Estimating the Effects of Credit Constraints on Productivity of Peruvian Agriculture

Tiemen Woutersen
Shahidur R. Khandker

The World Bank
Development Research Group
Agriculture and Rural Development Team
October 2013

Abstract

This paper proposes an estimator for the endogenous switching regression models with fixed effects. The estimator allows for endogenous selection and for conditional heteroscedasticity in the outcome equation.

Applying the estimator to a dataset on the productivity in agriculture substantially changes the conclusions compared to earlier analysis of the same dataset.

This paper is a product of the Agriculture and Rural Development Team, Development Research Group. It is part of a larger effort by the World Bank to provide open access to its research and make a contribution to development policy discussions around the world. Policy Research Working Papers are also posted on the Web at <http://econ.worldbank.org>. The authors may be contacted at skhandker@worldbank.org.

The Policy Research Working Paper Series disseminates the findings of work in progress to encourage the exchange of ideas about development issues. An objective of the series is to get the findings out quickly, even if the presentations are less than fully polished. The papers carry the names of the authors and should be cited accordingly. The findings, interpretations, and conclusions expressed in this paper are entirely those of the authors. They do not necessarily represent the views of the International Bank for Reconstruction and Development/World Bank and its affiliated organizations, or those of the Executive Directors of the World Bank or the governments they represent.

Estimating the Effects of Credit Constraints on Productivity of Peruvian Agriculture

TIEMEN WOUTERSEN AND SHAHIDUR R. KHANDKER*

KEYWORDS: Endogenous switching regression models, credit constraints.

JEL codes: C10, O16, O13, Q10, Q14.

*Correspondence addresses: Department of Economics, Eller College of Management, University of Arizona, P.O. Box 210108, Tucson, AZ 85721, woutersen@email.arizona.edu; Shahidur R. Khandker is lead economist at the Development Research Group of the World Bank; his email address is skhandker@worldbank.org. We thank Wesley Blundell, William J. Martin, Ming-sen Wang, and Roula Yazigi for helpful comments and discussions.

1. INTRODUCTION

THE ENDOGENOUS SWITCHING REGRESSION MODEL is useful when analyzing individuals and firms that switch between two regimes, for example, being credit constrained versus being credit unconstrained. A credit constrained business may not be able to make the necessary investments, which may lower the productivity of that business. Similarly, a credit constrained farmer may not be able to purchase fertilizer or tools, which may also cause the productivity to be lower. The decision to switch from one regime to the other could also depend on unobserved factors, which would cause the state, such as being credit constrained, to be endogenous.

A recent paper by Charlier, Melenberg and van Soest (2001) estimates several switching models. They find that the data rejects them all. The models that they reject include a fixed effects model and Kyriazidou's (1997) semi-parametric model. However, it might be argued that the fixed effects models that Charlier et al. (2001) and Guirking and Boucher (2008) estimate do not adequately take the endogenous switching decision into account. Also, Kyriazidou's (1997) model does not allow for conditional time-varying heteroscedasticity so it may not be that surprising that the data reject that model as well.

We propose to generalize the existing fixed effects and random effects models to allow for endogenous switching. This generalization will allow for conditional heteroscedasticity in the outcome equation, a feature of almost any dataset. In particular, Maddala and Nelson's (1974) switching model is a special case of the proposed model and so is the linear model with fixed effects and heteroscedastic errors. The application we are interested in is agriculture financing in developing countries and micro nonfarm financing. In particular, Guirking and Boucher (2008) estimate that removing the credit constraint from constrained farmers would increase productivity by 26%. The number is based on an estimate of a fixed effects model. We extend this model with a selection equation and find that the credit constraint has a much smaller impact on the estimated coefficients of the model, demonstrating the importance of having a selection equation. Two related papers are Feder, Onchan and Raparla (1988) and Feder, Lau, Lin, and Luo (1990). There papers

argue that being credit constraint may very well be endogenous. Also, these papers do not use individual effects to control for the heterogeneity of the quality of the farmland so that these papers differ from Guirking and Boucher (2008) and the present paper.

We note that switching models are not only useful for loan decisions but are also useful for labor supply and household expenditure decisions. For example, Lee (1978) and Adamchik and Bedi (2000) estimate a switching model to analyze wage differences between different sectors of the economy. We expect our extensions to be useful for such applications as well. The methodology of this paper differs from Wooldridge (1995), who does not consider endogenous switching models. It also differs from Semykina and Wooldridge (2010), who use the inverse Mills ratio as an instrumental variable.

This paper is organized as follows. Section 2 introduces the model and states the consistency and asymptotic normality result of our estimator. Section 3 applies the new estimator to data on productivity in Peruvian agriculture and section 4 concludes.

2. MODEL AND THEOREM

In our application, farmers can be credit constrained or credit unconstrained. Being credit constrained may reduce output of the farm since it may be more difficult to buy the relevant inputs such as fertilizer, machines, as well as to hire farm hands or specialized workers. Thus, being credit constrained may reduce productivity. However, being credit constrained and having low productivity could also be caused by a common unobserved shock such as illness of the farmer. Thus, it is important to account for this sample selection and that is what Maddala and Nelson's (1974) "switching regression model with endogenous switching" intends to do. In particular, their switching regression model has a selection equation and an outcome equation. Let W_{it} be equal to one if the farmer i is credit constrained in period t and zero otherwise. If the farmer i is not credit constrained in period t , $W_{it} = 0$, then the productivity of the farm is

$$Y_i^{(0)} = \beta' X_{it} + f_i^{(0)} + \kappa_t + u_{it}^{(0)}, \quad (1)$$

where X_{it} denotes the regressors, $f_i^{(0)}$ is an individual specific fixed effect, κ_t is a time dummy and $u_{it}^{(0)}$ is the error term. The models considered by Maddala and Nelson (1974)

or Maddala (1983) do not have fixed effects or time dummies but we use those here. Similar to the last equation, if the farmer i is credit constrained in period t , $W_{it} = 1$, then the productivity of the farm is

$$Y_i^{(1)} = \alpha' X_{it} + f_i^{(1)} + \tau_t + u_{it}^{(1)}, \quad (2)$$

where the fixed effect $f_i^{(1)}$ and error term $u_{it}^{(1)}$ are in general different from $f_i^{(0)}$ and $u_{it}^{(0)}$ in equation (1). Maddala and Nelson (1974) assume that the error terms in the selection equation and in the outcome equation are jointly normal. This implies that the error terms in the outcome equations do not have expectation zero conditional on the regressors. Therefore, Maddala and Nelson (1974) and Maddala (1983) subtract the inverse Mills ratio with a known coefficient from the outcome equations. We also subtract the inverse Mills ratio from the outcome equations. However, since we do not assume that the error terms in the outcome equation are normally distributed we need to estimate the coefficient of the inverse Mills ratio. In particular, we propose the following procedure. Let Q_{it} denote the regressors of the selection equation of individual i in period t and suppose we observe N individuals for T periods. Our procedure allows for predetermined regressors (step 1A) or for exogenous regressors and correlated random effects (step 1B).

Step 1A (Selection equation with predetermined regressors): Estimate a Probit model with predetermined regressors. Let $(\hat{\gamma}_1, \dots, \hat{\gamma}_T)$ denote the quasi maximum likelihood estimator, i.e.

$$(\hat{\gamma}_1, \dots, \hat{\gamma}_T) = \arg \max \frac{\sum_i}{N} \frac{\sum_t}{T} \ln [\{\Phi(\gamma'_t Q_{it})\}^{W_{it}} \{1 - \Phi(\gamma'_t Q_{it})\}^{1-W_{it}}]. \quad (3)$$

Using $(\hat{\gamma}_1, \dots, \hat{\gamma}_T)$, calculate $\hat{R}_{it} = \frac{\phi(\hat{\gamma}'_t Q_{it})}{1 - \Phi(\hat{\gamma}'_t Q_{it})}$ for $i = 1, \dots, N$ and $t = 1, \dots, T$.

Step 1B (Selection equation with correlated random effects): Estimate a Probit model with strictly exogenous regressors, constant slope coefficients and correlated random effects. Let $(\hat{\gamma}, \hat{\psi}_1, \dots, \hat{\psi}_T)$ denote the quasi maximum likelihood estimator, i.e.

$$(\hat{\gamma}, \hat{\psi}_1, \dots, \hat{\psi}_T) = \arg \max \frac{\sum_i}{N} \frac{\sum_t}{T} \ln [\{\Phi(\gamma' Q_{it} + \frac{\sum_{t=1}^T \psi'_t Q_{it}}{T})\}^{W_{it}} \{1 - \Phi(\gamma' Q_{it} + \frac{\sum_{t=1}^T \psi'_t Q_{it}}{T})\}^{1-W_{it}}]. \quad (4)$$

Using $(\hat{\gamma}, \hat{\psi}_1, \dots, \hat{\psi}_T)$, calculate $\hat{R}_{it} = \frac{\phi(\hat{\gamma}'Q_{it} + \frac{1}{T} \sum_{t=1}^T \hat{\psi}_t'Q_{it})}{1 - \Phi(\hat{\gamma}'Q_{it} + \frac{1}{T} \sum_{t=1}^T \hat{\psi}_t'Q_{it})}$ for $i = 1, \dots, N$ and $t = 1, \dots, T$.

Step 2: After step 1A or step 1B we need to difference out the fixed effect. For every time period and every individual for which $W_{it} = W_{i,t-1} = 0$, calculate $\Delta Y_{it}^{(0)} = Y_{it}^{(0)} - Y_{i,t-1}^{(0)}$, $\Delta X_{it} = X_{it} - X_{i,t-1}$, and $\Delta \hat{R}_{it} = \hat{R}_{it} - \hat{R}_{i,t-1}$. Next, regress $\Delta Y_{it}^{(0)}$ on a constant, ΔX_{it} , and $\Delta \hat{R}_{it}$. The constant takes care of the difference in time dummies, $\kappa_t - \kappa_{t-1}$. For every time period and every individual for which $W_{it} = W_{i,t-1} = 1$, calculate $\Delta Y_{it}^{(1)} = Y_{it}^{(1)} - Y_{i,t-1}^{(1)}$ and regress $\Delta Y_{it}^{(1)}$ on a constant, ΔX_{it} , and $\Delta \hat{R}_{it}$.

If the researcher is willing to make stronger assumptions, then other differences can be used as well. In particular, define $\Delta_l Y_{it}^{(1)} = Y_{it}^{(1)} - Y_{i,t-l}^{(1)}$, $\Delta_l Y_{it}^{(0)} = Y_{it}^{(0)} - Y_{i,t-l}^{(0)}$, $\Delta_l X_{it} = X_{it} - X_{i,t-l}$, and $\Delta_l \hat{R}_{it} = \hat{R}_{it} - \hat{R}_{i,t-l}$ for $l = 1, \dots, T$. Then define the moment

$$g^{(0)}(\theta) = \frac{1}{N} \begin{pmatrix} \sum_i \Delta Y_{iT}^{(0)} - \Delta X_{iT}\beta - \Delta \hat{R}_{iT}\delta - (\kappa_T - \kappa_{T-1}) \\ \sum_i \Delta X_{iT} \{ \Delta Y_{iT}^{(0)} - \Delta X_{iT}\beta - \Delta \hat{R}_{iT}\delta - (\kappa_T - \kappa_{T-1}) \} \\ \sum_i \Delta \hat{R}_{iT} \{ \Delta Y_{iT}^{(0)} - \Delta X_{iT}\beta - \Delta \hat{R}_{iT}\delta - (\kappa_T - \kappa_{T-1}) \} \\ \sum_i \Delta_2 Y_{iT}^{(0)} - \Delta_2 X_{iT}\beta - \Delta_2 \hat{R}_{iT}\delta - (\kappa_T - \kappa_{T-2}) \\ \sum_i \Delta_2 X_{iT} \{ \Delta_2 Y_{iT}^{(0)} - \Delta_2 X_{iT}\beta - \Delta_2 \hat{R}_{iT}\delta - (\kappa_T - \kappa_{T-2}) \} \\ \sum_i \Delta_2 \hat{R}_{iT} \{ \Delta_2 Y_{iT}^{(0)} - \Delta_2 X_{iT}\beta - \Delta_2 \hat{R}_{iT}\delta - (\kappa_T - \kappa_{T-2}) \} \\ \dots \end{pmatrix}$$

where κ_1 is normalized to be zero and $\theta = \{\beta, \delta, \kappa_2, \dots, \kappa_T\}'$. Let the moment for the other outcome, $g^{(1)}(\bar{\theta})$, be similarly defined, where $\tau_1 = 0$, and $\bar{\theta} = \{\alpha, \varphi, \tau_2, \dots, \tau_T\}'$. One can use this general method of moment procedure instead of the least squares method in step 2 above but we do not consider this in further detail here. Also, in the application we use a regressor in step 1A that is not used in step 2. This is usually called an exclusion restriction.

We now state the assumptions.

Assumption 1 (Selection equation with predetermined regressors): Let $E(W_{it}|Q_{it}) = \Phi(\gamma_t'Q_{it})$. Let $E\{Q_{it}Q_{it}'\}$ be nonsingular for $t = 1, \dots, T$. Let the parameter space Θ_A be compact; define $\omega = \{\alpha, \beta, \gamma_1, \dots, \gamma_T, \delta, \kappa, \tau, \varphi\}'$ and let the true value ω_0 be in the interior of Θ_A .

Assumption 1 allows for arbitrary correlation of the error in the selection equation and also allows the variance of this equation to vary with time. De Jong and Woutersen (2011) discuss dynamic binary choice models in more detail. An example that satisfies assumption 1 is

$$W_{it} = Health_{i1} + Harvest_{it} \cdot (t + 1) + \eta_{it}\sqrt{t + 1} \quad (5)$$

where $Health_{i1}$ is the health of farmer i in period 1, $Harvest_{it}$ is the harvest of farmer i in period t , and η_{it} is a standard normal error term that is i.i.d. conditional on the regressors.

Assumption 2 (Selection equation with correlated random effects): Let $E(W_{it}|Q_{i1}, \dots, Q_{iT}, v_i) = \Phi(\gamma'Q_{it} + v_i)$ where $v_i = \frac{1}{T} \sum_{t=1}^T \psi'_t Q_{it}$. Let the parameter space, Θ_B , be compact; define $\varpi = \{\alpha, \beta, \gamma_1, \dots, \gamma_T, \psi_1, \dots, \psi_T, \delta, \kappa, \tau, \varphi\}'$, and let the true value ϖ_0 be in the interior of Θ_B . Let $E\{(Q_{it} - v_i)(Q_{it} - v_i)'\}$ be nonsingular for $t = 1, \dots, T$ and all $\psi_t \in \Theta_B$.

This assumption allows for correlated random effects since v_i can depend on the regressors. Such correlated random effects were proposed by Chamberlain (1980). Mundlak (1978) lets the random effect depend on the averages of the regressors, $v_i = \psi' \frac{1}{T} \sum_{t=1}^T Q_{it}$, and the last assumption also allows for that.

The next assumption is about an error term being uncorrelated with regressors, after differencing out the fixed effect. Define

$$\begin{aligned} \Delta\varepsilon_{it}^{(0)} &= \Delta Y_i^{(0)} - (\beta' \Delta X_{it} + \Delta R_{it} \delta + \Delta \kappa_t), \\ \Delta\varepsilon_{it}^{(1)} &= \Delta Y_i^{(1)} - (\alpha' \Delta X_{it} + \Delta R_{it} \varphi + \Delta \tau_t), \end{aligned} \quad (6)$$

where $R_{it} = \frac{\phi(\gamma'_t Q_{it})}{1 - \Phi(\gamma'_t Q_{it})}$ is calculated if step 1A (predetermined regressors) is used and $R_{it} = \frac{\phi(\gamma' Q_{it} + \frac{1}{T} \sum_{t=1}^T \psi'_t Q_{it})}{1 - \Phi(\gamma' Q_{it} + \frac{1}{T} \sum_{t=1}^T \psi'_t Q_{it})}$ if step 1B (correlated random effects) is used.

Assumption 3: Let $Y_i^{(0)} = \beta' X_{it} + R_{it} \delta + \kappa_t + \varepsilon_{it}^{(0)}$ if and only if $W_{it} = 0$. Let $Y_i^{(1)} = \alpha' X_{it} + R_{it} \varphi + \tau_t + \varepsilon_{it}^{(1)}$ if and only if $W_{it} = 1$. Let $\Delta\varepsilon_{it}^{(0)}$ and $\Delta\varepsilon_{it}^{(1)}$ be uncorrelated with ΔX_{it} , ΔW_{it} , and ΔR_{it} and let $E(\Delta\varepsilon_{it}^{(0)}) = E(\Delta\varepsilon_{it}^{(1)}) = 0$. Let $var(\Delta\varepsilon_{it}^{(0)} | \Delta X_{it}, \Delta W_{it}, \Delta R_{it}) > 0$, and $var(\Delta\varepsilon_{it}^{(1)} | \Delta X_{it}, \Delta W_{it}, \Delta R_{it}) > 0$ for all ΔX_{it} , ΔW_{it} , and ΔR_{it} . Let $E[\{(1, \Delta X_{it},$

$\Delta W_{it}, \Delta R_{it})(1, \Delta X_{it}, \Delta W_{it}, \Delta R_{it})'|W_{it} = W_{i,t-1} = 0]$ be nonsingular for some $t \in \{1, \dots, T\}$ and $P(W_{it} = W_{i,t-1} = 0) > 0$. Let $E[\{(1, \Delta X_{is}, \Delta W_{is}, \Delta R_{is})(1, \Delta X_{is}, \Delta W_{is}, \Delta R_{is})'\}|W_{is} = W_{i,s-1} = 1]$ be nonsingular for some $s \in \{1, \dots, T\}$ and $P(W_{is} = W_{i,s-1} = 1) > 0$.

Assumption 4: Let $\{Q_{it}, W_{it}, X_{it}, Y_{it}, t = 1, \dots, T\}, i = 1, \dots, T$ be i.i.d. across individuals. Let $E(|Q_{it}|^4) < M, E(|W_{it}|^4) < M, E(|X_{it}|^4) < M, E(|Y_{it}|^4) < M$ for all $t = 1, \dots, T$, and $i = 1, \dots, T$ where $M < \infty$.

Kyriazidou (1997) imposes exchangability of the error terms. This implies that the error terms are homoscedastic while the outcome equation in this paper allows for conditional heteroscedasticity and the selection equation of assumption 1 allows for time-varying variances as in the example above.¹ In order to correct the standard errors for our two step estimator, it is convenient to write the estimator as the maximum of an objective function. This is similar to Heckman's (1979) sample selection estimator. In the application, however, we bootstrap the estimators. That is, we sample the data with replacement and go through step 1A and step 2 for every dataset that we generated. The objective function that is used to prove asymptotic normality of our estimator as well as the asymptotic variance-covariance matrix is presented in the technical appendix.

Theorem 1 (Consistency and Asymptotic Normality):

Let assumptions 1 and 3-4 hold. Then

$$\begin{aligned} \hat{\omega} &\xrightarrow[p]{p} \omega_0 \text{ and} \\ \sqrt{NT}(\hat{\omega} - \omega_0) &\xrightarrow[p]{p} N(0, \Omega_A) \end{aligned}$$

where Ω_A is positive semidefinite.

Let assumptions 2-4 hold. Then

$$\begin{aligned} \hat{\varpi} &\xrightarrow[p]{p} \varpi_0 \text{ and} \\ \sqrt{NT}(\hat{\varpi} - \varpi_0) &\xrightarrow[p]{p} N(0, \Omega_B) \end{aligned}$$

where Ω_B is positive semidefinite.

¹Dustmann and Rochina-Barrachina (2007) discuss empirical identification issues with Kyriazidou's estimator.

As Horowitz (2001, theorem 2.2) shows, bootstrapping an asymptotically normally distributed estimator that can be represented by an influence function yields a consistent variance-covariance matrix and consistent confidence intervals.² In the application, we bootstrap the estimator.

3. PRODUCTIVITY IN PERUVIAN AGRICULTURE

We apply the approach proposed above to a dataset about Peruvian agriculture. This dataset contains information about whether the farmer was credit constrained, which crops were grown on her/his farm, the amount of labor used, etc. The data was data was collected in 1997 and 2003 and were previously analyzed by Guirkinger and Boucher (2008). We follow their specification and use regression coefficients that are constant over time. The only difference between our specification and Guirkinger and Boucher (2008) is that we apply the methodology of the previous sections and use a selection equation to account for the endogeneity. We report the results in Table (1). The first column reports the first stage Probit regression. The second and third columns report columns (D) and (E) in the original paper. The fourth and fifth columns report our estimates corresponding to their original columns (D) and (E). The sixth and seventh columns are columns (F) and (G) from the original paper. The last two columns are the corresponding estimates using our method. The standard errors are calculated by bootstrapping the two-step estimation together. We also implement a Wald test to compare the differences between the credit constrained farmers and the credit unconstrained farmers. We find that adding a selection equation to the model shrinks the differences between the coefficients of the constrained and unconstrained farmers. This finding has important policy implications since the effect of being credit constrained is not as large as previously thought. Our statistical test confirms these findings. In particular, we used a Wald test with 11 degrees of freedom to test the statistical difference between the credit constrained farmers and the credit unconstrained farmers. Using the specification of the original paper, we find that the value of the Wald test is 56.62 (comparing columns D and E) and 61.17 (comparing columns F and G). Adding the selection equation reduces these numbers to 22.92 (comparing

²Horowitz (2001, Theorem 2.2) averages $g_n(X_i)$.

columns D' and E') and 25.96 (comparing columns F' and G').³ Thus, the differences do not completely go away but are substantially reduced.

Next, we estimate that if credit constrained farmers were to become unconstrained then productivity would increase by 10.6%, which is significantly smaller than the estimate of 26% from Guirking and Boucher (2008).⁴ Moreover, this 10.6% increase is not statistically significant. A full comparison between our estimates and Guirking and Boucher's can be seen in Table 2. The second and third columns (B and B') are estimates for absolute productivity change by credit constrained type if the credit constraints were removed. The fourth and fifth columns (C and C') are relative productivity increases for each group if they were to become unconstrained. The total estimated impact on productivity for removing credit constraints is found in the seventh and eighth columns (E and E'), which includes our overall estimate of 10.6%.

Thus, applying the estimator to a dataset on the productivity in Peruvian agriculture shows that the new estimator changes the conclusions compared to earlier results on the same dataset. In particular, adding a selection equation to a model with fixed effects causes the coefficients of the credit constrained farmers to be quite similar to the coefficients of the unconstrained farmers.

4. CONCLUSION

We propose an estimator for the endogenous switching regression models with fixed effects. The estimator allows for endogenous selection and for conditional heteroscedasticity in the outcome equation. Applying the estimator to a dataset on the productivity in Peruvian agriculture shows that the new, more general estimator substantially changes the conclusions compared to the earlier analysis of the same dataset. In particular, adding a selection equation to a model with fixed effects causes the coefficients of the credit constrained farmers to be similar to those of the unconstrained farmers, demonstrating the importance of having a selection equation. Relaxing the single index assumption of the selection equation by extending Altonji and Matzkin (2005) is left for future research and

³The p-values of these for test were 0.000, 0.000, 0.0181, and 0.0066.

⁴We use the same methodology to calculate the change in productivity as Guirking and Boucher (2008) but we use the new estimator for the parameter value, using step 1A and step 2.

so is developing a model that allows for selection between more than two regimes.

5. APPENDIX: CONSISTENCY AND ASYMPTOTIC NORMALITY RESULTS

In order to prove consistency and asymptotic normality of our two step estimator, it is convenient to write the estimator as the maximum of an objective function. Let assumptions 1 and 3-4 hold. Define

$$S(\omega) = \frac{1}{N} \sum_i S_i(\omega) \text{ where}$$

$$\begin{aligned} S_i(\omega) = & \frac{\sum_t}{T} \ln[\{\Phi(\gamma'Q_{it} + \frac{1}{T} \sum_{t=1}^T \psi'_t Q_{it})\}^{W_{it}} \{1 - \Phi(\gamma'Q_{it} + \frac{1}{T} \sum_{t=1}^T \psi'_t Q_{it})\}^{1-W_{it}}] \\ & - \frac{\sum_t}{T} 1(\Delta Y_i^{(0)} \text{ observed}) \cdot \{\Delta Y_i^{(0)} - (\beta' \Delta X_{it} + \Delta R_{it} \delta + \Delta \kappa_t)\}^2 \\ & - \frac{\sum_t}{T} 1(\Delta Y_i^{(1)} \text{ observed}) \cdot \{\Delta Y_i^{(1)} - (\alpha' \Delta X_{it} + \Delta R_{it} \varphi + \Delta \tau_t)\}^2. \end{aligned}$$

Let $\frac{\partial S(\omega)}{\partial \omega}|_{\omega=\omega_0}$ denote the derivative of $S(\omega)$ with respect to ω and evaluated at the true value ω_0 . Let Λ denote the variance-covariance matrix of $\sqrt{N} \frac{\partial S(\omega)}{\partial \omega}|_{\omega=\omega_0}$. Remember that $\omega = \{\alpha, \beta, \gamma_1, \dots, \gamma_T, \delta, \kappa, \tau, \varphi\}'$. Let H_1 denote the second derivative of $E\{S(\omega)\}$ with respect to the vector $\gamma = \{\gamma_1, \dots, \gamma_T\}'$, evaluated at the true value of γ , i.e.

$$H_1 = \partial^2 E\{S(\omega)\} / \partial \gamma \partial \gamma' |_{\gamma=\gamma_0}.$$

Similarly, let $\omega_{subset} = \{\alpha, \beta, \delta, \kappa, \tau, \varphi\}'$ and $H_2 = \partial^2 E\{S_2(\omega_{subset})\} / \partial \omega_{subset} \partial \omega'_{subset} |_{\omega_{subset}=\omega_{subset,0}}$. Thus, the Hessian of $E\{S(\omega)\}$ has the following form,

$$\partial^2 E\{S(\omega)\} / \partial \omega \partial \omega' |_{\omega=\omega_0} = \begin{pmatrix} H_1 & 0 \\ 0 & H_2 \end{pmatrix}.$$

Note maximizing $S(\omega)$ with respect to ω is the same as maximizing

$$\frac{\sum_i}{N} \frac{\sum_t}{T} \ln[\{\Phi(\gamma'Q_{it} + \frac{1}{T} \sum_{t=1}^T \psi'_t Q_{it})\}^{W_{it}} \{1 - \Phi(\gamma'Q_{it} + \frac{1}{T} \sum_{t=1}^T \psi'_t Q_{it})\}^{1-W_{it}}]$$

with respect to γ and maximizing

$$\begin{aligned} & - \frac{\sum_i}{N} \frac{\sum_t}{T} 1(\Delta Y_i^{(0)} \text{ observed}) \cdot \{\Delta Y_i^{(0)} - (\beta' \Delta X_{it} + \Delta R_{it} \delta + \Delta \kappa_t)\}^2 \\ & - \frac{\sum_i}{N} \frac{\sum_t}{T} 1(\Delta Y_i^{(1)} \text{ observed}) \cdot \{\Delta Y_i^{(1)} - (\alpha' \Delta X_{it} + \Delta R_{it} \varphi + \Delta \tau_t)\}^2. \end{aligned}$$

with respect to ω_{subset} . Thus, our ‘first step estimator’ is equivalent to the estimator considered by Newey and McFadden (1994, example 1.2). Their conditions are satisfied

so that $\hat{\gamma} - \gamma_0 = o_p(1)$. Next, note that

$$\begin{aligned} & - \frac{\sum_i}{N} \frac{\sum_t}{T} 1(\Delta Y_i^{(0)} \text{ observed}) \cdot \{\Delta Y_i^{(0)} - (\beta' \Delta X_{it} + \Delta R_{it} \delta + \Delta \kappa_t)\}^2 \\ & - \frac{\sum_i}{N} \frac{\sum_t}{T} 1(\Delta Y_i^{(1)} \text{ observed}) \cdot \{\Delta Y_i^{(1)} - (\alpha' \Delta X_{it} + \Delta R_{it} \varphi + \Delta \tau_t)\}^2. \end{aligned}$$

is a concave function. The i.i.d. assumption and the assumption on the moments ensure that this expression converges in probability to its expectation. The full rank assumption then ensures that this expectation has a unique maximum at the true value. Thus, all the assumptions of Newey and McFadden (1994, theorem 2.7) are satisfied so that $\hat{\omega}_{subset} - \omega_{subset} = o_p(1)$.

Concerning the asymptotic normality: Note that $S(\omega)$ is twice continuously differentiable. Interpret the first derivative as a moment and note that all the assumptions of Newey and McFadden (1994, theorem 3.4) are satisfied. The asymptotic normality follows and

$$\Omega = \begin{pmatrix} H_1^{-1} & 0 \\ 0 & H_2^{-1} \end{pmatrix} \Lambda \begin{pmatrix} H_1^{-1} & 0 \\ 0 & H_2^{-1} \end{pmatrix}.$$

The nonzero variation follows from the strictly positive variation of the error terms. The matrix Ω can be estimated using a sample analogue. In particular, define

$$\begin{aligned} \hat{H}_1 &= \frac{\partial^2 S(\omega)}{\partial \gamma \partial \gamma'} \Big|_{\omega=\hat{\omega}}, \\ \hat{H}_2 &= \frac{\partial^2 S(\omega)}{\partial \omega_{subset} \partial \omega'_{subset}} \Big|_{\omega=\hat{\omega}}, \\ \hat{\Lambda} &= \begin{pmatrix} \frac{\sum_i}{N} \left\{ \frac{\partial S_i(\gamma)}{\partial \gamma} \right\} \frac{\partial S_i(\gamma)}{\partial \gamma'} & \frac{\sum_i}{N} \left\{ \frac{\partial S_i(\omega_{subset})}{\partial \omega_{subset}} \right\} \frac{\partial S_i(\gamma)}{\partial \gamma'} \\ \frac{\sum_i}{N} \left\{ \frac{\partial S_i(\gamma)}{\partial \gamma} \right\} \frac{\partial S_i(\omega_{subset})}{\partial \omega'_{subset}} & \frac{\sum_i}{N} \left\{ \frac{\partial S_{2,i}(\omega_{subset})}{\partial \omega_{subset}} \right\} \frac{\partial S_{2,i}(\omega_{subset})}{\partial \omega_{subset}} / \end{pmatrix} \Big|_{\omega=\hat{\omega}}, \end{aligned}$$

and

$$\hat{\Omega} = \begin{pmatrix} \hat{H}_1^{-1} & 0 \\ 0 & \hat{H}_2^{-1} \end{pmatrix} \hat{\Lambda} \begin{pmatrix} \hat{H}_1^{-1} & 0 \\ 0 & \hat{H}_2^{-1} \end{pmatrix}.$$

Theorem 4.5 by Newey and McFadden (1994) yields that $\hat{\Omega} = \Omega + o_p(1)$.

Now suppose that assumption 2-4 hold. Define

$$\bar{S}(\varpi) = \frac{1}{N} \sum_i \bar{S}_i(\varpi) \text{ where}$$

$$\begin{aligned} \bar{S}_i(\varpi) &= \frac{\sum_t}{T} \ln \left[\left\{ \Phi(\gamma' Q_{it} + \frac{\sum_{t=1}^T \psi'_t Q_{it}}{T}) \right\}^{W_{it}} \left\{ 1 - \Phi(\gamma' Q_{it} + \frac{\sum_{t=1}^T \psi'_t Q_{it}}{T}) \right\}^{1-W_{it}} \right] \\ & - \frac{\sum_t}{T} 1(\Delta Y_i^{(0)} \text{ observed}) \cdot \{\Delta Y_i^{(0)} - (\beta' \Delta X_{it} + \Delta R_{it} \delta + \Delta \kappa_t)\}^2 \\ & - \frac{\sum_t}{T} 1(\Delta Y_i^{(1)} \text{ observed}) \cdot \{\Delta Y_i^{(1)} - (\alpha' \Delta X_{it} + \Delta R_{it} \varphi + \Delta \tau_t)\}^2. \end{aligned}$$

Remember that $\varpi = \{\alpha, \beta, \gamma_1, \dots, \gamma_T, \psi_1, \dots, \psi_T, \delta, \kappa, \tau, \varphi\}'$. The same reasoning as above holds but now the consistency of the first step follows from Chamberlain (1980). The function $\bar{S}(\varpi)$ is twice continuously differentiable and Ω can again be consistently estimated by its sample analogue. Also, the conditions of Horowitz (2001, theorem 2.2) are satisfied since the estimator is asymptotically normally distributed and this normality follows from an averaging operator, in particular from applying the Lindeberg–Lévy central limit theorem to $\frac{1}{\sqrt{N}} \sum_i \frac{\partial S_i(\omega)}{\partial \omega}$ and $\frac{1}{\sqrt{N}} \sum_i \frac{\partial \bar{S}_i(\bar{\omega})}{\partial \bar{\omega}}$.

REFERENCES

- [1] Adamchik, V., and V. Bedi (2000): “Wage differentials between the public and the private sectors: Evidence from an economy in transition,” *Labour Economics*, 7: 203-224.
- [2] Altonji, J. G., Matzkin, R. L. (2005): “Cross section and panel data estimators for nonseparable models with endogenous regressors,” *Econometrica*, 73: 1053-1102.
- [3] Boucher, S. R., C. Guirking, and C. Trivelli (2009): “Direct Elicitation of Credit Constraints: Conceptual and Practical Issues with an Application to Peruvian Agriculture,” *Economic Development and Cultural Change*, 57, 4: 609-640 .
- [4] Charlier, E., B. Melenberg, and A. van Soest (2001): “An analysis of housing expenditure using semiparametric models and panel data,” *Journal of Econometrics*, 101, 71-107.
- [5] Chamberlain, G. (1980): “Analysis with qualitative data,” *Review of Economic Studies*, 47: 225-238.
- [6] De Jong, R. M., and T. Woutersen (2011): “Dynamic Time Series Binary Choice,” *Econometric Theory*, 27: 673-702.
- [7] Dustmann, C., M. E. Rochina-Barrachina (2007): “Selection correction in panel data models: an application to the estimation of females’ wage equations,” *Econometric Journal*, 10: 263-293.
- [8] Feder, G., T. Onchan, T. Raparla (1988): “Collateral, guaranties and rural credit in developing countries: evidence from Asia,” *Agricultural Economics*, Vol. 2, 3: 231–245.
- [9] Feder, G., Lau, L. J., Lin, J. Y., & Luo, X. (1990): “The relationship between credit and productivity in Chinese agriculture: a microeconomic model of disequilibrium,” *American Journal of Agricultural Economics*, 72(5), 1151-1157.
- [10] Guirking, C. and S. R. Boucher (2008) “Credit Constraints and Productivity in Peruvian Agriculture”, *Agricultural Economics*, 39: 295-308.

- [11] Heckman, J. J. (1979): "Sample Selection Bias as a Specification Error," *Econometrica*, 47: 153-161.
- [12] Horowitz, J. L. (2001): "The Bootstrap," in *Handbook of Econometrics*, Vol. 5, ed. by J. J. Heckman and E. Leamer. Amsterdam: North-Holland.
- [13] Kyriazidou, E. (1997): "Estimation of a Panel Data Sample Selection Model," *Econometrica*, 65: 1335-1364.
- [14] Lee, L. (1978): "Unionism and Wage Rates: A Simultaneous Equations Model with Qualitative and Limited Dependent Variables," *International Economic Review*, 19: 415-433.
- [15] Maddala, G. (1983): *Limited-Dependent and Qualitative Variables in Econometrics*, Econometric Society Monographs No. 3, Cambridge University Press, New York.
- [16] Maddala, G. S., and F. D. Nelson (1974): "Maximum likelihood methods for models of markets in disequilibrium," *Econometrica*, 42: 1013-1030.
- [17] Mundlak, Y. (1978): "On the pooling of time series and cross section data," *Econometrica*, 46: 69-85.
- [18] Newey, W. K., and D. McFadden (1994): "Large Sample Estimation and Hypothesis Testing," in *Handbook of Econometrics*, Vol. 4, ed. by R. F. Engle and D. McFadden. Amsterdam: North-Holland.
- [19] Semykina, A., and J. M. Wooldridge (2010): "Estimating panel data models in the presence of endogeneity and selection," *Journal of Econometrics*, 157: 375-380.
- [20] Wooldridge, J. M. (1995): "Selection corrections for panel data models under conditional mean independence assumptions," *Journal of Econometrics*, 68: 115-132.

Table 1: Estimation results for productivity

	Probit (constant)	(D)	(E)	(D')	(E')	(F)	(G)	(F')	(G')
	b/se	Unconstrained productivity b/se	Constrained productivity b/se	Unconstrained productivity b/se	Constrained productivity b/se	Unconstrained productivity b/se	Constrained productivity b/se	Unconstrained productivity b/se	Constrained productivity b/se
A	0.00 (0.01)	-85.57 (58.78) 14.45 (13.26)	-130.62** (48.65) 182.67* (82.27)	-131.03 (96.25) 14.61 (38.59)	-134.65 (94.43) 133.58 (118.76)	-88.65 (59.57)	-116.20* (48.20)	-137.92 (93.92)	-122.15 (86.10)
K									
K/A						-24.28 (127.11) 7.72 (42.06)	645.98** (236.42) -34.93 (33.08)	158.88 (210.46) -18.42 (98.43)	586.26 (394.83) -19.27 (86.17)
Labor/A									
Labor	-0.01 (0.02)	-61.50 (48.74)	2.33 (38.08)	-45.55 (69.31)	-15.40 (58.54)				
Dependency ratio	0.11 (0.23)	490.90 (418.87)	10.34 (330.09)	397.34 (610.32)	-67.39 (404.30)	696.21 (404.46)	29.14 (318.23)	472.43 (596.20)	14.82 (386.36)
Reg income	-0.00 (0.13)	263.85 (167.80)	-14.79 (214.76)	49.97 (346.38)	-30.66 (283.65)	244.34 (171.22)	32.22 (202.05)	52.85 (356.85)	-37.26 (265.71)
Herd size	0.01 (0.01)	53.66** (20.12)	40.89 (21.76)	42.73 (55.64)	52.57 (53.32)	56.55** (20.31)	38.94 (21.03)	44.45 (53.65)	54.61 (49.71)
Rice		632.30* (252.99)	93.30 (147.59)	763.05* (361.37)	74.44 (276.05)	672.66* (268.99)	115.48 (146.88)	713.98 (381.22)	114.22 (268.85)
Cotton		-279.51 (223.06)	-27.99 (153.51)	-616.28 (340.81)	-292.07 (235.61)	-236.15 (226.56)	-23.86 (147.37)	-681.16 (363.86)	-323.61 (217.68)
Banana		-374.69 (267.39)	754.11** (275.51)	-368.99 (575.70)	669.92 (681.77)	-395.52 (272.64)	759.00** (271.02)	-366.49 (589.59)	688.66 (691.07)
Corn		61.04 (186.70)	-64.55 (117.97)	-0.29 (280.17)	-164.73 (188.64)	12.66 (185.75)	-38.69 (116.87)	-25.66 (280.70)	-137.48 (193.14)
Durables	-0.02* (0.01)	5.03 (2.67)	4.83 (28.11)	24.98 (29.03)	-28.85 (44.61)	5.92* (2.64)	6.38 (26.34)	23.62 (26.47)	-37.18 (45.08)
Constant	0.66*** (0.14)	1493.80*** (344.02)	977.59*** (262.72)	406.41** (150.74)	357.68* (158.21)	1220.67*** (344.68)	948.34*** (248.31)	461.99** (157.09)	374.78* (157.97)
Title	-0.41*** (0.10)								
Network	-1.25*** (0.19)								
δ				-532.09 (667.13)	421.59 (498.24)			-555.28 (693.75)	497.45 (464.47)

Table 2: Counterfactuals: The impact of eliminating credit constraints on productivity and regional output

Type of credit constraint	A	Frequency in sample	B	Productivity change	B'	Productivity change	C	Relative change	C'	D	Land controlled	E	Impact on regional output	E'	Impact on regional output
Quantity rationed		23.50%		516.28		256.37	58.3%	28.90%		20.5%		11.90%	6.41%		
Risk rationed		15.50%		[176] 477.71		[667.73] 129.96				16.0%		[4.5%] 10.90%	[15.7%] 2.53%		
Transaction cost rationed		10.50%		[175] 412.75		[697.25] 160.29				7.8%		[4.7%] 3.8%	[16.2%] 1.52%		
Constrained		49.50%		[216] 482.24		[620.77] 196.41				44.2%		[2.1%] 26%	[6.0%] 10.60%		
				[149]		[657.27]		58.9%	24%			[8.4%]	[35.7%]		