The institutional dimension of illicit behavior:
Evidence from the land property rights and coca crops in Colombia

-Appendix -

Juan Carlos Muñoz-Mora[*]
ECARES - Université Libre de Bruxelles

Santiago Tobón-Zapata [†]
Université Catholique de Louvain

Jesse Willem d'Anjou[‡]
CERMI - Université Libre de Bruxelles

Draft Version
This version: May, 2013

[*]European Center for Advanced Research in Economics and Statistics -ECARES . Université Libre de Bruxelles. 50, Avenue Roosevelt CP 139, 1050 Brussels, Belgium e-mail:*juan.carlos.munoz.mora@ulb.ac.be*.

[†] Université Catholique de Louvain. Department of Economics, Collége L. H. Dupriez, 3 Place Montesquieu B-1348 Louvain-la-Neuve. Email: *santiagotobon@gmail.com*

[‡]Center for European Research in Microfinance, Université Libre de Bruxelles. 50, Avenue Roosevelt CP 114, 1050 Brussels, Belgium, e-mail:*jesse.w.d.anjou@gmail.com*.

CONTENTS

1. DATA PREPARATION

1.1. **Formality Index.**

1.1.1. *Cadastral information in Colombia.*

1.1.2. *Data cleaning.* The criteria to drop some municipalities due to cadastral updating is based in the follows

If the last cadastral updating was before 1998 and

2. EXTRA DESCRIPTIVE STATISTICS

3. STATIC MODEL

4. SYSTEM GMM

We estimate a dynamic panel data model by using the System GMM estimator developed by ? (?) and (?, ?)lundell and Bond (1998) based on the Difference GMM estimator proposed earlier by Holtz-Eakin et al. (1988) and Arellano and Bond (1991). Both, Difference and System GMM estimators are specific applications of the generalized-method-of-moments (Hansen, 1982). In this section we explain the underlying details of the estimator in five steps. First, we outline a simple autoregressive panel data model and review inconsistency issues of common estimators. Second, considering the same simple autoregressive model, we explain the Difference GMM estimator. Third, we go further to construct the System GMM estimator. Fourth, we extend the explanation to multivariate models. Finally, we outline some issues about the conditions for optimality of the System GMM estimator and relevant tests. In most of the explanation we follow rigorous surveys of Difference and System GMM estimators made by Arellano and Honore (2001), Blundell et al. (2000) and Bond (2002).

4.1. **Biases of standard panel data estimators.** We consider a simple autoregressive panel data model of the form:

$$
\begin{aligned}
y_{it} &= \alpha y_{it-1} + \varepsilon_{it} \\
\varepsilon_{it} &= \mu_i + v_{it}
\end{aligned}
\tag{1'}
$$

where the following assumptions hold for the structure of the fixed effect and the idiosyncratic shock:

$$
E\left[\mu_i\right] = E\left[v_i\right] = E\left[\mu_i v_i\right] = 0; \text{ for } i = 1, ..., N \text{ and } t = 2, ..., T
\tag{2'}
$$

and

$$
E\left[v_{it} v_{is}\right] = 0; \text{ for } i = 1, ..., N \text{ and } \forall t \neq s
\tag{3'}
$$

Additionally, a further assumption regarding the initial conditions is also valid (Ahn and Schmidt, 1995):

$$
E\left[y_{i1} v_{it}\right] = 0; \text{ for } i = 1, ..., N \text{ and } t = 2, ..., T
\tag{4'}
$$

The introduction of a lag of the dependent variable in (**??**) yields asymptotic bias in standard estimators such as OLS and Fixed Effects, which are both biased and inconsistent (see Baltagi (2008) and Blundell et al. (2000) for further details). In the OLS case, since y_{it} is a function of μ_i, y_{it-1} is also a function of μ_i and thus it is correlated with the error term. Furthermore, in the case of the Fixed Effects estimator, the within transformation removes the fixed effects by taking the equation in levels $y_{it} = \alpha y_{it-1} + \mu_i + v_{it}$ and subtracting out the average over time $\bar{y}_{i.} = \alpha \bar{y}_{i.-1} + \mu_i + \bar{v}_{i.}$. The resulting model is of the form $(y_{it} - \bar{y}_{i.}) = \alpha(y_{it-1} - \bar{y}_{i.-1}) + (v_{it} - \bar{v}_{i.})$. However, the idiosyncratic shock v_{it-1} is correlated with y_{it-1} and therefore in the transformed model $(y_{it-1} - \bar{y}_{i.-1})$ will be correlated with $(v_{it} - \bar{v}_{i.})$ (Nickel, 1981).

[4] Under an additional restriction of *covariance stationarity*[§] the asymptotic bias of the OLS estimator is given by (Blundell et al., 2000; Sevestre and Trognon, 1985):

$$\text{plim}\,(\hat{\alpha}_{OLS} - \alpha) = (1 - \alpha)\,\frac{E[\mu_i^2]/E[v_{it}^2]}{E[\mu_i^2]/E[v_{it}^2] + \frac{1-\alpha}{1+\alpha}}$$

and respectively, the asymptotic bias of the Fixed Effects estimator is (Nickel, 1981):

$$\text{plim}\,(\hat{\alpha}_{FE} - \alpha) = -\frac{\frac{1+\alpha}{T-1}\left(1 - \frac{1-\alpha^T}{T(1-\alpha)}\right)}{1 - \frac{2\alpha}{(1-\alpha)(T-1)}\left(1 - \frac{1-\alpha^T}{T(1-\alpha)}\right)}$$

4.2. **Difference GMM estimator.** The autoregressive panel data model described by (1') to (4') imply the following $0.5(T-1)(T-2)$ orthogonality conditions, which are equivalent to those given by (??) in the paper except for the exclusion of other covariates different than the lag of the dependent variable:

(5') $$E\left[y_{it-s}\Delta\varepsilon_{it}\right] = 0; \text{ for } t = 3, ..., T,\ 2 \leq s \leq t - 1$$

The usefulness of these orthogonality conditions is straightforward when looking at the first-difference transformation of the model, aimed at eliminating the fixed effects. This is done by multiplying (1') by $\mathbf{I}_N \otimes \mathbf{M}_\Delta$ with $\mathbf{I}_N$ being an identity matrix of order N and $\mathbf{M}_\Delta$ a matrix with diagonal of $-1s$ and a sub-diagonal of $1s$ to the right. The transformation yields a model of the form[¶]:

$$\Delta y_{it} = \alpha\Delta y_{it-1} + \Delta v_{it}$$

The conditions in (5') can also be expressed as:

$$E\left[\mathbf{Z}'_{di}\mathbf{\Delta}\varepsilon_i\right] = 0$$

where $\mathbf{\Delta}\varepsilon_i$ is the $T-2$ vector $(\Delta\varepsilon_{i3}, \Delta\varepsilon_{i4}, ..., \Delta\varepsilon_{iT})'$ and $\mathbf{Z}_{di}$ is the $(T-2) \times 0.5(T-1)(T-2)$ instruments matrix given by:

$$\mathbf{Z}_{di} = \begin{bmatrix} y_{i1} & 0 & 0 & ... & 0 & ... & 0 \\ 0 & y_{i1} & y_{i2} & ... & 0 & ... & 0 \\ . & . & . & ... & . & ... & . \\ 0 & 0 & 0 & ... & y_{i1} & ... & y_{iT-2} \end{bmatrix}$$

or, when collapsing the matrix in order to reduce the instrument count, $\mathbf{Z}_{di}$ becomes a $(T-2) \times (T-2)$ matrix given by:

$$\mathbf{Z}_{di} = \begin{bmatrix} y_{i1} & 0 & ... & 0 \\ y_{i1} & y_{i2} & ... & 0 \\ . & . & ... & . \\ y_{i1} & y_{i2} & ... & y_{iT-2} \end{bmatrix}$$

The Difference GMM estimator solves:

$$\hat{\alpha}_{diff-GMM} = \arg\min_{\alpha\in\Re} \mathbf{\Delta}\varepsilon'\mathbf{Z}_d\mathbf{W}_N^d\mathbf{Z}'_d\mathbf{\Delta}\varepsilon$$

for some symmetric, positive semi definite matrix $\mathbf{W}_N^d$, $\mathbf{Z}'_d$ the $0.5(T-1)(T-2) \times N(T-2)$ instruments matrix $(\mathbf{Z}'_{d1}, \mathbf{Z}'_{d2}, ..., \mathbf{Z}'_{dN})$ or its $(T-2) \times N(T-2)$ collapsed counterpart and $\mathbf{\Delta}\varepsilon'$

[§]*Covariance stationarity* requires further assumptions on the structure of the error components, namely $E[v_{it}^2] = \sigma_v^2$ for $i = 1, ..., N$ and $E[\lambda_{i1}^2] = \sigma_v^2/(1-\alpha^2)$ for $i = 1, ..., N$, $t = 2, ..., T$ where λ_{i1} is such that the process is *mean stationary* with $y_{i1} = \mu_i/(1-\alpha) + \lambda_{i1}$

[¶]An alternative to first-differencing is the forward-orthogonal-deviations transformation in which the fixed effects are removed by subtracting the average of all available future observations instead of the previous one (Arellano and Bover, 1995). This alternative yields better results whenever there are gaps in unbalanced panels

the $N(T-2)$ vector $(\mathbf{\Delta}\varepsilon_1', \mathbf{\Delta}\varepsilon_2', ..., \mathbf{\Delta}\varepsilon_N')$. Finally, by letting $\mathbf{\Delta y}$ and $\mathbf{\Delta y}_{-1}$ be $N(T-2)$ vectors^{5} constructed in the same way as $\mathbf{\Delta}\varepsilon$ by stacking the $T-2$ vectors $\mathbf{\Delta y}_i' = (\Delta y_{i3}, \Delta y_{i4}, ..., \Delta y_{iT})$ and $\mathbf{\Delta y}_{i,-1}' = (\Delta y_{i2}, \Delta y_{i4}, ..., \Delta y_{iT-1})$ respectively across individuals, the estimator for α takes the form:

$$\hat{\alpha}_{diff-GMM} = \left(\mathbf{\Delta y}_{-1}'\mathbf{Z}_d\mathbf{W}_N^d\mathbf{Z}_d'\mathbf{\Delta y}_{-1}\right)^{-1}\mathbf{\Delta y}_{-1}'\mathbf{Z}_d\mathbf{W}_N^d\mathbf{Z}_d'\mathbf{\Delta y}$$

The optimal two-step generalized-method-of-moments estimation is achieved with a choice of the weighting matrix $\mathbf{W}_N^d$ given by:

$$\mathbf{W}_N^d = \left(\frac{1}{N}\sum_{i=1}^{N}\mathbf{Z}_{di}'\hat{\mathbf{\Delta}\varepsilon}_i\hat{\mathbf{\Delta}\varepsilon}_i'\mathbf{Z}_{di}\right)^{-1}$$

The first step comprises on using an identity weighting matrix to estimate $\hat{\alpha}$ consistently. Then, in a second step, the residuals from this estimation (i.e. $\hat{\mathbf{\Delta}\varepsilon}_i$) are used in the optimal choice of $\mathbf{W}_N^d$. This renders the two-step estimation of α.

4.3. **System GMM estimator.**

I now consider an additional assumption by imposing a restriction on the initial conditions process generating y_{i1} (Arellano and Bover, 1995; Blundell and Bond, 1998):

(6') $$E[\Delta y_{i2}\mu_i] = 0; \text{ for } i = 1, ..., N$$

Together with (1') to (4'), assumption (6') imply an additional set of $T-2$ orthogonality conditions given by:

(7') $$E[\Delta y_{it-1}\varepsilon_{it}] = 0; \text{ for } t = 3, ..., T$$

As in the case for the Difference GMM estimator, the conditions in (7') can be expressed as:

$$E[\mathbf{Z}_{li}'\varepsilon_i] = 0$$

where ε_i is the $T-2$ vector $(\varepsilon_{i3}, \varepsilon_{i4}, ..., \varepsilon_{iT})'$ and $\mathbf{Z}_{li}$ is the $(T-2)\times(T-2)$ instruments matrix given by:

$$\mathbf{Z}_{li} = \begin{bmatrix} \Delta y_{i2} & 0 & ... & 0 \\ 0 & \Delta y_{i3} & ... & 0 \\ . & . & ... & . \\ 0 & 0 & ... & \Delta y_{iT-1} \end{bmatrix}$$

or, when collapsing the matrix in order to reduce the instrument count, $\mathbf{Z}_{li}$ becomes a $T-2$ column vector such that:

$$\mathbf{Z}_{li} = \begin{bmatrix} \Delta y_{i2} \\ \Delta y_{i3} \\ ... \\ \Delta y_{iT-1} \end{bmatrix}$$

Blundell and Bond (1998) propose to exploit jointly the moments conditions in (5') and (7'). To do so, the dataset is stacked with transformed and untransformed observations by multiplying it by $\mathbf{M} = (\mathbf{M}_\Delta \ \mathbf{I})'$. This yields an augmented dataset of the form $\mathbf{Y}_i = (\mathbf{\Delta y}_i \ \mathbf{y}_i)'$ for each individual i, where $\mathbf{M}_\Delta$ is the matrix used in the first-difference transformation, a matrix with diagonal of $-1s$ and a sub-diagonal of $1s$ to the right.

By letting $\mathbf{E}_i$ be $(\mathbf{\Delta}\varepsilon_i \ \varepsilon_i)'$, conditions (5') and (7') can be expressed as a system of equations of the form:

$$E[\mathbf{Z}_i'\mathbf{E}_i] = 0$$

where $\mathbf{Z}_i$ considers both sets of instruments corresponding to (5') and (7') respectively and so it is defined by the $2(T-2)\times(T-2)(0.5(T-1)+1)$ matrix:

$$\mathbf{Z}_i = \begin{bmatrix} \mathbf{Z}_{di} & 0 \\ 0 & \mathbf{Z}_{li} \end{bmatrix}$$

The two-step System GMM estimator is then analogous to the one described for the Difference GMM. It solves:

$$\hat{\alpha}_{sys-GMM} = \underset{\alpha \in \Re}{\arg \min} \, \mathbf{E}'\mathbf{Z}\mathbf{W}_N^s\mathbf{Z}'\mathbf{E}$$

for some symmetric, positive semi definite weighting matrix $\mathbf{W}_N^s$, $\mathbf{Z}'$ the $(T-2)(0.5(T-1)+1) \times 2N(T-2)$ or its $(T-1) \times 2N(T-2)$ counterpart for collapsed instruments matrix $(\mathbf{Z}'_1, \mathbf{Z}'_2, ..., \mathbf{Z}'_N)$ and $\mathbf{E}'$ the $N(T-2)$ vector $(\mathbf{E}'_1, \mathbf{E}'_2, ..., \mathbf{E}'_N)$. The System GMM estimator is given by:

$$\hat{\alpha}_{sys-GMM} = \left(\mathbf{Y}'_{-1}\mathbf{Z}\mathbf{W}_N^s\mathbf{Z}'\mathbf{Y}_{-1}\right)^{-1}\mathbf{Y}'_{-1}\mathbf{Z}\mathbf{W}_N^s\mathbf{Z}'\mathbf{Y}$$

where $\mathbf{Y}$ and $\mathbf{Y}_{-1}$ are $N(T-2)$ vectors constructed as $\mathbf{\Delta y}$ and $\mathbf{\Delta y}_{-1}$ for the Difference GMM estimator.

Finally, the optimal two-step System GMM is obtained with a choice of the weighting matrix $\mathbf{W}_N^s$ given by:

$$\mathbf{W}_N^s = \left(\frac{1}{N}\sum_{i=1}^{N}\mathbf{Z}'_i\hat{\mathbf{E}}_i\hat{\mathbf{E}}'_i\mathbf{Z}_i\right)^{-1}$$

where the residuals $\hat{\mathbf{E}}_i$ are obtained from a first-step consistent estimation of α.

4.4. Multivariate models.
We now move to an autoregressive panel data model of the form:

(8')
$$y_{it} = \alpha y_{it-1} + \beta x_{it} + \varepsilon_{it}$$
$$\varepsilon_{it} = \mu_i + v_{it}$$

where μ_i and v_{it} satisfy (2') to (4').

In addition to the orthogonality conditions given by (5') and (7'), which remain valid, the presence of further explanatory variables in the model allows to increase the number of moment conditions depending on whether the x_{it} process is strictly exogenous, predetermined or endogenous (see Blundell et al. (2000) for additional details). Recall the x_{it} process is exogenous if:

(9') $$E\left[x_{is}v_{it}\right] = 0; \text{ for } s = 1, ..., T; \, t = 2, ..., T,$$

it is predetermined (or weakly exogenous) if:

(10')
$$E\left[x_{is}v_{it}\right] = 0; \text{ for } s = 1, ..., T; \, t = 2, ..., T$$
$$E\left[x_{is}v_{it}\right] \neq 0; \text{ for } s = t+1, ..., T; \, t = 2, ..., T$$

and it is endogenous if:

(11')
$$E\left[x_{is}v_{it}\right] = 0; \text{ for } s = 1, ..., t-1; \, t = 2, ..., T$$
$$E\left[x_{is}v_{it}\right] \neq 0; \text{ for } s = t, ..., T; \, t = 2, ..., T$$

The additional orthogonality conditions are implied by a combination of the errors component structure given by (2') to (4') and respectively (9'), (10') and (11'). For the first-difference transformation of the model, if the x_{it} process is strictly exogenous, the additional moment conditions are:

(12') $$E\left[x_{is}\Delta\varepsilon_{it}\right] = 0; \text{ for } t = 3, ..., T \text{ and } 1 \leq s \leq T,$$

on the other hand, if the x_{it} process is predetermined:

(13') $$E\left[x_{it-s}\Delta\varepsilon_{it}\right] = 0; \text{ for } t = 3, ..., T \text{ and } 1 \leq s \leq t-1,$$

and if the x_{it} process is endogenous, the respective additional orthogonality conditions are:

(14') $$E\left[x_{it-s}\Delta\varepsilon_{it}\right] = 0; \text{ for } t = 3, ..., T \text{ and } 2 \leq s \leq t-1$$

Likewise, there are additional moment conditions for the equation in levels. If the x_{it} process is strictly exogenous or predetermined, these conditions are:

(15') $$E\left[\Delta x_{it}\varepsilon_{it}\right] = 0; \text{ for } t = 2, ..., T,$$

and if the x_{it} process is endogenous:

$$(16') \hspace{5cm} E\left[\Delta x_{it-1}\varepsilon_{it}\right] = 0; \text{ for } t = 3, ..., T$$

4.5. Optimal GMM and relevant tests.

Although the Difference and System GMM estimators are both consistent under (5'), (7') and (12') to (16'), provided the orthogonality conditions in (7'), (15') and (16') are valid, System GMM is more efficient than Difference GMM. This issue has been supported with Montecarlo simulations studying the finite sample behavior of both estimators (see Blundell and Bond (1998), Blundell et al. (2000) for further details). The main reason for this gain in efficiency is that when explanatory variables show persistence or the number of time series observations is small relative to the number of individuals, the set of instruments of the Difference GMM estimator becomes weak since past levels of the variables carry out little information about future changes. Therefore, the informative power of additional instruments in the System GMM estimator takes relevance when performing the estimation (Arellano and Bover, 1995; Blundell and Bond, 1998; Blundell et al., 2000).

Furthermore, the validity of the additional moment conditions given by (7'), (15') and (16') can be regarded as an empirical issue as they correspond to overidentifying restrictions (Blundell and Bond, 1998). Therefore, the choice of System over Difference GMM needs to be substantiated with overidentification tests not only for the full set of instruments but for the suspicious set of instruments included in the System GMM estimation as well. Hence the Hansen and Difference-in-Hansen tests, specifying the latter so as to focus exclusively on the extra moment conditions, are of remarkable importance when assessing the use of the System GMM estimator and both tests should be reported in every estimation obtained by using this approach. In particular, section ?? of the paper discusses the robustness of my results according to overidentifying restrictions.

On the other hand, the presence of serial autocorrelation in the idiosyncratic disturbance term v_{it} renders some specific lags, both levels and differences, invalid as instruments (Arellano and Bond, 1991). As pointed out by Roodman (2009b), the full disturbance term ε_{it} is presumed to be autocorrelated since it contains the individual specific effect μ_i, which is handled by the design of the Difference and System GMM estimators. However, if for instance the idiosyncratic disturbances exhibit serial autocorrelation of order 1, then the first lag of the endogenous variable y_{it-1}, i.e. y_{it-2} is also endogenous. To control for this situation, Arellano and Bond (1991) develop a test for serial autocorrelation of the idiosyncratic disturbance aside from the fixed effects. Specifically, the test is applied to residuals of the equations in differences. Hence, the validity of lags 2 and further as instruments for endogenous variables requires no serial autocorrelation of order 2. All results in section ?? report Arellano and Bond (1992) tests for serial correlation of order 1 and 2 for all estimates and support the specification of the model as discussed in detail in section ??.

Moreover, a relevant issue to consider when following the generalized-method-of-moments approach is instrument proliferation. This is particularly important in the Difference or System GMM estimators as the number of instruments tends to increase largely. A first concern is a bias in coefficient estimates because instruments over-fit instrumented variables (Ziliak, 1997; Windmeijer, 2005). Second, Andersen and Sørensen (1996) and Bowsher (2002) show that a large number of instruments undermines the Hansen test of overidentification restrictions by increasing the likelihood of false-positive results. There are two main techniques used to reduce the instrument count and overcome these difficulties. Either to use some out of all available lags as instruments or to combine them by collapsing the instruments matrix without dropping any lags to retain more information. Roodman (2009b) provides evidence in favor of the second alternative by using Montecarlo simulations. Collapsing the instruments matrix has been increasingly used in empirical studies and therefore We apply this technique in all estimations (see for instance Christiaensen et al. (2011), Heid et al. (2012) and Beck and Levine (2004) for applications using System GMM estimators).

A final concern is related to a severe downward bias in standard errors when performing the two-step generalized-method-of-moments estimation in small samples. The reason for this bias

is that the estimated variance does not take into consideration an extra variation due to the parameters used in the first-step weighting matrix (Windmeijer, 2005). This issue can be handled by implementing a correction proposed by Windmeijer (2005) in which the extra variation is estimated to adjust the variance. All results reported in section **??** include the Windmeijer (2005) correction.