RIETI Discussion Paper Series 08-E-032

Econometric Analysis of Irreversible Investment with Financial Constraints: Comparison of Parametric and Semiparametric Estimations

ASANO Hirokatsu
Asia University

Research Institute of Economy, Trade & Industry, IAA
The Research Institute of Economy, Trade and Industry
http://www.rieti.go.jp/en/

RIETI Discussion Paper Series 08-E -032

Econometric Analysis of Irreversible Investment with Financial Constraints: Comparison of Parametric and Semiparametric Estimations

May 2008

Hirokatsu Asano
Department of Economics
Asia University
Tokyo, Japan

Abstract: This analysis investigates irreversible investment with financial constraints by parametric and semiparametric estimations. The analysis examines four U.S. industries, employing a sample selection model as it develops its econometric model in accordance with real options theory. The analysis finds that liquidity positively affects capital investment, which is compatible with the theory. In addition, while investment is insensitive to sales revenue and operating costs, capital stock negatively affects investment. The analysis also finds that the sample selection bias is large and that a biased OLS estimator underestimates the coefficients of interest. The analysis' model selection is inconclusive.

JEL Codes: C23, C24, E22, G31, G35
Keywords: Real Options Theory, Sample Selection Models, Two-Step Estimations, Fixed Effects

Econometric Analysis of Irreversible Investment with Financial Constraints: Comparison of Parametric and Semiparametric Estimations

This analysis examines irreversible investment based on real options theory of capital investment. Capital investment is regarded as irreversible if a firm cannot sell its used capital. Thus, by irreversible investment, the firm can adjust its capital stock upward but not downward. As a result, the firm becomes concerned with the possibilities of its level of capital stock becoming too high in an economic recession. In this case, real options theory demonstrates that the firm becomes conservative toward investing (see, for example, Dixit and Pindyck, 1994). A possible econometric model appropriate for real options theory is one of the sample selection models. The analysis estimates the econometric model by semiparametric or distribution-free estimators as well as by parametric estimators.

The analysis focuses on effects of financial constraints on capital investment. When a firm has a promising investment project but its internal funds are insufficient, it seeks external funds. However, if the firm has limited access to external funds due to asymmetric information between the borrowing firm and a lending bank, it faces financial constraints. In Tobin's *q* theory, financially constrained investment shows so-called cash-flow sensitivities, as Fazzari, Hubbert and Peterson (1989) first pointed out. Firms paying low dividends are likely to face financial constraints, and their investment is sensitive to their cash flow. Some textbooks (for example, Romer 2006 and Tirole 2005) now include discussions about the cash-flow sensitivities of financially constrained investment. On the other hand, being based on real options theory, Holt (2003) examines irreversible investment and shows that, for financially constrained firms, investment is sensitive to their liquidity or cash holdings. This present analysis empirically examines the liquidity sensitivities of investment.

Real option theory is an economic application of stochastic dynamic programming. The optimal investment is conditional on the current level of capital stock, and the current investment raises the capital stock's future level. Thus, current investment affects future investment, which means that investments are intertemporarily related. Real options theory incorporates this inter-temporal relationship of investment into its theoretical analyses. When a firm contemplates a new investment project, it will acquire more information about the prospects of the project by waiting. After this waiting period, the firm can make an appropriate decision. Real options theory theorizes the value of waiting, which corresponds to an analogy of financial options. This analysis incorporates the properties of real options theory into its econometric model.

The solution of real options theory is characterized as stationary even though the setup of the theory is dynamic. The solution has a time-invariant function whose arguments include only current variables but no past variables. Therefore, explanatory variables in the analysis contain no lagged variables. This is in contrast with Tobin's *q* models which often show that estimated coefficients for lagged variables are significant. Therefore, lagged variables are indispensable for the q models. It is also well-known that the residuals in the q models show strong and long-lasting autocorrelation. Excluding lagged variables may cause a different dynamics in residuals, so that this analysis examines the autocorrelation of residuals. Asano (2002) estimated a similar investment model by the method of maximum likelihood and showed that the lag length of residuals' autocorrelation was likely to be one year, contrasting to the long-lasting autocorrelation in the q models.

When investment is irreversible, a firm alternately shows positive investment and zero investment. This corresponds to the so-called barrier control of real options theory. In the coordinate of state variables, there is the so-called continuation region whose boundary is called a barrier. When the point presenting the current state is located within the continuation region, control variables remain unchanged. In the case of capital investment, zero investment is optimal in the continuation region. When the point of the current state reaches the barrier the control variables change in such a way as to make the point of the current state move along the barrier. As a result, the optimal investment becomes strictly positive. However, this analysis focuses on positive investment observations, discarding zero investment observations. The data of the analysis is, therefore, not a random sample so that the sample selection is an econometric issue.

In order to correct bias caused by sample selection, the analysis relies on a principle proposed by Heckman (1979). The econometric model proposed by Heckman, which is usually known as the Heckit model, has a two-step method: first, an estimation of a binary choice model and, second, an estimation of a regression model with a correction term. The binary choice model sets up the selection rule that sorts out observations for the second step estimation. Parametric estimators of the binary choice model require a distributional assumption while semiparametric estimators do not. In the analysis, the semiparametric estimator of the binary choice model is the one proposed by Ichimura (1990). Then, for the second step estimation, the parametric estimations can calculate a correction term by the estimates of the binary choice model, in accordance with the distributional assumption. Under the normality assumption, the correction term is equal to the inverse Mills ratio. However, the semiparametric estimators, which do not assume any distributional assumption, need to figure out the functional form of the correction term. The analysis employs two estimators: the one proposed by Newey (1999) and the other proposed by Cosslett (1991). The comparison of parametric and semiparametric estimators may reveal the validity of the distributional assumption.

Abel and Eberly (1996) developed a theoretical model of capital investment based on real options theory. They assumed an iso-elastic demand curve and a Cobb-Douglas production function with stochastic coefficients. In their model, stochastic economic conditions were a product of sales revenue, operating costs and capital stock. Then, Abel and Eberly (1998) investigated another theoretical model based on real options theory. In this model, capacity utilization measured the stochastic economic conditions. Data for capacity utilization, however, are difficult to obtain. In an analysis without capacity utilization data, the capacity utilization turns into an example of omitted variables in estimations, and they are eventually added to a disturbance term in a regression equation. If they are correlated with some explanatory variables, they are called fixed effects and cause endogeneity bias. In order to deal with fixed effects, the analysis employs the procedure proposed by Chamberlain (1987), who took advantage of panel data econometrics.

One advantage of panel data is to increase the sample size by accumulating the data over a period of many years. However, because the analysis investigates capital investment by financially constrained firms, the analysis chooses a short time period. Long-surviving firms are likely to be large and reputable, but unlikely to be financially constrained. Therefore, the analysis chooses two for the time dimension of the panel data. The data are firm-level data from selected industries (NAICS four-digit industry-group level) rather than the entire manufacturing sector because differences in technologies or market conditions may cause different investment behaviors among industries. The selection criterion of industries is the number of member firms in one industry.

The analysis shows that capital investment of examined industries is actually sensitive to liquidity. The sample selection bias is sizable although the analysis sometimes fails to reject the no-selection-bias hypothesis, and a biased ordinary-least-squares estimator underestimates the coefficients of interest. Section 1 describes the econometric models of the sample selection, Section 2 discusses estimation results and section 3 contains the conclusion.

1. Econometric Models

Although a firm shows positive investment and zero investment alternately, the econometric analysis in this paper focuses only on positive investment, discarding zero investment. Due to this, the econometric model is one of the sample selection models. The analysis follows the principle proposed by Heckman (1995). The model is a two-step model which requires an adjustment of the second step standard errors with the first step standard errors. The analysis employs panel data and deals with the fixed effects by using Chamberlain's procedure (1980). The semiparametric estimators in the analysis are Ichimura's semiparametric least squares estimator of the single-index model (1990) for the first step, and Newey's series estimator (1999) and Cosslett's estimator of the dummy variables model (1991) for the second step. The analysis compares its sample selection models by three criteria of model selection: the adjusted R^2, the Akaike Information Criterion and the Bayesian Information Criterion.

A firm invests only when economic conditions are favorable. Or, the firm invests when the following condition holds:

$$z'_{it}\eta + \gamma_i + a_{it} > 0 \qquad (1)$$

where *z* is a vector of explanatory variables, η is a coefficient vector, γ is the fixed effects, and *a* is a zero-mean disturbance term. Subscripts *i* and *t* index firm and time, respectively, with $i \in [1, N]$ and $t \in [1, T]$. The variable vector *z* contains financial data measuring the economic conditions. For dealing with the fixed effects, the analysis relies on Chamberlain's procedure (1980). The procedure assumes the following relation:

$$\gamma_i = \gamma_0 + z'_{i1}\gamma_1 + \cdots + z'_{iT}\gamma_T + b_i \qquad (2)$$

where γ_0 is a constant, γ's are coefficient vectors and *b* is a zero-mean disturbance term. By combining equations (1) and (2), the selection equation of the analysis becomes as follows:

$$z'_{it}\eta + (\gamma_0 + z'_{i1}\gamma_1 + \cdots + z'_{iT}\gamma_T) + v_{it} > 0 \qquad (3)$$

where $v_{it} = a_{it} + b_i$. Chamberlain's procedure was originally developed to deal with the random effects, but Wooldridge (1995) showed that the procedure was also applicable for the fixed effects. Estimating equation (3) yields estimates necessary to a calculate correction term for the second step.

When the firm invests, the amount of investment is a function of financial data affecting investment. The investment function can be written as follows:

$$y_{it} = x'_{it}\beta + \theta_i + c_{it} \qquad (4)$$

where y is the measure of investment, x is another vector of explanatory variables, β is a coefficient vector of interest, θ is the fixed effects, and c is a zero-mean disturbance term. The variable vector x contains financial data which are also contained in variable z, i.e. the variable vector x is a subset of the variable vector z. Similarly to equation (1), the analysis applies Chamberlain's procedure. It assumes the following relation:

$$\theta_i = \theta_0 + x'_{i1}\theta_1 + \cdots + x'_{iT}\theta_T + d_i \tag{5}$$

where θ_0 is a constant, θ's are coefficient vectors and d is a zero-mean disturbance term. By combining equations (4) and (5), the regression equation becomes as follows:

$$y_{it} = x'_{it}\beta + \left(\theta_0 + x'_{i1}\theta_1 + \cdots + x'_{iT}\theta_T\right) + u_{it} \tag{6}$$

where $u_{it} = c_{it} + d_i$. The analysis employs only positive investment observations but discards zero investment observations. Thus, the econometrics model for the analysis is the following sample selection model:

$$\begin{aligned} &y_{it} = x'_{it}\beta + x'_i\theta + u_{it} \qquad \text{if } z'_{it}\eta + z'_i\gamma + v_{it} > 0 \\ &y_{it} \text{ is discarded} \qquad\qquad \text{otherwise} \end{aligned} \tag{7}$$

where $x_i = \left(1 \quad x'_{i1} \quad \cdots \quad x'_{iT}\right)'$, $z_i = \left(1 \quad z'_{i1} \quad \cdots \quad z'_{iT}\right)'$, $\theta = \left(\theta_0 \quad \theta'_1 \quad \cdots \quad \theta'_T\right)'$ and $\gamma = \left(\gamma_0 \quad \gamma'_1 \quad \cdots \quad \gamma'_T\right)'$.

Then, the expected value of y conditional on the selection can be written as follows:

$$E\left[y_{it} \middle| x_i, z'_{it}\eta + z'_i\gamma + v_{it} > 0\right] = x'_{it}\beta + x'_i\theta + E\left[u_{it} \middle| x_i, z'_{it}\eta + z'_i\gamma + v_{it} > 0\right] \tag{8}$$

where E denotes the expected value. Instead of assuming a bivariate normal distribution for the disturbances, u and v, the analysis assumes the following conditional expectation:

$$E[u_{it} | x_i, v_i] = E[u_{it} | v_{it}] = m(v_{it}). \tag{9}$$

Then, the conditional expectation in equation (8) can be written as follows:

$$E[u_{it} | v_{it} > -z_{it}'\eta - z_i'\gamma] = g(z_{it}'\eta + z_i'\gamma). \tag{10}$$

By assuming that the disturbance v is normally distributed and the function m is linear, the function g is equal to the inverse Mills ratio. This is Heckman's two-step estimator, also known as the Heckit estimator. In addition, the analysis estimates equation (10) by assuming a logistic distribution for the disturbance v.

By dropping the distributional assumption on the disturbance v, the analysis resorts to semiparametric estimators. The first step is to estimate the coefficient vectors η and γ by Ichimura's semiparametric least squares (SLS) estimator of the single-index model (1993). The second step is to estimate the functional form of the function g, and this analysis employs two estimators: Newey's series estimator (1999) and Cosslett's estimator of the dummy variables model (1991).

Ichimura's estimator combines the kernel method and the method of nonlinear least squares. Ichimura's weighted semiparametric least squares (WSLS) estimator incorporates the heteroskedasticity of the disturbance term v into estimations. Its weight is equal to the square of the residuals which are obtained by Ichimura's (non-weighted) SLS estimator of the same model. For comparison, the analysis also estimates the selection equation by three parametric methods: the nonlinear least squares (NLSQ) estimator with the normality assumption, and the maximum likelihood estimators of the probit and logit models.

The second-step semiparametric estimations are Newey's series estimator and Cosslett's estimator of the dummy variables model. Newey's estimator approximates the function g by the power series, and Cosslett's approximates the function by a step function. For Newey's estimator, the analysis employs the following approximation (Pagan and Ullah, 1999):

$$\hat{g}\left(z_{it}'\eta + z_i'\gamma\right) \approx \sum_{l=1}^{L} \psi_l \left\{2\Phi\left(z_{it}'\hat{\eta} + z_i'\hat{\gamma}\right) - 1\right\}^l \tag{11}$$

where Φ is the cumulative distribution function of the standard normal distribution and ψ's are coefficients, and the variable L takes values of three and five. Newey's estimator asymptotically converges to a normal distribution. The explanatory variables of Cosslett's estimator include dummy variables which are determined by the value of the function g's argument. The range of the argument is split into several intervals and each dummy variable corresponds to one of the intervals. However, Cosslett's estimator does not converge to a normal distribution asymptotically. As a result, hypothesis testing is problematic and the adjustment of the standard errors is, therefore, unnecessary. For comparison, the analysis estimates equation (8) without the conditional expectation term by the method of ordinary least squares (OLS). This OLS estimator is likely to be biased due to the sample selection.

The analysis employs three criteria of model selection in order to compare its sample selection models: the adjusted R^2, the Akaike Information Criterion and the Bayesian Information Criterion. The BIC penalizes the loss of degree of freedom more heavily than the AIC and tends to choose a simple model.

The data used by the analysis is panel data from four U.S. industry groups: pharmaceutical and medicine manufacturing (NAICS 3254), computer and peripheral equipment manufacturing (NAICS 3341), semiconductor and other electronic component manufacturing (NAICS 3344), and navigational, measuring, electromedical, and control instruments manufacturing (NAICS 3345). The analysis uses the data from these industries because of the number of their member firms. As Table 1 indicates, all four industries contain about one hundred or more firms. The largest firm is about one million times larger than the smallest firm in each industry. Furthermore, the largest firm is five to one hundred times larger than the average firm. The data set of the analysis contains many small firms. These small firms are likely to face financial constraints for investing.

Standard & Poor's Compustat provides financial data for the analysis. The items are sales revenue (*Re*, item 12), operating costs (*Co*, item 41), capital stock (*K*, item 8), liquidity (*F*, item 1 + item 2) and current liabilities (*Li*, item 5). Capital stock is normalized by multiplying the ratio of the real stock to the historical cost of the tangible assets for each industry. The Bureau of Economic Analysis reports the tangible assets data on an annual basis. Other variables except *K* are normalized by the Producer Price Index. The variable *x* contains *Re*, *Co*, *K* and *F*, while the variable *z* contains *Re*, *Co*, *K*, *F* and *Li*. The analysis predicts the positive sign for the variables *Re* and *F*, while predicting the negative sign for the variables *Co*, *K* and *Li*. If Acquisitions (item 129) exceeds five percent of capital stock, *K*, the corresponding data are removed from the data set.

The dependent variable measuring investment is the ratio of the real stock of capital between two consecutive years which is adjusted by the depreciation rate, as the following equation shows:

$$y_{it} = Log\left[\frac{K_{i,t+1}}{K_{it}}\right] + \hat{\delta} \tag{12}$$

where $\hat{\delta}$ is the estimated rate of depreciation. Equation (12) is approximately equal to the ratio of investment to capital stock. The estimated rate of depreciation is the fifteen-year average of the depreciation rate, and the depreciation rate is the ratio of depreciation to real stock of capital for the relevant industry. When y_{it} is below one standard error, the corresponding observation is regarded as zero investment. As Table 2 shows, one quarter to one half of observations are classified as zero investment.

The time dimension of the panel data is two. The analysis chooses the smallest dimension because it focuses on financially constrained investment. When the authors of this paper chose a high dimension such as ten or fifteen years, firms are chosen with at least eight years of data out of a ten-year period or ten years of data out of a fifteen-year period. Consequently, the firms in the analysis were likely to be well-established and unlikely to face financial constraints. On the other hand, variables employed in the analysis are strongly autocorrelated so that data of two consecutive years show little variations. Therefore, the analysis chooses years which are three or four years apart, i.e., 2000 and 2003 or 2000 and 2004.

2. Results

The analysis of this paper finds that liquidity positively affects financially constrained investment. The analysis also detects some sample selection bias. However, estimates are similar between semiparametric estimators and parametric estimators, and the model selection of the analysis is inconclusive. Thus, more research is required for model selection. At the same time, the analysis shows that standard errors of semiparametric estimators are as small as those of parametric estimators even without any distributional assumptions.

Table 3 shows the estimates for the semiparametric and parametric estimators of the selection equation. In this analysis, most of the probit estimates are about sixty percent of the corresponding logit estimates, which is a well-known fact (for example, Greene 2008). The differences between NLSQ estimates and probit estimates, both of which are based on the normality assumption, are small or less than one standard error. In addition, the signs of estimated coefficients are predicted ones. Thus, the parametric estimators of the analysis show reasonable results. The WSLS and SLS estimates are also similar to the estimates of the corresponding parametric models. The residual sum of square is comparable between the NLSQ estimator and SLS estimator for every industry. The WSLS estimator that takes heteroskedasticity into account shows similar estimates but greater standard errors than the SLS estimator. Nonetheless, significant estimates remain significant when switching the SLS estimator to the WSLS estimator. The WSLS estimates are used to calculate the correction term for the second step.

Table 4 shows the estimates for the regression equation. The estimates of semiparametric estimators are similar for each examined industry. Estimated coefficients of the variables *Log K* and *Log F* are significant. In addition, the estimated coefficients for the variable *Log K* are negative and the ones for the variable *Log F* are positive, which are compatible with theory. However, estimated coefficients of the variable *Log Re* and *Log Co* are often insignificant and show wrong signs for some insignificant estimates. The estimators of the sample selection model sometimes fail to reject the hypothesis of no selection bias. The OLS estimator, however, which is likely biased because of the sample selection, always underestimates the coefficients of interest.

For model selection, three criteria fail to find any agreeable model. Only for NAICS 3341, the adjusted R^2, Akaike Information Criterion (AIC), and Bayesian Information Criterion (BIC) agree to conclude that the most favorable model is the Heckit model with the logistic distribution and the least favorable model is Cosslet's dummy variable model. For the other three industries, each criterion concludes differently. The adjusted R^2 concludes that Cosslett's dummy variables model is the most favorable model and the Heckit model with the normal distribution is the least favorable model. AIC chooses Cosslett's model as the most favorable model, while BIC chooses the Heckit model with the normal distribution for NAICS 3254 and 3344m and Newey's model with $L = 5$.

For the pharmaceutical and medicine manufacturing industry (NAICS 3254), the estimated coefficients for the variables *Log K* and *Log F* are significant and their signs are as predicted. The estimated coefficients for the variables *Log Re* and *Log Co*, on the other hand, are insignificant. Thus, investment is sensitive to capital stock and liquidity but insensitive to sales revenue and operating costs. Furthermore, the estimates and their standard errors of two semiparametric estimators are comparable with those of the Heckit estimator. Three estimators of the sample selection model reject the hypothesis of no sample selection bias at the ten-percent significance level. The OLS estimates for the variables *Log K* and *Log F* are less in absolute value than those of the sample selection models, although they are significant. Therefore, the sample selection bias yields underestimations of the coefficients.

For the computer and peripheral equipment manufacturing industry (NIACS 3341), however, Newey's estimator fails to yield significant estimates. Also, it fails to detect the sample selection bias. On the other hand, the Heckit estimator shows some significant estimates. Namely, the estimated coefficients for the variables *Log K* and *Log F* are significant and show the predicted signs. Although the hypothesis of no selection bias is rejected, the OLS estimates are less in absolute value than the Heckit estimates.

For the semiconductor and other electronic component manufacturing industry (NAICS 3344), all three estimators of the sample selection model yield similar estimates and standard errors to each other. The estimated coefficients for the variable *Log K* are negative and significant, while those for the variable *Log F* are positive but insignificant. Although the estimators of the sample selection model fail to reject any selection bias hypothesis, the OLS estimates are less in absolute value than the estimates for the sample selection model.

For the navigational, measuring, electromedical, and control instruments manufacturing industry (NAICS 3345), all three estimators of the sample selection model again show similar estimates and standard errors to each other. They yield significant estimates for the variables *Log K* and *Log F* with the predicted signs. They also reject the no sample bias hypothesis at the five-percent significance level. The OLS estimates are again less in absolute value than the estimates for the sample selection model.

Table 5 shows the estimated coefficients of correlation in residuals. The pharmaceutical and medicine manufacturing industry shows significant estimates. However, the estimated correlation coefficient is less than 0.2, which is weak. The other three industries show insignificant estimates for the correlation coefficient. This demonstrates that autocorrelation in residuals is not problematic in the analysis.

Figure 1 shows curves of four estimated functions for the correction term. Two of them are power functions estimated by Newey's series estimator and another is a step function estimated by Cosslett's dummy variables estimator. The analysis does not estimate the constant term for these two estimators which results in the vertical positions of these curves not being determined. Two other curves are functions of the Heckit models calculated by distributional assumptions: normal and logistic distributions. For all four industries, the curves of the semiparametric models approximately overlap, except the one that is the function of Newey's estimator with the order of five for NAICS 3341. Furthermore, most curves of the semiparametric models show a similar curve regardless of the industry, suggesting that the distribution of the disturbance term is identical for each industry. These three curves of the semiparametric models seem to be closer to the curve assuming the logistic distribution than the normality assumption.

3. Conclusions

This paper investigates irreversible investment with financial constraints by parametric and semiparametric estimations. The analysis in the paper examines four U.S. industries: pharmaceutical and medicine manufacturing, computer and peripheral equipment manufacturing, semiconductor and other electronic component manufacturing and navigational, measuring, electromedical, and control instruments manufacturing. The econometric model is developed in accordance with real options theory so that it is a sample selection model without lagged variables.

The semiparametric estimators of the sample selection model yield similar estimates and standard errors to each other and, often, to the parametric Heckit estimator. The analysis found that liquidity positively affects capital investment, which is compatible with the theory. It also found that capital stock negatively affects investment, while investment is insensitive to sales revenue and operating costs.

The analysis focuses only on positive investment, discarding zero investment. Therefore, the sample selection bias is an econometric issue. The analysis is also concerned with the fixed effects. The econometric model is developed to deal with the sample selection and the fixed effects. The analysis finds that the sample selection bias is large although the no-selection-bias hypothesis is sometimes accepted. The biased OLS estimator always underestimates the coefficients of interest. Moreover, the parametric and semiparametric estimators of the sample selection model yield similar estimates and standard errors. The curves of the correction term by the three semiparametric models seem to be closer to the correction term assuming the logistic distribution than the normality assumption. However, more analyses are required for model selection.

References

Abel, Andrew B. and Eberly, Janice C., "The Mix and Scale of Factors with Irreversibility and Fixed Costs of Investment," *Carnegie-Rochester Conference Series of Public Policy*, June 1998, vol. 48, pp. 101-135.

———, "Optimal Investment with Costly Reversibility," *Review of Economic Studies*, October 1996, vol. 63, no. 4, pp. 581-593.

Arellano, Manuel, "Computing Robust Standard Errors for Within-Groups Estimator," *Oxford Bulletin of Economics and Statistics*, November 1987, vol. 49, no. 4, pp. 431-434.

Asano, Hirokatsu, "Costly Reversible Investment with Fixed Costs: An Empirical Study," *Journal of Business and Economic Statistics*, April 2002, vol. 20, no. 2, pp. 227-240.

Chamberlain, Gary, "Analysis on Covariance with Qualitative Data," *Review of Economic Studies*, January 1980, vol. 47, no. 1, pp. 225-238.

Cosslett, Stephen R., "Semiparametric Estimation of a Regression Model with Sample Selection," in *Nonparametric and Semiparametric Methods in Econometrics and Statistics*, ed. Barnett, A. A., Powell, J. and Tauchen, G. E., 1991, Cambridge University Press, pp. 175-197.

Fazzari, Steven M., Hubbert, Robert G. and Peterson, Bruce C., "Financing Constraints and Corporate Investment," *Brookings Papers on Economic Activity*, 1988, no. 1, pp. 141-195.

Greene, William, H., *Econometric Analysis*, 6th ed. Prentice Hall, 2008.

Heckman, James J., "Sample Selection Bias as a Specification Error," *Econometrica*, January 1979, vol. 47, no. 1, pp. 153-161.

Holt, Richard W.P., "Investment and Dividends under Irreversibility and Financial Constraints," *Journal of Economic Dynamics and Control*, 2003, vol. 27, no. 3, pp. 467-502.

Ichimura, Hidehiko, "Semiparametric Least Squares (SLS) and Weighted SLS Estimation of

Single-Index Models," *Journal of Econometrics*, July 1993, vol. 58, no. 1-2, pp. 71-120.

Newey, Whitney, "Two Step Series Estimation of Sample Selection Models," MIT Working *Paper*, 1999.

Pagen, Adrian and Ullah, Aman, *Nonparametric Econometrics*, Cambridge University Press, 1999.

Romer, David, *Advanced Macroeconomics*, 3rd ed., McGraw-Hill/Irwin, 2006.

Wooldridge, Jeffrey M., "Selection Correction for Panel Data Models under Conditional Mean Independence Assumptions," *Journal of Econometrics*, July 1995, vol. 68, no. 1, pp. 115-132.

Table 1 Data Statistics, by Industry

(a) Pharmaceutical and Medicine Manufacturing (NAICS 3254)

Number of Firms (*N*)		212	
Examined Years (*T* = 2)		2000, 2003	
	Mean	Minimum	Maximum
Sales Revenue (*Re*)	652	0.022	40,363
Operating Costs (*Co*)	230	0.024	21,538
Liquidity (*F*)	380	0.005	16,857

(b) Computer and Peripheral Equipment Manufacturing (NAICS 3341)

Number of Firms (*N*)		76	
Examined Years (*T* = 2)		2000, 2003	
	Mean	Minimum	Maximum
Sales Revenue (*Re*)	5,073	0.333	31,888
Operating Costs (*Co*)	3,445	0.463	25,205
Liquidity (*F*)	1,771	0.244	9,119

(c) Semiconductor and Other Electronic Component Manufacturing (NAICS 3344)

Number of Firms (*N*)		92	
Examined Years (*T* = 2)		2000, 2004	
	Mean	Minimum	Maximum
Sales Revenue (*Re*)	1,206	0.493	33,726
Operating Costs (*Co*)	472	2.043	9,429
Liquidity (*F*)	650	2.402	17,952

(d) Navigational, Measuring, Electromedical, and Control Instruments Manufacturing (NAICS 3345)

Number of Firms (*N*)		120	
Examined Years (*T* = 2)		2000, 2003	
	Mean	Minimum	Maximum
Sales Revenue (*Re*)	279	0.001	16,895
Operating Costs (*Co*)	175	0.035	12,836
Liquidity (*F*)	99	0.008	2,716

Note: Sales revenue, operating costs and liquidity are data from the year 2000 in million $

Table 2 Number of observations

NAICS	3254	3341	3344	3345
Firms investing in both years	124	21	29	47
Firms investing only in first year	39	15	29	21
Firms investing only in second year	35	13	16	26
Firms not investing at all	14	27	18	26

Table 3 (part 1) Estimates for Selection Equation, by Industry

(a) Pharmaceutical and Medicine Manufacturing (NAICS 3254)

	Semiparametric Estimators		Parametric Estimators		
	WSLS	SLS	NLSQ	Probit	Logit
Log *Re*	0.040	0.042	0.034	0.028	0.044
	(0.134)	(0.111)	(0.101)	(0.110)	(0.189)
Log *Co*	-0.514	-0.554	-0.395	-0.439	-0.757
	(0.153)	(0.179)	(0.160)	(0.173)	(0.303)
Log *K*	-0.365	-0.391	-0.286	-0.276	-0.485
	(0.162)	(0.149)	(0.145)	(0.153)	(0.270)
Log *F*	0.796	0.885	0.620	0.656	1.116
	(0.127)	(0.198)	(0.145)	(0.157)	(0.273)
Log *CL*	-0.272	-0.290	-0.218	-0.235	-0.380
	(0.170)	(0.158)	(0.176)	(0.196)	(0.338)
SSR / LL	353.6	57.2	58.1	-178.2	-178.8

(b) Computer and Peripheral Equipment Manufacturing (NAICS 3341)

	Semiparametric Estimators		Parametric Estimators		
	WSLS	SLS	NLSQ	Probit	Logit
Log *Re*	0.419	0.317	0.702	0.117	-0.167
	(0.906)	(0.303)	(0.765)	(0.396)	(0.666)
Log *Co*	-0.350	-0.336	-0.504	-0.392	-0.616
	(1.001)	(0.346)	(0.449)	(0.349)	(0.577)
Log *K*	-0.691	-0.639	-1.155	-0.739	-1.355
	(0.994)	(0.338)	(0.614)	(0.307)	(0.604)
Log *F*	0.863	0.830	1.309	1.045	1.739
	(1.498)	(0.536)	(0.506)	(0.409)	(0.694)
Log *CL*	0.105	0.056	0.270	-0.124	-0.090
	(1.401)	(0.482)	(0.606)	(0.475)	(0.831)
SSR / LL	144.4	25.6	25.1	-77.3	-77.3

Notes: (1) standard errors in parentheses

(2) SSR: Residual Sum of Squares for WSLS, SLS and NLSQ estimators

(3) LL: Log Likelihood for Probit and Logit Models

(4) Some estimates are omitted from the table.

Table 3 (part 2) Estimates for Selection Equation, by Industry

(c) Semiconductor and Other Electronic Component Manufacturing (NAICS 3344)

	Semiparametric Estimators		Parametric Estimators		
	WSLS	SLS	NLSQ	Probit	Logit
Log *Re*	1.066	1.223	0.723	0.827	1.354
	(0.986)	(0.335)	(0.791)	(0.699)	(1.215)
Log *Co*	-0.550	-0.718	-0.302	-0.649	-0.976
	(0.740)	(0.252)	(0.601)	(0.582)	(0.990)
Log *K*	-0.911	-0.937	-0.720	-0.465	-0.822
	(0.551)	(0.189)	(0.377)	(0.329)	(0.561)
Log *F*	1.571	1.600	1.257	0.776	1.374
	(0.792)	(0.258)	(0.445)	(0.342)	(0.599)
Log *CL*	0.257	0.286	0.180	0.172	-0.297
	(0.746)	(0.266)	(0.569)	(0.529)	(0.895)
SSR / LL	176.4	34.1	35.0	-104.2	-103.9

(d) Navigational, Measuring, Electromedical, and Control Instruments Manufacturing (NAICS 3345)

	Semiparametric Estimators		Parametric Estimators		
	WSLS	SLS	NLSQ	Probit	Logit
Log *Re*	-0.150	-0.150	-0.236	0.032	0.062
	(0.307)	(0.152)	(0.272)	(0.209)	(0.349)
Log *Co*	-0.384	-0.384	-0.275	-0.242	-0.430
	(0.264)	(0.161)	(0.283)	(0.248)	(0.414)
Log *K*	-1.869	-1.869	-1.815	-0.662	-1.209
	(0.410)	(0.219)	(0.458)	(0.222)	(0.403)
Log *F*	1.462	1.462	1.146	0.858	1.480
	(0.327)	(0.132)	(0.305)	(0.223)	(0.393)
Log *CL*	0.224	0.224	0.375	-0.075	-0.136
	(0.393)	(0.174)	(0.416)	(0.316)	(0.534)
SSR / LL	229.8	43.1	43.7	-133.1	-132.7

Notes: (1) standard errors in parentheses

(2) SSR: Residual Sum of Squares for WSLS, SLS and NLSQ estimator

(3) LL: Log Likelihood for Probit and Logit Models

(4) Some estimates are omitted from the table.

Table 4 (part 1) Estimates for Regression Equation, by Industry

(a) Pharmaceutical and Medicine Manufacturing (NAICS 3254)

	Sample Selection Model					OLS
	Newey (3)	Newey (5)	Cosslett	Heckit (N)	Heckit (L)	
Log *Re*	0.021	0.026	0.030	0.020	0.019	0.026
	(0.042)	(0.051)	(0.033)	(0.034)	(0.034)	(0.034)
Log *Co*	0.044	0.013	0.011	-0.019	-0.021	0.083
	(0.107)	(0.135)	(0.064)	(0.099)	(0.098)	(0.089)
Log *K*	-0.421	-0.452	-0.451	-0.463	-0.464	-0.392
	(0.076)	(0.089)	(0.056)	(0.064)	(0.064)	(0.056)
Log *F*	0.196	0.244	0.246	0.295	0.296	0.124
	(0.089)	(0.121)	(0.073)	(0.077)	(0.076)	(0.044)
$\bar{R}^2$	0.366	0.364	0.376	0.356	0.360	0.327
AIC	-1.712	-1.704	-1.720	-1.703	-1.709	-1.661
BIC	-1.525	-1.493	-1.497	-1.539	-1.545	-1.509
Pr[CT = 0]	0.275	0.428	0.000	0.026	0.024	N/A

(b) Computer and Peripheral Equipment Manufacturing (NAICS 3341)

	Sample Selection Model					OLS
	Newey (3)	Newey (5)	Cosslett	Heckit (N)	Heckit (L)	
Log *Re*	0.617	0.402	0.625	0.399	0.466	0.195
	(0.744)	(0.728)	(0.296)	(0.314)	(0.307)	(0.368)
Log *Co*	0.018	0.178	-0.006	-0.339	-0.241	0.096
	(0.451)	(0.477)	(0.211)	(0.235)	(0.206)	(0.188)
Log *K*	-0.602	-0.212	-0.742	-1.104	-1.021	-0.545
	(0.836)	(0.784)	(0.160)	(0.222)	(0.184)	(0.134)
Log *F*	0.365	-0.019	0.428	1.108	0.924	0.213
	(0.912)	(1.333)	(0.193)	(0.342)	(0.262)	(0.168)
$\bar{R}^2$	0.531	0.543	0.508	0.526	0.544	0.389
AIC	-2.069	-2.077	-2.002	-2.079	-2.120	-1.836
BIC	-1.555	-1.499	-1.424	-1.630	-1.670	-1.419
Pr[CT = 0]	0.311	0.448	0.005	0.002	0.000	N/A

Notes: (1) standard errors in parentheses
(2) The limiting distribution of Cosslett's dummy variables estimator is not normal.
(3) Some estimates are omitted from the table.
(4) Pr[CT=0]: the *p* value of hypothesis testing with the null that all estimated coefficients of correction terms are equal to zero
(5) Newey (3) and Newey (5): Newey's Series Estimator with L = 3 and 5, Heckit (N) and Heckit (L): Heckman's procedure with normal and logistic distribution assumptions, N/A: not applicable

Table 4 (part 2) Estimates for Regression Equation, by Industry

(c) Semiconductor and Other Electronic Component Manufacturing (NAICS 3344)

	Sample Selection Model					OLS
	Newey (3)	Newey (5)	Cosslett	Heckit (N)	Heckit (L)	
Log *Re*	0.270	0.282	0.342	0.238	0.264	0.143
	(0.190)	(0.201)	(0.184)	(0.154)	(0.155)	(0.130)
Log *Co*	0.094	0.079	0.085	0.079	0.067	0.139
	(0.130)	(0.126)	(0.138)	(0.120)	(0.110)	(0.111)
Log *K*	-0.508	-0.537	-0.488	-0.423	-0.445	-0.366
	(0.174)	(0.173)	(0.124)	(0.132)	(0.133)	(0.134)
Log *F*	0.214	0.221	0.100	0.156	0.181	0.075
	(0.277)	(0.286)	(0.143)	(0.194)	(0.205)	(0.170)
$\bar{R}^2$	0.283	0.276	0.343	0.275	0.280	0.276
AIC	-2.434	-2.408	-2.497	-2.438	-2.445	-2.449
BIC	-2.024	-1.948	-2.011	-2.080	-2.087	-2.116
Pr[CT = 0]	0.770	0.836	0.007	0.245	0.144	N/A

(d) Navigational, Measuring, Electromedical, and Control Instruments Manufacturing (NAICS 3345)

	Sample Selection Model					OLS
	Newey (3)	Newey (5)	Cosslett	Heckit (N)	Heckit (L)	
Log *Re*	-0.321	-0.297	-0.172	-0.382	-0.370	-0.379
	(0.074)	(0.092)	(0.072)	(0.092)	(0.090)	(0.091)
Log *Co*	-0.045	-0.030	-0.005	-0.092	-0.100	-0.021
	(0.087)	(0.119)	(0.071)	(0.098)	(0.096)	(0.092)
Log *K*	-0.522	-0.641	-0.400	-0.433	-0.473	-0.169
	(0.136)	(0.173)	(0.172)	(0.117)	(0.117)	(0.116)
Log *F*	0.172	0.223	0.103	0.242	0.264	-0.042
	(0.116)	(0.154)	(0.125)	(0.115)	(0.109)	(0.071)
$\bar{R}^2$	0.755	0.770	0.782	0.735	0.741	0.722
AIC	-2.297	-2.348	-2.378	-2.234	-2.256	-2.192
BIC	-1.962	-1.972	-1.918	-1.941	-1.963	-1.920
Pr[CT = 0]	0.037	0.002	0.000	0.009	0.003	N/A

Notes: (1) standard errors in parentheses
(2) The limiting distribution of Cosslett's dummy variables estimator is not normal.
(3) Some estimates are omitted from the table.
(4) Pr[CT=0]: the *p* value of hypothesis testing with the null that all estimated coefficients of correction terms are equal to zero
(5) Newey (3) and Newey (5): Newey's Series Estimator with L = 3 and 5, Heckit (N) and Heckit (L): Heckman's procedure with normal and logistic distribution assumptions, N/A: not applicable

Table 5 Estimated Correlation Coefficients of Residuals

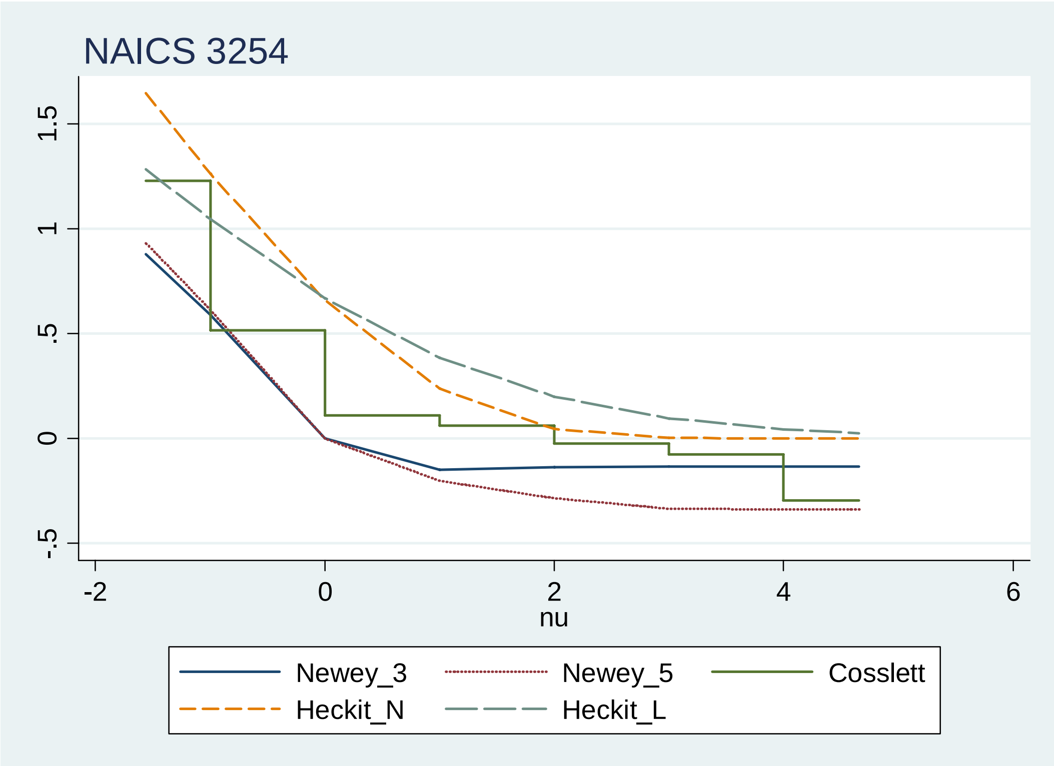
	Newey (3)	Newey (5)	Cosslett	Heckit (N)	Heckit (L)	OLS
NAICS	-0.153	-0.159	-0.137	-0.102	-0.095	-0.095
3254	(0.053)	(0.053)	(0.055)	(0.075)	(0.077)	(0.077)
NAICS	0.018	-0.006	0.019	-0.050	0.094	0.094
3341	(0.082)	(0.075)	(0.083)	(0.154)	(0.135)	(0.135)
NAICS	0.012	-0.015	-0.038	0.040	0.157	0.157
3344	(0.174)	(0.185)	(0.151)	(0.405)	(0.414)	(0.414)
NAICS	-0.067	-0.067	0.020	-0.363	-0.260	-0.260
3345	(0.057)	(0.055)	(0.058)	(0.113)	(0.111)	(0.111)

Notes: (1) Standard errors in parentheses

(2) Newey (3) and Newey (5): Newey's Series Estimator with L = 3 and 5, Heckit (N) and Heckit (L): Heckman's procedure with normal and logistic distribution assumptions

Figure 1 (part 1) Graphical Form of Function *g* by Industry

(a) Pharmaceutical and Medicine Manufacturing



(b) Computer and Peripheral Equipment Manufacturing

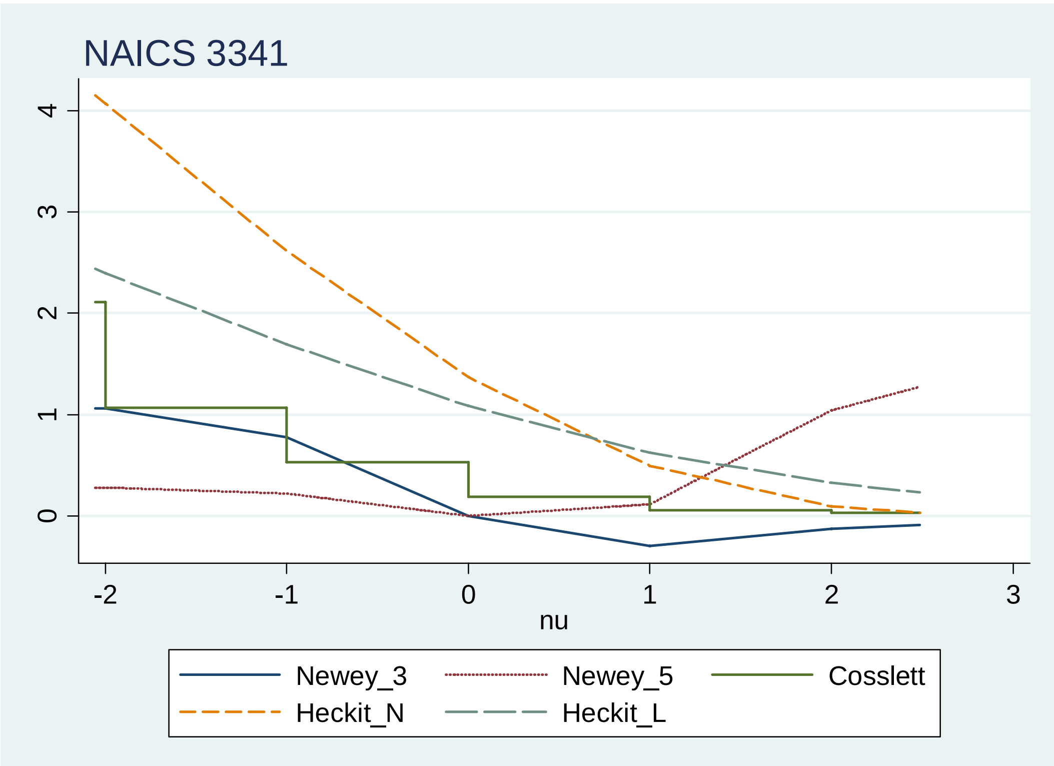
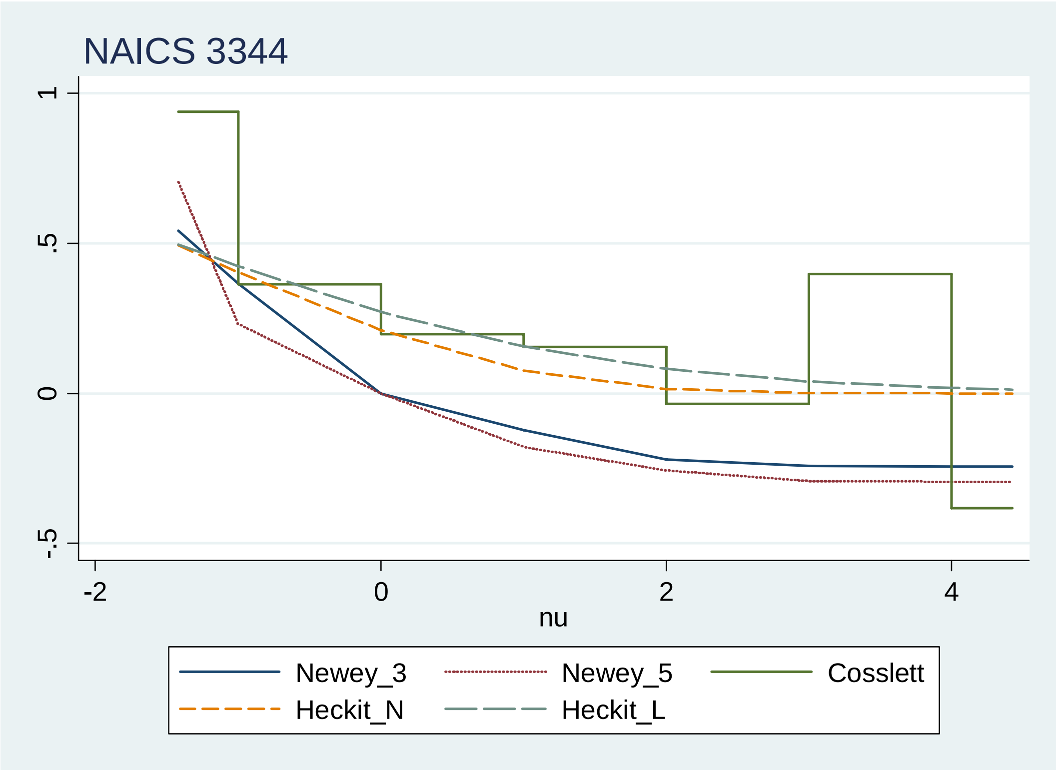


Figure 1 (part 2) Graphical Form of Function *g* by Industry

(c) Semiconductor and Other Electronic Component Manufacturing



(d) Navigational, Measuring, Electromedical, and Control Instruments Manufacturing