Dealing with endogenous treatment conflated with sample selection and measurement error

Alfonso Miranda (alfonso.miranda@cide.edu)
Yu Zhu (y.u.zhu@dundee.ac.uk)

This presentation is based on material published in the online appendix of Miranda, A., Zhu, Yu. (2020). The Effect of Deficiency at English on Female Immigrants' Wage in the UK: correcting for measurement error, endogenous treatment, and sample selection bias. *Applied Economics Letters* **28 (5):349-353.**

Motivation: the female language wage gap example

We investigate how English as Additional Language (EAL) affects the wage gap among foreign-born female immigrants in the UK.

$$
\begin{aligned}
EAL_i^* &= \mathbf{x}_{i,EAL}\boldsymbol{\beta}_{EAL} + u_{i,EAL} && (1)\\
s_i^* &= \mathbf{x}_{i,s}\boldsymbol{\beta}_s + \theta_s EAL_i + u_{i,s} && (2)\\
\log w_i^* &= \mathbf{x}_{i,logw}\boldsymbol{\beta}_{logw} + \theta_{logw} EAL_i + u_{i,logw} && (3)
\end{aligned}
$$

with,

$$
\begin{aligned}
EAL^{**} &= EAL^* + \upsilon && (4)\\
EAL_i &= 1\left(EAL_i^{**} > 0\right) && (5)\\
s_i &= 1\left(s_i^* > 0\right) && (6)\\
\log w_i &= \begin{cases} \log w_i^* \text{ if } s_i = 1 \\ \text{missing otherwise} \end{cases} && (7)
\end{aligned}
$$

θ_{logw} is the parameter of interest. Three potential problems: (i) response variable is subject to sample selection bias, (ii) binary endogenous treatment (EAL); (ii) measurement error in EAL.

Motivation

This econometric model, or "structure", is common to many applications of interest in economics. However, today many applied economists are unsure what must be done to obtain a consistent estimator and give up.

Available consistent estimators

- **Traditional alternative**: estimate joinly (1)-(6) by Maximum likelihood (ML).
  - Involves imposing assumptions about $D(\mathbf{u}|\mathbf{x})$. Multivariate normality $\mathbf{u}|\mathbf{x} \sim \mathcal{N}(\boldsymbol{\mu}, \boldsymbol{\Sigma})$ is a popular choice.
  - Inconsistent if $D(\mathbf{u}|\mathbf{x})$ is misspecified. Departures from join normality are common in applications.
- **Go bayesian**: also imposes restrictive assumptions about $D(\mathbf{u}|\mathbf{x}) \Rightarrow$ inconsistent estimator if the distribution of $\mathbf{u}$ is misspecified.

- **Reize (2001) 2SE** : A control function approach that delivers a two-stage consistent estimator that requires less computer power and is easier to implement... But still imposes $\mathbf{u}|\mathbf{x} \sim \mathcal{N}(\boldsymbol{\mu}, \boldsymbol{\Sigma})$.
- **Wooldridge (2002) 3SE**: A control function approach that delivers a consistent three-stage estimator easy to implement. Wooldridge approach, unlike Raize's, does not requires $\mathbf{u}|\mathbf{x} \sim \mathcal{N}(\boldsymbol{\mu}, \boldsymbol{\Sigma}) \Rightarrow$ easier to relax the normality assumption using semiparametric index models of Gallant and Nychka (1987).

Raize estimator

Uses results from the multivariate truncated normal distribution discussed in detail by Tallis (1961) and Poirer (1980), and indirectly related to the double selection problem of Poirier (1980), De Luca and Perachchi (2007) and Rosenman et. al. (2010). The objective is to get a consistent estimator of equation (3)

$$\log w_i^* = \mathbf{x}_{i,logw}\boldsymbol{\beta}_{logw} + \theta_{logw} EAL_i + u_{i,logw}$$

but matters are complicated by the fact that *logw* is only observed when $s = 1$ and by the endogeneity of the treament,

$$\begin{aligned} E\left(u_{logw}|\mathbf{x}_{logw}, EAL = 0, s = 1\right) &\neq 0 \\ E\left(u_{logw}|\mathbf{x}_{logw}, EAL = 1, s = 1\right) &\neq 0 \end{aligned}$$

because of this, fitting (3) by OLS delivers an inconsistent estimator.

If one could condition on $E\left(u_{logw} | \mathbf{x}_{logw}, s=1\right)$ it is possible to implement a control function approach that eliminates the OLS bias caused the mixture of endogeneity and sample selection (in unobservable heterogeneity).

Under a multivariate normal assumption, Tallis (1961) shows that

$$E\left(u_{logw}|\mathbf{x}_{logw}, EAL=1, s=1\right) = E\left[u_{logw}|\mathbf{x}_{logw}, u_{EAL} > -\mathbf{x}_{EAL}\boldsymbol{\beta}_{EAL}, u_s > -\left(\mathbf{x}_s\boldsymbol{\beta}_s + \theta_s EAL\right)\right]$$

$$= \rho_{logw,s}\left\{\frac{\phi\left(\mathbf{x}_s\boldsymbol{\beta}_s + \theta_s EAL\right)\Phi\left(\frac{\mathbf{x}_{EAL}\boldsymbol{\beta}_{EAL} - \rho_{EAL,s}(\mathbf{x}_s\boldsymbol{\beta}_s + \theta_s EAL)}{\sqrt{1-\rho^2_{EAL,s}}}\right)}{\Phi_2\left(\mathbf{x}_{EAL}\boldsymbol{\beta}_{EAL}, \mathbf{x}_s\boldsymbol{\beta}_s + \theta_s EAL, \rho_{EAL,s}\right)}\right\}$$

$$+ \rho_{logw,EAL}\left\{\frac{\phi\left(\mathbf{x}_{EAL}\boldsymbol{\beta}_{EAL}\right)\Phi\left(\frac{\mathbf{x}_s\boldsymbol{\beta}_s + \theta_s EAL - \rho_{EAL,s}(\mathbf{x}_{EAL}\boldsymbol{\beta}_{EAL})}{\sqrt{1-\rho^2_{EAL,s}}}\right)}{\Phi_2\left(\mathbf{x}_{EAL}\boldsymbol{\beta}_{EAL}, \mathbf{x}_s\boldsymbol{\beta}_s + \theta_s EAL, \rho_{EAL,s}\right)}\right\}$$

And

$$E\left(u_{logw}|\mathbf{x}_{logw}, EAL=0, s=1\right) = E\left[u_{logw}|\mathbf{x}_{logw}, u_{EAL} \leq -\mathbf{x}_{EAL}\boldsymbol{\beta}_{EAL}, u_s > -\left(\mathbf{x}_s\boldsymbol{\beta}_s + \theta_s EAL\right)\right]$$

$$= \rho_{logw,s}\left\{\frac{\phi\left(\mathbf{x}_s\boldsymbol{\beta}_s + \theta_s EAL\right)\Phi\left(\frac{-\mathbf{x}_{EAL}\boldsymbol{\beta}_{EAL}-\rho_{EAL,s}(\mathbf{x}_s\boldsymbol{\beta}_s+\theta_s EAL)}{\sqrt{1-\rho^2_{EAL,s}}}\right)}{\Phi_2\left(-\mathbf{x}_{EAL}\boldsymbol{\beta}_{EAL}, \mathbf{x}_s\boldsymbol{\beta}_s + \theta_s EAL, -\rho_{EAL,s}\right)}\right\}$$

$$- \rho_{logw,EAL}\left\{\frac{\phi\left(-\mathbf{x}_{EAL}\boldsymbol{\beta}_{EAL}\right)\Phi\left(\frac{\mathbf{x}_s\boldsymbol{\beta}_s+\theta_s EAL+\rho_{EAL,s}(\mathbf{x}_{EAL}\boldsymbol{\beta}_{EAL})}{\sqrt{1-\rho^2_{EAL,s}}}\right)}{\Phi_2\left(-\mathbf{x}_{EAL}\boldsymbol{\beta}_{EAL}, \mathbf{x}_s\boldsymbol{\beta}_s + \theta_s EAL, -\rho_{EAL,s}\right)}\right\}$$

Hence, after some rearrangement, Raize's first stage consist in estimating equations (1) and (2) by bivariate probit and calculating the generalized inverse mills rations given by:

$$\widehat{\lambda}_s = EAL \left\{ \frac{\phi\left(\mathbf{x}_s\widehat{\boldsymbol{\beta}}_s + \widehat{\theta}_s EAL\right)\Phi\left(\frac{\mathbf{x}_{EAL}\widehat{\boldsymbol{\beta}}_{EAL} - \widehat{\rho}_{EAL,s}\left(\mathbf{x}_s\widehat{\boldsymbol{\beta}}_s + \widehat{\theta}_s EAL\right)}{\sqrt{1-\widehat{\rho}^2_{EAL,s}}}\right)}{\Phi_2\left(\mathbf{x}_{EAL}\widehat{\boldsymbol{\beta}}_{EAL}, \mathbf{x}_s\widehat{\boldsymbol{\beta}}_s + \widehat{\theta}_s EAL, \widehat{\rho}_{EAL,s}\right)} \right\}$$

$$+ (1 - EAL) \left\{ \frac{\phi\left(\mathbf{x}_s\widehat{\boldsymbol{\beta}}_s + \widehat{\theta}_s EAL\right)\Phi\left(\frac{-\mathbf{x}_{EAL}\widehat{\boldsymbol{\beta}}_{EAL} - \widehat{\rho}_{EAL,s}\left(\mathbf{x}_s\widehat{\boldsymbol{\beta}}_s + \widehat{\theta}_s EAL\right)}{\sqrt{1-\widehat{\rho}^2_{EAL,s}}}\right)}{\Phi_2\left(-\mathbf{x}_{EAL}\widehat{\boldsymbol{\beta}}_{EAL}, \mathbf{x}_s\widehat{\boldsymbol{\beta}}_s + \widehat{\theta}_s EAL, -\rho_{EAL,s}\right)} \right\}$$

And

$$\widehat{\lambda}_{EAL} = EAL \left\{ \frac{\phi\left(\mathbf{x}_{EAL}\widehat{\boldsymbol{\beta}}_{EAL}\right)\Phi\left(\frac{\mathbf{x}_s\widehat{\boldsymbol{\beta}}_s + \widehat{\theta}_s EAL - \widehat{\rho}_{EAL,s}\left(\mathbf{x}_{EAL}\widehat{\boldsymbol{\beta}}_{EAL}\right)}{\sqrt{1-\widehat{\rho}^2_{EAL,s}}}\right)}{\Phi_2\left(\mathbf{x}_{EAL}\widehat{\boldsymbol{\beta}}_{EAL}, \mathbf{x}_s\widehat{\boldsymbol{\beta}}_s + \widehat{\theta}_s EAL, \widehat{\rho}_{EAL,s}\right)} \right\}$$

$$-(1-EAL)\left\{ \frac{\phi\left(-\mathbf{x}_{EAL}\widehat{\boldsymbol{\beta}}_{EAL}\right)\Phi\left(\frac{\mathbf{x}_s\widehat{\boldsymbol{\beta}}_s + \widehat{\theta}_s EAL + \widehat{\rho}_{EAL,s}\left(\mathbf{x}_{EAL}\widehat{\boldsymbol{\beta}}_{EAL}\right)}{\sqrt{1-\widehat{\rho}^2_{EAL,s}}}\right)}{\Phi_2\left(-\mathbf{x}_{EAL}\widehat{\boldsymbol{\beta}}_{EAL}, \mathbf{x}_s\widehat{\boldsymbol{\beta}}_s + \widehat{\theta}_s EAL, -\widehat{\rho}_{EAL,s}\right)} \right\}$$

The sencond step involves estimating

$$\log w^* = \mathbf{x}_{logw}\boldsymbol{\beta}_{logw} + \theta_{logw} EAL + \tau_1\widehat{\lambda}_s + \tau_2\widehat{\lambda}_{EAL} + \epsilon_{logw}$$

by OLS on the $s = 1$ sample. SEs can be obtained using the bootstrap to take into account 1st step parameter variation.

Notice that

$$E\left(u_{logw} | \mathbf{x}_{logw}, s=1\right) = \tau_1 \lambda_s + \tau_2 \lambda_{EAL}.$$

As a consequence,

$$\epsilon_{logw} = u_{logw} - E\left(u_{logw} | \mathbf{x}_{logw}, s=1\right) = \tau_1 \lambda_s + \tau_2 \lambda_{EAL}.$$

So that

$$E\left(\epsilon_{logw} | \mathbf{x}_{logw}, EAL, s=1\right) = 0$$

as needed. The instrument that deals with endogeneity of EAL is also likely to deal with the problem of measurement error; so, Raize's estimator is likely to deal with all three problems we need to address.

Wooldridge estimator

- In the context of a model with a continuous response subject to sample selection bias and a continuous endogenous explanatory variable, Wooldridge (2002) recommends using a two-step Heckman sample selection approach to correct for selection bias, while explicitly addressing the problems caused by an endogenous explanatory variable in a second step.
- He recommends fitting the second step of a Heckman model by 2SLS (Wooldridge 2002, p. 567).
- This is a control function approach that delivers a consistent estimator.
- The method is more general than it appears at first, it works regardless of the nature of the endogenous variable with minor modifications.

- In the present context, there are two extra complications: (1) the endogenous variable is not continuous, rather a endogenous binary treatment; (2) the endogenous treatment enters the sample selection model.
- Our proposed solution: fit the second stage of the Heckman model by 2SLS instrumenting EAL with fitted EAL probability from a 1st stage OLS of EAL on controls to avoid the "forbidden regression problem" (Wooldridge 2010, p. 265-268.).

This suggest a three-step estimator

- First stage: Fit a OLS regression of *EAL* on $\mathbf{z}$ (all instruments in the system).
- Second stage: fit a reduced form selection equation

$$\begin{aligned} s^* &= \mathbf{z}\boldsymbol{\pi}_s + \varepsilon_s & (8) \\ s &= 1\left(s^{**} > 0\right) & (9) \end{aligned}$$

  by probit and get the usual inverse mills ratio $\widehat{\lambda} = \phi\left(\cdot\right)/\Phi\left(\cdot\right)$.
- Third stage: Fit

$$\log w^* = \mathbf{x}_{logw}\boldsymbol{\beta}_{logw} + \theta_{logw}EAL + \delta\widehat{\lambda} + \epsilon_{logw}$$

  by 2SLS using $\mathbf{z}$, fitted EAL from the 1st step, and $\widehat{\lambda}$ from the 2nd stage as instruments. SEs can be obtained using the bootstrap to take into account 1st and 2nd step parameter variation.

Notice that by taking a ‘reduced form’ approach in the estimation of the selection mechanism, join modelling of EAL and s is not needed because $\epsilon_{logw} = u_{logw} - E\left(u_{logw}|\mathbf{z}, s = 1\right)$. Thus $E\left(\epsilon_{logw}|\mathbf{z}, s = 1\right) = 0$ by construction and adding $\widehat{\lambda}$ deals with sample selection bias. Moreover, because $\widehat{\lambda}$ is only a function of the instruments $\mathbf{z}$ we have that $\widehat{\lambda}$ is exogenous. So, after controlling for $\widehat{\lambda}$ the only challenge that remains is the fact that ϵ_{logw} and EAL are still partially correlated even after controlling for $\mathbf{z}$. This is why we need to fit the 3rd step by 2SLS to get a consistent estimator.

In Wooldridge's approach, unlike Raize's, modelling of the selection mechanism does not need to be 'structural' but rather a 'reduced form' that simply projects s into the space spanned by $\mathbf{z}$. This approach makes it easier relaxing the normality assumption using a semiparametric index model (Gallant and Nychka 1987) in the 1st and 2nd steps, and then add powers of the EAL and selection indexes as instruments in the 2SLS fitted in the 3rd step to implement a flexible control function.

Monte Carlo simulation study

$r = 1, \ldots, 10000$ simulated data sets with sample size of $1,000$. Let y the main continuous response, by *treat* the endogenous treatment, and s the sample selection dummy. At each replication two independent standard normal variables (x_1, x_2) and two Bernoulli variates (d_1, d_2) with $p = 0.5$ are simulated to play the role of explanatory variables.

- x_1, x_2, d_1, d_2 enter all treatment, selection, and main response equations.
- Three independent standard normal variables *zyvar*, *ztreat*, and *zsel* are generated at each replication to play the role of instruments.
- Error terms u^r_y, u^r_{treat}, u^r_s are drawn from a multivariate normal distribution with $sd(u_y) = 0.7$, $sd(u_{treat}) = sd(u_s) = 1$ and correlations $\mathrm{Cor}(u_{treat}, u_s) = Cor(u_y, u_{treat}) = -0.2$ and $Cor(u_y, u_s) = 0.8$.
- **Contrasted estimators**: OLS, 2SLS (ignoring selection), Naïve two-stage (1st stage probit for EAL, 2nd stage Heckman controlling for $\widehat{p}(EAL)$), Reize 2SE, and Wooldridge 3SE.
- Standard errors are bootstrapped 50 times in each replication.

Table TAC.1. Monte Carlo simulation study - estimated bias and standard deviation of points estimates for coefficients in the equation for y_i

Coefficient	True Value	25% missing		50% missing		75% missing	
		Bias	SD	Bias	SD	Bias	SD
Ordinary Least Squares							
treat	1.00	-0.199	0.059	-0.228	0.072	-0.249	0.107
x1	1.00	-0.004	0.029	0.000	0.035	0.002	0.049
x2	-1.00	0.004	0.028	0.000	0.035	-0.002	0.049
d1	1.00	-0.003	0.056	0.001	0.068	0.003	0.095
d2	-1.00	0.005	0.056	0.001	0.068	0.000	0.095
zyvar	1.00	0.000	0.028	0.000	0.034	0.000	0.048
IV no selection							
treat	1.00	-0.083	0.088	-0.114	0.110	-0.127	0.165
x1	1.00	0.007	0.029	0.011	0.036	0.013	0.051
x2	-1.00	-0.008	0.029	-0.011	0.036	-0.013	0.050
d1	1.00	0.008	0.057	0.011	0.069	0.014	0.096
d2	-1.00	-0.006	0.056	-0.010	0.069	-0.011	0.097
zyvar	1.00	0.000	0.028	0.001	0.034	0.000	0.048
Naive 2SE							
treat	1.00	-0.157	0.078	-0.265	0.094	-0.373	0.132
x1	1.00	-0.016	0.030	-0.027	0.036	-0.037	0.050
x2	-1.00	0.016	0.030	0.027	0.036	0.037	0.050
d1	1.00	-0.016	0.057	-0.026	0.070	-0.036	0.096
d2	-1.00	0.017	0.057	0.028	0.070	0.039	0.096
zyvar	1.00	0.000	0.028	0.000	0.033	0.000	0.044

Note: Statistics calculated over 10,000 Monte Carlo replications with sample size of 1,000 Standard errors bootstrapped 50 times in each Monte Carlo replication. Mean Probability of treatment is 0.5 in all cases. Simulated error terms of equations (A1)-(A3) are multivariate normal with mean vector zero and $sd(u_y)$=0.79, $sd(u_{treat})=sd(u_s)=1$, $Cor(u_{treat},u_s)=Cor(u_y u_{treat})$=-0.2 and $Cor(u_y,u_s)=0.8$. Noise/signal ratio is 0.25 in main response and treatment equations and 0.3 in the selection equations. True parameters in the treatment equation are: x1=-0.58, x2=0.58, d1=-0.58, d2=0.58, ztreat=1.8. True parameters in the selection equation are: treat=1.2, x1=-0.12, x2=0.12, d1=-0.12, d2=0.12, zsel=-1.75.

Table TAC.1 (Cont.)

Coefficient	True Value	25% missing		50% missing		75% missing	
		Bias	SD	Bias	SD	Bias	SD
Reize 2SE							
treat	1.00	-0.002	0.082	-0.008	0.098	-0.034	0.135
x1	1.00	-0.001	0.029	-0.001	0.036	-0.003	0.049
x2	-1.00	0.001	0.029	0.001	0.035	0.002	0.048
d1	1.00	0.000	0.057	-0.001	0.068	-0.001	0.094
d2	-1.00	0.002	0.057	0.002	0.068	0.004	0.094
zyvar	1.00	0.000	0.027	0.000	0.032	0.000	0.043
Wooldridge 3SE							
treat	1.00	-0.006	0.088	-0.001	0.108	0.016	0.159
x1	1.00	0.000	0.030	0.000	0.036	-0.001	0.050
x2	-1.00	0.000	0.030	0.000	0.036	0.000	0.050
d1	1.00	0.000	0.057	0.000	0.069	0.001	0.096
d2	-1.00	0.001	0.057	0.001	0.069	0.001	0.095
zyvar	1.00	0.000	0.027	0.000	0.032	0.000	0.044

Note: Statistics calculated over 10,000 Monte Carlo replications with sample size of 1,000 Standard errors bootstrapped 50 times in each Monte Carlo replication. Mean Probability of treatment is 0.5 in all cases. Simulated error terms of equations (A1)-(A3) are multivariate normal with mean vector zero and $sd(u_y)$=0.79, $sd(u_{treat})=sd(u_s)=1$, $Cor(u_{treat}, u_s)=Cor(u_y u_{treat})$=-0.2 and $Cor(u_y, u_s)=0.8$. Noise/signal ratio is 0.25 in main response and treatment equations and 0.3 in the selection equations. True parameters in the treatment equation are: x1=-0.58, x2=0.58, d1=-0.58, d2=0.58, ztreat=1.8. True parameters in the selection equation are: treat=1.2, x1=-0.12, x2=0.12, d1=-0.12, d2=0.12, zsel=-1.75.

Table TAB.2: Monte Carlo simulation study – average standard error divided by standard deviation of estimates (ASE/SD) and coverage of estimated 95% confidence intervals

Coefficient	25% missing		50% missing		75% missing	
	ASE/SD	Covrg (%)	ASE/SD	Covrg (%)	ASE/SD	Covrg (%)
Ordinary Least Squares						
treat	1.00	8	1.01	12	1.00	35
x1	1.00	94	0.99	95	0.99	95
x2	1.00	95	1.00	95	1.00	95
d1	1.00	95	1.00	95	1.00	95
d2	1.00	95	0.99	95	1.00	95
zyvar	1.00	95	1.00	95	0.99	95
IV no selection						
treat	1.00	84	1.00	82	0.99	87
x1	0.99	94	0.98	94	0.98	94
x2	1.00	94	0.99	94	0.98	94
d1	1.00	95	0.99	94	0.99	95
d2	1.00	95	0.99	95	0.99	95
zyvar	0.99	95	1.00	95	0.98	94
Naive 2SE						
treat	1.00	48	1.00	20	1.00	20
x1	0.99	91	0.98	87	0.99	87
x2	1.00	91	1.00	88	1.00	88
d1	0.99	94	0.99	92	0.99	93
d2	0.99	94	0.99	92	0.99	92
zyvar	0.99	94	0.99	94	0.99	94

Note. Statistics calculated over 10,000 Monte Carlo replications with sample size of 1,000. Standard errors bootstrapped 50 times in each Monte Carlo replication.

Table TAC.2 (Cont.)

Coefficient	25% missing		50% missing		75% missing	
	ASE/SD	Covrg (%)	ASE/SD	Covrg (%)	ASE/SD	Covrg (%)
Reize 2SE						
treat	1.00	95	1.00	94	1.01	94
x1	0.99	94	0.98	94	0.99	94
x2	0.99	94	1.00	94	1.00	94
d1	0.99	94	0.99	94	1.00	94
d2	1.00	94	0.99	94	1.00	94
zyvar	1.00	94	1.00	94	0.99	94
Wooldridge 3SE						
treat	1.00	94	1.00	94	1.01	95
x1	0.99	94	0.98	94	0.99	94
x2	0.99	94	1.00	94	1.00	94
d1	1.00	94	0.99	94	1.00	94
d2	1.00	94	0.99	94	1.00	94
zyvar	1.00	94	1.00	94	1.00	94

Note. Statistics calculated over 10,000 Monte Carlo replications with sample size of 1,000. Standard errors bootstrapped 50 times in each Monte Carlo replication.

The end, many thanks!!!

References

- De Luca, G., Peracchi, F. (2007). A sample selection model for unit and item nonresponse in cross-sectional surveys. CEIS Working Paper No. 99.
- Gallant, A.R. and Nychka, D.W. (1987) Semi-nonparametric maximum likelihood estimation, Econometrica, 55, 363–390.
- Miranda, A., Zhu, Yu. (2020). The Effect of Deficiency at English on Female Immigrants' Wage in the UK: correcting for measurement error, endogenous treatment, and sample selection bias. *Applied Economics Letters* **28** (5):349-353.
- Poirier, D.J. (1980). Partial observability in bivariate probit models, Journal of Econometrics 12, 209-217.
- Reize, F. (2001). FIML estimation of a Bivariate Probit Selection Rule – An application on firm growth and subsidization. ZEW (Centre for European Economic Research) Discussion Paper No. 01-13.
- Rosenman, R., Mandal, B., Tennekoon, V., Hill, L.G. (2010). Estimating treatment effectiveness with sample selection. School of Economic Sciences, Washington State University, WP2010-5.

- Tallis, G.M. (1961) The moment generating function of the truncated multi-normal distribution, Journal of the Royal Statistical Society Series B 23, 223-229.
- Wooldridge, J. (2002). Econometric Analysis of Cross Section and Panel Data. Cambridge: MIT Press, Massachusetts.
- Wooldridge, J. (2010). Econometric Analysis of Cross Section and Panel Data (2ed). Cambridge: MIT Press, Massachusetts.