Selection bias on the dependent variable and an endogenous covariate in a non-linear framework: a corrected inverse Mills ratio approach

Esther Devilliers, A. Carpentier

Selection bias on the dependent variable and an endogenous covariate in a non-linear framework: a corrected inverse Mills ratio approach

Esther Devilliers, Alain Carpentier

August 28, 2023

1 Introduction

Heckman (1976, 1979) selection model and its endogenous regime switching extension as proposed by Lee (1978) are widely used in econometrics. First used to investigate wages and labor supply functions (*e.g.*, Heckman, 1978; Lee, 1978), the approach was then introduced to social sciences Berk (1983) and to the agricultural production literature Pitt (1983). The original framework benefited from several extensions: allowing for heteroskedasticity (e.g., Donald, 1995; Shively, 1998), considering panel data (Kyriazidou, 1997), considering a multinomial – instead of binary – selection variable (Di Falco and Veronesi, 2013; Wu and Babcock, 1998), introducing a double selectivity model with sequential adoption (Khanna, 2001), *etc.* More recently, extensions of the endogenous regime switching model to the case of endogenous covariates were suggested (*e.g.*, Murtazashvili and Wooldridge, 2016; Schwiebert, 2015; Takeshima and Winter-Nelson, 2012). While allowing for the endogenous covariates to be correlated with the selection model, these extensions do not consider the case where the selection model affects both the response variable and the endogenous covariates. However, one could argue that the selection process might not only affect the interest variable but also the endogenous covariate. For instance, in agricultural economics, potential yield is often consider as being part of a technology adoption decision (*e.g.*, Kumbhakar et al., 2009). When considering a primal production function, not only input use levels are endogenous covariates in the production function but the level of potential input savings, because it affects the economic return of the technology, can also be consider as being part of a technology adoption decision. Hence, technology selection bias affects both the production function and the input demand equations.

We aim at contributing to the literature by allowing the selection process to affect both the response variable and its endogenous covariates. Additionally, we consider an ERS framework with a non-linear response function making both the simultaneous estimation "full information" and the instrumental variable "limited information" approaches hardly tractable. We use a control function approach (see Wooldridge, 2015) relying on two sets of control functions. The first one controls for the endogenous sample selection issues implied by ERS models whereas the second one controls for input use endogeneity. Our estimation approach can be considered as an extension of the widely used two-step approach that was initially proposed by Heckman (1976, 1979) to account for endogenous sample selection and later adapted by Lee (1978) to the case of Gaussian ERS models. In particular, we show that the expression of the so-called inverse Mills ratio used in Heckman's two step approach for estimating regression models under

endogenous sample selection needs to be adapted accordingly. The remainder of the article is structured as follows. First, we present the endogenous regime switching framework in the case of endogenous covariates. Second, we present our multistep estimation approach.

2 An endogenous regime switching framework with endogenous covariates

The standard endogenous regime switching (ERS) model can be defined with the following equations:

$$\begin{aligned} y^r &= s^r(\mathbf{x}; \boldsymbol{\beta}^r) + v^r \\ y &= ry^1 + (1-r)y^0 \\ r &= 1\{m(\mathbf{z}; \boldsymbol{\gamma}) + e \geq 0\} \end{aligned} \quad , \tag{1}$$

where $r \in \{0, 1\}$ represents the regimen variable, y is the latent interest variable and (y^0, y^1) its observed counterpart. $s^r(\cdot)$ and $m(\cdot)$ denote functions that are known to the analyst. Standard assumption for error term e is that it has zero means and follows either a logistic or a normal distribution (see, *e.g.*, Greene, 2020). In the standard ERS framework, $\mathbf{x}$ and $\mathbf{z}$ are assumed to be exogenous while v^r and e are correlated. In particular, from e and v^r correlation derives the fact that $E[v^r|\mathbf{x}, \mathbf{z}, r] \neq 0$, even if exogeneity condition $E[v^r|\mathbf{x}] = 0$ holds. We can write the conditional expectation of y on $\mathbf{x}$ and r as:

$$E[y^r|\mathbf{x}, \mathbf{z}, r = r] = s^r(\mathbf{x}; \boldsymbol{\beta}^r) + E[v^r|\mathbf{z}, r], \tag{2}$$

where $E[v^r|\mathbf{z}, r]$ is a non-trivial function of $\mathbf{z}$ that generates an estimation bias when it is ignored.

Schwiebert (2015) and Murtazashvili and Wooldridge (2016) relax the exogeneity condition $E[v^r|\mathbf{x}] = 0$ and assume that among $\mathbf{x}$ are endogenous covariates correlated with the selection variable r. Let denote $\mathbf{x} = (\mathbf{x}_1, \mathbf{x}_2)$ the vector of covariates where $\mathbf{x}_1$ and $\mathbf{x}_2$ represent respectively the exogenous and endogenous covariates. Assuming the following model for $\mathbf{x}_2$:

$$x_{2,k} = d(\mathbf{w}; \alpha_k) + u_k,$$

where $\mathbf{x}_2 = (x_{2,k} : k \in K)$, we have $E[u_k|r] \neq 0$. Instead of considering that $\mathbf{x}_2$ and r are "simply"

correlated, we assume here that the endogenous selection process not only affects y but also $\mathbf{x}_2$. Thus, in addition to the ERS model defined in Equation (1), we have:

$$\begin{aligned} x_{2,k}^r &= d^r(\mathbf{w}; \alpha_k^r) + u_k^r \\ x_{2,k} &= r x_{2,k}^1 + (1-r) x_{2,k}^0 \end{aligned} . \tag{3}$$

While $\mathbf{x}_2$ endogeneity invalidates the conditional expectation of y on $(\mathbf{x}, r)$, we can rewrite Equation (2) for $\mathbf{x}_2$:

$$E[x_{2,k}^r | \mathbf{w}, r = r] = d^r(\mathbf{w}; \alpha_k^r) + E[u_k^r | r], \tag{4}$$

where $E[u_k^r|r]$ is a non-trivial function of $\mathbf{z}$ that generates an estimation bias when it is ignored.

As in a standard ERS framework, we assume that variables $\mathbf{z}$ are exogenous with respect to error terms v^r and $\mathbf{u}^r$. From that we derive that $\mathbf{x}_2^r$ and v^r are correlated conditional on $\mathbf{z}$ only if error terms v^r and $\mathbf{u}^r$ are correlated. We assume that error terms vector $(v^r, \mathbf{u}^r, e)$ is jointly normal and independent of control and instrumental variables $\mathbf{z}$[1], with:

$$(\mathbf{u}^r, v^r, e)|\mathbf{z} \sim \mathcal{N}(\mathbf{0}, \mathbf{\Omega}^r) \text{ with } \mathbf{\Omega}^r = \begin{bmatrix} \mathbf{\Omega}_{uu}^r & \mathbf{\Omega}_{uv}^r & \mathbf{\Omega}_{ue}^r \\ (\mathbf{\Omega}_{uv}^r)' & \omega_{vv}^r & \omega_{ve}^r \\ (\mathbf{\Omega}_{ue}^r) & \omega_{ve}^r & 1 \end{bmatrix} \text{ for } r \in \{0, 1\}, \tag{5}$$

where

$$\mathbf{\Omega}_{uv}^r = (\omega_{k,uv}^r : k \in K),\ \mathbf{\Omega}_{ue}^r = (\omega_{k,ue}^r : k \in K) \text{ and } \mathbf{\Omega}_{uu}^r = [\omega_{k\ell,uu}^r : (k, \ell) \in K \times K].$$

3 A multistep estimation procedure

We consider a fully parametric endogenous regime switching model. We assume that $m(\cdot)$, the model for regimen variable r, is a Probit link function and that $d^r(\cdot)$, the model for endogenous covariates $\mathbf{x}_2$, is a linear function. Finally, we suppose that s^r, the model for y^r, is non-linear in its parameters β^r. Both the simultaneous estimation with "full information" maximum likelihood and the instrumental variable "limited information" least squares approaches considered in Schwiebert (2015) and Murtazashvili and Wooldridge (2016) are hardly tractable in a non-linear framework. Our estimation approach relies on the

[1]The joint normality is usually imposed in an ERS framework. In our case, because of the well-known properties of multivariate normal variables, this assumption is convenient to deal with the multiple endogeneity issues we face.

distributional assumptions given in Equation (5). It consists in a sequence of estimation problems that are easy to solve, *i.e.* Probit model and least squares estimation problems. It relies on two sets of control functions. The first one is used to deal with the input use endogeneity issue in the production function while the second one is used to deal with the sample selection issues due to the production practice choice.

As for Heckman two-step estimation procedure, first step consists in estimating the Probit model of $r|\mathbf{z}$ by maximum likelihood to obtain consistent estimates of parameters γ. This estimates can then be used for obtaining consistent estimates of Mills ratio terms $\lambda^r\left(m(\mathbf{z},\gamma)\right)$. Second step requires Heckman's standard result Heckman (1976, 1979):

$$E[u_k^r|\mathbf{z}, r=r] = \omega_{k,ue}^r \lambda^r(m(\mathbf{z},\gamma)),$$

to consistently estimate using standard linear least squares the model for $\mathbf{x}_2$ with:

$$x_{2,k}^r = d^r(\mathbf{w};\alpha_k) + \omega_{k,ue}^r \lambda^r(m(\mathbf{z},\gamma)) + \eta_k^r, \tag{6}$$

where $E[\eta_k^r|\mathbf{z}, r=r] = 0$. Term $\omega_{k,ue}^r \lambda^r(m(\mathbf{z},\gamma))$ defines a control function for endogenous selection of the observation characterized by $r=r$ in the considered sub-sample. Error term η_k^r is defined by $\eta_k^r = u_k^r - \omega_{k,ue}^r \lambda^r(m(\mathbf{z},\gamma))$, which implies that exogeneity condition $E[\eta_k^r|\mathbf{z}, r=r] = 0$ necessarily holds (in the sub-sample characterized by $r=r$).

Because of $\mathbf{x}_2$ endogeneity, standard Heckman's result $E[v^r|\mathbf{x},\mathbf{z},r=r] = \omega_{k,ve}^r \lambda^r(m(\mathbf{z},\gamma))$ does not hold for y. First, we can rewrite $E[v^r|\mathbf{x},\mathbf{z},r=r]$ as $E[v^r|\mathbf{u}^r,\mathbf{z},r=r]$. The joint normality distribution of error terms vector $(\mathbf{u}^r, v^r, e)$ yields

$$(v^r, e)|(\mathbf{z},\mathbf{u}^r) \sim \mathcal{N}(\mathbf{R}^r\mathbf{u}^r, \mathbf{\Psi}^r) \tag{7}$$

where

$$\mathbf{R}^r = \begin{bmatrix} (\boldsymbol{\rho}_{uv}^r)' \\ (\boldsymbol{\rho}_{ue}^r)' \end{bmatrix} = \begin{bmatrix} (\boldsymbol{\omega}_{uv}^r)'(\mathbf{\Omega}_{uu}^r)^{-1} \\ (\boldsymbol{\omega}_{ue}^r)'(\mathbf{\Omega}_{uu}^r)^{-1} \end{bmatrix}$$

$$\mathbf{\Psi}^r = \begin{bmatrix} \psi_{vv}^r & \psi_{ve}^r \\ \psi_{ve}^r & \psi_{ee}^r \end{bmatrix} = \begin{bmatrix} \omega_{vv}^r - (\boldsymbol{\omega}_{uv}^r)'(\mathbf{\Omega}_{uu}^r)^{-1}\boldsymbol{\omega}_{uv}^r & \omega_{ve}^r - (\boldsymbol{\omega}_{uv}^r)'(\mathbf{\Omega}_{uu}^r)^{-1}\boldsymbol{\omega}_{ue}^r \\ \omega_{ve}^r - (\boldsymbol{\omega}_{uv}^r)'(\mathbf{\Omega}_{uu}^r)^{-1}\boldsymbol{\omega}_{ue}^r & 1 - (\boldsymbol{\omega}_{ue}^r)'(\mathbf{\Omega}_{uu}^r)^{-1}\boldsymbol{\omega}_{ue}^r \end{bmatrix}.$$

These results imply that terms v^r and e can be decomposed as the following:

$$\begin{aligned} v^r &= (\boldsymbol{\rho}^r_{uv})'\mathbf{u}^r + \varepsilon^r_v \\ e &= (\boldsymbol{\rho}^r_{ue})'\mathbf{u}^r + \varepsilon^r_e \end{aligned},$$

where $(\varepsilon^r_v, \varepsilon^r_e)|(\mathbf{z}, \mathbf{u}^r) \sim \mathcal{N}(\mathbf{0}, \boldsymbol{\Psi}^r)$. This also implies that:

$$E[v^r|\mathbf{z}, \mathbf{u}^r, r = r] = (\boldsymbol{\rho}^r_{vu})'\mathbf{u}^r + E[\varepsilon^r_v|\mathbf{z}, \mathbf{u}^r, r = r].$$

Observing that $r = 1[m(\mathbf{z}, \gamma) + (\boldsymbol{\rho}^r_{ue})'\mathbf{u}^r + \varepsilon^r_{e=} \geq 0]$, it suffices to apply standard results on the means of truncated normal variables for obtaining:

$$E[\varepsilon^r_v|\mathbf{z}, \mathbf{u}^r, r = r] = \psi^r_{ve}(\psi^r_{ee})^{-1/2}\lambda^r\left((\psi^r_{ee})^{-1/2}(m(\mathbf{z}, \gamma) + (\boldsymbol{\rho}^r_{eu})'\mathbf{u}^r)\right) = \psi^r_{ve}\lambda^r_y,$$

where $(\psi^r_{ee})^{-1/2}$ is a scale parameter to account for $\mathbf{u}^r$ being an argument of λ^r. This corrected version of the standard inverse Mills ratio account for the double selection process, *i.e.* on $\mathbf{x}_2$ and y. From this result we can derive that:

$$E[y|\mathbf{x}, \mathbf{z}, \mathbf{u}^r, r = r] = s^r(\mathbf{x}; \boldsymbol{\beta}^r) + (\boldsymbol{\rho}^r_{vu})'\mathbf{u}^r + \psi^r_{ve}\lambda^r_y,$$

and

$$y = s^r(\mathbf{x}; \boldsymbol{\beta}^r) + (\boldsymbol{\rho}^r_{vu})'\mathbf{u}^r + \psi^r_{ve}\lambda^r_y + \mu^r \tag{8}$$

where error term μ^r is defined by $\mu^r = v^r - (\boldsymbol{\rho}^r_{vu})'\mathbf{u}^r - \psi^r_{ve}\lambda^r$ and satisfies $E[\mu^r|\mathbf{z}, \mathbf{u}^r, \lambda^r, r = r] = 0$. Term $(\boldsymbol{\rho}^r_{vu})'\mathbf{u}^r$ is a control function for the endogeneity of $\mathbf{x}_2$ that plays the role of instrumental variables. Consistent estimates of this vector can be obtained by using the residual terms of the model of $\mathbf{x}^r_2$ as described by Equation (6). Term $\psi^r_{ve}\lambda^r_y$ is a control function for the regimen choice accounting for x^r_2 endogeneity. From λ^r_y, we already have consistent estimates of $\mathbf{u}^r$ and $m(\mathbf{z}, \gamma)$. To obtain consistent estimates of λ^r_y, we need to get consistent estimates of ψ^r_{ee} and $\boldsymbol{\rho}^r_{eu}$. From the definition given in (7), estimating ψ^r_{ee} and $\boldsymbol{\rho}^r_{eu}$ consists in estimating $\boldsymbol{\omega}^r_{ue}$ and $\boldsymbol{\Omega}^r_{uu} = \left[\omega^r_{k\ell,uu} : (k, \ell) \in K \times K\right]$. Consistent estimates of $\boldsymbol{\omega}^r_{ue}$ are given by Equation (6). $\omega^r_{k\ell,uu}$ are marginal covariance parameters between u^r_k and

u^r_ℓ. To get an expression of $\omega^r_{k\ell,uu}$, we need first to use the normality assumption on $e|\mathbf{z}$ to write:

$$\begin{aligned} E[e|\mathbf{z}, r = r] &= \lambda^r(m(\mathbf{z},\gamma)) \\ V[e|\mathbf{z}, r = r] &= 1 - m(\mathbf{z},\gamma)\lambda^r(m(\mathbf{z},\gamma)) - \lambda^r(m(\mathbf{z},\gamma))^2 \end{aligned},$$

which yields:

$$E[(e)^2|\mathbf{z}, r = r] = 1 - m(\mathbf{z},\gamma)\lambda^r(m(\mathbf{z},\gamma)).$$

Then, from the joint normality of vectors $(\mathbf{u}^r, v^r, e)|\mathbf{z}$, we can derive that:

$$\mathbf{u}^r = \boldsymbol{\omega}^r_{ue} e + \boldsymbol{\varepsilon}^r_u$$

where $\boldsymbol{\varepsilon}^r_u|(\mathbf{z}, e) \sim \mathcal{N}(\mathbf{0}, \boldsymbol{\Omega}^r_{uu} - \boldsymbol{\omega}^r_{ue}(\boldsymbol{\omega}^r_{ue})')$. This result allows us to draw the conditional independence of $\boldsymbol{\varepsilon}^r_u$ and e on $\mathbf{z}_{it}$, implying in turn that residual terms $\boldsymbol{\varepsilon}^r_u = (\varepsilon^r_{k,u} : k \in K)$ do not depend on r conditionally on $\mathbf{z}$ which yields to:

$$\begin{aligned} u^r_{k,it}u^r_{\ell,it} &= \omega^r_{k,ue}\omega^r_{\ell,ue}(e_{it})^2 + \omega^r_{k,ue}e_{it}\varepsilon^r_{\ell,u,it} + \omega^r_{\ell,ue}e_{it}\varepsilon^r_{k,u,it} + \varepsilon^r_{k,u,it}\varepsilon^r_{\ell,u,it} \\ E[e_{it}\varepsilon^r_{\ell,u,it}|\mathbf{z}_{it}, r_{it} = r] &= E[\varepsilon^r_{\ell,u,it}]E[e_{it}|\mathbf{z}_{it}, r_{it} = r] = 0 \quad . \\ E[\varepsilon^r_{k,u,it}\varepsilon^r_{\ell,u,it}|\mathbf{z}_{it}, r_{it} = r] &= E[\varepsilon^r_{k,u,it}\varepsilon^r_{\ell,u,it}] = \omega^r_{k\ell,uu} - \omega^r_{k,ue}\omega^r_{\ell,ue} \end{aligned}$$

Collecting these results permits to write:

$$E[u^r_k u^r_\ell|\mathbf{z}, r = r] = \omega^r_{k,ue}\omega^r_{\ell,ue}E[(e)^2|\mathbf{z}, r = r] + \omega^r_{k\ell,uu} - \omega^r_{k,ue}\omega^r_{\ell,ue},$$

and, finally

$$E[u^r_k u^r_\ell|\mathbf{z}, r = r] = \omega^r_{k\ell,uu} - \omega^r_{k,ue}\omega^r_{\ell,ue}m(\mathbf{z},\gamma)\lambda^r(m(\mathbf{z},\gamma)). \tag{9}$$

From Equation (9), we derive $\omega^r_{k\ell,uu} = E[u^r_k u^r_\ell|\mathbf{z}, r = r] + \omega^r_{k,ue}\omega^r_{\ell,ue}m(\mathbf{z},\gamma)\lambda^r(m(\mathbf{z},\gamma))$. Consistent estimates of $\omega^r_{k,ue}\omega^r_{\ell,ue}m(\mathbf{z},\gamma)\lambda^r(m(\mathbf{z},\gamma))$ are already available and $E[u^r_k u^r_\ell|\mathbf{z}, r = r]$ can be consistently estimated by its empirical counterpart from Equation (6). As a result, we obtain a consistent estimator of parameter $\omega^r_{k\ell,uu}$ and thus of term λ^r_y. Finally, we can estimate Equation (8) by applying non-linear least squares and get consistent estimates of parameters $(\boldsymbol{\beta}^r, \boldsymbol{\rho}^r_{vu}, \psi^r_{ve})$. The detailed steps of this estimation procedure can be found in Appendix 5.1.

4 Discussion/Conclusion

This paper presents a novel estimation approach for an "extended" endogenous regime switching model with endogenous covariates. Previous approaches were unsatisfactory when considering a non-linear framework. We thus consider an approach relying on control functions for both endogenous covariates and regime. We develop an estimation approach inspired by Heckman two-step approach for selection model with a corrected inverse Mills ratio. The great asset of such estimation procedure is its simple implementation combined with its parcimony. Yet it relies on restrictive assumptions: the conditional joint normality assumption given in Equation (5) are necessary to ensure the consistency of the estimation procedure results. Relaxing the independence assumption of error term vectors $(v^r, \mathbf{u}^r, e)$ and $\mathbf{z}$ is very difficult, excepted for allowing heteroscedasticity of the error terms conditionally on $\mathbf{z}$. Relaxing the normality assumptions for vectors $(v^r, \mathbf{u}^r, e)$ would require substantial adjustments in the estimation procedure presented above.

Another consistent estimation process for CMP specific yield models can be derived from Wooldridge (2010). The estimation procedures relies on (i) a set of conditional mean linearity conditions given by $E[v^r|\mathbf{z}, e] = \theta_v^r e$ and $E[\mathbf{u}^r|\mathbf{z}, e] = \theta_u^r e$ for $r \in \{0, 1\}$ and (ii) the normality assumption for the CMP choice model *i.e.* $e|\mathbf{z} \sim \mathcal{N}(0, 1)$. Under these assumptions, augmented yield model with control function $\omega_{ve}^r \lambda^r(\gamma_0 + \boldsymbol{\gamma'}_z \mathbf{z})$ can be estimated by two-stage least squares. Mills ratio terms $\lambda^r(\gamma_0 + \boldsymbol{\gamma'}_z \mathbf{z})$ are included in the instrument set of the two-stage least squares estimator and permit to deal with the endogenous sample selection issues. Yet, because we consider non-linear yield models, estimators based on orthogonality conditions, such as non-linear two-stage least squares, are more complex to estimate.[2] The generalized method of moments can be a solution to get better estimation results. Indeed, the generalized method of moments relies on estimated instruments that are designed for making better use of the information content of instrumental variables than standard non-linear two-stage least squares estimators.[3] An interesting extension of this work would be to evaluate how much estimation results differ when using the different estimation approaches.

[2]For instance, Latruffe et al. (2017) report estimation results that document this point. Standard non-linear two-stage least squares estimators perform poorly when estimating their non-linear stochastic production frontier models.

[3]See, *e.g.*, Chamberlain (1987) and Newey (1990, 1993). The considered instruments need to be sufficiently close to the efficient instrument of the considered estimation problem, the form of which was determined by Chamberlain (1987). They can be built based on preliminary estimation steps. Latruffe et al. (2017) report that the generalized method of moments estimators based on suitably designed instruments substantially outperform standard two-stage least squares estimators when estimating non-linear stochastic production frontiers.

References

Berk, R. A. (1983). "An Introduction to Sample Selection Bias in Sociological Data". In: *American Sociological Review* 48.3, pp. 386–398. ISSN: 00031224.

Chamberlain, G. (1987). "Asymptotic efficiency in estimation with conditional moment restrictions". In: *Journal of econometrics* 34.3, pp. 305–334.

Di Falco, S. and M. Veronesi (2013). "How can African agriculture adapt to climate change? A counterfactual analysis from Ethiopia". In: *Land Economics* 89.4, pp. 743–766.

Donald, S. G. (1995). "Two-step estimation of heteroskedastic sample selection models". In: *Journal of Econometrics* 65.2, pp. 347–380. ISSN: 0304-4076.

Greene, W. H. (2020). *Econometric Analysis*. Pearson Education Limited.

Heckman, J. J. (1976). "The common structure of statistical models of truncation, sample selection and limited dependent variables and a simple estimator for such models". In: *Annals of economic and social measurement, volume 5, number 4*. NBER, pp. 475–492.

— (1978). "Dummy endogenous variables in a simultaneous equation system". In: *Econometrica: Journal of the Econometric Society*, pp. 931–959.

— (1979). "Sample selection bias as a specification error". In: *Econometrica: Journal of the econometric society*, pp. 153–161.

Khanna, M. (2001). "Sequential adoption of site-specific technologies and its implications for nitrogen productivity: A double selectivity model". In: *American journal of agricultural economics* 83.1, pp. 35–51.

Kumbhakar, S. C., E. G. Tsionas, and T. Sipiläinen (2009). "Joint estimation of technology choice and technical efficiency: an application to organic and conventional dairy farming". In: *Journal of Productivity Analysis* 31.3, pp. 151–161.

Kyriazidou, E. (1997). "Estimation of a Panel Data Sample Selection Model". In: *Econometrica* 65.6, pp. 1335–1364.

Latruffe, L. et al. (2017). "Subsidies and technical efficiency in agriculture: Evidence from European dairy farms". In: *American Journal of Agricultural Economics* 99.3, pp. 783–799.

Lee, L.-F. (1978). "Unionism and wage rates: A simultaneous equations model with qualitative and limited dependent variables". In: *International economic review*, pp. 415–433.

Murtazashvili, I. and J. M. Wooldridge (2016). "A control function approach to estimating switching regression models with endogenous explanatory variables and endogenous switching". In: *Journal of Econometrics* 190.2, pp. 252–266.

Newey, W. K. (1990). "Efficient instrumental variables estimation of nonlinear models". In: *Econometrica: Journal of the Econometric Society*, pp. 809–837.

— (1993). "16 Efficient estimation of models with conditional moment restrictions". In: *Econometrics*. Vol. 11. Handbook of Statistics. Elsevier, pp. 419–454.

Pitt, M. M. (1983). "Farm-level fertilizer demand in Java: a meta-production function approach". In: *American Journal of Agricultural Economics* 65.3, pp. 502–508.

Schwiebert, J. (2015). "Estimation and interpretation of a Heckman selection model with endogenous covariates". In: *Empirical Economics* 49.2, pp. 675–703.

Shively, G. E. (1998). "Modeling impacts of soil conservation on productivity and yield variability: evidence from a Heteroskedastic Switching Regression". In: *Selected paper at the Annual Meeting of American Agricultural Economics Association*. Salt Lake City, Utah.

Takeshima, H. and A. Winter-Nelson (2012). "Sales location among semi-subsistence cassava farmers in Benin: a heteroskedastic double selection model". In: *Agricultural Economics* 43.6, pp. 655–670.

Wooldridge, J. M. (2010). *Econometric analysis of cross section and panel data*. MIT press.

— (2015). "Control function methods in applied econometrics". In: *Journal of Human Resources* 50.2, pp. 420–445.

Wu, J. and B. A. Babcock (1998). "The choice of tillage, rotation, and soil testing practices: Economic and environmental implications". In: *American Journal of Agricultural Economics* 80.3, pp. 494–511.

5 Appendices

5.1 Detailed estimation procedure

(A.1) Compute the maximum likelihood estimate of $\boldsymbol{\gamma}$, $\hat{\boldsymbol{\gamma}}$, by estimating the Probit model of $r|\mathbf{z}$ based on the full sample.

(A.2) Compute the estimates of the Mills ratio terms $\lambda^r(m(\mathbf{z},\boldsymbol{\gamma}))$, $\hat{\lambda}^{r(0)} = \lambda^r(m(\mathbf{z},\hat{\boldsymbol{\gamma}}))$, for the subsample with regimen r, $r \in \{0,1\}$.

(B.1) Compute the least squares estimates of $(\boldsymbol{\alpha}_k^r, \omega_{k,ue}^r)$, $(\hat{\boldsymbol{\alpha}}_k^r, \hat{\omega}_{k,ue}^r)$, by regressing $x_{2,k}$ on $(\mathbf{w}, \hat{\lambda}^{r(0)})$ based on the sup-sample with regimen r, for $k \in K$ and $r \in \{0,1\}$.

(B.2) Compute the estimates of error terms u_k^r, $\hat{u}_k^r = x_{2,k} - d^r(\mathbf{w}, \hat{\boldsymbol{\alpha}}_k^r)$, for the sampled observations with regimen r, for $k \in K$ and $r \in \{0,1\}$. Construct the estimate of vector $\boldsymbol{\omega}_{ue}^r$, $\hat{\boldsymbol{\omega}}_{ue}^r = (\hat{\omega}_{k,ue}^r : k \in K)$, for $r \in \{0,1\}$.

(C.1) Compute the estimates of parameters $\omega_{k\ell,u}^r$, with

$$\hat{\omega}_{k\ell,uu}^r = (\textstyle\sum 1(r=r))^{-1} \Big(\sum 1(r=r)\hat{u}_k^r\hat{u}_\ell^r + \hat{\omega}_{k,ue}^r\hat{\omega}_{\ell,ue}^r \sum m(\mathbf{z},\hat{\boldsymbol{\gamma}})\hat{\lambda}^{r(0)}(m(\mathbf{z},\hat{\boldsymbol{\gamma}}))\Big)$$

for $(k,\ell) \in K \times K$ and $r \in \{0,1\}$.

(C.2) Construct the estimate of matrix $\boldsymbol{\Omega}_{uu}^r$, $\hat{\boldsymbol{\Omega}}_{uu}^r = [\hat{\omega}_{k\ell,uu}^r : (k,\ell) \in K \times K]$, that of term ψ_{ee}^r, $\hat{\psi}_{ee}^r = 1 - (\hat{\boldsymbol{\omega}}_{ue}^r)'(\hat{\boldsymbol{\Omega}}_{uu}^r)^{-1}\hat{\boldsymbol{\omega}}_{ue}^r$, and that of vector $\boldsymbol{\rho}_{eu}^r$, $\hat{\boldsymbol{\rho}}_{eu}^r = (\hat{\boldsymbol{\Omega}}_{uu}^r)^{-1}\hat{\boldsymbol{\omega}}_{ue}^r$, for $r \in \{0,1\}$.

(C.3) Compute the estimates of control terms λ_y^r, $\hat{\lambda}^r = (\hat{\psi}_{ee}^r)^{-1/2}\lambda^r\left((\hat{\psi}_{ee}^r)^{-1/2}[m(\mathbf{z},\hat{\boldsymbol{\gamma}}) + (\hat{\boldsymbol{\rho}}_{eu}^r)'\hat{\boldsymbol{u}}^r]\right)$, for the sampled observations with regime r, for $r \in \{0,1\}$.

(D) Compute the non-linear least suqares estimates of $(\boldsymbol{\beta}^r, \boldsymbol{\rho}_{vu}^r, \psi_{ve}^r)$, $(\hat{\boldsymbol{\beta}}^r, \hat{\boldsymbol{\rho}}_{vu}^r, \hat{\psi}_{ve}^r)$, by considering the approximate "doubly augmented" model of y,

$y = s^r(\mathbf{x};\boldsymbol{\beta}^r) + (\boldsymbol{\rho}_{vu}^r)'\hat{\boldsymbol{u}}^r + \psi_{ve}^r\hat{\lambda}_y^r + \hat{\mu}^r$

with $E[\hat{\mu}^r] = 0$, based on the sub-sample of observations with regime r, for $r \in \{0,1\}$.

(E) Use resampling techniques for computing the empirical distribution of the corresponding estimators of parameter vector $(\boldsymbol{\beta}^r, \boldsymbol{\rho}_{vu}^r, \psi_{ve}^r)$, as well as of parameter vector $(\boldsymbol{\alpha}_k^r, \omega_{k,ue}^r)$ for $k \in K$, for $r \in \{0,1\}$.

5.2 Insight on a generalized method of moments estimation approach for our endogenous regime switching model with endogenous covariates

Let's consider the yield functions from the second step of the approach proposed by Wooldridge (2010):

$$y = s^r(\mathbf{x};\boldsymbol{\beta}^r) + \omega_{ve}^r \lambda^r(m(\mathbf{z},\gamma)) + \mu^r,$$

where $E[\mu^r|\mathbf{z}, r = r] = 0$. The efficient instrument corresponding to this model is given by:

$$\boldsymbol{\zeta}^r(\mathbf{z}) = E[(\mu^r)^2|\mathbf{z}, r = r]^{-1} \tfrac{\partial}{\partial(\boldsymbol{\beta}^r, \omega_{ve}^r)} E[s^r(\mathbf{x};\boldsymbol{\beta}^r) + \omega_{ve}^r \lambda^r(m(\mathbf{z},\gamma))|\mathbf{z}, r = r],$$

or, equivalently by:

$$\boldsymbol{\zeta}^r(\mathbf{z}) = V[\mu^r|\mathbf{z}, r = r]^{-1} \begin{bmatrix} \mathbf{s^r}(\mathbf{z};\boldsymbol{\beta}^r) \\ \lambda^r(m(\mathbf{z},\gamma)) \end{bmatrix},$$

where $\mathbf{s}^r(\mathbf{z};\boldsymbol{\beta}^r) = E\left[\frac{\partial}{\partial \boldsymbol{\beta}^r} s^r(\mathbf{x};\boldsymbol{\beta}^r)|\mathbf{z}, r = r\right]$. In what follows, the conditional heteroskedasticity correction term $V[\mu^r|\mathbf{z}, r = r]^{-1}$ is ignored, as it is usually the case in practice. When $s^r(\cdot)$ is linear in $\mathbf{x}$, we have for the following formula for the gradient term:

$$\mathbf{s}^r(\mathbf{z};\boldsymbol{\beta}^r) = (1, E[\mathbf{x}|\mathbf{z}, r = r]).$$

In this case, (near)-efficient instrument $\boldsymbol{\zeta}^r(\mathbf{z})$ can easily be estimated *a priori*. It suffices to observe that the considered ERS model yields $E[x_k|\mathbf{z}, r = r] = \alpha_{k,0}^r + \mathbf{w}'\boldsymbol{\alpha}_{k,w}^r + \omega_{k,ue}^r \lambda^r(m(\mathbf{z},\gamma))$. Yet, in cases where $s^r(\cdot)$ is nonlinear in $\mathbf{x}$, computing gradient term $\mathbf{s}^r(\mathbf{z};\boldsymbol{\beta}^r)$ is much more challenging. Given that $E[\mathbf{x}|\mathbf{z}, r = r]$ can be computed, a possible solution consists of giving a rough approximation for the gradient term as

$$\mathbf{s}^r(\mathbf{z};\boldsymbol{\beta}^r) \approx \tfrac{\partial}{\partial \boldsymbol{\beta}^r} s^r\left(E[\mathbf{x}|\mathbf{z}, r = r];\boldsymbol{\beta}^r\right),$$

for determining instruments for estimating the model of y in the GMM framework. The structure of efficient instruments depending heavily on the functional form of the model, we cannot give a general result here.