTRY;
model RDLPS RDLPS2 RES2 = COUNTRY/ p nouni;
output out=NEW2 p=PSHAT S2HAT R2HAT;
run;

data REG1;
set NEW2;
if R2HAT=0 or R2HAT=. then delete;
RHAT = SQRT(R2HAT);
Y  = RDLPS2/RHAT;
X0  = 1/RHAT;
X1  = S2HAT/RHAT;
X2  = PSHAT/RHAT;
keep Y X0 X1 X2 TYPE;
run;

proc reg
data=REG1;
model Y = X0 X1 X2/noint covb; /* THETA/star */
run;

data REG2;
set NEW2;
if R2HAT = . then delete;
RHAT = SQRT(R2HAT);
Y2  = RDLPS2*RHAT;
X20  = RHAT;
X21  = S2HAT*RHAT;
X22  = PSHAT*RHAT;
keep Y2 X20 X21 X22 TYPE;
run;

proc reg
data=REG2;
model Y2 = X20 X21 X22/noint covb; /* Artificial Regression*/
run;