No Mangos in the Tundra: Spatial Heterogeneity in Agricultural Productivity Analysis*

Markus EBERHARDT[a,b] and Francis TEAL[b,c]

[a] *School of Economics, University of Nottingham*

b Centre for the Study of African Economies, Department of Economics, University of Oxford

c Institute for the Study of Labor (IZA), Bonn

Oxford Bulletin of Economics and Statistics
(accepted for publication)

July 2012

Abstract: In line with the wider macro productivity literature existing studies of agricultural production largely neglect technology heterogeneity, variable time-series properties and the potential for heterogeneous but correlated Total Factor Productivity (TFP) across countries. Our empirical approach accommodates these difficulties and seeks to model the nature of the cross-section dependence in a sample of 128 countries (1961-2002). Our results suggest that agro-climatic environment drives similarity in TFP evolution across countries with heterogeneous production technology. This provides a possible explanation for the failure of technology transfer from advanced countries of the temperate 'North' to arid and/or equatorial developing countries of the 'South'.

Keywords: agriculture, cross-country productivity analysis, heterogeneous technology, non-stationary panel econometrics

JEL classification: C23, O13, Q10

*We are grateful to Anindya Banerjee, Marco Barassi, Steven Block, Stephen Bond, Marcel Fafchamps, Michael Klein, Rocco Macchiavello, John Muellbauer, Hashem Pesaran, James Reade, Ron Smith, Måns Söderbom, Jon Temple, Joakim Westerlund and two anonymous referees for helpful comments and suggestions. We also thank seminar/session participants at workshops in Oxford, the CSAE Conference, the International Panel Data Conference, the World SEA Congress, Gothenburg, the NEUDC, Brown, Birmingham and the NAPW for their input. All remaining errors are our own. We are indebted to Peter Pedroni and Takashi Yamagata for making their computer code freely available on the internet. A Technical Appendix to the paper as well as the data and Stata routines can be found at http://sites.google.com/site/medevecon. The first author gratefully acknowledges financial support from the UK Economic and Social Research Council [grant numbers PTA-031-2004-00345, PTA-026-27-2048], the Bill & Melinda Gates Foundation and the James Baker III Institute for Public Policy at Rice University. Correspondence: markus.eberhardt@nottingham.ac.uk

I Introduction

'[A]ssumptions of a common production function, and perfect and competitive factor markets … get in the way of understanding international differences in productivity — particularly differences between advanced and underdeveloped economies.' (Nelson, 1968, p.1229)

Ever since Hayami and Ruttan (1970) introduced the use of panel data to estimate cross-country production functions for the agriculture sector, academic studies have emphasised the conceptual desirability of technology heterogeneity across countries to reflect the sizeable differences in agro-climatic environment, agricultural output mix and level of commercialisation observed.[1] The literature further highlighted the potential for barriers to technology transfer between countries that are specific to the agricultural sector, in particular the problems of transfers between the developed countries of the temperate 'North' and the developing countries of the arid or equatorial 'South'. From a theoretical standpoint the assumption of differential technology across countries therefore seems almost a given, rather than merely an extension worth considering. In practice, however, perhaps due to the constraints imposed by estimation techniques or data availability, the empirical investigation of agricultural productivity was typically based on models which imposed technology homogeneity across countries, or only allowed for heterogeneity by splitting the sample into crude geographical groups. We highlight this seeming mismatch between theory and empirics in our title: if these concerns were unimportant and a single common production function could be applied to characterise global production, then this would be somewhat akin to saying we would expect to see the cultivation of mangos in the Tundra.[2]

The assumption of a homogeneous production function may mask or distort important insights into development, as the above comment by Nelson (1968) suggests. Hayami and Ruttan (1970) argue that Nelson's comment (which referred to manufacturing) equally applies to the analysis of agricultural production. Their approach further highlights the implications of and anecdotal evidence for technology being endogenous to the prevailing 'economic system', namely the

> *'differential diffusion of agricultural technology…[and] of the scientific and technical capacity to invent and develop new mechanical, biological, and chemical technology specifically adapted to the factor endowments and prices in a particular country or region.'* (Hayami and Ruttan, 1970, p.898)

Alternative conceptual arguments in favour of a heterogeneous production function point to the difference in output structure (wheat *vs.* rice *vs.* livestock) and in the commercialisation of agriculture (subsistence *vs.* industrialised farming), both of which are functions of the level of development and productive specialisation across countries. This aside, the commonly applied proxies for production inputs represent a mix of very broad (land, labour) and very specific inputs (fertilizer, tractors for capital stock). One could therefore argue that attempts at capturing this observed and induced heterogeneity in an aggregate agricultural production function cannot succeed if merely unobservable Total Factor Productivity (TFP) is allowed to differ across countries. In addition to technology heterogeneity, these considerations highlight potential limits to technological spillover from innovations between countries with very different agro-climatic makeup and resource endowment. Much of agricultural technology has to be viewed as location-specific, with attempts at direct technology transfer from one

[1] Here and in the remainder of the paper we refer to heterogeneity in 'technology parameters' to indicate differential production function parameters on observable factor inputs across countries.

[2] Tundra refers to the agro-climatic zone bordering the Arctic (the far north of North America and Eurasia) characterised by permafrost and low levels of natural vegetation. According to the US National Mango Board mangos originated in Southeast Asia and India, and the latter still dominates world production today.

agro-climatic region to another largely doomed to failure (Ruttan, 2002).

In this paper we extend the insights gained from the emerging literature on multi-factor models in nonstationary panels to cross-country empirical productivity analysis in the agricultural sector. We adopt a common factor model approach and estimate production functions for a panel of 128 developing and developed countries using annual data from 1961 to 2002 (FAO, 2007). The common factor approach is uniquely suited to macro productivity analysis (Bai, 2009; Chudik, Pesaran and Tosetti, 2011) where the impact of unobservables on output is conceptualised as TFP and treated as a 'measure of our ignorance' (Abramowitz, 1956). The framework allows for a rich set of cross-section correlations which are deemed pervasive in macro panel data, arising from common shocks and/or spillover effects, the latter determined for instance by trade, policy or shared cultural/social heritage. Our focus in the empirical part of this paper is on the changes in the parameter estimates and diagnostic test results when we move between pooled and heterogeneous estimators, between methods which ignore cross-section dependence and those which accommodate it, and between approaches that put different emphasis on the time-series properties of 'long' panels.

The relevance of concerning ourselves with an appropriate empirical representation of agricultural productivity at the macro level is clear given the renewed interest in sectoral development as well as the link between agricultural productivity and aggregate economic growth and/or poverty reduction in the literature (e.g. Block, 2010; Gollin, 2010). This aside, studies of the structural transformation within countries and its link to economic development, building on the dual economy model by Lewis (1954), crucially rely on a valid representation of agricultural technology and productivity in cross-country empirical analysis (Vollrath, 2009a; Eberhardt and Teal, forthcoming).

Our paper makes three contributions to the literature on agricultural productivity at the macro level: (i) we demonstrates that failure to account for technology heterogeneity leads to misspecified empirical models, with serious implications for any TFP estimates obtained; (ii) we show that previous findings of decreasing returns to scale at the global level are an artefact of this empirical misspecification; (iii) we open up the black box of unobserved heterogeneity and analyse a number of hypotheses for the processes driving cross-country correlation and spillovers. Our results suggest that agro-climatic environment provides more convincing evidence than purely spatial concerns of neighbourhood or distance.

Our study also speaks to the much-maligned cross-country growth literature (Durlauf, Johnson and Temple, 2005) by providing a step-by-step approach to the empirical analysis of macro panel data. This approach is concerned with (i) a more flexible framework to allow heterogeneity across countries, (ii) a concern for the underlying data properties (nonstationarity, cross-section correlation) and thus for diagnostic testing of rival empirical models, and (iii) the robustness of the estimation to variable endogeneity and reverse causality, which has rightly been a focus of the existing literature albeit to the detriment of any other matters related to empirical specification (Eberhardt and Teal, 2011).

The remainder of the paper is organised as follows: Section II sketches the existing applied literature. Section III introduces the empirical model, our extension to the CCE estimators and the data. Section IV presents and discusses the main empirical results; a number of important questions related to the direction of causation are investigated in Section V. Section VI concludes.

II Related Studies on Agricultural Productivity

Following the seminal contribution by Hayami and Ruttan (1970, 1985) the literature on agricultural productivity analysis across countries using panel data or 'repeated cross-sections' can be broadly distinguished by two aspects. The first relates to the data used, the second to the empirical restrictions placed on production technology: whether countries are allowed to have differential technology parameters, TFP levels and evolution, and whether constant returns to scale are imposed. While our review is by no means exhaustive, we believe that the studies in Table A-2 of the Appendix capture the breadth of the empirical field at the present time.

Most cross-country studies on agriculture use the data provided by the Food and Agriculture Organisation (FAO) which contains output and input variables for a large number of countries from 1961 onwards, but relies on tractors and agricultural machinery as a proxy for agricultural capital (e.g. Craig, Pardey and Roseboom, 1997; Cermeño, Maddala and Trueblood, 2003; Bravo-Ortega and Lederman, 2004). The alternative to this is a dataset developed by World Bank researchers (Larson, Butzer, Mundlak and Crego, 2000) which provides agricultural fixed capital stock data for up to 57 developing and developed countries from 1967 to 1992 (an update covers 1972-2000 for 30 countries). We highlight this difference since with the notable exception of Martin and Mitra (2002) *all* empirical studies which use the World Bank dataset(s) obtain very high capital coefficients, typically between .35 and .6.[3] In contrast, all studies using the FAO data with tractors proxying for fixed capital stock obtain capital coefficients in the range .05 to .2.

The second major aspect, relating to technology heterogeneity, has commonly been limited to the modelling of TFP. Technology heterogeneity across countries has either been ignored (Hayami and Ruttan, 1970; Craig et al., 1997; Martin and Mitra, 2002) or approached by splitting the sample into 'homogeneous groups', e.g. by level of development (Hayami and Ruttan, 1985; Cermeño et al., 2003; Gutierrez and Gutierrez, 2003). The work by Mundlak, Larson and Butzer (1999, 2008) as well as Vollrath (2009b) represents an exception in this regard, in that they recognise the potential for technology differences across countries and attempt to highlight their importance for empirical analysis. Although many of these studies stress the importance of allowing for production technology to differ across countries, none of them investigates this in a manner which allows for full heterogeneity for technology parameters *as well as* TFP.

Closely linked to the empirical specification of production technology is the treatment of returns to scale.[4] The underlying returns to scale of agricultural production affect the size-distribution of farms within an economy, however, we can also think of a number of constraints, e.g. insecure legal environment or variations in land tenure arrangements, that influence both of these processes in a similar fashion. For cross-country analysis findings of increasing, decreasing or constant returns to scale (all of which are present in Table A-2) are typically justified with reference to micro-econometric studies or the structural change within countries witnessed over the sample period. Hayami and Ruttan (1985) argue that increasing returns in developed economies are linked to the indivisibility of fixed capital,

[3]The Martin and Mitra (2002) paper arrives at a much lower coefficient of .12. Given that the methods applied are similar this discrepancy may be caused by the alternative deflation strategy applied in Martin and Mitra (2002). Similar to the practice in the FAO data, the authors advocate the use of a single LCU-US$ exchange rate (base year 1990) in favour of using annual exchange rates as implemented in the Larson et al. (2000) data.

[4]Imposition of constant returns is at times argued to be inherently problematic given the presence of a fixed factor input (land) in the agricultural production function. However, this characteristic does not prevent constant returns to all (fixed and variable) inputs, which implies decreasing returns to variable inputs (Graham and Temple, 2006).

which has played an increasingly important role in the substitution of labour in these countries. The constant returns in their developing country sample is said to be the outcome of increasing population pressure on land, with efforts to increase productivity directed toward saving land by applying more fertilizer, chemicals or improved seeds. Since these inputs are highly divisible, the authors argue, it is not surprising to encounter constant returns in this subsample.

The paper by Gutierrez and Gutierrez (2003) to the best of our knowledge represents the only study which accounts for time-series properties of the data, using nonstationary panel econometric methods, while most other studies favour pooled OLS or LSDV/within estimators. As in common with the vast majority of cross-country empirical analysis, none of the studies reviewed considers the impact of cross-section dependence in the data on empirical estimates. The presence of such dependence can result in misleading inference and even inconsistency in standard fixed effects panel estimators favoured in this literature (Phillips and Sul, 2003; Bai, 2009; Kapetanios, Pesaran and Yamagata, 2011).

In recent work Vollrath (2009b) develops the idea of technology heterogeneity in agriculture by highlighting differential agro-climatic environment and output structure (e.g. crops *vs.* livestock). His results — focused on the labour coefficient — highlight considerable technology heterogeneity across continents, climate zones and country groups. Note that with the exception of Gutierrez and Gutierrez (2003) and Vollrath (2009b) this literature largely sidesteps concerns over variable endogeneity and the bias induced.

This concludes our brief sketch of the literature on the inter-country production function in agriculture, where we limited ourselves to studies employing linear parametric regressions. Frontier estimation and nonparametric efficiency estimation strategies have also been used in this literature (e.g. Coelli and Rao, 2005), however this approach neglects two of the fundamental issues we highlight in this study, namely cross-section correlation and the salient time-series properties of the data (Schmidt, 2009). The adoption of growth/productivity accounting is another strategy (e.g. Restuccia, Yang and Zhu, 2008), however this methodology cannot disentangle the underlying endogeneity problem, such that inputs cannot be argued to *cause* output (Gollin, 2010). Our empirical strategy will attempt to address these concerns.

III Econometric Strategy and Implementation

Empirical Model

We model production in country i at time t employing a Cobb-Douglas production function $Y_{it} = A_{it} X_{it}^{\beta_i}$ where Y is agricultural output, X is a set of observed inputs and A is unobserved TFP. The technology parameters β_i are constant over time but can differ across countries.[5]

Our empirical framework builds on a common factor representation of the log-linearised production function. For $i = 1, \ldots, N$, $t = 1, \ldots, T$ and $m = 1, \ldots, k$ let

$$
\begin{aligned}
y_{it} &= \beta_i' x_{it} + u_{it} \qquad u_{it} = \alpha_i + \lambda_i' f_t + \varepsilon_{it} && (1)\\
x_{mit} &= \pi_{mi} + \delta_{mi}' g_{mt} + \rho_{1mi} f_{1mt} + \ldots + \rho_{nmi} f_{nmt} + v_{mit} && (2)\\
& f_t = \varrho' f_{t-1} + \epsilon_t \qquad \text{and} \qquad g_t = \kappa' g_{t-1} + \epsilon_t && (3)
\end{aligned}
$$

where $f_{\cdot mt} \subset f_t$. We follow the existing literature and include (proxies for) labour, agricultural capital stock, livestock, fertilizer and land under cultivation as the m observed inputs x_{it} in our model for observed output y_{it} (all variables in logarithms). Unobserved agricultural

[5] In order to address concerns about parameter nonconstancy we carry out a robustness test using recursive estimation — see Figure 2 in the Technical Appendix.

TFP is represented by a combination of country-specific TFP levels α_i and a set of common factors f_t with factor loadings that can differ across countries (λ_i).[6]

We also introduce an empirical representation of the observed inputs in equation (2) in order to indicate the possibility for endogeneity: the input variables x_{it} are driven by sets of common factors g_{mt} and f_{nmt}, whereby the latter may represent a subset of the factors driving output. This overlap of common factors creates severe difficulties for the identification of the technology parameters β_i (Kapetanios et al., 2011, Remark 5).[7]

Equation (3) indicates that the factors are persistent over time, which allows for nonstationarity in the factors ($\varrho = 1$, $\kappa = 1$) and thus the observables. This implies various combinations of cointegration: between output y and inputs x; between output y, inputs x and (some of) the unobserved factors f; or none of the above (noncointegration). Note that we allow for the possibility of nonstationary TFP, which seems desirable given the nature of technological progress (Bond, Leblebicioglu and Schiantarelli, 2010).

The use of annual time-series data commonly raises concerns regarding the distorting influence of business cycles on estimation results (Durlauf et al., 2005), such that analysis is typically carried out on time-averaged data. The common factor model approach is uniquely suitable to deal with *any* business cycle or resulting capacity utilization effects, whether they represent idiosyncrasies of a small number of economies, or global business cycles: in the former case we can appeal to the Chudik et al. (2011) result whereby an infinite number of 'weak factors' may be introduced to the model to capture local spillover effects. In the latter case a 'strong factor' (i.e. of the nature we have assumed throughout) can be used to model the heterogeneous impact of a global shock.[8]

The common factor model framework is arguably ideally suited for the analysis of cross-country productivity (Bai, 2009; Chudik et al., 2011) but has thus far not been applied very widely (e.g. Cavalcanti, Mohaddes and Raissi, 2011; Eberhardt, Helmers and Strauss, forthcoming). All of the above concerns are founded in econometric theory and empirical observation (nonstationarity, cross-section correlation), with the specific concerns over technology heterogeneity in agriculture motivated by our discussion in the introduction as well as more generally by the 'new growth theory' (see Temple, 1999).

Empirical implementation

Our empirical implementation follows a philosophy whereby various regression models with differing restrictions on parameters and assumptions about residual distribution are compared and contrasted. With reference to our empirical model in equations (1) to (3) this implies different assumptions regarding β_i, λ_i, α_i as well as the persistence of the underlying common factors. In the pooled model we estimate OLS with year dummies, two-way fixed effects (2FE) and the Pesaran (2006) common correlated effects (CCE) pooled estimator; these all assume common slope parameters $\beta_i = \beta$ and impose different restrictions on TFP evolution via fixed effects (α_i) and the impact of common factors (λ_i). In the heterogeneous models we implement the Pesaran and Smith (1995) Mean Group (MG) and the heterogeneous version of the CCE estimators (CCEMG), both of which allow for heterogeneous slopes (β_i) but

[6] In contrast to other applications where they are seen as a nuisance these factors with heterogeneous impact are a central interest in productivity analysis since they collectively represent TFP evolution over time.

[7] The literature on productivity analysis at the firm-level refers to this as 'transmission bias' and offers a number of estimation techniques to address this issue (see Eberhardt and Helmers, 2010, for a recent survey). None of these techniques were designed for long-T integrated panel data and they furthermore cannot address the concerns of technology heterogeneity and cross-section dependence at the heart of this paper.

[8] In terms of empirical implementation the Pesaran (2006) CCE estimators were already shown to be robust to both types of data dependencies (Pesaran and Tosetti, 2011; Chudik et al., 2011).

again differ in their assumptions about common factors. The performance of all of these estimators in nonstationary panels with cross-section dependence has been discussed in great detail elsewhere (Coakley, Fuertes and Smith, 2006), so that we focus on the Pesaran (2006) CCE estimators implemented and extended in the present analysis.

Econometric theory and simulation studies (Pesaran, 2006; Coakley et al., 2006; Kapetanios et al., 2011; Westerlund and Urbain, 2011) indicate that the newly-developed CCE estimators are able to accommodate the type of endogeneity we introduced in the empirical equation to arrive at consistent estimates for common β coefficients or the means of heterogeneous β_i. This result is robust even when the cross-section dimension N is small, when variables are nonstationary, cointegrated or not, subject to structural breaks and/or in the presence of 'weak' unobserved common factors (spatial spillovers) and global/local business cycles (Chudik et al., 2011; Kapetanios et al., 2011; Pesaran and Tosetti, 2011).

The CCE estimators account for the presence of unobserved common factors with heterogeneous factor loadings by introducing cross-section averages for the dependent and independent variables into the regression model, each with a country-specific parameter. Pesaran (2006) shows that the asymptotic consistency of the estimators is based on any weighted cross-section aggregates provided the weights w_i satisfy the conditions

$$w_i = O\left(\frac{1}{N}\right) \qquad \sum_{i=1}^{N} w_i \lambda_i \neq 0 \qquad \sum_{i=1}^{N} |w_i| < K \tag{4}$$

where K is some finite positive constant (Coakley et al., 2006). In the standard CCE estimator the weights are the same for all countries ($1/N$). The economic interpretation could be that the unobserved factors which influence productivity are common to all countries. In a simple empirical extension we experiment with a number of weight-matrices prior to taking the cross-section average to implement alternative scenarios:

(i) The *'Neighbourhood Effect'*: many empirical studies have argued that the performance of neighbours to country i has a significant effect on the latter's TFP, and tried to measure this spillover empirically using spatial econometric methods (e.g. Ertur and Koch, 2007). In our implementation we construct cross-section averages of y and x for country i from the values of i's *contiguous* neighbours.

(ii) The *'Gravity Effect'*: gravity models suggest that geographical distance is a powerful determinant of the magnitude of economic exchange between countries (e.g. Frankel and Romer, 1999). We adopt this approach to see whether distance between countries (a proxy for climatic, soil, cultural and socio-economic differences) can explain the effects of unobserved heterogeneity. For country i the observations for countries $j = 1, \ldots, N-1$ are weighted by the inverse of the population-weighted distance between i and j before computing the cross-section aggregates.[9]

(iii) The *'Agro-Climatic Distance Effect'*: much of the existing literature highlights the differential agro-climatic characteristics across countries and links the failure of technology transfer to heterogeneity in climate and resource endowment. For country i we weight the observations for countries $j = 1, \ldots, N-1$ by a measure for agricultural distance between countries i and j before computing the cross-section aggregate.[10]

[9]Following convention in the spatial econometric literature we set the diagonals on the contiguity and distance matrices to zero. For all three weight matrices we 'row-normalise' the weights (across i).

[10]The distance measure is computed for every country-pair using data on the share of cultivated land within each of twelve climatic zones (following Jaffe, 1986). This can be interpreted as a multi-variate correlation coefficient which varies between zero and unity: a low (high) value indicates little (high) similarity in the distribution of cultivated land across climatic zones in the two countries (Pardey, James, Alston, Wood, Koo, Binenbaum, Hurley

Our conceptual justification for these variants of the CCE estimators is the situation where the average of factor loadings across countries is non-zero, but systematic patterns are present in the data. In the distance case we implicitly test the hypothesis that country i is driven by unobserved common factors which are the same in countries in close proximity, but is much less affected by other factors which drive countries further away. The neighbourhood case represents an extreme extension of the same argument. In the final case we test the hypothesis that countries with similar agro-climatic environment (e.g. tropics) are affected by a shared set of common factors, but that they are not (or much less) affected by a separate set of common factors which in turn influence countries in a very different agro-climatic environment (e.g. temperate zone).

While we are able to address endogeneity arising from common factors, the results need to be tested for reverse causality. In the case of the production function the empirical results could represent a labour demand equation or investment equation in disguise. In order to rule out this possibility we carry out tests for the direction of causation in Section V below.

All empirical implementation is subject to explicit or implicit assumptions which can be reflected in the presence or absence of 'well-behaved' residuals, displaying stationarity, cross-section independence, and lack of autocorrelation. Although the applied macro panel literature is strangely devoid of residual diagnostics (Banerjee, Eberhardt and Reade, 2010) we base our decision on which model emerges as the preferred specification on this approach.[11]

Data

The principal data source for our empirical analysis is the Food and Agriculture Organisation's *FAOSTAT* panel database (FAO, 2007), from which we obtain annual observations for agricultural net output, economically active labour force in agriculture, number of tractors used in agriculture, arable and permanent crop land, and fertilizer use in 128 countries from 1961 to 2002 (average $T = 40.3$ observations).

Additional time-invariant data on geographical distance between countries and contiguity is taken from CEPII (2006), and data on the share of agricultural land by climatic zone from Matthews (1983), available in Gallup, Mellinger and Sachs (1999). Data construction is discussed in the Appendix, which also contains descriptive statistics.

IV Empirical Results

Time-Series Properties and Cross-Section Dependence

We carry out a set of stationarity and nonstationarity tests for individual country time-series as well as first (Maddala and Wu, 1999) and second generation (Pesaran, 2007; Pesaran, Smith and Yamagata, 2009) panel unit root tests, results for which are reported in a Technical Appendix. Ultimately, in case of the present data dimensions and characteristics, and given all the problems and caveats of individual country and panel unit root tests, we can suggest *most*

and Glewwe, 2007). Details and an illustration can be found in a Technical Appendix.

[11] A number of alternative nonstationary panel estimators for the case of stationary factors are available in the literature (Pedroni, 2000), however given our emphasis on cross-section dependence and the nonstationary evolution of TFP we do not consider them here. We also do not adopt any methods where unobserved factors and the production function are estimated jointly (e.g. Bai, Kao and Ng, 2009) since these rely on identifying the right number of 'relevant' factors and have been argued to be unable to capture cross-section dependence of the 'weak' type (Chudik et al., 2011). Recent theoretical work furthermore concluded that 'in practice one is unlikely to do better than when using the relatively simple CA [CCE] approach' (Westerlund and Urbain, 2011).

conservatively that nonstationarity cannot be ruled out in this dataset. Below we indicate residual stationarity for each empirical model, analysed using Pesaran (2007), which we interpret as informal tests for cointegration (Banerjee and Carrion-i-Silvestre, 2011).[12]

The results for the cross-section dependence (CSD) analysis are again presented in a Technical Appendix. These provide strong evidence for the presence of cross-section correlation within the sample, based on average variable cross-country correlation coefficients, principal component analysis and the Pesaran (2004) CD test. Formal CD test results and mean absolute correlation coefficients for model residuals are reported below.

Pooled estimation results

We present the estimation results for the pooled specifications in Table 1.[13] In a lower panel of the table we report the implied returns to scale and labour coefficients as well as residual diagnostics. For models where CRS cannot be rejected we also estimate a restricted version of the model (in columns marked [b]). Recall for the following discussion that the 2FE estimator represents the empirical implementation of choice in the applied literature.

We first discuss parameter coefficients: in common with many studies using the FAO data, the coefficients on capital (tractors) are relatively low across all models, ranging from .056 (agro-climate CCEP) to .13 (POLS). The land coefficients are high and relatively stable across specifications (.24 to .36), whereas the fertilizer coefficients range from .025 to .17. Livestock again has a rather large coefficient across all specifications (.22 to .42). Regarding the implied labour coefficients, we find very low magnitudes across all specifications, with the standard CCEP even providing nonsensical negative values. The most striking pattern in these results is the magnitude of the implied decreasing returns to scale: based on this analysis we can conclude that in a pooled specification the data in all but two of the alternative CCEP estimators rejects constant returns emphatically, with input elasticities in the commonly favoured 2FE estimator adding up to around .8. This finding may reflect a global production function with substantial decreasing returns to scale; alternatively, it may reflect empirical misspecification.

Turning to the diagnostics, unit root tests indicate that the CCEP estimators yield stationary residuals, in contrast to the standard panel estimators in levels (POLS, 2FE) for which nonstationary residuals cannot be rejected. In the presence of nonstationary errors t-statistics are invalid (Kao, 1999) and tend to vastly overstate the precision of the parameter estimates (Bond and Eberhardt, 2009). Mean absolute residual correlations for POLS and 2FE are relatively high, at around .4, whereas this measure drops to around .16 in all other regression models. The formal CD tests for cross-section dependence yield very mixed results: only the residuals for the distance and agro-climate CCEP versions suggest cross-section independence. Further specification tests emphatically reject residual normality, serial independence and homoskedasticity in *all* models (results not reported). These diagnostics indicate that the commonly preferred 2FE estimator has serially correlated errors, which are nonstationary, non-normal, heteroskedastic and correlated across countries. Note that input parameter estimates for this estimator are reasonably close to those in Craig et al. (1997), the closest match for this dataset and specification.

If constant returns are imposed in *all* models (not reported), the estimates for land per worker rise dramatically with diagnostics virtually unchanged. The erroneous imposition of CRS leads to larger magnitudes for either the land or the implied labour coefficient albeit with

[12] Formal cointegration analysis is discussed in Section V. Further residual diagnostics are available on request.

[13] The dependent variable and the independent variables are expressed in per worker terms, such that the addition of the labour variable indicates deviation from constant returns to scale.

Table 1: Pooled regressions

	[1] POLS	[2] 2FE	[3] CCEP	[4] CCEP		[5] CCEP	[6] CCEP	
weight matrix♯			none	neighbour		distance	agro-climate	
				[a]	[b]		[a]	[b]
labour	-0.059	-0.191	-0.319	-0.015	–	-0.265	-0.195	–
	[17.50]**	[14.79]**	[2.35]**	[0.10]		[2.13]*	[1.72]	
tractors pw	0.131	0.058	0.074	0.089	0.090	0.061	0.056	0.083
	[24.12]**	[13.76]**	[4.86]**	[4.37]**	[4.69]**	[4.69]**	[3.61]**	[5.40]**
livestock pw	0.219	0.358	0.360	0.313	0.317	0.357	0.386	0.421
	[28.61]**	[30.85]**	[7.00]**	[4.16]**	[4.01]**	[6.32]**	[8.10]**	[9.23]**
fertilizer pw	0.169	0.073	0.025	0.049	0.049	0.029	0.027	0.032
	[28.08]**	[23.63]**	[4.67]**	[4.98]**	[4.88]**	[4.34]**	[4.41]**	[5.58]**
land pw	0.253	0.294	0.241	0.316	0.321	0.239	0.301	0.364
	[30.42]**	[21.07]**	[2.34]*	[3.56]**	[4.39]**	[2.81]**	[3.07]**	[4.73]**
Implied β_L	0.169	0.027	-0.020	0.217	0.222	0.050	0.035	0.100
Returns♭	DRS	DRS	DRS	CRS	–	DRS	CRS	–
RMSE	0.435	0.148	0.075	0.091	0.091	0.076	0.079	0.082
Stationarity †	I(1)	I(1)	I(0)	I(0)	I(0)	I(0)	I(0)	I(0)
Mean $\lvert\rho_{ij}\rvert$ ‡	0.42	0.41	0.18	0.15	0.16	0.16	0.16	0.17
CD (p)	-2.49 (.01)	9.64 (.00)	4.97 (.00)	5.07 (.00)	5.02 (.00)	1.58 (.11)	0.58 (.56)	-1.31 (.19)

Notes: Dependent variable: [1] & [3]-[6] log output per worker, [2] dto. in 2FE transformation as described in Coakley et al. (2006). We include $(T-1)$ year dummies in equation [1]. Independent variables: all variables are in log per worker terms with the exception of log labour. Constant term included but not reported. In models [4] and [6] where unrestricted models in column [a] indicate constant returns we report the restricted models (CRS imposed) in column [b]. Sample: $n = 5,162$ observations, $N = 128$ countries. The values in square brackets are absolute t-statistics, based on heteroskedasticity-robust standard errors. * and ** indicate statistical significance at the 5% and 1% level respectively. ♯ We apply different $N \times N$ weight matrices to construct the cross-section averages as described in the main text. ♭ The implied returns to scale are labeled decreasing (DRS) if the coefficient on labour is negative significant, and constant (CRS) if this coefficient is insignificant. The implied labour coefficient is computed by adding up all the coefficients on the RHS variables (except for labour), subtracting them from unity and then adding the coefficient on labour (the result is the implied labour coefficient *if constant returns were to hold*). RMSE reports the root mean squared error.
Residual Diagnostics: † Pesaran (2007) CIPS test results: I(0) — stationary, I(1) — nonstationary. Full results available on request. ‡ Mean Absolute Correlation Coefficient & Pesaran (2004) CD test, H_0: no CSD.

diagnostics that *reject* constant returns as well as the most common regression assumptions for this estimator (stationary, serially uncorrelated, normal and homoskedastic residuals).

Regarding the Pesaran (2006) CCEP and our extensions, our results in Table 1 show that the former emphatically rejects CRS and yields summed input elasticities of around .68, with a *negative* implied labour coefficient; of the other CCE estimators the distance version behaves in a similar fashion, whereas the other two cannot reject CRS. Only the distance and agro-climatic CCEP variants cannot reject cross-section independence in the residuals.[14]

In conclusion, our pooled models largely reject constant returns to scale, yield very low values for the implied coefficient on labour and over a range of specification tests indicate a combination of non-normality and heteroskedasticity, cross-section dependence, nonstationarity and/or serial correlation in the residuals. Allowing for heterogeneity in the unobserved common factors (as in the CCEP), although alleviating a potential identification problem, does not seem to provide an overall panacea. Our next analytical step is therefore to allow for full technology heterogeneity across countries.

[14] One may speculate that the other CCEP models fail to account for all cross-section dependence in the data since sets of unobservables differ across country groups (as developed above).

Averaged country regression estimates

We present the results from Mean Group-type estimators in Table 2; for all estimators we present the robust coefficient mean across N country parameters.[15] Following Pesaran and Smith (1995) the t-ratios reported for each average estimate test whether the average parameter is statistically different from zero.

The MG model displays large decreasing returns. The standard, neighbour and agro-climate CCEMG in contrast have insignificant coefficients on labour, indicating constant returns in the *average* country regression. Around 25 to 35 countries reject CRS at the 5% level of significance in each of these models. The distance CCEMG indicates very large and highly significant decreasing returns to scale.

Table 2: Mean Group type estimators

	[1]	[2]		[3]		[4]	[5]	
	MG	**CCEMG**		**CCEMG**		**CCEMG**	**CCEMG**	
weight matrix$^\sharp$		none		neighbour		distance	agro-climate	
		[a]	[b]	[a]	[b]		[a]	[b]
labour	-0.357	-0.126	–	-0.193	–	-0.311	-0.039	–
	[2.23]*	[1.04]		[1.67]		[2.62]**	[0.36]	
tractors pw	0.075	0.064	0.109	0.051	0.096	0.078	0.058	0.086
	[3.31]**	[3.26]**	[5.13]**	[2.14]*	[4.17]**	[3.60]**	[2.45]*	[3.82]**
livestock pw	0.246	0.324	0.321	0.306	0.321	0.278	0.292	0.339
	[8.07]**	[9.44]**	[9.47]**	[7.22]**	[8.22]**	[7.24]**	[7.51]**	[9.97]**
fertilizer pw	0.030	0.031	0.036	0.026	0.035	0.029	0.029	0.035
	[4.86]**	[5.02]**	[5.63]**	[4.51]**	[5.19]**	[5.11]**	[4.77]**	[5.63]**
land pw	0.210	0.200	0.201	0.111	0.237	0.081	0.214	0.190
	[2.79]**	[2.68]**	[3.57]**	[1.41]	[4.14]**	[1.14]	[3.35]**	[3.63]**
implied β_L	0.082	0.256	0.333	0.312	0.311	0.223	0.223	0.353
Returns	DRS	CRS	–	CRS	–	DRS	CRS	–
# reject CRS	48	26	–	36	–	35	32	–
RMSE	0.066	0.055	0.059	0.054	0.060	0.053	0.056	0.060
Stationarity †	I(0)	I(0)	I(0)	I(0)	I(0)	I(0)	I(0)	I(0)
Mean $\lvert\rho_{ij}\rvert$ ‡	0.15	0.15	0.15	0.14	0.14	0.14	0.14	0.14
CD statistic (p)	9.16 (.00)	0.21 (.84)	-0.23 (.82)	2.04 (.04)	2.02 (.04)	-0.49 (.62)	-0.11 (.92)	-1.01 (.31)

Notes: $n = 5,162$, N=128. We present robust sample means for all model parameters. Terms in brackets are absolute t-statistics (H_0: $(1/N)\sum_i \hat{\beta}_i = 0$) following Pesaran and Smith (1995). We report the number of countries rejecting CRS. For all other details see Table 1.

Regarding average parameter estimates on the factor inputs, the MG estimator yields coefficients for capital around .08, for livestock around .25, for fertilizer around .03 and for land around .20. In comparison to the parameter differences across models in the *pooled* specifications, these estimates are relatively similar to those in the CCEMG models, with the crucial exception of the returns to scale coefficient and thus the implied labour elasticity: for the MG this is .08, whereas in the latter group it is around .30.[16] Within the CCEMG group

[15]This procedure employs weights based on the absolute residuals to reduce the impact of outliers on the mean estimate Hamilton (1992). Median estimates (not reported) are very close to the robust means presented.

[16]There are two ways of reporting the variation in the implied labour parameter to illustrate heterogeneity: *firstly*, we can report the 95% confidence interval of the coefficient computed from the robust means reported here, which yields the interval [.23, .48]. *Second*, we can compute the implied labour coefficient in each country and create a robust mean (.24), which yields the 95% confidence interval [.13, .36]. Note that country-specific parameter coefficients should not be viewed in isolation (Pedroni, 2007, p.440), frequently yielding economically implausible magnitudes (Boyd and Smith, 2002), as is illustrated by the extreme values of the distribution in the present case, -2.30 and 3.05. In Eberhardt and Teal (forthcoming) we provide a formal explanation for the

the standard and the agro-climate estimators yield very similar results, in particular for the restricted models.[17] For these livestock has a larger coefficient than in the MG benchmark.

Turning to the diagnostics, all models reject nonstationary residuals in the Pesaran (2007) CIPS tests. Mean absolute error correlation is uniformly low at .15, but cross-section independence is rejected in the MG as well as in the neighbour CCEMG estimator. Taking the results for other diagnostic tests into account,[18] we suggest that evidence in favour of the standard CCEMG and agro-climate CCEMG seems most convincing: compared to the other models these provide some evidence of serially uncorrelated and normal residuals.

Summary of the findings

In conclusion to our production function estimations we find that the pooled models favoured by the existing applied literature result in regression residuals which seriously question the validity of the empirical estimates, while heterogeneous parameter models yield more favourable diagnostic results, in particular for the standard and agro-climate CCEMG (CRS). In order to emphasise the stark differences in the coefficient estimates between the 2FE and our preferred CCEMG estimates as well as the 'importance' attached to the respective input a direct comparison reveals that (i) the magnitude of the tractor coefficient next to doubles, (ii) that of the fertilizer coefficient halves, (iii) that of the land coefficient drops by a third, while (iv) the (implied) labour coefficient increases eight-fold when moving from a 2FE-based to an agro-climate CCEMG-based analysis — these numbers speak for themselves in justifying the more flexible approach taken in the present analysis.[19]

V Cointegration and the Direction of Causation

A central focal point of the cross-country growth literature over the past decades has been the concern over variable endogeneity, to the extent that a credible instrumentation strategy represents the dominant criterion for the validity of empirical results. In the present analysis we have emphasised one type of endogeneity, whereby common factors drive inputs as well as output, leading to identification issues unless these factors are accounted for. In addition we need to be concerned about a type of endogeneity which more fundamentally implies 'reverse causality'. In a simplified version of equations (1) and (2) we can express this as

$$y_{it} = \beta_i x_{it} + u_{it} \qquad u_{it} = \alpha_i + \lambda_i f_t + \varepsilon_{it} \tag{5}$$

$$x_{it} = \pi_i + \delta_i g_t + \rho_i f_t + \psi_i \varepsilon_{it} + v_{it} \tag{6}$$

for a single covariate x and single factor f contained in both the y- and x-equations. Due to the presence of $\psi_i \varepsilon_{it}$ in the second equation we should be concerned over whether y 'causes' x or the reverse being the case or both. In the case of a production function we may be

occurrence of these implausible country estimates alongside the much more sensible average values.

[17]If CRS is imposed in all models it can be seen that this restriction does not create the same dramatic changes to the land estimates as in the pooled estimator case.

[18]The cross-section correlation test suggested by Bai (2009) yields similar results across estimators (not reported). A number of additional residual tests (results not reported) based on the combination of time-series tests yield favourable diagnostics for the CCEMG models.

[19]So as to assure the robustness of the results discussed above we carried out a number of additional regressions and tests, all of which are available on request or contained in a Technical Appendix: dynamic versions of the above empirical models vindicate the results discussed; parameter constancy over time is suggested to be a valid assumption; correlation coefficients between the means of the variable series and the heterogeneous parameter estimates indicate limited correlation in the preferred CCEMG models, which implies that there is no systematic difference in the technology estimates associated with the levels of inputs or output.

concerned that our estimation equation is a labour demand equation or investment equation in disguise. The standard approach has therefore been to instrument for x using one or a set of variables z which satisfy the conditions of informativeness ($\mathbb{E}[zx] \neq 0$) and validity ($\mathbb{E}[z\varepsilon] = 0$). Many researchers have expressed doubt whether it is possible to obtain credible instruments in the macro data context (e.g. Durlauf et al., 2005; Clemens and Bazzi, 2009) and it is furthermore well-established amongst econometricians that instrumentation in a pooled model is impossible if the underlying relationship is heterogeneous across countries (Pesaran and Smith, 1995).

Having adopted a panel time series approach the issues of endogeneity and direction of causation can here take an alternative pathway: given that our variables are nonstationary, we first need to show that they are cointegrated; in a second step we can then apply a test for direction of causation to provide evidence that our above results can indeed be interpreted as empirical production functions. Cointegration tests are commonly carried out as a pre-estimation testing procedure, however we have delayed these until after estimation since we hypothesise that unobservables (TFP) form part of the cointegrating vector (Banerjee and Carrion-i-Silvestre, 2011).

We carry out panel cointegration tests based on the error correction model representation, following Gengenbach, Urbain and Westerlund (2009). Continuing with the simple model in (5) and (6) we estimate the following equation for each country i

$$\Delta y_{it} = \alpha_i y_{i,t-1} + \gamma'_{1i} x_{i,t-1} + \gamma'_{2i} f_{i,t-1} + \sum_{s=1}^{p_i} \pi'_{1is} \Delta y_{i,t-s} + \sum_{s=0}^{p_i} \pi'_{2is} \Delta x_{i,t-s} + \sum_{s=0}^{p_i} \pi'_{3is} \Delta f_{i,t-s} + \varepsilon_{it} \quad (7)$$

In order to capture the common factors f we use the cross-section averages of all variables in the model, including lagged Δy_{it} and Δx_{it}, depending on the lag-length p_i. We run this test for each CCEMG model with different deterministics (none, intercept, intercept and trend). In each case the test statistic $\bar{\tau}^*$ is a simple average of the t-ratios for $\hat{\alpha}_i$ from the country regressions. Since the individual t-ratios as well as their averages have non-standard distributions under the null we provide a number of critical values.

Table 3: Gengenbach, Urbain & Westerlund (2009) cointegration tests

$\bar{\tau}^*$	Standard	Neighbour	Distance	Agro-Clim	10%	5%	1%
Model 1	-3.653***	-3.446***	-3.631***	-3.681***	-3.152	-3.204	-3.297
Model 2	-3.697***	-3.752***	-3.633***	-3.957***	-3.459	-3.507	-3.584
Model 3	-3.765*	-3.885***	-3.631	-3.879**	-3.749	-3.793	-3.881

Notes: The $\bar{\tau}^*$ statistics are averages of the N t-ratios from the country ECM regressions, where extreme t-ratios have been replaced by bounds (truncated; we used $\varepsilon = .000001$) following the strategy devised in Gengenbach et al. (2009). This paper also provides simulated critical values we present in the right-most 3 columns (we pick $N = 100$ as this is the largest dimension for which critical values were reported — our judgment is thus more conservative). Model 1-3 refers to the ECM specification without deterministics, with an intercept and with intercept and trend, respectively. H_0: no error correction, i.e. no cointegration. Lag-length p_i determined using Bayesian/Schwarz IC; empirical specification: static model, CRS imposed.

As can be seen in Table 3 in most cases the ECM equations reject the null of 'no error correction' at the 1% level (thus implying cointegration).[20]

We now turn to the issue of causality, where we follow the example and discussion in Canning and Pedroni (2008, Section 4). Provided there exists a cointegrating relationship

[20] We further carried out a set of panel cointegration tests following Pedroni (1999, 2004) where an estimate of country-specific TFP is included in the cointegrating vector. Results (available on request) support the above finding of heterogeneous cointegration.

between variables the Granger Representation Theorem (Engle and Granger, 1987) states that these series can be represented in the form of a dynamic Error Correction Model (ECM). Generically, for a pair of cointegrated variables x and y we can write

$$\Delta y_{it} = c_{1i} + \lambda_{1i}\hat{e}_{i,t-1} + \sum_{j=1}^{K} \psi_{11ij}\Delta y_{i,t-j} + \sum_{j=1}^{K} \psi_{12ij}\Delta x_{i,t-j} + \varepsilon_{1it} \tag{8}$$

$$\Delta x_{it} = c_{2i} + \lambda_{2i}\hat{e}_{i,t-1} + \sum_{j=1}^{K} \psi_{21ij}\Delta y_{i,t-j} + \sum_{j=1}^{K} \psi_{22ij}\Delta x_{i,t-j} + \varepsilon_{2it} \tag{9}$$

where $\hat{e}_{i,t-1}$ represents the 'disequilibrium term' $\hat{e} = y - \hat{\beta}_i x - \hat{d}$ constructed using the estimated cointegrating relationship between these two variables (d represents deterministic terms). Equations (8) and (9) further include lagged differences of the variables in the cointegrating relationship. In the above example there are only two equations, since we have two variables in the cointegrating relationship. The Granger Representation Theorem implies that for a long-run equilibrium relationship to exist between y and x at least one of λ_{1i} and λ_{2i} must be non-zero: if (and only if) $\lambda_{1i} \neq 0$ then x has a causal impact on y (our notation: $x \to y$), if (and only if) $\lambda_{2i} \neq 0$ then the causal impact is reversed ($y \to x$). If both λ_{1i} and λ_{2i} are non-zero they determine each other jointly.

We carried out the above regressions for the pooled 2FE and heterogeneous MG and CCEMG specifications, where in each case the disequilibrium term $\hat{e}$ represents the model residuals of static regression(s) with CRS imposed. In keeping with the Pesaran (2006, 2007) approach in case of the CCEMG estimators we include cross-section averages of all variables in the ECM (including lags) to the regression — note again that this implies TFP is part of the cointegrating relationship.[21]

The ECM regressions are estimated at the country-level and empirical estimates for λ_i are investigated using standard t-ratios, given that all variables in the ECM regression are stationary — this property is confirmed for $\hat{e}$ in the CCE regressions but rejected for 2FE (not presented). We present the results of this analysis in Table 4: in the row for the 'output equation' we report the results for a test whether the set of inputs (and in the CCEMG models: TFP) have a causal impact on output, with the null hypothesis of 'no causal impact' and in analogy for the other variables in the subsequent rows. We report the averaged t-ratio (GM= $N^{-1}\sum_i t_{\lambda_i}$) and its probability value, as well as a Fisher-type statistic (constructed from the p-values of the t-ratios in each ECM regression) and its probability value. In either case the null hypothesis is that $\lambda_i = 0$. It is important to note that the interpretation of the GM and Fisher statistics differ in the case where the λ coefficients *differ* across countries: the former is a two-sided tests and can take on positive or negative values under the non-zero alternative hypothesis; the latter in contrast is a one-sided test and only takes positive values. The difference is therefore one between $\hat{\lambda}_i$ being *on average* zero compared with being *pervasively* zero (Canning and Pedroni, 2008). If the two tests disagree we learn about the heterogeneity of $\hat{\lambda}_i$ across countries. We also report the robust mean parameter for $\hat{\lambda}_i$. The final column summarises our analysis to reach a verdict on direction of causation.

We first investigate the 2FE results: the difference between the Fisher and GM test results suggests that there *do* exist long-run causal relations between these variables and that they differ across countries. However, we have to conclude that *all* variables are endogeneously

[21] Due to the limited time-series dimension of our data we present results with one and two lags ($K = 2$, implying 24 covariates), which reduces the sample to $N = 122$ countries. We carried out all of the tests with $K = 3$ (implying 34 covariates), which reduces the sample to $N = 103$ countries, and found similar patterns of rejection and non-rejection across estimators and ECM equations.

determined, such that the empirical equation cannot be interpreted as a production function. The averaged $\hat{\lambda}_i$ parameters confirm this interpretation. An alternative view is that the empirical misspecification has led to nonstationary residuals, such that there is noncointegration between the variables in this model.

Table 4: Canning & Pedroni (2008) tests for direction of causation

2FE	GM	(p)	Fisher	(p)	mean $\hat{\lambda}_i$	t-stats	Verdict		
output equation	-0.97	0.00	485.2	0.00	-0.142	-7.91	x	$\rightarrow$	y
tractor equation	0.18	0.17	456.2	0.00	0.024	1.81	x_{-tr}, y	$\rightarrow$	x_{tr}
livestock equation	0.38	0.00	351.0	0.00	0.043	3.77	x_{-live}, y	$\rightarrow$	x_{live}
fertilizer equation	0.10	0.42	432.3	0.00	0.141	1.82	x_{-f}, y	$\rightarrow$	x_f
land equation	0.37	0.00	395.2	0.00	0.011	1.91	x_{-n}, y	$\rightarrow$	x_n
MG	GM	(p)	Fisher	(p)	mean $\hat{\lambda}_i$	t-stats	Verdict		
output equation	-2.93	0.00	1,612.1	0.00	-0.976	-24.00	x	$\rightarrow$	y
tractor equation	-0.16	0.87	274.7	0.20	-0.029	-0.98	x_{-tr}, y	$\nrightarrow$	x_{tr}
livestock equation	0.03	0.98	307.6	0.01	0.015	0.55	x_{-live}, y	$\rightarrow$	x_{live}
fertilizer equation	-0.06	0.96	257.2	0.47	-0.116	-0.85	x_{-f}, y	$\nrightarrow$	x_f
land equation	-0.06	0.95	286.5	0.09	-0.004	-0.33	x_{-n}, y	$\rightarrow$	x_n
CCEMG	GM	(p)	Fisher	(p)	mean $\hat{\lambda}_i$	t-stats	Verdict		
output equation	-2.09	0.04	938.4	0.00	-0.925	-18.63	x, TFP	$\rightarrow$	y
tractor equation	-0.08	0.94	189.3	1.00	-0.045	-1.28	x_{-tr}, y, TFP	$\nrightarrow$	x_{tr}
livestock equation	0.12	0.91	331.1	0.00	0.011	0.27	x_{-live}, y, TFP	$\rightarrow$	x_{live}
fertilizer equation	-0.04	0.97	232.7	0.69	0.023	0.14	x_{-f}, y, TFP	$\nrightarrow$	x_f
land equation	0.06	0.95	225.7	0.79	0.010	0.68	x_{-n}, y, TFP	$\nrightarrow$	x_n
CCEMG neighbour	GM	(p)	Fisher	(p)	mean $\hat{\lambda}_i$	t-stats	Verdict		
output equation	-1.88	0.06	856.0	0.00	-0.784	-16.09	x, TFP	$\rightarrow$	y
tractor equation	0.03	0.97	260.4	0.22	0.032	0.79	x_{-tr}, y, TFP	$\nrightarrow$	x_{tr}
livestock equation	0.07	0.94	305.7	0.00	0.004	0.11	x_{-live}, y, TFP	$\rightarrow$	x_{live}
fertilizer equation	0.16	0.87	299.1	0.01	0.179	1.02	x_{-f}, y, TFP	$\rightarrow$	x_f
land equation	0.07	0.95	312.3	0.00	0.002	0.11	x_{-n}, y, TFP	$\rightarrow$	x_n
CCEMG distance	GM	(p)	Fisher	(p)	mean $\hat{\lambda}_i$	t-stats	Verdict		
output equation	-2.04	0.04	932.4	0.00	-0.997	-21.42	x, TFP	$\rightarrow$	y
tractor equation	-0.12	0.90	243.8	0.49	-0.041	-1.10	x_{-tr}, y, TFP	$\nrightarrow$	x_{tr}
livestock equation	0.10	0.92	301.9	0.01	0.050	1.21	x_{-live}, y, TFP	$\rightarrow$	x_{live}
fertilizer equation	-0.02	0.98	257.5	0.26	-0.059	-0.31	x_{-f}, y, TFP	$\nrightarrow$	x_f
land equation	-0.12	0.91	244.2	0.48	-0.001	-0.06	x_{-n}, y, TFP	$\nrightarrow$	x_n
CCEMG agro-climate	GM	(p)	Fisher	(p)	mean $\hat{\lambda}_i$	t-stats	Verdict		
output equation	-2.25	0.02	1,035.5	0.00	-0.935	-20.16	x, TFP	$\rightarrow$	y
tractor equation	-0.02	0.98	241.8	0.53	-0.013	-0.42	x_{-tr}, y, TFP	$\nrightarrow$	x_{tr}
livestock equation	0.15	0.88	380.0	0.00	0.048	1.23	x_{-live}, y, TFP	$\rightarrow$	x_{live}
fertilizer equation	0.07	0.94	242.5	0.52	-0.004	-0.02	x_{-f}, y, TFP	$\nrightarrow$	x_f
land equation	-0.09	0.93	227.3	0.77	-0.001	-0.08	x_{-n}, y, TFP	$\nrightarrow$	x_n

Notes: We report various test statistics for the null of no long-run causal impact between sets of different variables. In each case '*variable* equation' refers to the ECM regression with 'variable' on the LHS. GM gives the group-mean average of country-specific t-ratios for the coefficient on the disequilibrium term ($\hat{\lambda}_i$) which is distributed $N(0,1)$. Fisher gives $-2\sum_i \log \pi_i$ where π_i is the probability value of the country-specific t-ratio on the disequilibrium term. The Fisher statistic is distributed $\chi^2(2N)$. The final two columns but one report the mean estimate for $\hat{\lambda}_i$ and the associated t-ratio. In the 'Verdict' column we summarise the analysis, using x_{-tr} for 'all inputs other than tractors' etc. and $\rightarrow$ as short-hand for 'does cause' and $\nrightarrow$ for 'does not cause'. TFP is included implicitly via cross-section averages (see main text). Models: static with CRS imposed.

Turning to the MG results, it can be seen that the GM and Fisher statistics agree that there is a causal relationship from the inputs to output. However, the remaining ECM results suggest that there are heterogeneous causal relationships in the land and livestock equations, meaning that the production function may not be the only relationship represented by our empirical results. The four CCEMG results have two characteristic features: *firstly*, we find strong evidence for a causal relationship from inputs and implied TFP to output in all four models based on GM and Fisher statistics as well as mean $\hat{\lambda}_i$. *Secondly*, in the standard, distance and agro-climate CCEMG results we cannot reject the null of no long-run causal relationship in the other ECM equations (tractor, fertilizer, land), with the exception of the livestock equation, whereas this not the case for the neighbour CCEMG results (livestock, fertilizer and land reject). The evidence for uni-directional causation from inputs to output is thus somewhat mixed for these heterogeneous parameter models. Having said that, the share of countries which reject the absence of a long-run, causal relationship for the output equations is uniformly high in all these models (not reported). For the other equations less than 15% of countries (less than 20% in the neighbour, less than 7% in the agro-climate case) reject the null of no long-run causal relationship.

The overarching concern of this analysis was to show that our empirical results can be reasonably assumed to represent a causal relationship from inputs to output, and not vice versa. Our results suggest that we can be comfortable with this interpretation in the standard, distance and agro-climatic CCEMG models, and to a lesser extent in the neighbourhood CCEMG and MG models. The misspecified 2FE model was shown to fail this test entirely.

VI Conclusions

In this paper we investigated the determinants of agricultural productivity in a large panel of developing and developed countries, allowing for technology heterogeneity, variable non-stationarity, cross-section dependence and flexibility in the returns to scale. A review of the literature indicated that empirical implementation in existing studies is dominated by the pooled OLS (POLS) and two-way fixed effects (2FE) models, with constant returns to scale often imposed without formal testing. Our results display considerable differences in the estimated parameters when moving from pooled models to averaged country regressions, and between equations ignoring and accounting for cross-section dependence, and we use diagnostic testing to determine favourable specification(s) and estimator(s). We draw four conclusions from our analysis:

Firstly, parameter heterogeneity plays a crucial role in cross-country productivity investigation for agriculture. The use of pooled models is largely rejected by the diagnostic tests we apply. These tests further indicate that the preferred pooled estimator applied in the literature (2FE) is seriously misspecified. The implications for the interpretation of technology parameter and by extension any TFP estimates derived from these are shown to be very serious in our sample for the agricultural sector. Beyond the application to agricultural production, it is important to reiterate that if technology is heterogeneous across countries then none of the standard instrumentation strategies applied in the cross-country empirical literature (instrumentation using z variables or lags) are valid (Pesaran and Smith, 1995), since the empirical specifications in these cases assume technology homogeneity (e.g. Acemoglu, Johnson and Robinson, 2001; Miguel, Satyanath and Sergenti, 2004; Rajan and Subramanian, 2008).

Secondly, even once technology heterogeneity with regards to observable inputs is accounted for, our results show that the presence of time-varying unobserved heterogeneity matters for empirical modelling. We employed the common factor structure to model variable endogeneity and tested all models for reverse causality. In contrast, standard instrumentation strategies in cross-country empirical analysis assume that variable series are cross-sectionally independent and do not allow for the heterogeneous impact of global shocks such as the recent financial crisis across countries. The importance of observed and unobserved heterogeneity in macro panel empirics has long been recognised in the econometric theory literature, however few applied studies have followed the call toward an integrated treatment of the production technology *in its entirety*, including heterogeneity and its source (Durlauf, Kourtellos and Minkin, 2001). We feel this opens up interesting avenues of enquiry away from the hankering after natural experiments and the quest for instrumental variables which have dominated the recent cross-country literature.

Thirdly, the imposition of constant returns to agricultural production on pooled regression equations is rejected by the data and leads to qualitatively different empirical results. In contrast, the imposition of CRS on individual country regressions does not change average parameter coefficients considerably compared with the unrestricted results: thus the observed decreasing returns at the global level evident in the pooled models is an artefact of empirical misspecification.

Fourthly, our extension to the CCEMG estimator of applying (exogenous) weight matrices before computing the cross-section averages and in effect imposing more structure on the nature of cross-section correlation in the data has provided further interesting insights. Based on our results we suggest that the common correlated effects driving the cross-section dependence in agricultural production are closely proxied by our measure for agro-climatic distance. The implication of our findings is that agricultural TFP is affected by different factors and has different levels of responsiveness across geographical regions of the world due to agro-climatic diversity. Furthermore, technology transfer between countries is limited by whether technology can be adapted to the local environment — both of these statements are widely accepted in the literature but prior to this study were ignored in empirical analyses of cross-country productivity. They merit further investigation which we are currently pursuing.

References

Abramowitz, Moses (1956). “Resource and output trend in the United States since 1870.” *American Economic Review*, Vol. 46(2): 5–23.

Acemoglu, Daron, Johnson, Simon and Robinson, James A. (2001). “The Colonial Origins of Comparative Development: An Empirical Investigation.” *American Economic Review*, Vol. 91(5): 1369–1401.

Bai, Jushan (2009). “Panel Data Models with Interactive Fixed Effects.” *Econometrica*, Vol. 77(4): 1229–1279.

Bai, Jushan, Kao, Chihwa and Ng, Serena (2009). “Panel cointegration with global stochastic trends.” *Journal of Econometrics*, Vol. 149(1): 82–99.

Banerjee, Anindya, Eberhardt, Markus and Reade, James (2010). “Panel Estimation for Worriers.” Oxford University, Department of Economics Discussion Paper Series #514.

Banerjee, Anindya and Carrion-i-Silvestre, Josep Lluís (2011). “Testing for Panel Cointegration using Common Correlated Effects Estimators.” University of Birmingham, Department of Economics Discussion Paper Series 11-16.

Block, Steven (2010). “The Decline and Rise of Agricultural Productivity in Sub-Saharan Africa Since 1961.” NBER Working Papers 16481, National Bureau of Economic Research.

Bond, Stephen R. and Eberhardt, Markus (2009). “Cross-section dependence in nonstationary panel models: a novel estimator.” Paper presented at the Nordic Econometrics Meeting, Lund/Sweden, October 2009.

Bond, Stephen R., Leblebicioglu, Asli and Schiantarelli, Fabio (2010). "Capital Accumulation and Growth: A New Look at the Empirical Evidence." *Journal of Applied Econometrics*, Vol. 15(7): 1073–1099.

Boyd, Derick A.C. and Smith, Ron P. (2002). "Some econometric issues in measuring the monetary transmission mechanism, with an application to developing countries." In: Lavan Mahadeva and Peter Sinclair (Editors), "Monetary Transmission in Diverse Economies," (Cambridge University Press), pp. 68–99.

Bravo-Ortega, Claudio and Lederman, Daniel (2004). "Agricultural Productivity and its Determinants: Revisiting International Experiences." *Estudios de Economia*, Vol. 31(2): 133–163.

Canning, David and Pedroni, Peter (2008). "Infrastructure, Long-Run Economic Growth And Causality Tests For Cointegrated Panels." *Manchester School*, Vol. 76(5): 504–527.

Cavalcanti, Tiago, Mohaddes, Kamiar and Raissi, Mehdi (2011). "Growth, development and natural resources: New evidence using a heterogeneous panel analysis." *The Quarterly Review of Economics and Finance*, Vol. 51(4): 305–318.

CEPII (2006). "Datasets for Geodesic Distances." Centre D'Etudes Prospectives et D'Information Internationales. Note by Thierry Mayer and Soledad Zignago.

Cermeño, Rodolfo, Maddala, G.S. and Trueblood, Michael A. (2003). "Modelling technology as a dynamic error components process: The case of the inter-country agricultural production function." *Econometric Reviews*, Vol. 22: 289–306.

Chudik, Alexander, Pesaran, M. Hashem and Tosetti, Elisa (2011). "Weak and Strong Cross Section Dependence and Estimation of Large Panels." *Econometrics Journal*, Vol. 14(1): C45–C90.

Clemens, Michael and Bazzi, Samuel (2009). "Blunt Instruments: On Establishing the Causes of Economic Growth." Center for Global Development Working Papers #171.

Coakley, Jerry, Fuertes, Ana-Maria and Smith, Ron P. (2006). "Unobserved heterogeneity in panel time series models." *Computational Statistics & Data Analysis*, Vol. 50(9): 2361–2380.

Coelli, Tim J. and Rao, D. S. Prasada (2005). "Total factor productivity growth in agriculture: a Malmquist index analysis of 93 countries, 1980-2000." *Agricultural Economics*, Vol. 32(S1): 115–134.

Craig, Barbara J, Pardey, Philip G and Roseboom, Johannes (1997). "International Productivity Patterns: Accounting for Input Quality, Infrastructure, and Research." *American Journal of Agricultural Economics*, Vol. 79(4): 1064–1076.

Durlauf, Steven N., Johnson, Paul A. and Temple, Jonathan R.W. (2005). "Growth Econometrics." In: Philippe Aghion and Steven Durlauf (Editors), "Handbook of Economic Growth," Vol. 1 of *Handbook of Economic Growth*, chapter 8 (Elsevier), pp. 555–677.

Durlauf, Steven N., Kourtellos, Andros and Minkin, Artur (2001). "The local Solow growth model." *European Economic Review*, Vol. 45(4-6): 928–940.

Eberhardt, Markus and Helmers, Christian (2010). "Untested Assumptions and Data Slicing: A Critical Review of Firm-Level Production Function Estimators." Oxford University, Department of Economics Discussion Paper Series #513.

Eberhardt, Markus, Helmers, Christian and Strauss, Hubert (forthcoming). "Do spillovers matter when estimating private returns to R&D?" *The Review of Economics and Statistics*.

Eberhardt, Markus and Teal, Francis (2011). "Econometrics for Grumblers: A New Look at the Literature on Cross-Country Growth Empirics." *Journal of Economic Surveys*, Vol. 25(1): 109–155.

Eberhardt, Markus and Teal, Francis (forthcoming). "Structural Change and Cross-Country Growth Empirics." *World Bank Economic Review*.

Engle, Robert F. and Granger, Clive W.J. (1987). "Cointegration and error correction: representations, estimation and testing." *Econometrica*, Vol. 55(2): 252–276.

Ertur, Cem and Koch, Wilfried (2007). "Growth, technological interdependence and spatial externalities: theory and evidence." *Journal of Applied Econometrics*, Vol. 22(6): 1033–1062.

FAO (2007). "FAOSTAT." Online database, Rome: UN Food and Agriculture Organisation.

Frankel, Jeffrey A. and Romer, David (1999). "Does Trade Cause Growth?" *American Economic Review*, Vol. 89(3): 379–399.

Gallup, John L., Mellinger, Andrew D. and Sachs, Jeffrey D. (1999). "Geography Datasets: Agricultural Measures." Available online at CID, Harvard University.

Gengenbach, Christian, Urbain, Jean-Pierre and Westerlund, Joakim (2009). "Panel Error Correction Testing with Global Stochastic Trends." Working paper, Maastricht: METEOR.

Gollin, Doug (2010). "Agricultural productivity and economic growth." In: Prabhu Pingali and Robert Evenson (Editors), "Handbook of Agricultural Economics," Vol. 18 of *Handbook of Agricultural Economics*, chapter 71 (Elsevier North-Holland), pp. 3825–3866.

Graham, Bryan S. and Temple, Jonathan R.W. (2006). "Rich nations, poor nations: how much can multiple equilibria explain?" *Journal of Economic Growth*, Vol. 11(1): 5–41.

Gutierrez, Luciano and Gutierrez, M. M. (2003). "International R&D spillovers and productivity growth in the agricultural sector. A panel cointegration approach." *European Review of Agriculture Economics*, Vol. 30(3): 281–303.

Hamilton, Lawrence C. (1992). "How Robust is Robust Regression?" *Stata Technical Bulletin*, Vol. 1(2).

Hayami, Yujiro and Ruttan, Vernon W. (1970). "Agricultural Productivity Differences Among Countries." *American Economic Review*, Vol. 60(5): 895–911.

Hayami, Yujiro and Ruttan, Vernon W. (1985). *Agricultural development: An international perspective* (Baltimore: Johns Hopkins University Press).

Jaffe, Adam (1986). "Technological Opportunity and Spillovers of R&D: Evidence from Firms' Patents, Profits and Market Value." NBER Working Paper Series No. 1815, NBER.

Kao, Chihwa (1999). "Spurious regression and residual-based tests for cointegration in panel data." *Journal of Econometrics*, Vol. 65(1): 9–15.

Kao, Chihwa, Chiang, Min-Hsien and Chen, Bangtian (1999). "International R&D spillovers: An application of estimation and inference in panel cointegration." *Oxford Bulletin of Economics and Statistics*, Vol. 61(S1): 691–709.

Kapetanios, George, Pesaran, M. Hashem and Yamagata, Takashi (2011). "Panels with Nonstationary Multifactor Error Structures." *Journal of Econometrics*, Vol. 160(2): 326–348.

Larson, Donald F., Butzer, Rita, Mundlak, Yair and Crego, Al (2000). "A cross-country database for sector investment and capital." *World Bank Economic Review*, Vol. 14(2): 371–391.

Levinsohn, James and Petrin, Amil (2003). "Estimating Production Functions Using Inputs to Control for Unobservables." *Review of Economic Studies*, Vol. 70(2): 317–341.

Lewis, W. Arthur (1954). "Economic Development with Unlimited Supplies of Labour." *The Manchester School*, Vol. 22: 139–191.

Maddala, G. S. and Wu, Shaowen (1999). "A comparative study of unit root tests with panel data and a new simple test." *Oxford Bulletin of Economics and Statistics*, Vol. 61(Special Issue): 631–652.

Martin, Will and Mitra, Devashish (2002). "Productivity Growth and Convergence in Agriculture versus Manufacturing." *Economic Development and Cultural Change*, Vol. 49(2): 403–422.

Matthews, Elaine (1983). "Global Vegetation and Land Use: New High Resolution Databases for Climate Studies." *Journal of Climate and Applied Meteorology*, Vol. 22: 474–487.

Miguel, Edward, Satyanath, Shanker and Sergenti, Ernest (2004). "Economic Shocks and Civil Conflict: An Instrumental Variables Approach." *Journal of Political Economy*, Vol. 112(4): 725–753.

Mundlak, Yair, Larson, Donald and Butzer, Rita (1999). "Rethinking within and between regressions: The case of agricultural production functions." *Annales D'Economie et de Statistique*, Vol. 55/56: 475–501.

Mundlak, Yair, Larson, Donald F. and Butzer, Rita (2008). "Heterogeneous technology and panel data: the case of the agricultural production function." Policy Research Working Paper Series #4536.

Nelson, Richard (1968). "A 'Diffusion' Model of International Productivity Differences in Manufacturing Industry." *American Economic Review*, Vol. 58(5): 1219–1248.

Pardey, Philip, James, Jennifer, Alston, Julian, Wood, Stanley, Koo, Bonwoo, Binenbaum, Eran, Hurley, Terrance and Glewwe, Paul (2007). "Science, Technology and Skills." Background Paper for the World Bank 'World Development Report 2008: Agriculture for Development'.

Pedroni, Peter (1999). "Critical values for cointegration tests in heterogeneous panels with multiple regressors." *Oxford Bulletin of Economics and Statistics*, Vol. 61(S1): 653–670.

Pedroni, Peter (2000). "Fully modified OLS for heterogeneous cointegrated panels." In: Badi H. Baltagi (Editor), "Nonstationary panels, cointegration in panels and dynamic panels," (Amsterdam: Elsevier).

Pedroni, Peter (2004). "Panel Cointegration: Asymptotic and Finite Sample Properties of Pooled Time Series Tests with an Application to the PPP Hypothesis." *Econometric Theory*, Vol. 20(3): 597–625.

Pedroni, Peter (2007). "Social capital, barriers to production and capital shares: implications for the importance of parameter heterogeneity from a nonstationary panel approach." *Journal of Applied Econometrics*, Vol. 22(2): 429–451.

Pesaran, M. Hashem (2004). "General Diagnostic Tests for Cross Section Dependence in Panels." IZA Discussion Paper No. 1240.

Pesaran, M. Hashem (2006). "Estimation and inference in large heterogeneous panels with a multifactor error structure." *Econometrica*, Vol. 74(4): 967–1012.

Pesaran, M. Hashem (2007). "A simple panel unit root test in the presence of cross-section dependence." *Journal of Applied Econometrics*, Vol. 22(2): 265–312.

Pesaran, M. Hashem and Smith, Ron P. (1995). "Estimating long-run relationships from dynamic heterogeneous panels." *Journal of Econometrics*, Vol. 68(1): 79–113.

Pesaran, M. Hashem, Smith, Vanessa and Yamagata, Takashi (2009). "Panel Unit Root Tests in the Presence of a Multifactor Error Structure." Cambridge University, unpublished working paper, September.

Pesaran, M. Hashem and Tosetti, Elisa (2011). "Large Panels with Common Factors and Spatial Correlations." *Journal of Econometrics*, Vol. 161(2): 182–202.

Phillips, Peter C. B. and Sul, Donggyu (2003). "Dynamic panel estimation and homogeneity testing under cross section dependence." *Econometrics Journal*, Vol. 6(1): 217–259.

Rajan, Raghuram G. and Subramanian, Arvind (2008). "Aid and Growth: What Does the Cross-Country Evidence Really Show?" *The Review of Economics and Statistics*, Vol. 90(4): 643–665.

Restuccia, Diego, Yang, Dennis Tao and Zhu, Xiaodong (2008). "Agriculture and Aggregate Productivity: A Quantitative Cross-Country Analysis." *Journal of Monetary Economics*, Vol. 55(2): 234–250.

Ruttan, Vernon W. (2002). "Productivity Growth in World Agriculture: Sources and Constraints." *Journal of Economic Perspectives*, Vol. 16(4): 161–184.

Schmidt, Peter (2009). "Estimates of Technical Inefficiency in Stochastic Frontier Models with Panel Data: Generalized Panel Jackknife Estimation." Keynote speech, Invited Session in Honour of Cheng Hsiao, 15th Conference on Panel Data, Bonn, July 2009.

Temple, Jonathan (1999). "The New Growth Evidence." *Journal of Economic Literature*, Vol. 37(1): 112–156.

Vollrath, Dietrich (2009a). "How important are dual economy effects for aggregate productivity?" *Journal of Development Economics*, Vol. 88(2): 325–334.

Vollrath, Dietrich (2009b). "The Agricultural Basis of Comparative Development." Paper presented at NEUDC, Boston, November 2009.

Westerlund, Joakim and Urbain, Jean-Pierre (2011). "Cross sectional averages or principal components?" Maastricht: METEOR, Maastricht Research School of Economics of Technology and Organization, Working Paper #53.

Appendix: Data, descriptives and synthetic literature review

The principal data source for our empirical analysis is the Food and Agriculture Organisation's *FAOSTAT* database (FAO, 2007), from which we obtain annual observations for agricultural net output, economically active labour force in agriculture, number of tractors used in agriculture, arable and permanent crop land and fertilizer use in 128 countries from 1961 to 2002. The total number of observations is 5,162 with an average T of 40.3. Real agricultural net output (in thousand International $) is based on all crops and livestock products originating in each country. Intermediate primary inputs of agricultural origin are deducted, including fodder and seed. The quantities for each commodity are weighted by the respective 1999-2001 average international commodity prices and then summed for each year by country. The prices are in international dollars, derived using a Geary-Khamis formula for the agricultural sector.[a] The labour variable represents the annual time series for total economically active population in agriculture. For capital stock in agriculture we follow a common convention and use total number of agricultural tractors in use as a proxy. The livestock variable is constructed from the data for asses (donkeys), buffalos, camels, cattle, chickens, ducks, horses, mules, pigs, sheep & goats and turkeys. Following convention we use a conversion detailed in Hayami and Ruttan (1970) to convert the numbers for individual animal species into the livestock variable. The fertilizer variable represents agricultural fertilizer consumed in metric tons, which includes 'crude' and 'manufactured' fertilizers. The land variable represents arable and permanent crop land (in 1000 hectare). Descriptive statistics are presented in Table A-1. The countries in our sample are listed in a Technical Appendix.

Analysing agricultural production for a large number of countries inevitably raises concerns over data reliability. In contrast to macro data provided in for instance the Penn World Table (CIC, University of Pennsylvania) FAOSTAT does not offer a data quality grade for each country, but instead labels each observation. Most output data used in our analysis carries the note '[a]ggregates may include official, semi-official or estimates'. For inputs we obtain more details, which suggest that tractor data is least reliable, with around 45% of observations estimated.[b] Thus data is far from perfect for cross-country comparison, although estimating production functions country by country and accounting for unobserved common factors should go some way to ward against systematic over-/underreporting of variable magnitudes. We find that there is no statistically significant relationship between the share of tractor data estimated and the coefficients on tractor, livestock, fertilizer or land in our preferred agro-climatic CCEMG model, which could be viewed as evidence to that end.

Additional time-invariant data on geographical distance between countries and contiguity (neighbourhood) is taken from CEPII (2006), and data on the share of agricultural land by climatic zone from Matthews (1983), available in Gallup et al. (1999).

[a] Refer to the Technical Appendix to Restuccia et al. (2008), available on Restuccia's website.

[b] We report Penn World Table country quality grade and share of non-estimated tractor data in a Technical Appendix.

Table A-1: Descriptive statistics

Variables in untransformed level terms					
Variable	**mean**	**median**	**std. dev.**	**min.**	**max.**
logs					
output	14.24	14.24	1.71	8.07	19.57
labour	14.01	14.09	1.84	8.01	20.05
tractors	9.01	8.87	2.79	0.69	15.51
livestock	14.90	14.92	1.71	8.80	19.51
fertilizer	10.82	10.97	2.69	1.61	17.49
land	14.69	14.78	1.80	6.91	19.07
annual growth rate					
output	2.3%	2.4%	8.8%	-83.0%	87.6%
labour	0.3%	0.8%	2.6%	-28.8%	28.8%
tractors	4.4%	2.0%	9.9%	-121.8%	138.6%
livestock	1.4%	1.6%	6.4%	-93.3%	182.9%
fertilizer	5.6%	3.5%	40.1%	-626.3%	393.2%
land	0.8%	0.1%	3.6%	-41.8%	79.0%

Variables in per worker terms					
Variable	**mean**	**median**	**std. dev.**	**min.**	**max.**
logs					
output	0.23	-0.03	1.42	-2.22	4.00
tractors	-5.00	-4.97	3.01	-13.67	0.68
livestock	0.89	0.81	1.38	-2.77	4.63
fertilizer	-3.19	-2.87	2.67	-11.56	1.95
land	0.68	0.67	1.15	-2.20	4.95
annual growth rate					
output	2.0%	2.0%	9.0%	-80.3%	109.9%
tractors	4.1%	2.1%	10.1%	-120.2%	136.5%
livestock	1.2%	1.2%	6.6%	-93.5%	182.9%
fertilizer	5.4%	4.2%	40.0%	-627.8%	390.8%
land	0.5%	0.0%	4.1%	-43.0%	81.6%

Notes: We report the descriptive statistics for output (in I$1,000), labour (headcount), tractors (number), livestock (cattle-equivalent numbers), fertilizer (in metric tonnes) and land (in hectare) for the full regression sample ($n = 5,162$; $N = 128$).

Table A-2: Selective Review of the Literature

	Hayami & Ruttan (1970)	Hayami & Ruttan (1985)	Craig et al (1997)	Mundlak et al (1999)	Martin & Mitra (2002)	Cermeño et al (2003)	Gutierrez & Gutierrez (2003)	Bravo-Ortega & Lederman (2004)	Mundlak et al (2008)	Vollrath (2009)
Data										
Countries	36	43	98	37	49	84	47	86	30	100
Observations	108	129	588	777	1,248	2,604	1,081	2,993	870	3,532
Year(s)/Period	1955, 1960, 1965	1960, 1970, 1980	1961-1990	1970-1990	1967-1992	1961-1991	1970-1992	1961-1997	1972-2000	1961-1999
Data source	own, annual	own, annual	FAO, 5-year avg.	WB, annual	WB, annual	FAO, annual	WB, annual	FAO, annual	WB, annual	FAO, annual
Specification										
Model	pooled	pooled	pooled	pooled	pooled	pooled by region	pooled	pooled	pooled	group
Estimation method	POLS	POLS	LSDV	2FE	LSDV	LSDV	panel DOLS	LSDV	2FE	LP
Covariates:										
'standard'	L, N, Live, F, K	L, N, Live, F, K †	L, N, Live, F, K	L, N, K, F	L, N, K	L, N, K, Live, F	L, N, K	L, N, P, K, F, Live	L, N, Live, K, F	L, N, Live, K, F
'non-standard'	School, Tech Edu	School, Tech Edu	Lit, Life, Infra (quality-adjusted)	School, Dev, Yield, Prices				irrigated land		aggr. GDPpc
Year dummies	✓	✓		implicit	*N* trends	✓	✓	✓	implicit	✓
Country dummies			✓	implicit	✓	✓	✓	✓	implicit	✓
Returns to scale	CRS imposed	unrestricted	unrestricted	unrestricted	CRS imposed	unrestricted	unrestricted	CRS imposed	CRS imposed	unrestricted
Findings										
labour	.40	.50	.25	.08	.64		.29	.11	.12	.07 – .98
land	.07	.03	.40	.42	.24	.02 – .47	.14	.04	.33	
pastures								.53		
irrigated land								.03		
capital/machinery	.11	.07	.05	.36	.12	.02 – .08	.58	.04	.34	
livestock	.29	.31	.35			.03 – .40		.25	.08	
traction animals			-.06							
fertilizer	.14	.15	.04	.08		.00 – .10		.05	.13	
TFP			no difference across *i* or over *t*	accounted via 2FE transformation	sign. diff. across *i*, 1.4-3.5% growth pa	sign. diff. across *i* and over *t*	accounted via DOLS method	sign. common trend term (1%pa)	accounted via 2FE transformation	accounted via Lev-Pet
Returns to scale	(CRS imposed)	not tested	DRS	CRS not rejected	not tested	differs across regions	CRS	(CRS imposed)	(CRS imposed)	not tested
Reference	Table 2, (17)	Table 6-2, (Q14)	Table 1, (1)	Table 4	Table 1, footnote	Tables 1-4, (7)	Table 2, DOLS	Table 6(A), (2)	Table 2, Within	Table 2, various

Notes: The table compares Cobb-Douglas production function estimates from studies as indicated. Since many studies report results for more than one specification we concentrate on the most general model for 'standard' production inputs with panel data. The specific reference for each study's entry is stated at the bottom of the table. The studies by Hayami and Ruttan (1985) and Gutierrez and Gutierrez (2003) also provide separate estimations by country groups (developed, developing), finding production technology differences to be considerable and negligible. †For this study the reported coefficients are derived from data deflated by the number of farms.
Datasets: Hayami and Ruttan (1970) and Hayami and Ruttan (1985) use data from a wider range of sources which includes the FAO, OECD and other international institutions. FAO – various databases provided by the Food and Agriculture Organisation, WB – the Larson et al. (2000) dataset and its updated version.
Output: agricultural output (expressed in some type of constant 'international' monetary terms) in all studies covered here.
Inputs: L – agricultural labour, N – (crop) land, Live – livestock, F – fertilizer, K – capital stock/tractors, P – pastures, School – school enrolment, Tech Edu – technical education, Lit – literacy, Life – life expectancy, Dev – development level, Yield – peak yield, Infra – infrastructure variables, Prices – various.
Estimators: POLS – pooled OLS, LSDV – Least Squares Dummy Variable, 2FE – Two-way Fixed Effects, DOLS – Dynamic OLS, following Kao et al. (1999). This estimator imposes a homogeneous cointegration vector but allows for country-specific short-run dynamics. LP – Levinsohn and Petrin (2003), first step for output elasticity wrt labour only (polynomial of order three in inputs included).