Shocks or Measurement Error?
Evidence from asset accumulation in Ethiopia

Melinda Vigh*
Tinbergen Institute and VU University Amsterdam

June 7, 2011

Abstract

Large fluctuations in survey data on the income and asset holdings of rural households in Ethiopia suggest that these households are forced to cope with an enormous amount of uninsured risk. However, these observed fluctuations could be substantially exaggerated due to measurement errors. To address this issue, we decompose the error term of dynamic panel data models into true shocks and measurement error. Using survey data from Ethiopia we estimate a linearized optimal livestock accumulation rule. Using these data we find that we cannot reject the hypothesis that random fluctuations in the observed livestock holdings are driven by true shocks.

Keywords: measurement error, asset accumulation, dynamic panel data
JEL: D14

1 Introduction

Large fluctuations in survey data on the income and asset holdings of rural households in Ethiopia suggest that these households are forced to cope with a substantial amount of uninsured risk. For example, the mean coefficient of variation of households' reported livestock holdings is 0.71, while that of their real income is 0.64 over a 10 year period for the surveyed households in the Ethiopian Rural Household Survey (ERHS). Looking at this magnitude of variation in the income and livestock holdings of households, the question arises that how large are the shocks to the livelihood of these households, and what is the magnitude of their impact on their growth and welfare in the absence of functioning insurance and credit mechanisms? In an attempt to address the first part of this question we analyze the asset accumulation dynamics of rural households using household survey data.

Obviously, large fluctuations in observed asset holdings can be caused by the combination of two factors: uninsured risks to households' economic and social environment, and imprecise measurement of asset ownership in household surveys.

Regarding risk, heavy reliance on agriculture coupled with the rather limited availability of insurance and credit services leave these households at the mercy of good weather and hope. For example, Dercon (2002) reports that 78 percent of the households in the Ethiopian Rural Household Survey suffered from crop failures due to flood, drought, frost and other climatic events, while 39 percent of them were affected by animal diseases, lack

*I would like to thank prof. Chris Elbers and prof. Jan Willem Gunning for their helpful comments and suggestions at various stages of this project. All remaining errors are my own.

of water and grazing land in the past 20 years. Others like Townsend (1994) and Kinsey et al. (1998) in India and Zimbabwe, respectively, also find large fluctuations in harvest due to erratic weather conditions.

Addressing the impact of risk on asset accumulation in such an environment, Elbers et al. (2007) and Pan (2008) estimate the structural parameters of a Ramsey model using data from Zimbabwe and the ERHS data, respectively. Their results imply that households are forced to cope with a huge amount of income and asset risks. Elbers et al. (2007) report that if households could insure against all income and asset shocks, their livestock holdings would be on average 46 percent higher over a 50-year period. Pan (2008) finds similar results over a 90-year period for a median household in the Ethiopian dataset.

The above results suggest that uninsured risks loom large in rural villages. However, noting the huge variance of livestock holdings over time in the ERHS dataset, we can suspect that measurement error in the self-reported measure of livestock holdings magnifies the size of true shocks in the estimation. Since the estimation method used in Elbers et al. (2007) and Pan (2008) does not allow for measurement error, their estimates on risk are upward bounds on the true magnitude of risk.

The present paper develops a method that can separate measurement error from the true shocks in a linear dynamic panel model and applies it on the livestock accumulation dynamics of rural households in the ERHS dataset. The method is based on the fact that measurement error does not affect the underlying dynamics of the true but unobserved dependent variable; it only has a short-term effect on the observed level of the variable, while true shocks have a persistent effect. Using the difference in the structure of the two error components we can disentangle the two and find consistent estimators for the parameters of interest.

The estimation procedure is implemented in two steps. First, the autoregressive coefficient and the coefficients of the other variables are estimated using generalized method of moments (GMM). The commonly used Arellano-Bond and system estimators have to be adapted to the inclusion of measurement error because measurement error induces serial correlation in the error components. Bond et al. (2001) discusses consistent estimators for this case. In the second step, the error term is decomposed into true shocks and measurement error using the covariance matrix of the residuals and the parameter estimates from the first step.

The method outlined above can be easily implemented if the observations are from consecutive periods or there are no additional variables in the regression. However, a large part of the (household) survey data are collected several years apart like the ERHS dataset. In this case, standard estimation procedures of multivariate dynamic panel data models will yield inconsistent parameter estimates. This happens because lagged values of the additional regressors, which determine the unobserved lagged dependent variables in the unobserved years, are swept into the residuals. In this paper, I discuss that a possible solution for this case is using nonlinear GMM with coefficient restrictions on the lagged additional variables. This method, however, works only if the dynamics of the additional variables are known or can be estimated.

Another issue is when observations are not equally spaced over time. In this case, first differencing over the available observations does not eliminate the fixed effects, and therefore the Arellano-Bond and system estimators become inconsistent. However, if the initial conditions are satisfied the level (nonlinear) GMM estimator remains consistent. In the estimation procedure, I rely on this approach because there are not sufficient equally spaced observation available in the ERHS dataset.

Using this method, I estimate a log-linearized version of the optimal livestock accumulation rule of Pan (2008) using self-reported livestock data from the ERHS dataset from 1994-2004. The estimation results of the univariate model suggest high persistence of the livestock holdings and limited growth. However, the estimate of the magnitude of shocks is sizable. In the multivariate regression controlling for land and labor size, the estimates are less precise but the results remain similar. Based on the Sargan statistics of the level GMM estimator, I find that I cannot reject the hypothesis that measurement error does not drive the random fluctuations in data.

The structure of the paper is as follows. Section 2 describes the economic model. Section 3 discusses the estimation strategy. Section 4 introduces the ERHS dataset used in the estimation. Section 5 presents the estimation results. Finally, section 6 concludes.

2 The model

The analysis of asset accumulation decisions of the households is based on a general model which is also used in Elbers et al. (2007) and Pan (2008). In this model, we can write the optimal asset accumulation rule for asset K and household i as

$$K_{i,t} = \psi(\omega(K_{i,t-1}, X_{it}, s_{it})), \tag{1}$$

where $\psi(\cdot)$ denotes the policy function; and $\omega(\cdot)$ describes the wealth at hand the households in period t. Wealth at hand is the sum of household income and accumulated asset holdings, therefore it is determined by the level of asset holdings K_{it}, household production factors X_{it} and shocks to wealth s_{it} that have a variance of σ_s^2 and mean zero. A limitation of this formulation is that it does not allow for serially correlated shocks. However, given the data used in this paper this is not a serious restriction as the observations are several years apart.

I avoid using household income directly in the model because the measures of income from household surveys are notoriously prone to measurement error, even more so than asset stock data. This is the reason why Elbers et al. (2007) uses a model of livestock accumulation rather than income dynamics to assess the size of wealth shocks. However, the most important reason for analyzing the asset accumulation over the income process is that the identification strategy of measurement error and true shocks relies on the dynamic formulation of the model. We assume that household income is a function of production factors such as land, labor, lagged asset holdings, other household specific characteristics and uninsured income and wealth shocks that are contained in s.

To estimate the model, I log-linearize (1). This yields the following linear dynamic panel model:

$$\tilde{k}_{it} = \alpha_i + \beta_i \tilde{k}_{i,t-1} + \gamma_i \tilde{x}_{it} + e_{it}, \tag{2}$$

where $\tilde{k}_{it} = \log(K_{it})$, $\tilde{x}_{it} = \log(X_{it})$, and $e_{it} \sim N(0, \sigma_{e,i}^2)$. I assume that $E(e_{i,t}e_{i,s}) = 0$ and $E(e_{i,t}e_{j,t}) = 0$ for all $t \neq s$ and $j \neq i$. The coefficients of (2) are individual specific if the long-run level of k and x differs among households. In the estimation, I assume common coefficients for β, γ and σ_e^2, however I let α_i to be individual specific with mean α and individual specific level η_i. Hence, I estimate the following model

$$\tilde{k}_{it} = \alpha + \beta \tilde{k}_{i,t-1} + \gamma \tilde{x}_{it} + \eta_i + e_{it}, \tag{3}$$

Note that σ_e^2 can no longer be interpreted as the magnitude of wealth shocks but as the size of fluctuation in asset holdings due to wealth shocks.

3 Econometric method

I assume that asset holdings are measured with an additive error but the other production factors are observed accurately.[1] Hence,

$$\begin{aligned} k_{it} &= \tilde{k}_{it} + m_{it}, &(4)\\ x_{it} &= \tilde{x}_{it}, &(5)\end{aligned}$$

where k_{it} and x_{it} are the log of observed livestock and other production factor holdings, respectively; m_{it} is the measurement error and it is distributed independently across households and over time with $m_{it} \sim N(0, \sigma_m^2)$ for all i, t. Combining (3) and (4) yields the equation to estimate:

$$k_{it} = \alpha + \beta k_{i,t-1} + \gamma x_{it} + \eta_i + e_{it} + \mu_{it}, \tag{6}$$

where $\mu_{it} = m_{it} - \beta m_{i,t-1}$ is the measurement error component of the error term, η_i is the individual specific fixed effect and e_{it} is the shock to livestock holdings.

Correlation between $k_{i,t-1}$ and the latent error terms, μ_{it} and η_i, complicate the consistent estimation of the model parameters. Assuming that x_{it} is uncorrelated with the error terms, observe that the long-run correlation of the lagged asset holdings with the latent error terms are

$$\begin{aligned} E(k_{i,t-1}\eta_i) &= \sigma_\eta^2/(1-\beta) > 0, &(7)\\ E(k_{i,t-1}\mu_{it}) &= -\beta\sigma_m^2 < 0, &(8)\end{aligned}$$

where the first inequality is due to the recursive nature of k, and the second inequality arises due to the fact that the measurement error component of the observed livestock holdings in the previous period does not drive the asset dynamics, and hence it ends up in the error term. In addition, x_{it} can be also correlated with e_{it}, η_i or m_{it}.

In the following, I first address the estimation of β in the absence of other regressors, x_{it}. Then, I discuss the inclusion of additional regressors and the estimation of β and γ, specification testing, and the decomposition of the error terms. Subsequently I examine the case where observations are more than one period apart, possibly at different intervals.

GMM estimation without additional regressors

In the absence of additional regressors the model becomes[2]

$$\begin{aligned} k_{it} &= \alpha + \beta k_{i,t-1} + u_{it}, &(9)\\ u_{it} &= \eta_i + e_{it} + \mu_{it}. &(10)\end{aligned}$$

I assume that

$$\begin{aligned} E(e_{it}e_{is}) = E(m_{it}m_{is}) &= 0 \quad \text{for all } t \neq s, &(11)\\ E(e_{it}m_{is}) &= 0 \quad \text{for all } t, s. &(12)\end{aligned}$$

and $E(e_{it}) = E(m_{it}) = 0$, but $E(\eta_i) \neq 0$.

[1] Measurement error in the additional regressors does not affect the consistency of the estimates of β and γ, however, without further information it complicates the decomposition of the error term.

[2] The following section relies on Blundell et al. (2000) and Bond et al. (2001).

Due to (7) and (8) the ordinary least squares (OLS) estimator of β is inconsistent.[3] Taking first differences (6):

$$\begin{aligned} \Delta k_{it} &= \beta \Delta k_{i,t-1} + \Delta u_{it}, \qquad (13) \\ \Delta u_{it} &= \Delta e_{it} + \Delta \mu_{it} \qquad (14) \end{aligned}$$

eliminates the individual specific fixed effects, however the correlation between the lagged livestock holdings and measurement error becomes more pronounced:

$$E(\Delta k_{i,t-1} \Delta \mu_{it}) = -(1+\beta)^2 \sigma_m^2 < 0, \qquad (15)$$

and it creates dependency between the livestock holdings and the shocks:

$$E(\Delta k_{i,t-1} \Delta e_{it}) = -\sigma_e^2 < 0. \qquad (16)$$

Therefore, the OLS estimator over the first differenced regression is also inconsistent.

Arellano and Bond (1991) suggested using lagged dependent variables in levels as instruments for the differenced lagged dependent variable in (13) to estimate β consistently because

$$E(k_{i,t-s} \Delta e_{it}) = E(k_{i,t-s}(e_{it} - e_{i,t-1})) = 0 \quad \text{for } 2 \leq s \leq t-1, \qquad (17)$$

and

$$E(k_{i,t-s} \Delta \mu_{it}) = E(k_{i,t-s}(m_{it} - (1+\beta) m_{i,t-1} + \beta m_{i,t-2})) = 0 \quad \text{for } 3 \leq s \leq t-1. \qquad (18)$$

In the absence of measurement error, that implies using y_{i1} as instrument for Δy_{i3}; y_{i1}, y_{i2} for Δy_{i4}, and so forth. Hence, we need at least $T = 3$ observations on the households to be able to estimate the model parameters consistently in this case. However, when the dependent variable is measured with error, in (18) we observe that y_{i1} can only be used as instrument for Δk_{i4} and above. Therefore, in this case we need at least $T = 4$ observations over the households for consistent estimators for the model parameters. The moment restrictions in (17) and (18) can be expressed more compactly as

$$E(Z'_{di} \Delta u_i) = 0 \qquad (19)$$

where Z_{di} is the $(T-3) \times 0.5(T-1)(T-3)$ matrix

$$Z_{di} = \begin{bmatrix} k_{i1} & 0 & 0 & \cdots & 0 & \cdots & 0 \\ 0 & k_{i1} & k_{i2} & \cdots & 0 & \cdots & 0 \\ . & . & . & \cdots & . & \cdots & . \\ 0 & 0 & 0 & \cdots & k_{i1} & \cdots & k_{i,T-3} \end{bmatrix}. \qquad (20)$$

Above, I have looked at the validity of these instruments, whether they are uncorrelated with the error terms. However, it is equally important that they are also relevant, i.e. they have explanatory power over the dependent variable, otherwise the estimator will have poor finite sample properties in terms of precision and bias. In the absence of additional regressors and measurement error, Blundell and Bond (1998) show that the level instruments suggested by Arellano and Bond (1991) become weak if β is close to 1 or the relative variance of the fixed effects to the true shocks is large. To detect serious finite sample bias using the Arellano-Bond estimator, Bond et al. (2001) suggest comparing the

[3] Note that the two sources of inconsistency drive the OLS estimate of β in different directions: fixed effects cause an overestimation of the autoregressive parameter, while measurement error causes a downward bias.

estimate of β to the estimates using the Within Group[4] estimator and the OLS estimator. They argue that the Arellano-Bond estimator of β should lie between the Within Group and the OLS estimates of β, because the first is downward and the latter is upward biased. Otherwise, the Arellano-Bond estimator increases the bias of the estimate of β.

The problem of weak instruments is aggravated by the presence of measurement error, as a further lag is required for a valid instrument. This decreases the correlation between the endogenous regressor, $\Delta k_{i,t-1}$, and the instrument. In addition, measurement error adds noise to all observations of k, therefore the correlation between $\Delta k_{i,t-1}$ and $k_{i,t-3}$ is further reduced.

Arellano and Bover (1995) propose additional moment conditions for the regression in level

$$E(\Delta k_{i,t-s} u_{it}) = 0 \quad \text{for } 1 \leq s \leq t-2 \tag{21}$$

in the absence of measurement error, if the initial conditions of the model satisfy the assumption:

$$E(\eta_i \Delta k_{i2}) = 0 \tag{22}$$

for all i. This initial condition restriction is satisfied if k_{it} is a *mean stationary* process[5] or the activity described by the model has been going on for a sufficiently long time prior to the sample period, so that the influence of the true initial conditions are negligible. This is likely to be the case for asset accumulation dynamics, as the households have to make the same decisions over and over again.[6]

In the presence of measurement error (21) holds only for $2 \leq s \leq t-2$ assuming (22). Therefore, Δk_{i2} can be used as an instrument for k_{i3} in the regression in levels, which imply the moment conditions $E(\Delta k_{i2} u_{i4}) = 0$. The moment conditions for estimating the level regression can be written compactly as

$$E(Z'_{li} u_i) = 0 \tag{25}$$

where Z_{li} is the $(T-3) \times 0.5(T-1)(T-3)$ matrix

$$Z_{li} = \begin{bmatrix} \Delta k_{i2} & 0 & 0 & \cdots & 0 & \cdots & 0 \\ 0 & \Delta k_{i2} & \Delta k_{i3} & \cdots & 0 & \cdots & 0 \\ . & . & . & \cdots & . & \cdots & . \\ 0 & 0 & 0 & \cdots & \Delta k_{i2} & \cdots & \Delta k_{i,T-2} \end{bmatrix}. \tag{26}$$

Note that (9) includes a constant, α, which can be identified using the moment condition

$$E(\iota' u_{it}) = 0, \tag{27}$$

where ι denotes the N-vector of ones.

[4]The Within Group estimator transforms the variables in the original regression in levels by taking the deviations from the individual specific means, thereby getting rid of the individual specific fixed effects. For further details see for example Arellano (2003).

[5]Conditions for mean stationarity are the following:

$$y_{i1} = \frac{\alpha + \eta_i}{1-\beta} + e_{i1} + m_{i1} \quad \text{and} \tag{23}$$

$$E(e_{i1}) = E(\eta_i e_{i1}) = E(m_{i1}) = E(\eta_i m_{i1}) = 0 \quad \text{for all } i. \tag{24}$$

[6]This assumption can be violated in the case of households that just started to accumulate assets, and they do not have a steady amount of assets.

Moment conditions (25)-(27) for the regression in level can be also used in combination with the Arellano-Bond moment conditions as suggested by Arellano and Bover (1995). In this case, from the regression in level there are (T-1) additional moment conditions

$$E(\Delta k_{i,t-2} u_{it}) = 0 \tag{28}$$

for the estimation because $E(\Delta k_{i,t-s} u_{it}) = 0$ for $s \geq 3$ is implied by the identity

$$E(\Delta k_{i,t-s} u_{it}) - E(\Delta k_{i,t-s} u_{i,t-1}) = E(k_{i,t-s} \Delta u_{it}) - E(k_{i,t-s-1} \Delta u_{it}), \tag{29}$$

where the terms on the right hand side are moment conditions of the Arellano-Bond estimator and the second term on the left hand side follows from moment condition (28).

Stacking the first differenced and the level regression in the *system estimator* implies the following formulation of the moment conditions:

$$E(Z'_{si} \omega_i) = 0, \tag{30}$$

where

$$\omega_i = \begin{bmatrix} \Delta u_i \\ u_i \end{bmatrix} \tag{31}$$

and

$$Z_{si} = \begin{bmatrix} Z_{di} & 0 & 0 & \cdots & 0 & 0 & 0 & \cdots & 0 \\ 0 & \Delta k_{i2} & 0 & \cdots & 0 & \iota & 0 & \cdots & 0 \\ 0 & 0 & \Delta k_{i3} & \cdots & 0 & 0 & \iota & \cdots & 0 \\ \cdot & \cdot & \cdot & \cdots & \cdot & \cdot & \cdot & \cdots & \cdot \\ 0 & 0 & 0 & \cdots & \Delta k_{i,T-2} & 0 & 0 & \cdots & \iota \end{bmatrix}. \tag{32}$$

Multivariate regression

In the previous section, I have only considered the estimation of β and α. However, model (6) also includes other regressors. The treatment of these variables in the estimation procedure depends on the assumptions about its correlation with the error components of (6). Blundell et al. (2000) discuss the possible cases in detail. First, I consider cases where x_{it} is allowed to be correlated with the individual fixed effects, i.e. $E(x_{it}\eta_i) \neq 0$, then I briefly discuss the case where x_{it} is independent of the fixed effects. Concerning the true shocks and the measurement error, I consider two cases:

1. x_{it} is *predetermined*:

$$E(x_{i,t-s} e_{it}) = 0 \quad \text{for } s \geq 0 \quad \text{and} \quad E(x_{i,t+s} e_{it}) \neq 0 \quad \text{for } s > 0, \tag{33}$$

$$E(x_{is} \mu_{it}) = 0 \quad \text{for all } s, t. \tag{34}$$

In this case, the following moment conditions can be added to the first-differenced GMM estimator:

$$E(x_{i,t-s} \Delta u_{it}) = 0 \quad \text{for } 1 \leq s \leq t-1. \tag{35}$$

Hence, x_{i3}, x_{i2}, x_{i1} can be used as instruments for $\Delta x_{i4}, \Delta y_{i3}$.

2. x_{it} is *endogenously determined*:

$$E(x_{i,t-s} e_{it}) = 0 \quad \text{for } s \geq 1 \quad \text{and} \quad E(x_{i,t+s} e_{it}) \neq 0 \quad \text{for } s \geq 0, \tag{36}$$

$$E(x_{is} \mu_{it}) = 0 \quad \text{for all } s, t. \tag{37}$$

Correlation between x_{it} and e_{it} can arise due to shocks that directly affect both livestock holdings and the other production factors. It is important to note, that allowing for contemporaneous correlation between x_{it} and the error components can accommodate measurement error in these regressors, as the measurement error of x_{it} would be swept into the error components with the term $-\gamma m_{x,it}$. This formulation allows for the measurement error of x_{it} and k_{it} to be correlated contemporaneously but not serially.

Allowing for contemporaneous correlation induces the following additional moment conditions for the first-differenced GMM estimator:

$$E(x_{i,t-s}\Delta u_{it}) = 0 \quad \text{for } 2 \leq s \leq t-1. \tag{38}$$

Hence, x_{i2}, x_{i1} can be used as instruments for $\Delta x_{i4}, \Delta y_{i3}$.

In case of the level and system GMM estimators, both Δx_{it} and Δk_{it} have to be uncorrelated to the individual fixed effects, hence

$$E(\Delta x_{it}\eta_i) = E(\Delta k_{it}\eta_i) = 0 \quad \text{for all } t \tag{39}$$

has to hold. Blundell et al. (2000) show that this condition is satisfied if both x_{it} and k_{it} are a mean stationary processes. Further, they argue that it is sufficient for (39) to hold if both processes have been going on for a sufficiently long time prior to the sample period such that the true initial conditions are negligible for both processes. In this case, provided that Δx_{it} is uncorrelated with η_i, Δy_{it} will be uncorrelated with η_i even if the mean of x_{it}, and therefore y_{it} is time-varying.

Note that in case x is predetermined, $T = 3$ is sufficient to identify β and γ in the multivariate regression using the Arellano-Bond estimator because x_{i2}, x_{i1} can be used as instruments for $\Delta k_{i2}, \Delta x_{i3}$. Hence, the additional regressors in the dynamic specification yield valid instruments for the autoregressive parameter β, if x is predetermined and its the lagged values are observed.

If both Δx_{it} and Δy_{it} are uncorrelated with η_i, then in addition to (21) the following moment conditions are valid:

$$E(\Delta x_{i,t-s}u_{it}) = 0 \quad \text{for } 0 \leq s < t-2, \tag{40}$$

if x_{it} is predetermined, and

$$E(\Delta x_{i,t-s}u_{it}) = 0 \quad \text{for } 1 \leq s < t-2, \tag{41}$$

if x_{it} is endogenously determined. However, in the case of the system estimator only $s = 0$ and $s = 1$ contains additional information to the Arellano-Bond moment conditions for predetermined and endogenous x, respectively.

Table 3 summarizes the list of instruments for the different GMM estimators discussed above.

Finally, note that the constant term with coefficient α can also be thought of as an additional regressor in the model. However, due to its time-invariant nature it is only possible to identify α using a regression in levels but not using the first differenced Arellano-Bond estimator. It is also a special regressor, since it is by construction orthogonal to the error terms, in particular to the fixed effects. Therefore, it is an exogenous regressor and it can be instrumented by itself, as done in (27).

Instruments	$\sigma_m^2 = 0$	$\sigma_m^2 \neq 0$
Level estimator		
Δk_{t-s}	$1 \leq s \leq t-2$	$2 \leq s \leq t-2$
Δx_{t-s} (predetermined)	$0 \leq s \leq t-2$	$0 \leq s \leq t-2$
Δx_{t-s} (endogenous)	$1 \leq s \leq t-2$	$1 \leq s \leq t-2$
FD estimator		
k_{t-s}	$2 \leq s \leq t-1$	$3 \leq s \leq t-1$
x_{t-s} (predetermined)	$1 \leq s \leq t-1$	$1 \leq s \leq t-1$
x_{t-s} (endogenous)	$2 \leq s \leq t-1$	$2 \leq s \leq t-1$
System estimator		
k_{t-s}	$2 \leq s \leq t-1$	$3 \leq s \leq t-1$
x_{t-s} (predetermined)	$1 \leq s \leq t-1$	$1 \leq s \leq t-1$
x_{t-s} (endogenous)	$2 \leq s \leq t-1$	$2 \leq s \leq t-1$
Δk_{t-s}	$s = 1$	$s = 2$
Δx_{t-s} (predetermined)	$s = 0$	$s = 0$
Δx_{t-s} (endogenous)	$s = 1$	$s = 1$
Min number of obs.		
x endogenous or absent	$T = 3$	$T = 4$
x predetermined	$T = 3$	$T = 3$

Table 1: List of instruments for GMM estimation

Specification tests

The standard test for testing the validity of moment conditions used in the GMM estimation procedure is the Sargan test of overidentifying restrictions (see Hansen (1982)). The Sargan test statistic is given by

$$S = \frac{1}{N} \hat{r}' Z W_n Z' \hat{r} \tag{42}$$

where $\hat{r}$ are the estimated residuals, Z is the matrix of instrumental variables used in the estimation, and W_n is the weighing matrix in the estimation:

$$W_n = \left(\frac{1}{N} \sum_{i=1}^{N} Z_i' \hat{r}_i \hat{r}_i' Z_i \right)^{-1} . \tag{43}$$

Under the null hypothesis that the moment conditions are valid, the test statistic is asymptotically distributed chi-squared with $m - k$ degrees of freedom, where m is the number of moment conditions and k is the number of estimated parameters.

A test for the validity of the level moment conditions can be obtained as a difference between the S-statistic of the system estimator and the S-statistic of the first-difference estimator. Under the null hypothesis that the level moment conditions are valid, the test statistic is asymptotically distributed chi-squared with $m_s - m_d$, where m_s and m_d are the number of moment conditions used in the system and first-difference estimation, respectively.

Similarly, it can be tested whether x_{it} is predetermined or endogenously determined. Under the null hypothesis that x_{it} is predetermined, the statistics is the difference between the S-statistic including x_{it} or Δx_{it} as instruments in the estimation and the S-statistic excluding x_{it} or Δx_{it} from the list of instruments. Under the null hypothesis, this test is distributed chi-squared with the dimension of x_{it} as the degrees of freedom in case of the first-differenced or level GMM estimator, and twice the dimension of x_{it} in case of the system estimator.

Decomposing shocks and measurement error

After estimating β and γ and testing the consistency of the estimates using the specification tests described above, I turn to the main interest of the paper, namely, the decomposition of the error components into true shocks and measurement error. To eliminate the individual fixed effects we take the first-differenced estimated residuals from (6):

$$\widehat{\Delta\omega}_{it} = \Delta k_{it} - \hat{\beta}\Delta k_{i,t-1} - \hat{\gamma}\Delta x_{it}. \tag{44}$$

If $T = 4$, it is possible to calculate the covariance matrix of the estimated residuals for two periods:

$$Cov(\widehat{\Delta\omega}_{i3}, \widehat{\Delta\omega}_{i4}) = \begin{bmatrix} \hat{\sigma}_3^2 & \hat{\sigma}_{34} \\ \hat{\sigma}_{34} & \hat{\sigma}_4^2 \end{bmatrix}. \tag{45}$$

From (6) follows, that if the estimates of β and γ are consistent, and the process has been going on for a sufficiently long time, $\widehat{\Delta\omega}$ has a stationary variance, hence both $\hat{\sigma}_3^2$ and $\hat{\sigma}_4^2$ tend to the same probability limit. Therefore, we estimate $\sigma_{\Delta\omega}^2$ by the mean of $\hat{\sigma}_3^2$ and $\hat{\sigma}_4^2$. The first-order auto-covariance between $\Delta\omega$, $\rho_{\Delta\omega}$, is estimated by $\hat{\sigma}_{34}$.

Now, let us consider the components of $\Delta\omega$:

$$\Delta\omega_{it} = \Delta e_{it} + \Delta m_{it} - \beta\Delta m_{i,t-1}. \tag{46}$$

From here, it follows that

$$\begin{aligned} \sigma_{\Delta\omega}^2 &= 2\sigma_e^2 + 2(1 + \beta + \beta^2)\sigma_m^2, && (47) \\ \rho_{\Delta\omega} &= -\sigma_e^2 - (1 + \beta)^2\sigma_m^2. && (48) \end{aligned}$$

Solving this system of equations, the estimates of the error components are

$$\begin{aligned} \hat{\sigma}_m^2 &= -\frac{\hat{\sigma}_{\Delta\omega}^2 + 2\hat{\rho}_{\Delta\omega}}{2\hat{\beta}}, && (49) \\ \hat{\sigma}_e^2 &= -\hat{\rho}_{\Delta\omega} - (1 + \hat{\beta})^2\hat{\sigma}_m^2. && (50) \end{aligned}$$

This approach can be easily extended for larger T, as well. As a practical note, minimizing the squared deviations in (49) instead of solving the system can be useful to guarantee the non-negativity of the variance estimates. Finally, it is also possible to recover the sample variance of the fixed effects η_i from

$$\begin{aligned} \hat{\omega}_{it} &= k_{it} - \hat{\beta}k_{i,t-1} - \hat{\gamma}x_{it}, && (51) \\ Var(\omega_{it}) &= \sigma_e^2 + Var(\eta) + (1 + \beta^2)\sigma_m^2. && (52) \end{aligned}$$

Tests of variance

I use bootstrapping to calculate the standard errors and the 95 percent confidence interval of the parameter estimates. Then, testing whether one of the coefficients is equal to θ_0 can be implemented by observing whether θ_0 is included in the confidence interval of the coefficient estimate. This method can be applied to transformations of the parameter estimates, as well.

Hence, we can test whether σ_m^2 or σ_e^2 is significantly different from zero by observing whether $\varepsilon > 0$ is included in the confidence interval of the coefficient for a small ε. Note that if x_{it} is measured with error, its variance is going to inflate σ_e^2.

The presence of measurement error can also be tested directly using a Sargan test. The test involves testing the validity of the additional instruments that can be used in the estimation in the absence of measurement error. In case of the Arellano-Bond estimator it is testing the validity of k_{t-2}, and in the case of the level GMM estimator it is testing the validity of Δk_{t-1}. The Sargan statistic in both cases (and also in the case of the system estimator) is distributed at chi-squared under the null hypothesis that the additional instruments are valid with one degrees of freedom.

Estimation with distant observations

Data availability can limit the analysis to observations that are more than one period apart.[7] In this section I investigate the consequences of distant but equally spaced observations on the estimation procedure. In the next section, I look at estimators with observations at different intervals. For simplicity, I omit the use of the individual index in the following sections.

To analyze (6) with distant observations, it is instructive to rewrite k_t as the accumulation of previous values of regressors and shocks, fixed effects and measurement error since an initial period $(t-\tau)$ with asset holdings $k_{t-\tau}$:

$$k_t = \alpha \sum_{s=0}^{\tau-1} \beta^s + \beta^\tau k_{t-\tau} + \gamma \sum_{s=0}^{\tau-1} \beta^s x_{t-s} + \eta \sum_{s=0}^{\tau-1} \beta^s + \sum_{s=0}^{\tau-1} \beta^s e_{t-s} + m_t - \beta^\tau m_{t-\tau}. \quad (53)$$

First, assuming that there are no additional regressors ($\gamma = 0$) the model becomes

$$k_t = \alpha \sum_{s=0}^{\tau-1} \beta^s + \beta^\tau k_{t-\tau} + \eta \sum_{s=0}^{\tau-1} \beta^s + e_t + \beta e_{t-1} + \ldots + \beta^{\tau-1} e_{t-\tau+1} + m_t - \beta^\tau m_{t-\tau}. \quad (54)$$

In this case the level instruments $\Delta k_{t-2\tau}, \Delta k_{t-3\tau}, \ldots$ are valid if the first observation of Δk is uncorrelated with the fixed effects. For the order τ differences of (54)

$$\Delta^\tau k_t = \beta^\tau \Delta^\tau k_{t-\tau} + \Delta^\tau e_t + \beta \Delta^\tau e_{t-1} + \ldots + \beta^{\tau-1} \Delta^\tau e_{t-\tau+1} + \Delta^\tau m_t - \beta^\tau \Delta^\tau m_{t-\tau}, \quad (55)$$

instruments $k_{t-3\tau}, k_{t-4\tau}, \ldots$ are also valid. Hence, in the absence of additional regressors, the methods outlined in the previous sections remain valid. Note that the interpretation of the parameter estimates change, as the estimate of β becomes the estimate of β^τ, the estimate of α becomes $\alpha \sum_{s=0}^{\tau-1} \beta^s$, and the estimate of σ_e^2 becomes $\tau \sigma_e^2$.

In case $\gamma \neq 0$, the additional regressors have to be considered. It is visible in (53) that k_t on $k_{t-\tau}$ and x_t is also a function of all x's between periods $(t-1)$ and $(t-\tau+1)$. If these terms end up among the residuals and if instruments $x_{t-\tau}, x_{t-2\tau}, \ldots$ are correlated with regressor x_t (i.e. they are relevant), then they are also going to be correlated with the x terms in the residual $(x_{t-1}, \ldots, x_{t-\tau+1})$. Hence, instruments $x_{t-\tau}, x_{t-2\tau}, \ldots$ are either invalid (correlated with the residual) or irrelevant (do not explain future x's).

[7] If β is close to one, it is possible to decrease the bias of the Arellano-Bond estimator by using higher order differences in the data.

A solution for this problem is specifying the dynamics of the variables in x. Let's assume that the variables in x also follow an autoregressive process:

$$x_t = \mu_0 + \delta x_{t-1} + \mu + v_t, \tag{56}$$

where μ is an individual fixed effect possibly correlated with η and $E(\mu) = 0$, μ_0 is the constant, and shocks v_t are distributed $N(0, \sigma_v^2)$. Note that I assume that x's are not affected by measurement error for simplicity, but it would be easy to extend the specification to allow for measurement error.

Then, we can write (53) conditioning on $x_{t-\tau}$ as

$$\begin{aligned} k_t \quad = \quad & \alpha \sum_{s=0}^{\tau-1} \beta^s + \beta^\tau k_{t-\tau} + \gamma x_t + \gamma \left(\sum_{s=1}^{\tau-1} \beta^s \delta^{\tau-s} \right) x_{t-\tau} \\ & + \gamma(\mu_0 + \mu) \left(\sum_{s=1}^{\tau-1} \sum_{j=0}^{\tau-s-1} \beta^s \delta^j \right) + \gamma \left(\sum_{s=1}^{\tau-1} \sum_{j=1}^{s} \beta^j \delta^{s-j} v_{t-s} \right) \\ & + \eta \sum_{s=0}^{\tau-1} \beta^s + \sum_{s=0}^{\tau-1} \beta^s e_{t-s} + m_t - \beta^\tau m_{t-\tau} \end{aligned} \tag{57}$$

Here, only the first line of (57) is observable. The coefficient of $x_{t-\tau}$ assumes the knowledge of δ, therefore as the first step, the dynamics of x has to be estimated. Then, the estimation of β and γ can be done conditional on the estimated dynamics of x. Notice that the coefficient of $x_{t-\tau}$ is a nonlinear function of β and γ, hence a nonlinear GMM is required to estimate these parameters. Concerning the instruments, moment conditions involving $\Delta^\tau k_{t-2\tau}$, $\Delta^\tau x_{t-\tau}$ and their higher lags are valid for the estimator in levels. Note that the estimate of α can only be interpreted conditional on δ_0, as well. For the first differenced estimator instruments $k_{t-3\tau}, x_{t-2\tau}$ and their higher order lags can be used.

The decomposition of the error terms is also dependent on the estimate of σ_v^2. Taking order τ differences of (57) the residuals from the nonlinear GMM estimation have variance

$$Var(\Delta^\tau \omega_t) = 2\gamma^2 \sum_{s=1}^{\tau-1} \left(\sum_{j=1}^{s} \beta^j \delta^{s-j} \right)^2 \sigma_v^2 + 2 \sum_{s=0}^{\tau-1} \beta^{2s} \sigma_e^2 + 2 \left(1 + \beta^\tau + \beta^{2\tau}\right) \sigma_m^2, \tag{58}$$

and order τ covariance

$$Cov(\Delta^\tau \omega_t, \Delta^\tau \omega_{t-\tau}) = -\gamma^2 \sum_{s=1}^{\tau-1} \left(\sum_{j=1}^{s} \beta^j \delta^{s-j} \right)^2 \sigma_v^2 - \sum_{s=0}^{\tau-1} \beta^{2s} \sigma_e^2 - (1 + \beta^\tau)^2 \sigma_m^2. \tag{59}$$

Using (58) and (59) it is possible to solve for σ_m^2, and conditional on the estimates of σ_v^2 and δ also for σ_e^2.

Estimation with scattered data

Unfortunately, the dataset used in this paper does not have equally spaced data points over time. This complicates the estimation of the error components, however, it also has an unwanted effect on the validity of instruments, as taking first-differences over the available data points might not eliminate the fixed effects completely. I investigate this issue for a specific case below using (9)-(10) as the model specification. For simplicity, I omit the use of the individual index.

The observations in the dataset are from years 1994, 1997, 1999 and 2004. Let the data for 1994 be $t = 1$, then observations are available for $t = 1, 4, 6, 10$. First, I investigate whether the first-difference GMM estimator remains consistent in the absence of additional regressors. Given the observations, the first-difference estimator would be to regress $k_{11} - k_6$ on $k_4 - k_6$ and use k_1 to instrument the lagged dependent variable. Using (54) I can write $k_{11} - k_6$ in terms of $k_6 - k_4$ as

$$\begin{aligned} k_{11} - k_6 \;=\; & \beta^5(k_6 - k_4) - \beta^2(1-\beta^3)k_4 + \beta^2(1+\beta+\beta^2)(\alpha+\eta) \\ & + e_{11} + \beta e_{10} + \beta^2 e_9 + \beta^3 e_8 + \beta^4 e_7 - e_6 - \beta e_5 \\ & + m_{11} - (1+\beta^5)m_6 + \beta^2 m_4. \end{aligned} \tag{60}$$

Unfortunately, η does not drop out of (60) because the lag between the dependent variable, $k_{11} - k_6$, is different from the lag between the regressor, $k_6 - k_4$. Also, k_4 remains in (60) because of this reason. Using (54) to rewrite k_4 as a function of k_1, I get

$$\begin{aligned} k_{11} - k_6 \;=\; & \beta^5(k_6 - k_4) - \beta^5(1-\beta^3)k_1 + \beta^5(1+\beta+\beta^2)(\alpha+\eta) \\ & + e_{11} + \beta e_{10} + \beta^2 e_9 + \beta^3 e_8 + \beta^4 e_7 - e_6 - \beta e_5 \\ & - \beta^2(1-\beta^3)e_4 - \beta^3(1-\beta^3)e_3 - \beta^4(1-\beta^3)e_2 \\ & + m_{11} - (1+\beta^5)m_6 + \beta^5 m_4 + \beta^5(1-\beta^3)m_1. \end{aligned} \tag{61}$$

Hence, in regression

$$k_{11} - k_6 = \pi_0 + \pi_1(k_6 - k_4) + \omega \tag{62}$$

the residual is correlated with k_1 making it an invalid instrument. The covariance between k_1 and error term $\Delta^5\omega_{11}$ is

$$Cov(k_1, \Delta^5\omega_{11}) = \beta^5(1-\beta^3)\sigma_m^2 - \beta^5(1-\beta^3)Var(k_1) + \beta^5(1+\beta+\beta^2)Var(\eta). \tag{63}$$

Assuming a covariance stationarity for k, it is possible to write the variance of k_1 as

$$Var(k_1) = Var(k) = \frac{\sigma_e^2}{1-\beta^2} + \frac{\sigma_\eta^2}{(1-\beta)^2} + \sigma_m^2. \tag{64}$$

Then, (63) becomes

$$Cov(k_1, \Delta^5\omega_{11}) = -\beta^6\frac{1-\beta^3}{(1-\beta)^2}Var(\eta) - \beta^5\frac{1-\beta^3}{1-\beta^2}\sigma_e^2. \tag{65}$$

Hence, the first-difference GMM estimator underestimates β using the given lag structure in the dataset.

On the other hand, the moment conditions in the level GMM estimator remains valid, because the instrument $k_4 - k_1$ remains uncorrelated with the fixed effects and the other error components in (9)-(10). Note that a Sargan test can be implemented to access the (in)validity of the first-difference moment conditions.

As a consequence of the above, in the estimation process I rely on the level GMM estimator:

$$k_{11} = \pi k_6 + \omega, \tag{66}$$

where $E(\pi) = \beta^5$ and ω includes the error components including the fixed effects. In the identification of π I rely on the independence of the instruments from the error components.

Next, I investigate the inclusion of additional variables in the model. Using (57) with $t = 11$ and $\tau = 5$ the model becomes

$$\begin{aligned} k_{11} &= (1+\beta+\beta^2+\beta^3+\beta^4)\alpha + \beta^5 k_6 + \gamma x_{11} + \gamma(\beta\delta^4+\beta^2\delta^3+\beta^3\delta^2+\beta^4\delta)x_6 \\ &+(1+\beta+\beta^2+\beta^3+\beta^4)\eta \\ &+\gamma\left(\beta(1+\delta+\delta^2+\delta^3)+\beta^2(1+\delta+\delta^2)+\beta^3(1+\delta)+\beta^4\right)\mu \\ &+\gamma\beta v_{10}+\gamma(\beta\delta+\beta^2)v_9+\gamma(\beta\delta^2+\beta^2\delta+\beta^3)v_8 \\ &+\gamma(\beta\delta^3+\beta^2\delta^2+\beta^3\delta+\beta^4)v_7 \\ &+e_{11}+\beta e_{10}+\beta^2 e_9+\beta^3 e_8+\beta^4 e_7+m_{11}-\beta^5 m_6. \end{aligned} \tag{67}$$

Only the first line of (67) is observable and a nonlinear GMM is required to estimate α, β and γ conditional on the dynamics of x as specified in (56). Moment conditions involving $(k_4 - k_1)$ and $(x_6 - x_4), (x_4 - x_1)$ can be used to estimate this equation.

The decomposition of the error term in (67) requires the estimates for δ, σ_v^2 and the observational variance of the fixed effects μ. The observational variance of fixed effects η also has to be estimated along with σ_e^2 and σ_m^2 in this case. The error components can be identified using the following residuals:

$$\begin{aligned} \widehat{\Delta^5\omega_{11}} &= (k_{11}-k_6)-\hat\beta^5(k_6-k_1)-\hat\gamma(x_{11}-x_6) \\ &-\hat\gamma(\hat\beta\hat\delta^4+\hat\beta^2\hat\delta^3+\hat\beta^3\hat\delta^2+\hat\beta^4\hat\delta)(x_6-x_1), \end{aligned} \tag{68}$$

$$\begin{aligned} \widehat{\Delta^2\omega_6} &= (k_6-k_4)-\hat\beta^2(k_4-k_1)-\hat\beta^2(1-\hat\beta)k_1-\hat\gamma(x_6-x_4) \\ &-\hat\gamma\hat\beta\hat\delta x_4+\hat\gamma\hat\delta(\hat\beta\hat\delta+\hat\beta^2)x_1, \end{aligned} \tag{69}$$

$$\widehat{\omega_6} = k_6-\hat\beta^2 k_4-\hat\gamma x_6-\hat\gamma\hat\beta\hat\delta x_4. \tag{70}$$

These imply

$$\begin{aligned} \widehat{\Delta^5\omega_{11}} &= \hat\gamma\hat\beta(v_{10}-v_5)+\hat\gamma(\hat\beta\hat\delta+\hat\beta^2)(v_9-v_4)+\hat\gamma(\hat\beta\hat\delta^2+\hat\beta^2\hat\delta+\hat\beta^3)(v_8-v_3) \\ &+\hat\gamma(\hat\beta\hat\delta^3+\hat\beta^2\hat\delta^2+\hat\beta^3\hat\delta+\hat\beta^4)(v_7-v_2) \\ &+(e_{11}-e_6)+\hat\beta(e_{10}-e_5)+\hat\beta^2(e_9-e_4)+\hat\beta^3(e_8-e_3)+\hat\beta^4(e_7-e_2) \\ &+(m_{11}-m_6)-\hat\beta^5(m_6-m_1), \end{aligned} \tag{71}$$

$$\begin{aligned} \widehat{\Delta^2\omega_6} &= -\hat\beta^2\eta+(e_6-e_4)+\hat\beta(e_5-e_3)-\hat\beta^2 e_2+(m_6-(1+\hat\beta^2)m_4+\hat\beta^3 m_1) \\ &-\hat\gamma\hat\beta^2\mu+\hat\gamma\hat\beta(v_5-v_3)-\hat\gamma\hat\beta^2 v_2, \end{aligned} \tag{72}$$

$$\widehat{\omega_6} = (1+\hat\beta)\eta+e_6+\hat\beta e_5+m_6-\hat\beta^2 m_4+\hat\gamma\hat\beta\mu+\hat\gamma\hat\beta v_5, \tag{73}$$

and

$$\begin{aligned} Var\left(\widehat{\Delta^5\omega_{11}}\right) &= 2\left((1+\hat\beta^2+\hat\beta^4+\hat\beta^6+\hat\beta^8)\hat\sigma_e^2+(1+\hat\beta^5+\hat\beta^{10})\hat\sigma_m^2\right) \\ &+2\hat\gamma^2\left(\hat\beta^2+(\hat\beta\hat\delta+\hat\beta^2)^2+(\hat\beta\hat\delta^2+\hat\beta^2\hat\delta+\hat\beta^3)^2\right) \\ &+2\hat\gamma^2(\hat\beta\hat\delta^3+\hat\beta^2\hat\delta^2+\hat\beta^3\hat\delta+\hat\beta^4)^2, \end{aligned} \tag{74}$$

$$\begin{aligned} Var\left(\widehat{\Delta^2\omega_6}\right) &= \hat\beta^4\widehat{Var}(\eta)+(2+2\hat\beta^2+\hat\beta^4)\hat\sigma_e^2+(2+2\hat\beta^2+\hat\beta^4+\hat\beta^6)\hat\sigma_m^2 \\ &+\hat\gamma^2\hat\beta^4\widehat{Var}(\mu)+\hat\gamma^2(2\hat\beta^2+\hat\beta^4)\hat\sigma_v^2, \end{aligned} \tag{75}$$

$$\begin{aligned} Var(\widehat{\omega_6}) &= (1+\hat\beta)^2\widehat{Var}(\eta)+(1+\hat\beta^2)\hat\sigma_e^2+(1+\hat\beta^4)\hat\sigma_m^2 \\ &+\hat\gamma^2\hat\beta^2\widehat{Var}(\mu)+\hat\gamma^2\hat\beta^2\hat\sigma_v^2. \end{aligned} \tag{76}$$

Hence, the estimates of σ_e^2 and σ_m^2 are conditional on the specification of x and the estimates of δ, σ_v^2 and the observational variance of μ. Note that it is also possible to derive the error composition of other regressions (for example, $k_1 1$ on k_6), however, the ones used above rely the least on the distribution of x and the precision of the parameter estimates. Again, it is possible to minimize the squared deviations from equations (74)-(76) to obtain non-negative estimates for the variances of the error components.

4 Data

The data are taken from the Ethiopian Rural Household Survey (ERHS) that follows rural households between 1989-2009. The survey collects information on livestock holdings, income, land and household composition of 1477 households. In the analysis we use data collected in 1994, 1997, 1999 and 2004. After dropping households with incomplete data[8] in these survey rounds 977 households remain in our sample.

As mentioned earlier, we analyze the dynamics of livestock holdings because it is the most important asset for most rural households as a storage of wealth. Livestock holdings in the ERHS dataset can be measured through the value of self-reported livestock holdings. The prices of the animals are calculated from self-reported livestock values and prices. Measuring livestock holdings this way increases the measurement errors in the data. An alternative would be to determine constants weights for each animal and calculate livestock unit holdings. However, this method has the disadvantage that it does not take into account the changes in relative prices that are potentially important factors in the dynamics of livestock holdings over a 10 year period. Therefore, we opt for using real livestock value to measure livestock holdings as in Pan (2008) despite its larger errors in measurement.

The other production factors in the structural model of Pan (2008) are land and labor. We have observations on the land use of households in the last two agricultural seasons. We measure labor input through household composition by constructing male equivalent labor units using the household production function estimates of Pan (2008).

Table 4 summarizes the distribution of the variables used in the analysis. We estimate the economic model using the log of the variables. Around 20 percent households have zero livestock holdings in some but not all periods. Since these households contain valuable information regarding the size of shocks and livestock accumulation, we keep these observations and use the transformation $\log(K^o_{h,t} + \lambda)$ to ensure that zero livestock holdings can be evaluated. After experimenting with different values of λ, we set its value to 0.01.

5 Results

In this section, I analyze the log-linearized optimal livestock accumulation dynamics of rural households in the ERHS dataset. First, I consider regressions of livestock accumulation dynamics without additional regressors. Then, I include land and labor in the regressions as determinants of livestock holdings. Standard errors in all regressions are obtained through bootstrapping. The 95 percent confidence interval of the parameter estimates are also reported, because particularly the variance estimates are non-symmetric. In the case of GMM estimators, 2-step estimators are used (Note: reference needed).

Univariate livestock accumulation dynamics

As a starting point, Table 5 reports the OLS regression results of the asset accumulation dynamics. This specification assumes no fixed effects and measurement error, therefore

[8] If annual income is reported to be zero we also treat that as a missing observation.

Variable	Year	Obs	Mean	Min[3]	Max	Std.dev.
Log real livestock value[1]	1994	977	5.566	-4.605	10.174	4.301
	1997	977	6.524	-4.605	10.398	3.404
	1999	977	5.780	-4.605	9.424	3.893
	2004	977	6.315	-4.605	10.095	3.537
Log real annual income[1]	1994	977	7.266	1.319	10.840	1.181
	1997	977	7.614	3.704	11.017	1.082
	1999	977	7.348	-0.672	10.592	1.058
	2004	977	7.221	2.720	10.628	1.058
Log land-labor composite[2]	1994	977	1.446	-0.610	3.069	0.549
	1997	977	1.531	-0.610	3.044	0.524
	1999	977	1.495	-0.462	2.564	0.466
	2004	977	1.579	-0.514	3.231	0.483

Notes: 1. Calculated from aggregate data in the ERHS survey.
2. Calculated using estimates from the production function.
3. Log livestock data is calculated using the transformation $\log(K+\lambda)$ with $\lambda = 0.01$.

Table 2: Descriptive statistics of the ERHS data

Variable	Estimate	Std. error	CI 2.5%	CI 97.5%
Constant (α)	0.992	0.068	0.862	1.112
Lagged livestock (β)	0.843	0.010	0.825	0.864
Shocks (σ_e^2)	10.838	0.647	9.498	12.097

Dependent variable: livestock value
Model: $k_{11} = \alpha + \beta^5 k_6 + e_{11} + \Lambda$

Table 3: OLS estimation results (benchmark)

it is inconsistent but can be useful as a benchmark. The estimate of the autoregressive coefficient is 0.84. It is not possible to determine the direction of its bias because fixed effects and measurement error bias its estimate in different directions. It is interesting to remember the value of the variance estimate, 10.8, and compare it to the estimates of true shock in the GMM estimators.

The estimation results of the level GMM estimator is shown in Table 5. This estimator controls for both fixed effects and measurement error. Its consistency is dependent on the initial condition assumption that the first differenced instruments are independent of the fixed effects. This condition is likely to be satisfied as households in the sample were keeping livestock for an extensive time already. The coefficient estimate of the autoregressive term is similar to the OLS regression, however, it becomes very imprecise and it is indistinguishable from zero at 5 percent significance level. Therefore, it is not surprising that the estimates of the error components are also quite imprecise, and also not significantly different from zero at 5 percent significance level. Looking at the point estimates of the error components, it appears that measurement error constitutes the largest part of the random variations, and the effect of true shocks is quite limited, around 10 percent of the value found in the OLS regression.

Based on Table 5 we cannot reject the hypothesis that either the variance of true

Variable	Estimate	Std. error	CI 2.5%	CI 97.5%
Constant (α)	0.614	4.245	-1.050	12.909
Lagged livestock (β)	0.916	0.708	-1.136	1.217
Shocks (σ_e^2)	1.176	1.004	0.000	4.278
Measurement error (σ_m^2)	5.337	2.126	0.000	7.074
Fixed effects ($Var(\eta)$)	0.094	14.228	0.000	66.691
Dependent variable:	livestock value			
Model:	$k_{11} = \alpha + \beta^5 k_6 + \eta + e_{11} + m_{11} - \beta^5 m_6 + \Lambda$			
Instruments:	$\iota, \Delta^3 k_4$			
Variance decomposition:	1. $k_{11} - k_6$ on $k_6 - k_1$			
	2. $(k_6 - k_4)$ on $(k_4 - k_1)$			
	3. k_6 on k_4			

Table 4: Level GMM estimation results with measurement error

shocks or measurement error is zero. However, this result is driven by the highly imprecise estimate of the autoregressive coefficient that has a crucial role in decomposing the error term. Therefore, this test is weak in determining the whether the fluctuations in observed livestock accumulation is driven by measurement error or true shocks.

The presence of measurement error can also be tested by the Sargan test using the level GMM estimator with the additional instrument $\Delta^2 k_6$, which is invalid if there is measurement error in the model but valid otherwise. The regression results of this estimator and the Sargan statistics are reported in Table 5. The Sargan statistic is 0.74, which implies that the null hypothesis that $\Delta^2 k_6$ is a valid instrument cannot be rejected. Therefore, $\Delta^2 k_6$ is not significantly correlated with the residual that includes m_6. Hence, this suggests that measurement error is not an important driving factor of the fluctuations in observed livestock holdings.

The coefficient estimates using the level GMM estimator without measurement error in Table 5 become more precise as $\Delta^2 k_6$ is a relevant instrument. In the table, the estimate of β decreases compared to its OLS estimate, which is in line with our expectation because fixed effects bias the estimate of the autoregressive coefficient upwards. After the error term decomposition, I find that around half of the residual variance is driven by shocks and the rest is due to fixed effects.

Next, in Table 5 I condider the Arellano-Bond estimator with measurement error, which is inconsistent because the fixed effects cannot be eliminated using first differences on the available data. The estimate of β is downward biased, however it is so imprecise that even the OLS estimate of β is included in the 95 percent confidence interval. Not reported here, in the Arellano-Bond estimator without measurement error the imprecision of the estimates remains even though the estimator is consistent (if there is no measurement error) using observations from periods 1, 6 and 11 (years 1994, 1999, 2004).

Table 5 shows the results of the estimation using the system GMM estimator with measurement error. Theoretically, k_1 is an invalid instrument in the estimation, however, the Sargan statistic does not reject the validity of any of the instruments used. The estimates are again very imprecise. On the other hand, the estimation results for the system GMM without measurement error, displayed in Table 5, are in line with the level GMM estimator in Table 5 in term of variance estimates and the implication of the Sargan test. However, the estimate of β remains very imprecise.

Overall, the univariate regressions above suggest that measurement error is not an

Variable	Estimate	Std. error	CI 2.5%	CI 97.5%
Constant (α)	1.925	0.691	1.386	3.192
Lagged livestock (β)	0.700	0.109	0.524	0.788
Shocks (σ_e^2)	4.649	0.580	3.819	5.947
Fixed effects ($Var(\eta)$)	6.168	0.829	4.518	7.779
Sargan stat	0.744	1.933	0.005	6.367
df=1, prob	0.388	0.285	0.016	0.953

Dependent variable:	livestock value
Model:	$k_{11} = \alpha + \beta^5 k_6 + \eta + e_{11} + \Lambda$
Instruments:	$\iota, \Delta^2 k_6, \Delta^3 k_4$
Variance decomposition:	1. $k_{11} - k_6$ on $k_6 - k_1$
	2. k_{11} on k_6

Table 5: Level GMM estimation results without measurement error

Variable	Estimate	Std. error	CI 2.5%	CI 97.5%
Lagged livestock (β)	-0.652	0.660	-0.884	0.858
Shocks (σ_e^2)	4.069	1.960	0.000	6.039
Measurement error (σ_m^2)	0.000	2.695	0.000	7.928
Fixed effects ($Var(\eta)$)	48.707	32.494	0.000	95.414

Dependent variable:	livestock value
Model:	$\Delta^5 k_{11} = \beta^5 \Delta^2 k_6 + \Delta^5 e_{11} + \Delta^5 m_{11} - \beta^5 \Delta^2 m_6 + \Lambda$
Instruments:	k_1
Variance decomposition:	1. $k_{11} - k_6$ on $k_6 - k_1$
	2. $(k_6 - k_4)$ on $(k_4 - k_1)$
	3. k_6 on k_4

Table 6: Arellano-Bond GMM estimation results with measurement error

Variable	Estimate	Std. error	CI 2.5%	CI 97.5%
Constant (α)	3.086	4.330	0.786	11.437
Lagged livestock (β)	0.512	0.699	-0.848	0.892
Shocks (σ_e^2)	4.539	1.837	0.000	6.075
Measurement error (σ_m^2)	1.588	2.546	0.000	7.541
Fixed effects ($Var(\eta)$)	2.127	31.463	0.000	105.293
Sargan stat	1.253	1.862	0.001	6.452
df=1, prob	0.263	0.306	0.013	0.978

Dependent variable:	livestock value
Model:	$\Delta^5 k_{11} = \beta^5 \Delta^2 k_6 + \Delta^5 e_{11} + \Delta^5 m_{11} - \beta^5 \Delta^2 m_6 + \Lambda$
	$k_{11} = \alpha + \beta^5 k_6 + \eta + e_{11} + m_{11} - \beta^5 m_6 + \Lambda$
Instruments:	k_1
	$\iota, \Delta^3 k_4$
Variance decomposition:	1. $k_{11} - k_6$ on $k_6 - k_1$
	2. $(k_6 - k_4)$ on $(k_4 - k_1)$
	3. k_6 on k_4

Table 7: System GMM estimation results with measurement error

Variable	Estimate	Std. error	CI 2.5%	CI 97.5%
Constant α	3.256	3.233	1.794	10.341
Lagged livestock β	0.484	0.514	-0.619	0.720
Shocks σ_e^2	5.925	0.793	4.157	6.983
Fixed effects $Var(\eta)$	6.251	1.194	4.677	9.238
Sargan stat	2.035	2.926	0.006	11.270
df=2, prob	0.154	0.293	0.001	0.942

Dependent variable:	livestock value
Model:	$\Delta^5 k_{11} = \beta^5 \Delta^5 k_6 + \Delta^5 e_{11} + \Lambda$
	$k_{11} = \alpha + \beta^5 k_6 + \eta + e_{11} + \Lambda$
Instruments:	k_4, k_1
	$\iota, \Delta^2 k_6$
Variance decomposition:	1. $k_{11} - k_6$ on $k_6 - k_1$
	2. k_{11} on k_6

Table 8: System GMM estimation results without measurement error

important driving factor behind the observed fluctuations in livestock holdings. However, not controlling for fixed effects double the estimate of the magnitude of the shocks. Regarding the dynamics of livestock accumulation, there is a high persistence of livestock holdings and the estimates of the constant term suggests growth in livestock holdings over time. A note of caution applies to the above results, as their validity relies on the validity of the initial condition assumption. Hence, if the data are not from a long-run equilibrium distribution, these results are inconsistent. Unfortunately, there is not enough observation to test the validity of the initial condition assumption of the level GMM estimator.[9]

Multivariate dynamic panel regression

Next, I turn to the analysis of livestock accumulation dynamics of households controlling for their land use and labor capacity. As a first step, I estimate the dynamics of these variables. This is necessary as the parameters of the livestock accumulation model can only be estimated conditional on the dynamics of the other variables due to the non-consecutive nature of the the available data.

In the estimation of the dynamics of land and labor, I assume that they are not affected by measurement error (they are easier to measure) and there is no other factor driving their dynamics. Hence, in this case using data from periods 1, 6, and 11 the Arellano-Bond estimator is consistent, and the system estimator, as well, if the initial conditions are satisfied for the system estimator. Now, this condition can be tested by the Sargan statistic.

In the case of labor, Table 5 reports the results from the system estimator. The Sargan test does not reject the validity of the instrument for the level regression, but results from the Arellano-Bond estimator are also not significantly different but less precise. The estimates suggest that labor capacity of the households is very persistent and stable over time.

Turning to the dynamics of land use, in Table 5 the Sargan test rejects the validity of the moment condition for the level regression, hence the Arellano-Bond estimator has

[9]If data become available from the 2009 round of the ERHS, it will be possible to test the validity of the initial condition assumption of the level GMM estimator.

Variable	Estimate	Std. error	CI 2.5%	CI 97.5%
Constant (μ_0)	0.063	0.015	0.034	0.092
Lagged labor (δ)	0.947	0.014	0.921	0.973
Shocks (σ_v^2)	0.040	0.002	0.036	0.045
Fixed effects ($Var(\mu)$)	0.087	0.009	0.071	0.108
Sargan stat	1.861	3.139	0.006	12.508
df=1, prob	0.172	0.278	0.001	0.970

Dependent variable:	male equivalent labor unit
Model:	$\Delta^5 x_{11} = \delta^5 \Delta^5 x_6 + \Delta^5 v_{11} + \Lambda$
	$x_{11} = \mu_0 + \delta^5 x_6 + \mu + v_{11} + \Lambda$
Instruments:	x_1
	$\iota, \Delta^5 x_6$
Variance decomposition:	1. $x_{11} - x_6$ on $x_6 - x_1$
	2. x_{11} on x_6

Table 9: First stage regression results for labor: system GMM without measurement error

to be used. This is reported in Table 5. The results suggest that land use is also quite persistent, however, in this case it is not possible to identify μ_0.

Next, I estimate the livestock accumulation dynamics using the nonlinear level GMM estimator (67). In the estimation, I use the results of Table **??** and 5. The estimation results from the level GMM with measurement are reported in Table 5. The coefficient estimates of land and labor are both not significantly different from zero, which suggests that this is an incorrect specification.

Table 5 shows the results of the level GMM estimator without measurement error. The Sargan test cannot reject the validity of the additional instruments. In this case, the coefficient estimates of land and labor are significantly different from zero. The estimate of β is close to the estimate of the univariate level GMM estimator of Table 5 but less precise. The estimate of shocks is also in line with that result.

Summarizing the results, both the univariate and the multivariate regressions suggest that measurement error is not an important driving factor behind the observed fluctuations in livestock holdings. However, not controlling for fixed effects double the estimate of the magnitude of the shocks. However the validity of these results depends on the validity of the initial condition assumption. Hence, if the data are not from a long-run equilibrium distribution, these results are inconsistent. Unfortunately, there is not enough observation to test the validity of the initial condition assumption of the level GMM estimator.[10]

6 Conclusion

The first part of the paper investigates consistent estimators of dynamic panel data models in the presence of measurement error, and the decomposition of the error term into true shocks and measurement error. The decomposition is based on the fact that true shocks have a long-lasting effect on the dynamics of the variable of interest, whereas, measurement error has only a transient effect. Hence, if the autoregressive parameter is sufficiently different from zero, the two components can be separated by analyzing the structure of

[10]If data become available from the 2009 round of the ERHS, it will be possible to test the validity of the initial condition assumption of the level GMM estimator.

Variable	Estimate	Std. error	CI 2.5%	CI 97.5%
Constant (μ_0)	0.169	0.068	0.035	0.267
Lagged land (δ_1)	-0.525	0.556	-0.706	0.709
Shocks (σ_v^2)	0.354	0.061	0.232	0.461
Fixed effects ($Var(\mu)$)	0.573	0.160	0.352	0.920
Sargan stat	46.026	13.852	27.167	79.463
prob	0.000	0.000	0.000	0.000

Dependent variable:	land
Model:	$\Delta^5 x_{11} = \delta^5 \Delta^5 x_6 + \Delta^5 v_{11} + \Lambda$
	$x_{11} = \mu_0 + \delta^5 x_6 + \mu + v_{11} + \Lambda$
Instruments:	x_1
	$\iota, \Delta^5 x_6$
Variance decomposition:	1. $x_{11} - x_6$ on $x_6 - x_1$
	2. x_{11} on x_6

Table 10: First stage regression results for land: system GMM without measurement error

Variable	Estimate	Std. error	CI 2.5%	CI 97.5%
Lagged land (δ_1)	0.859	0.045	0.763	0.948
Shocks (σ_v^2)	0.288	0.028	0.237	0.346
Fixed effects ($Var(\mu)$)	0.295	0.066	0.228	0.494

Dependent variable:	land
Model:	$\Delta^5 x_{11} = \delta^5 \Delta^5 x_6 + \Delta^5 v_{11} + \Lambda$
Instruments:	x_1
Variance decomposition:	1. $x_{11} - x_6$ on $x_6 - x_1$
	2. x_{11} on x_6

Table 11: First stage estimation results for land: Arellano-Bond estimator without measurement error

Variable	Estimate	Std. error	CI 2.5%	CI 97.5%
Constant (α)	4.205	2.686	-0.043	8.850
Lagged livestock (β)	0.001	0.645	-1.087	1.038
Land (γ_1)	1.892	1.287	-0.329	6.027
Labor (γ_{labor})	1.697	1.254	-0.100	5.782
Shocks (σ_e^2)	9.574	3.114	0.000	10.102
Measurement error (σ_m^2)	0.737	4.110	0.000	12.970
Fixed effects ($Var(\eta)$)	3.863	34.551	0.000	150.042

Dependent variable:	livestock value
Additional regressors:	land, labor
Model:	$k_{11} = \beta^5 k_6 + \gamma x_{11} + \mathit{\Gamma} x_6 + \eta + e_{11} + m_{11} - \beta^5 m_6 + \Lambda$
Instruments:	$\Delta^3 k_4, \Delta^2 x_6, \Delta^3 x_4$
Variance decomposition:	1. $(k_{11} - k_6)$ on $(k_6 - k_1), (x_{11} - x_6), (x_6 - x_4)$
	2. $(k_6 - k_4)$ on $(k_4 - k_1), (x_6 - x_4), (x_4 - x_1)$
	3. k_6 on k_4, x_6, x_4

Table 12: Level (nonlinear) GMM estimation results with measurement error

Variable	Estimate	Std. error	CI 2.5%	CI 97.5%
Constant (α)	1.216	1.508	0.585	5.540
Lagged livestock (β)	0.696	0.338	-0.002	0.778
Land (γ_1)	1.478	0.623	0.569	2.839
Labor (γ_{labor})	1.523	0.731	0.289	3.319
Shocks (σ_e^2)	7.810	10.310	2.614	10.211
Sargan stat	1.227	5.542	0.287	21.492
prob	0.268	0.167	0.000	0.600

Dependent variable:	livestock value
Additional regressors:	land, labor
Model:	$k_{11} = \beta^5 k_6 + \gamma x_{11} + \mathit{\Gamma} x_6 + \eta + e_{11} + m_{11} - \beta^5 m_6 + \Lambda$
Instruments:	$\Delta^2 k_6, \Delta^3 k_4, \Delta^2 x_6, \Delta^3 x_4$
Variance decomposition:	1. $(k_{11} - k_6)$ on $(k_6 - k_1), (x_{11} - x_6), (x_6 - x_4)$
	2. k_{11} on k_6, x_{11}, x_6

Table 13: Level (nonlinear) GMM estimation results without measurement error

the covariance matrix of the error term. However, for this method to be feasible at least four observations are needed over time controlling for individual heterogeneity.

A contribution of this paper to the literature on applied dynamic panel data models is the detailed discussion of estimators using data from non-consecutive periods. In this context, I find that if observations are distant but equally spaced, the inclusion of other regressors in the model make the standard (linear) estimators inconsistent, but the use of a nonlinear estimator can restore consistency. This nonlinear estimator adds the lagged value of the other variables to the regression with constrained coefficients. The method, however, requires the specification of the dynamics of the other variables in the model, and the coefficient estimates are conditional on the dynamics of this specification. Therefore, the estimator becomes less precise.

A further complication in the estimation arises if the distant observations are not equally spaced. In this case, first differencing over the available observations does not eliminate the fixed effect, therefore the (nonlinear) Arellano-Bond and system estimators become inconsistent. A solution is to use the (nonlinear) level GMM estimator if the initial conditions are satisfied.

This method is applied to analyze the livestock accumulation dynamics of rural households in the ERHS dataset because there is no sufficient equally spaced data available. The estimation results of the univariate model suggest high persistence of the livestock holdings and limited growth. However, the estimate of the magnitude of shocks is sizable. In the multivariate regression controlling for land and labor size, the estimates are less precise but the results remain similar. Based on the Sargan statistics of the level GMM estimator, I find that I cannot reject the hypothesis that measurement error does not drive the random fluctuations in data.

Reference

Arellano, M. (2003): Panel Data Econometrics: Advanced Texts in Econometrics. Oxford University Press.

Arellano, M. and Bond, S. (1991):

Bond, S. (2002): Dynamic Panel Data Models: A Guide to Micro Data Methods and Practice. CEMMAP Working Paper CWP09/02.

Bond, S.; Hoeffler, A.; Temple, J. (2001): GMM Estimation of Empirical Growth Models

Blundell, R. and Bond, S. (1998): Initial conditions and moment restrictions in dynamic panel data models. Journal of Econometrics 87, pp. 115-143.

Blundell, R.; Bond, S.; Windmeijer, F. (2000): Estimation in Dynamic Panel Data Models: Improving on the Performance of the Standard GMM Estimator. The Institute for Fiscal Studies Working Paper Series, WP 00/12.

Davidson, R. and MacKinnon J. (2004): Econometric Theory and Methods. Oxford University Press.

Dercon, S. (2002): Income Risk, Coping Strategies and Safety Nets. The World Bank Research Observer 17(2), pp. 141-166.

Elbers, C. (2009): Dealing with asset measurement error. Mimeo.

Elbers, C.; Gunning, J.W. and Kinsey, B. (2007): Growth and Risk: Methodology and Micro Evidence. The World Bank Economic Review 21(1), pp. 1-20.

Gourieroux, C; Monfort, A. and Renault, E. (1993): Indirect inference. Journal of Applied Econometrics 8, pp. 85-118.

Hansen (1982):

Kinsey, B.; Burger, K. and Gunning, J.W. (1998): Coping with Drought in Zimbabwe: Survey Evidence on Responses of Rural Households to Risk. World Development 26(1), pp. 89-110.

Pan, L. (2008): Poverty, Risk and Insurance: Evidence form Ethiopia and Yemen. PhD Thesis, Tinbergen Institute.

Townsend, R.M. (1994): Risk and Insurance in Village India. Econometrica 62(3), pp. 539-591.

Wansbeek, T.J. and Koning, R.H. (1991): Measurement Error and Panel Data. Statistica Neerlandica 45, pp. 85-92.