Identifying Agricultural Factor Productivity from Micro-data: A Review of Approaches with an Application to EU Countries

Martin Petrick
Leibniz Institute of Agricultural Development in Transition Economies (IAMO), Halle (Saale)

Mathias Kloss
University Hospital Halle, Halle (Saale)

Abstract

We examine the plausibility of four established and innovative identification strategies for agricultural production functions using farm-level panel datasets from five EU countries. Newly suggested proxy and dynamic panel approaches provide attractive conceptual improvements over received Within and duality models. Even so, empirical implementation of such advancements does not always live up to expectations. This is particularly true for the dynamic panel estimator, which mostly failed to identify reasonable elasticities for the (quasi-) fixed factors. Less demanding proxy approaches represent an interesting alternative for agricultural applications. In our EU sample, high production elasticities for materials prevail. Hence, improving the availability of working capital is the most promising way to increase agricultural productivity.

Key Words

agricultural factor productivity; production function estimation; EU; Farm Accountancy Data Network

1 Introduction

The aim of this article is to review the central identifying assumptions maintained by six traditional and recent approaches to the estimation of production functions, apply them to farm-level data and ask how plausible they are in an agricultural context. Our methodological contribution is that we provide the first comparative evaluation of a number of recently proposed production function estimators for agricultural data. Our empirical contribution is a unique and current set of estimated production elasticities for five firm-level datasets at the EU country level.

Our European database covers firm-level data from five EU member states that was collected following a harmonized procedure in all countries. This is one of the first micro studies of agricultural productivity that simultaneously uses firm-level data from several countries for comparative purposes.[1] This extensive data allows us to come up with new, country-specific estimates of production elasticities in agriculture that are potentially robust to endogeneity and collinearity issues. While agriculture is a classical field of productivity estimation, there has been surprisingly little systematic analysis using the production function approach recently. MUNDLAK (2001) attributes this to the emergence and widespread acceptance of duality theory in agricultural economics from the 1970s onwards. This approach typically recovers the price elasticity of factor demand but not the production elasticities. As MUNDLAK (2001) notes and as we discuss below, the dual approach is based on restrictive theoretical assumptions and far from being without methodological problems. One key expectation from duality was that it would allow a more flexible representation of technology, such as based on the Translog functional form (SHUMWAY, 1995). Interestingly, our results show that making the Cobb Douglas production function more flexible by adding quadratic and interaction terms does not add much insight. In the Ordinary Least Squares (OLS) and WOOLDRIDGE (2009) case, the results were highly implausible, whereas they differed little from the Cobb Douglas for the Within panel estimator.

Our empirical estimates suggest that output elasticities of labor, land and fixed capital are low throughout our European subsamples. This finding is in contrast to recent estimates by MUNDLAK et al. (2012), according to whom there are significant returns to land and fixed capital in a cross-country sample of developing and developed countries. On the other hand, our materials elasticity is quite high, around 0.7. This outcome is particularly prominent in the

[1] RIZOV et al. (2013) provide an EU-wide comparative analysis based on the OLLEY and PAKES (1996) estimator. The distinct literature studying frontier models has just recently begun to address endogeneity concerns, including an application to dairy farms in the EU by LATRUFFE et al. (2017).

LEVINSOHN and PETRIN (2003) (LP), WOOLDRIDGE (2009) (WLP) and BLUNDELL and BOND (2000) (BB) estimators. In the conceptual part, we argue that these estimators provide more plausible identification strategies than established Within or duality approaches. While the one-period control function models of LP and WLP are easier to implement empirically, the multiperiod adjustment process implied by the BB model is more compelling in an agricultural context. But BB failed to produce reasonable results for the fixed variables in most of our country subsamples. There is hence a trade-off among theoretical plausibility and empirical robustness of the different identification strategies.

In the following Section 2, we discuss the key identification problems that have motivated much of the methodological debate in production function estimation as well as the four main assumptions invoked in the literature to address them. Section 3 describes the dataset. Section 4 presents the empirical results. Section 5 concludes.

2 Identification Problems in Production Function Estimation and Approaches to their Solution

2.1 A Typology of Production Factors

The process of agricultural production serves as a useful illustration for the different nature of production factors. For the ensuing discussion, two characteristics of these factors are of particular importance:

- their variability or the ease with which they can be adjusted, and
- whether they are observed by the econometrician.

Table 1 differentiates three categories of variability. Among the highly variable factors are intermediate inputs such as seed, fertilizer or concentrate fodder. These factors are typically included in farm-level datasets and thus observed by the econometrician (type I factors). In economic parlance, they are also called "control variables" because the decision maker (the farmer) can manipulate their level to achieve his/her objectives. Other highly variable control variables may be hard to observe from the outside, such as work effort (type IV factors).

Other important factors are much less variable and are subject to adjustment costs (type II and V factors, depending on whether they are observed). For example, land is often available in limited quantities only and subject to long-term rental agreements. Agriculture in Europe is typically organized in family farms on which labor is often highly immobile and may be influenced significantly by life cycle considerations of the farm family (GLAUBEN et al., 2009). Agricultural credit markets suffer from informational asymmetries and may be characterized by rationing and high transaction costs (see e.g. PETRICK and LATRUFFE, 2006). Management has long been recognized as an important factor of production that is nevertheless difficult to measure (MUNDLAK, 1961).

A final group includes factors that are completely fixed in the long run, such as the geographic location of the farm or the quality of its soils (type III and VI factors). All the less variable factors type II, III, V and VI are called "state variables", as their value cannot be modified within a short-term planning horizon.

As indicated in Table 1, there is an important distinction between the highly variable and unobserved factors type IV and VII. Some of these also come as a surprise to the farmer. They represent exogenous states (shocks) of the environment (type VII factors). However, how the farmer reacts to these shocks will be endogenous (type IV factors).

Table 1. A typology of production factors in agriculture

	Highly variable	Subject to adjustment costs	Fixed
Observed by econometrician & farmer	**Type I** seed, fertilizer, chemicals, concentrate, livestock numbers	**Type II** land, labor, machinery, buildings	**Type III** geographical location
Typically unobserved by econometrician but known to the farmer	**Type IV** farmer's effort, reaction to environmental shocks	**Type V** management abilities, human capital of labor force, availability of a farm successor	**Type VI** soil quality, climatic conditions
Unobserved by econometrician & unanticipated by the farmer	**Type VII** weather events, rainfall, diseases, legal requirements	--	--

Source: authors

2.2 Two Problems of Identification

To illustrate the involved problems, we start with a simple model of a farmer wishing to produce an aggregate output. Denote y_{it} the natural logarithm of farm i's output Y at time t, A_{it} land use of this farm, L_{it} labor, K_{it} fixed capital and M_{it} materials or working capital. These four factors of production are observed by the econometrician. ω_{it} is an aggregate, farm-specific, time-varying factor that is anticipated by the farmer at the time of decision making about current production, but unobserved by the econometrician. Without further specification, it compounds the effects of factors categorized as type IV to VI in Table 1. ε_{it} is a productivity shock not anticipated by the farmer (and not observed, thus type VII), or simply measurement error. Assuming a linear structure of the model and the availability of panel data containing the observed output and inputs, the econometrician's problem is to recover farm productivity determined by the following equation:

$$y_{it} = f(A_{it}, L_{it}, K_{it}, M_{it}) + \omega_{it} + \varepsilon_{it}, \qquad (1)$$

where $f(\cdot)$ is the production function.

Because ω_{it} will likely be correlated with the other input choices, estimation of (1) is subject to an *endogeneity problem* (MARSCHAK and ANDREWS, 1944). The production elasticities of the observed factors are not identified as the compound error term $\omega_{it} + \varepsilon_{it}$ is not identically and independently distributed (i.i.d.). Regressing output on observed input levels using OLS and choosing an appropriate functional form for $f(\cdot)$ will produce biased estimates. In particular, input coefficients will be upward biased if there is serial correlation in ω_{it}. This effect will be stronger the easier it is to adjust input use (LEVINSOHN and PETRIN, 2003: 332). A typical OLS result may be that the coefficients of labor and materials are upward biased, while those of land and capital are downward biased. Much of the methodological literature on production function estimation is concerned with precisely this issue (see the instructive review in GRILICHES and MAIRESSE, 1998).

According to the implicit theoretical setup so far, all observed factors are assumed to be control variables and are treated as being fully flexible (as if they all belonged to type I). The typical assumption in the literature (e.g. CHAMBERS, 1988) is then that output and all factors are traded on perfectly competitive markets so that on each of the markets all farmers face the same one price for the traded good. If farmers maximize profits defined as revenues from the sale of output minus costs of all inputs and $f(\cdot)$ is a monotonous and concave function, the canonical decision rule for allocating inputs is identical for all inputs and says that the marginal revenue product of each factor should equal its factor price. For example, for materials this decision rule is as follows:

$$p^Y \frac{\partial f}{\partial M} = p^M, \qquad (2)$$

with p^Y denoting the price of output and p^M that of materials, respectively. Estimation of (1) requires the assumption that the technology represented by $f(\cdot)$ is identical for all farmers included in the estimating sample. If all farmers also face the same price on each of the output and input markets, there is nothing in the model that induces heterogeneous factor use across farms except for the unobserved ω_{it}. This is the *collinearity problem* pointed out recently by BOND and SÖDERBOM (2005) and ACKERBERG et al. (2007). Factor use across firms varies only with the unobserved ω_{it}, so that again the different production elasticities are not identified.

We now review the main approaches found in the literature to deal with either of these identification problems. The discussion is guided by Table 2, which summarizes the four approaches we distinguish. After introducing each approach, we ask how plausible the specific identifying assumption is in the context of agriculture. We then evaluate to what extent the two key identification problems presented before are addressed and how the resulting estimator can be applied in practice.

2.3 Additively Separable, Time-invariant Firm Characteristics

The key idea of this approach is that ω_{it} can be further decomposed into:

$$\omega_{it} = \gamma_t + \eta_i + v_{it}, \qquad (3)$$

where γ_t is a time-specific shock that is identical for all farms in t (likely a type VII event), η_i is a farm-specific fixed effect that does not vary over time (a type VI factor), and v_{it} is the remaining farm- and time-specific productivity shock (type VII). Think of γ_t representing common weather or policy shocks and η_i capturing soil quality or time-invariant preferences of the manager. In a farming context, v_{it} may represent local weather conditions that vary between farms and years. If they are not anticipated by the manager, v_{it} is subsumed into ε_{it}. This approach leads to the popular fixed effects approach. By applying a *"within" transformation,* the fixed effect (η_i) is "swept out" of the estimating equation. Suggested by MUNDLAK (1961) in a farming context to eliminate "management bias" from the equation, this model has found widespread application at different levels of

Table 2. Identifying assumptions in production function estimation

	(A) ω_{it} is additively separable & time invariant	(B) Profit maximization & perfect competition on product & factor markets	(C) Heterogeneous frictions in factor adjustments	(D) ω_{it} evolves monotonously with an observed characteristic of the firm
If correct, does the assumption solve the endogeneity problem?	Yes.	Yes if prices can be used as instruments.	Yes if adjustment costs are sufficiently heterogeneous across inputs.	Yes.
Does it solve the collinearity problem?	Not without further assumptions.	Yes if there is only one free input.	Yes if adjustment costs are sufficiently heterogeneous across inputs.	Not without further assumptions: ACKERBERG et al. (2015), WOOLDRIDGE (2009).
Practical implementation	"Within" regression to sweep out fixed effect.	Share regression, approaches based on duality.	Typically combined with assumption (A) in a dynamic panel data regression model using first differences.	Semiparametric control function approaches using investment or intermediate inputs as proxies.
Remaining problems	Remaining variance may be too small to allow precise parameter estimation.	Prices with sufficient variation may not be observed. Heterogeneous firm-specific prices may not be exogenous.	Weak instruments, small variance of differenced variables.	Zero observations for proxies (e.g., investment). Slowly changing unobserved effects are not captured.
Plausibility in agriculture	Limited plausibility as farm- & time-specific effects are likely, e.g. reactions to weather shocks.	Limited plausibility as market imperfections on labor, land & capital markets are widespread in agriculture.	Plausible for land, labor, fixed capital, less for seed, fertilizer, plant protection, concentrate, energy.	Plausible for annually fluctuating shocks, less for slowly changing unobservables such as soil or management quality.
Examples in the literature	Widely used. See MUNDLAK (1961); overview in GRILICHES and MAIRESSE (1998).	Widely used. See overview in MUNDLAK (2001).	BLUNDELL and BOND (2000), HEMPELL (2005). No agricultural applications so far.	OLLEY and PAKES (1996), LEVINSOHN and PETRIN (2003), KAZUKAUSKAS et al. (2010) on Irish dairy farms, PETRIN and LEVINSOHN (2012), RIZOV et al. (2013) on EU-15 agr.

Source: authors

aggregation. The effect of γ_t is typically taken into account by including time dummies into the model. An alternative to Within is to estimate the model in first differences, as discussed by WOOLDRIDGE (2010: 321-326).

MUNDLAK et al. (2012: 146) present a recent application to agricultural productivity at the country level where the fixed and year effects alone explained 98.5% of output variation. Even so, the question remains whether it is legitimate to assume that v_{it} is an innovation that is orthogonal to observed factor use so that all unobserved factors are indeed either time invariant or the same for all farms.

Table 1 suggests that farm- and time-specific unobserved effects *which the farmer still takes into account when making input decisions* (type IV and V) are very likely to be relevant. Examples include annual fluctuations in rainfall or pest occurrence as well as patterns of livestock health. Furthermore, applications in practice have found that the within transformation removes (too) much variance from some of the variables, particular those which display little variation over time. In agriculture, input levels of the type II production factors land, labor and fixed capital often vary only little in time. As a consequence, the signal-to-noise ratio with regard to these factors is reduced and the estimated coefficients are biased downwards (GRILICHES and MAIRESSE, 1998: 180-185). Finally, without further assumptions, the collinearity problem is not addressed at all by this approach.

2.4 Profit Maximization and Perfect Competition

This approach imposes further microeconomic theory upon the data, including its main assumptions of profit maximization and perfect competition on product and input markets. A key result of this theory is the first-order condition (2), which multiplied through with $\frac{M}{p^Y Y}$ yields (for the case of materials):

$$\frac{\partial f}{\partial M}\frac{M}{Y} = \frac{p^M M}{p^Y Y}. \tag{4}$$

If one further assumes constant returns to scale, (4) says that the production elasticity of each input (left hand side) is equal to its value share in revenue (right hand side). All value shares add up to one. Given these assumptions, revenue shares of inputs are valid estimators of production elasticities. For the simple Cobb Douglas technology, the problem of estimating production elasticities has thus been "solved" by the imposition of strong theoretical assumptions. However, production function estimates of elasticities in agriculture were often found to differ from observed revenue shares (MUNDLAK, 2001).

For more flexible functional forms, (4) has led to the widely applied share regression model. For example, if the production function is assumed to be Translog the first order condition yields the following *share regression* (again for the case of materials):

$$s_{it}^M = \alpha^M + \alpha^{MM} m_{it} + \alpha^{MA} a_{it} + \alpha^{ML} l_{it} + \alpha^{MK} k_{it} + \omega_{it}^M + \varepsilon_{it}^M, \tag{5}$$

with $s_{it}^M = \frac{p_{it}^M M_{it}}{p_{it}^Y Y_{it}}$ the revenue share of materials of firm *i* at time *t*, α^X the direct and cross-elasticities of the involved inputs, ω_{it}^M the part of the unobserved productivity characteristic that affects s_{it}^M, and ε_{it}^M an i.i.d. error term.

Note that (5) is still subject to the endogeneity and collinearity of factors. The way out of these problems typical to this approach is finding appropriate instruments for the input levels, such as factor prices. However, factor prices may not be exogenous and may depend on past and current decisions of the farmer. Under such conditions, the theoretical model underlying this approach is clearly too simplistic to allow straightforward identification of the production function. On the other hand, if factor markets were at least approximately working as postulated by the theoretical ideal, there should be little price variation across farms so that the value of prices for solving the endogeneity and collinearity problems is doubtful.

2.5 Heterogeneous Frictions in Factor Adjustment

If prices are problematic instruments, another option is to look for a different source of exogenous variation that has explanatory power for productivity analysis. One such source now routinely employed, which is based on the literature on dynamic panel data modeling, are past decisions on factor use (ARELLANO and BOND, 1991; BLUNDELL and BOND, 1998). This literature suggests that current variation in input use is caused by lagged adjustment to past productivity shocks. It thus introduces the history of input use as a source of identification. Such identification is plausible if modifications of input levels are subject to adjustment costs (BOND and SÖDERBOM, 2005). This approach effectively turns observed input levels into state variables (type II) and makes them subject to an intertemporal optimization problem. One way to account for costly adjustment is to allow serial correlation of the unobserved productivity characteristic (v_{it}) of the firm.

BLUNDELL and BOND (2000) use lagged levels and differences of inputs as instruments in a General Methods of Moments (GMM) framework to estimate this model. Note that the within transformation (Section 2.3) assumes *strict* exogeneity of inputs which means that ω_{it} must not be transmitted to any future period. First differencing (FD) to eliminate fixed effects only assumes that input levels are *sequentially* exogeneous, i.e. transmission of ω_{it} to the next but one and subsequent periods is allowed (CHAMBERLAIN, 1982; WOOLDRIDGE, 2010: 321-326). FD is thus the typical approach to eliminate time invariant heterogeneity in GMM applications, as it allows input levels lagged more than two periods to be used as instruments for contemporaneous differences (ARELLANO and BOND, 1991). Of course, these instruments will only have power if there actually *is* such a transmission (e.g. motivated by adjustment costs). To increase the power of the GMM approach, BLUNDELL and BOND (1998) have shown that in addition to past levels, also lagged differences of inputs can be used as instruments if they are orthogonal to the fixed effects (η_i) – an assumption which will hold if their variance is assumed to be, in the broadest sense, stationary (ROODMAN, 2009: 114-115). This leads to the systems GMM estimator for production functions presented in BLUNDELL and BOND (2000).

If factor levels can suitably be instrumented by this approach, it addresses both the endogeneity and the collinearity problems. Contrary to the duality approach presented in Section 2.4, it is much more plausible that the instruments proposed here are actually valid in an agricultural context. There are important production factors in agriculture which are subject to adjustment costs (or "transaction costs"; type II variables in Table 1) and such costs should be an element in any plausible theory of agricultural factor markets.

As the nature of these costs is likely to differ among factors (see Section 2.1), it is also plausible that different factors of production display different dynamic paths of adjustment. This is a favorable condition for identification (BOND and SÖDERBOM, 2005). It is only with regard to some intermediate inputs such as seed, fertilizer, plant protection, concentrate, or energy that factor use appears to be more flexible so that the assumption of adjustment costs may be harder to justify (type I factors). In sum, this estimator is a promising candidate for agricultural applications.

2.6 Monotonous Coevolution of Unobserved Productivity Shocks with Observed Firm Characteristics

The final method to be discussed here avoids the main disadvantage of any fixed effects approach to unobserved heterogeneity, which is the typically low variance of the transformed variables. However, it also does not rely on the strong a-priories about market structure of duality theory to identify the productivity parameters of interest. It rather attempts to proxy ω_{it} (as a compound type IV to VI production factor) by a *non-parametric control function* which itself contains only observed firm characteristics. OLLEY and PAKES (1996) were the first to suggest log investment (i_{it}) as an observed characteristic driven by ω_{it}:

$$i_{it} = i_t(\omega_{it}, k_{it}), \tag{6}$$

where k_{it} is the pre-determined level of capital use at time *t*. The latter is assumed to evolve according to $k_{it+1} = (1-\delta)k_{it} + i_{it}$, with δ the depreciation rate.

The function $i_t(\cdot)$ can vary over time and is not parametrically restricted except that it needs to be monotonous in ω_{it}. This latter trait allows inversion of this function, so that:

$$\omega_{it} = h_t(i_{it}, k_{it}),$$

where h_t is now potentially observable and acts as a proxy for ω_{it}. Furthermore, it is assumed that unobserved productivity follows a first-order Markov process:

$$\omega_{it} = E[\omega_{it}|\omega_{it-1}] + \xi_{it}, \tag{7}$$

where ξ_{it} is an innovation (a type VII factor) uncorrelated with k_{it}, but possibly correlated with the other factors in the production function. Because k_{it} is a type II factor, the moment condition $E[\,k_{it}\xi_{it}] = 0$ can be used to identify α^K.

Given this setup, estimation proceeds in two stages. The basic idea is to jointly control for the influence of *k* and ω in the first stage and to recover the true coefficient of *k* as well as ω in the second. In an agricultural application, KAZUKAUSKAS et al. (2010) found for Irish dairy farms that the materials coefficient estimated with an OP procedure was lower than the OLS result. One problem that arises from using investment as a proxy is zero observations for certain years and firms. LEVINSOHN and PETRIN (2003) therefore suggested materials instead of investment as a proxy of ω_{it} in the previous algorithm.

If the control function fully captures the influence of ω_{it}, it solves the endogeneity problem and provides a useful alternative to the fixed effects approaches described before. However, in agriculture, the assumptions on monotonicity and dynamic evolution of the productivity shock must be considered with caution. A key question is *what exactly ω_{it} is representing and whether investment or material use are good proxies for it*. If ω_{it} stands for annually fluctuating, unobserved factors (type IV) such as management effort or reaction to environmental conditions, there may be cases where the “right behaviour” of the farmer (i.e., positive ω_{it}) does not lead to more investment. The same is true for materials. The productivity enhancing reaction to environmental shocks in crop production may sometimes be less input use (fertilizer, chemicals) rather than more. In all these cases, neither investment nor materials will be good proxies of ω_{it}. Furthermore, the “memoryless” first-order Markov process appears unconvincing if ω_{it} actually represents unobserved type V factors which are subject to adjustment costs. They evolve slowly and will typically have implications for the intertemporal optimization problem, so that also k_{it} is affected by them and (6) is misspecified. Investment may not be a good proxy for ω_{it} if there are other important determinants of it beyond k_{it}. In a farming context, this is likely to be the case, because investment decisions are usually influenced by long term business strategies and/or the availability of a farm successor.

Another problem with the procedure suggested by OP and LP is that it does not solve the collinearity problem. As discussed at length by ACKERBERG et al. (2015), unless one is willing to make very unintuitive assumptions on measurement error or timing, there is no data generation process that separately identifies the coefficients of the type I factors in either of the two approaches. WOOLDRIDGE (2009) suggests a simple procedure that borrows the identification strategy from OP and LP and modifies as well as extends the moment conditions to resolve the collinearity problem. Hence, this approach is referred to as the WOOLDRIDGE/LEVINSOHN/PETRIN (WLP) estimator (PETRIN and LEVINSOHN, 2012). This model is usually estimated within an IV estimation framework (PETRIN and LEVINSOHN, 2012).

In our agricultural application, the intuition of this approach may be as follows (cf. LEVINSOHN and PETRIN, 2003: 322). Consider ω_{it} to represent a farm-specific stock of management knowledge. Any positive shift of ω_{it} assumedly increases the marginal productivity of m_{it} and possibly all other production factors. As m can be readily adjusted, a profit-maximising farmer increases the level of m_{it} in response to the shift, thus motivating our use of m as a proxy for ω_{it}. The same process may also work in the other direction, so that farms with negative shocks reduce material inputs. If ω is persistent, the farm-specific over- or under-application of material inputs is likely to be correlated over time, so that past levels can be used as proxies for current productivity shifts. Consistent with primarily positive shifts is the empirical observation that, on average, both farm output and materials input increase over the years. This is precisely what our data confirms.

The assumption of costly factor adjustment is a cornerstone of both the dynamic panel data approach described in Section 2.5 and the present one. In both cases, this assumption provides moment conditions necessary for consistent estimation of the parameters. The main difference is that the former approach allows time-invariant fixed effects, whereas the latter does not. The former imposes a linear structure on the dynamic process, while it can be arbitrary in the latter. Even so, factor adjustment is assumed to occur in a single period in OP and followers, whereas the process covers many periods in the dynamic panel data models. In the light of agricultural applications, this may be one key advantage of the dynamic panel data approach.[2]

2.7 Interim Evaluation of Estimation Approaches

The previous discussion has displayed the variety of assumptions invoked for addressing the endogeneity and collinearity problems inherent to production function estimation. In our opinion, the assumptions underlying Within regression and the duality approach are fairly strong and implausible for the case of agriculture. Perhaps not surprisingly, they often have also not performed well in estimation practice. This insight shifts our attention to the promising new approaches using heterogeneous frictions in factor adjustment. We regard the presence of adjustment costs as particular relevant for the production factors that are of key interest in agricultural applications. They also provide an interesting link to more sophisticated theories of business structures in agriculture, which usually embody some form of adjustment frictions in agricultural factor use (such as ALLEN and LUECK, 2002, or POLLAK, 1985). So far, there are almost no applications to agricultural data of these new estimators. The following sections aim to fill this void.

3 Data

The data used in this study comes from the EU's Farm Accountancy Data Network (FADN), which provides a stratified farm level data set that holds accountancy data for all 28 EU member states. In the present study, we only use field crop farms, to justify the assumption of a homogenous state of technology across farms. The sample of countries is selected to reflect the diverse farm sizes and structures in EU agriculture. The range is from small-scale family farms in Italy and West Germany up to medium-sized commercial farms in Denmark, France and the UK (EUROPEAN COMMISSION, 2012). West Germany contains the nine federal states Baden-Württemberg, Bavaria, Hamburg, Hesse, Lower Saxony, North Rhine-Westphalia, Rhineland-Palatinate, Saarland and Schleswig-Holstein. It does not include Berlin and Bremen, which are not represented in the FADN data. Therefore, we produce separate results for the following countries:

- Denmark (DK),
- France (FR),
- Germany West (DEW),
- Italy (IT) and
- the United Kingdom (UK).

For every country in the study, we created a panel data set covering the years from 2001 up to 2008. In total, 14,801 observations were included in the EU-wide sample.

The variables and their measurement are readily available in the codebooks provided by FADN (EUROPEAN COMMISSION, 2007, 2008). Output is measured as the total farm output in euros. Labor is measured by the time worked in hours by total labor input on the farm, including both hired and family labor. The total utilized agricultural area is our land input in

[2] Other subtle differences between the two approaches are discussed in ACKERBERG et al. (2015).

ha. It includes owned and rented land, and land in sharecropping.

In this study, the material or working capital input is proxied by total intermediate consumption in euros. It consists of total specific costs and overheads arising from production in the accounting year. Among others, it includes feed, fuel, lubricants, water, electricity and seed. We do not include fertilizer in our materials specification. As land and fertilizer are highly correlated in the data sample, they are applied in more or less fixed ratios on the average farm, which might induce a multicollinearity problem in the estimations.[3] We approximate fixed capital inputs by using the opening valuation of machinery and buildings from the FADN data. Table 3 summarizes the variable definitions and gives the actual FADN codes.

Table 3. Selection of variables

FADN code	Variable description
Outputs	
SE131	Total output (EUR)
Inputs	
SE011	Labor input (hours)
SE025	Total utilized agricultural area (ha) = land
F72 + SE300 + SE305 + SE336	Costs for seed and seedlings + crop protection + other crop specific costs + overheads (EUR) = materials
L.SE450 + L.SE455	Opening valuation of machinery and buildings (EUR) = fixed capital

Note: L. denotes the one-year lag.
Source: authors, FADN data

4 Results

4.1 Overview

For this study, we estimated nine models per country: Output shares, OLS Cobb Douglas, OLS Translog, Within Cobb-Douglas, Within Translog, LP Cobb Douglas, WLP Cobb Douglas, WLP Translog, BB Cobb Douglas. The Within Translog was obtained by interacting the groupwise demeaned logs of factors and using an appropriate degree of freedom correction. Other than by simply calling a built-in fixed effects panel estimation command with the interacted variables in logs, this procedure ensures that levels are effectively eliminated from the regression. Below we present a summary of results. The full results as well as an in-depth analysis of the same are given in our companion paper PETRICK and KLOSS (2018).

Table 4 displays a summary evaluation of the estimators with regard to the estimated production elasticities and returns to scale. The performance of the Translog specifications and the dynamic panel data model is given particular attention. Generally, the interest was to detect systematic differences across estimators and countries and to assess their practical implementation. Detailed results tables are presented in PETRICK and KLOSS (2018), which includes an overview table for each country containing the results for the eight models, plus an additional table for each country including more in-depth diagnostic results for the BB model.

As a general tendency, factor elasticities were found to be low for land and capital, high for materials and somewhere in between for labor (Table 4 and Table 5). WLP estimates for the first two of these factors are in the range of 0.2 and lower, sometimes not significantly different from zero. The production elasticity of materials is typically between 0.5 and 0.8. Labor elasticities usually fluctuate at around 0.2. These magnitudes are broadly in line with OP estimates reported by RIZOV et al. (2013: 549).[4]

The estimates support the conventional wisdom that OLS tends to be upward biased for particularly variable factors. In the present data, this primarily applies to materials, the OLS estimate of which is (except for Denmark) higher than its revenue share. It may be taken as evidence for the existence of serially correlated, unobservable factors (OLLEY and PAKES, 1996: 1274). The opposite bias is found for capital in the Within estimator, which is typically below the revenue share. This tendency is also in line with previous studies and can be attributed to the low variance of capital over time (GRILICHES and HAUSMAN, 1986).

The LP estimator commonly produces a lower elasticity for materials than OLS, the only exception being the United Kingdom. In case of the WLP estimator the only exception is France. LP and WLP estimates are typically very similar which makes us feel

3 Inclusion of fertilizer leads to results that display negative estimates of land coefficients in conjunction with relatively high materials coefficients for several countries in the sample. See KLOSS (2017: 50-53) for an in-depth analysis of the role of materials and land.

4 RIZOV et al. (2013) pool all farm types and they combine capital and land into a single production factor. Their first-stage estimation includes subsidy levels and other controls.

Table 4. Summary evaluation of estimator performance

	Denmark	France	Germany (West)	Italy	United Kingdom
Factor elasticities	All OLS below shares; materials below shares throughout CD, insignificant in LP and WLP (=0); Capital=0 in Within	Land and labour=0 in BB; materials above shares in OLS, LP, WLP, BB; capital<0.1 in shares, Within, BB	Materials above shares in OLS, LP WLP, BB, lower in Within; capital≈0 in Within and WLP, higher in LP&BB	Land=0 in OLS, LP, WLP, <0 in BB; materials above shares throughout CD; capital<0.1 in OLS, >0.1 BB, =0 in all other CD	Materials above shares in OLS, LP, BB; Capital≤0.1 throughout
Returns to scale	Shares add up to 2.07; OLS, LP, lower but still >1; Within, WLP, BB close to 1	OLS>1; close to 1.0 for the other estimators	1.1 in OLS, <1 in Within, BB; close to 1.0 in LP, WLP	Shares add up to 1.61; OLS≈1.1; Within, LP, WLP<0.9; BB=0.5	OLS, Within, LP≈1.2; WLP≈1.1; BB≈1.5
Performance of Translog	OLS unreasonable; Within close to CD; WLP unreasonable; interactions not sig. in Within and WLP	OLS unreasonable; Within close to CD; WLP unreasonable; Interactions not sig. in WLP	OLS unreasonable; Within close to CD; WLP unreasonable; interactions not sig. in Within and WLP	OLS unreasonable; Within in part close to CD; WLP unreasonable; interactions not sig. in WLP	OLS unreasonable; Within close to CD; WLP unreasonable; interactions not sig. in OLS and WLP
BLUNDELL/BOND estimator	Specification tests ok; levels better instrumented than diff.; relatively poor instrumentation	OID not passed; land, mat, capital, output highly persistent; levels better instrumented than diff.	OID not passed; labour, land, mat highly persistent; capital, output explosive; levels better instrumented than diff.	OID not passed; labour, capital highly persistent; land and output explosive; poor instrumentation	Specification tests ok; labour, land, capital highly persistent; materials and output explosive; poor instrumentation

Notes: BB: BLUNDELL/BOND, CD: Cobb Douglas, LP: LEVINSOHN/PETRIN, OID: Over-identification test, OLS: Ordinary Least Squares, WLP: WOOLDRIDGE/LEVINSOHN/PETRIN
Source: PETRICK and KLOSS (2018)

confident of the proxy variable identification strategy. These models may thus be taken as plausible alternatives to the received estimators. However, on theoretical grounds the WLP model further corrects for collinearity which gives this estimator an edge over the LP model. In addition, empirically the former is occasionally more successful in identifying the capital coefficient, i.e. with a higher precision as indicated by the standard errors.

Estimated elasticities of scale fluctuate around 1.0. Given the previous findings on production elasticities, OLS estimates tend to be higher than Within estimates. Overall, the scale elasticity in European crop farming appears to be close to one.

4.2 Validity of the Proxy Variable

According to the theoretical set-up of the control function approaches the materials proxy should be increasing in unobserved productivity (ω_{it}). To elaborate on this so-called monotonicity condition, we proceed in a similar fashion as LEVINSOHN and PETRIN (2003) by producing three-dimensional productivity surfaces of $\omega_{it} = f(m_{it}, k_{it})$. As ω_{it} is by definition unobserved, we need to come up with an estimate, $\widehat{\omega}_{it}$. To this end, we predict $\widehat{\omega}_{it}$ by using the parameter estimates of the production function. Based on data for the three dimensions – omega, materials and capital – we interpolate and smooth the original data using thin plate splines due to DUCHON (1976), a widely used data interpolation method for multidimensional data (see HASTIE et al., 2009: 162-167) for an overview). The processed data can then be used to draw three-dimensional surface plots and to visually inspect the monotonicity condition, as reported in PETRICK and KLOSS (2018) for our data. In Table 5, the results of this analysis are summarized which indicates that in general, the monotonicity condition holds throughout the sample of countries.

4.3 Functional Form: Cobb Douglas vs. Translog

The results on the Translog specification display remarkably uniform features across countries: the Within Translog elasticities at sample means were typically close to the Within Cobb Douglas, and the interaction terms of the Translog were often not jointly different from zero. The OLS Translog, on the other hand, produced unreasonable results throughout, e.g. reflected in the coexistence of negative production elasticities for some factors and elasticities bigger than one for others (at sample means). Similarly unreasonable results are observed for the WLP Translog. In this model, not a single country displayed interaction terms that were jointly significantly different from zero. Additionally, we applied the KLEIBERGEN and PAAP (2006) underidentification test to the WLP Translog model. Failing to reject the null hypothesis that the equation is unidentified implies an increased bias in the estimated coefficients. The bias is in the same direction as in the OLS estimator (BAUM et al., 2007). While we always rejected the null at the 5% and even the 1% significance level in the Cobb Douglas model, we could not do so in the cases of Denmark (p=0.41) and the United Kingdom (p=0.62) in the Translog model.

To sum up, the Translog specification does not perform well. Our findings are in line with other recent studies utilizing FADN data with this functional form (cf. ZHENGFEI et al., 2006; LATRUFFE and NAUGES, 2013). The prime reason for these difficulties is multicollinearity, which is supposedly even more severe in the Translog than in the Cobb Douglas, as many more parameters have to be estimated. While we cannot ultimately decide whether the true data generation process followed a Translog technology, we can say that farm-level data typically does not allow estimating its parameters. This makes the Translog a less credible functional form for applied work.

Table 5. Agricultural production elasticities in comparison

	Denmark	France	Germany (West)	Italy	United Kingdom
Labor	0.62	0.17	0.22	0.32	0.19
Land	0.23	0.04	-0.01#	-0.01#	0.17
Materials	0.00#	0.80	0.77	0.51	0.62#
Capital	0.10#	0.12	0.09	0.02#	0.10#
Ret. to Scale	0.95	1.13	1.08	0.84	1.09
Monotonicity	+	o	+	+	+

Notes: results for field crop farms in EU countries based on WOOLDRIDGE/LEVINSOHN/PETRIN (WLP) estimator. # not significantly different from zero at conventional confidence levels. Montonicity: + holds throughout, o holds partially.
Source: authors

4.4 Dynamic Panel Data Estimation

Our analysis of the BB estimator found labor and land to be highly persistent (PETRICK and KLOSS, 2018), which makes dynamic panel data estimation a natural option. Moreover, we regressed the differences of the latest available year on the lagged levels of all available previous years and the latest available levels on all available lagged differences of previous years. The reported *p*-values and coefficients of determination allow an insight into the explanatory power of the instrument sets. Generally, the instrument performance was better for levels (instrumented by differences) than for differences (instrumented by levels). System GMM approaches which do not only use differences but also levels for instrumentation (e.g. BLUNDELL and BOND, 1998) are thus warranted. Even so, the elasticities of the persistent factors labor, land and capital could often not be identified. Parameters were very sensitive to the selection of the sample and the precise specification of the estimator. Occasionally, dynamic factor evolution apparently followed an explosive process, as the AR(1) coefficient was estimated to be bigger than one. On the other hand, the estimates for materials appear very reasonable throughout, as they were typically somewhere between the OLS and Within results. It is here where the BB estimator can likely claim some superiority.

5 Conclusions

Within this study, we show that the assumptions underlying Within regression and the duality approach are fairly strong and implausible for the case of agriculture. Within approaches neglect the potentially important unobserved factors that vary over time. Duality relies on short-term profit maximization of agents and perfect competition on output and factor markets. In agriculture, these conditions are unlikely to be met, which may be a reason why these approaches have not performed well in estimation practice.

In light of the comprehensive literature on adjustment frictions on rural land, labor and capital markets, we regard the presence of adjustment costs as particular relevant for the production factors that are of key interest in agricultural applications. OLLEY and PAKES (1996), BLUNDELL and BOND (2000), LEVINSOHN and PETRIN (2003) and WOOLDRIDGE (2009) all base their identification strategy on adjustment frictions in factor allocation, which seems to be an a-priori plausible approach. The main difference is that BB allow time-invariant fixed effects, whereas OP, LP and WLP do not. The former impose a linear structure on the dynamic process, while it can be arbitrary in the latter. Even so, factor adjustment is assumed to occur in a single period in the proxy or control function approaches, while the process potentially covers many periods in the dynamic panel data models. In agricultural applications, this is a conceptual advantage of the BB approach. Adjustments of land, labor and capital are typically of an intertemporal nature, which is not appropriately covered by a one-year lag. Furthermore, OP and LP do not satisfactorily address the problem of collinearity in production function estimation. These approaches regard labor and land as fully flexible production factors for which there is no source of identifying variance across observations (ACKERBERG et al., 2015). However, WOOLDRIDGE (2009) proposes a solution to this issue by modifying and extending the central identifying assumptions of OP and LP.

In the empirical section, we show that OLS and Within display the biases expected from the literature. OLS typically overestimated the variable factor materials, while Within underestimated the relatively fixed factor capital. LP produced plausible results and may be taken as an easy-to-implement alternative to the received estimators. Given the conceptual problems in identifying the supposedly flexible inputs labor and land, which the other estimators except for BB and WLP share, this is only a second-best choice. Generally, LP and WLP produced very similar results which strengthens our confidence in the proxy approach on the whole. However, the theoretical advantage in identifying the land and labor coefficients gives the latter an edge over the former.

The combined first-difference and instrumental variable approach of the BB estimator goes a long way in trying to get rid of all the factors perturbing an unbiased estimation of productivity. Its assumptions on adjustment costs are theoretically very plausible and could be empirically supported for labor, land and capital. However, there is evidence that in agriculture this approach overshoots the mark. Adjustment costs are so high and factor evolution is so persistent that, despite using the systems GMM approach of BLUNDELL and BOND (1998), there is often too little variance left for identification. It is only with regard to materials that this estimator appeared to produce reasonable estimates.

Extending the received Cobb Douglas specification to a Translog generally did not generate meaningful results. Either the results were obviously implausible (OLS and WLP) or little different from Cobb Douglas (Within). These results are supposedly a direct consequence of multicollinearity. Hence, the more parsimonious parameterization of the Cobb Douglas remains a pragmatic, empirically well-supported alternative. We regard the analysis of alternative functional forms in conjunction with FADN data as an interesting starting point for future research. For instance, ZHENGFEI et al. (2006) proposed augmented Translog specifications that incorporate agronomic principles. However, so far, in applied empirical work there has been a trade-off between more flexible functional forms for production functions and methodological sophistication with regards to estimation methods.

Our estimates show a consistent picture of very low production elasticities for labor, land and fixed capital, whereas the elasticity of materials is around 0.7 throughout indicating that improving the availability of working capital is the most promising way to increasing agricultural productivity. This finding is in contrast to recent estimates by MUNDLAK et al. (2012), which report significant returns to land and fixed capital in a cross-country sample of developing and developed countries. However, our results are widely consistent with the OP estimates provided by RIZOV et al. (2013) on EU countries. Compared to other world regions, field crop technologies in the EU are characterized by a strong responsiveness to variable inputs such as fuel, fertilizer and chemicals. In a policy perspective, attempts to increase agricultural productivity in the EU in the short run, i.e. with given technology, should focus on this factor. Whether farmers actually exhaust the returns to such inputs should be analyzed in subsequent work, for example by calculating shadow prices of production factors based on the estimates provided in this article.

Summing up the methodological insights of this analysis, the recently suggested approaches to the estimation of production functions provide attractive conceptual improvements over the received Within and duality models. Using adjustment costs for identification of factor use seems particularly plausible in a sector like agriculture, in which long-lasting adjustment frictions in land, labor and capital have been recognized for a long time. Even so, empirical implementation of the conceptual sophistications built in these estimators does not always live up to expectations. This is particularly true for the dynamic panel estimator suggested by BLUNDELL and BOND (2000), which mostly failed to identify reasonable elasticities for the (quasi-) fixed factors. Less demanding proxy approaches such as due to LEVINSOHN and PETRIN (2003) and WOOLDRIDGE (2009) represent an interesting alternative for agricultural applications.

References

ACKERBERG, D., L. BENKARD, S. BERRY and A. PAKES (2007): Econometric tools for analyzing market outcomes. In: Heckman, J.J. and E.E. Leamer (eds.): Handbook of Econometrics. Volume 6A. Elsevier/North-Holland, Amsterdam: 4171-4276.

– (2015): Identification Properties of Recent Production Function Estimators. In: Econometrica 83 (6): 2411-2451. In: http://doi.org/10.3982/ECTA13408.

ALLEN, D.W. and D. LUECK (2002): The Nature of the Farm. Contracts, Risk, and Organization in Agriculture. MIT Press, Cambridge.

ARELLANO, M. and S. BOND (1991): Some Tests of Specification for Panel Data: Monte Carlo Evidence and an Application to Employment Equations. In: Review of Economic Studies 58 (2): 277-297.

BAUM, C., M. SCHAFFER and S. STILLMAN (2007): Enhanced routines for instrumental variables/generalized method of moments estimation and testing. In: Stata Journal 7 (4): 465-506.

BLUNDELL, R. and S. BOND (1998): Initial conditions and moment restrictions in dynamic panel data models. In: Journal of Econometrics 87 (1): 115-143.

– (2000): GMM Estimation with Persistent Panel Data: An Application to Production Functions. In: Econometric Reviews 19 (3): 321-340.

BOND, S. and M. SÖDERBOM (2005): Adjustment Costs and the Identification of Cobb Douglas Production Functions. Working Paper, Oxford. In: https://www.ifs.org.uk/publications/3259, Abruf: 22.04.2018.

CHAMBERLAIN, G. (1982): Multivariate Regression Models for Panel Data. In: Journal of Econometrics 18 (1): 5-46.

CHAMBERS, R.G. (1988): Applied production analysis. A dual approach. Cambridge University Press, Cambridge.

DUCHON, J. (1976): Splines minimizing rotation invariant semi-norms in Sobolev spaces. In: Schempp, W. and K. Zeller (eds.): Constructive Theory of Functions of Several Variables. Springer, Berlin.

EUROPEAN COMMISSION (2007): Community Committee for the Farm Accountancy Data Network -Definitions of Variables used in FADN standard results. Brussels.

– (2008): Community Committee for the Farm Accountancy Data Network - Farm Return Data Definitions. Brussels.

– (2012): Agriculture in the European Union. Statistical and economic information. Report 2012. Luxembourg.

GLAUBEN, T., M. PETRICK, H. TIETJE and C. WEISS (2009): Probability and timing of succession or closure in family firms: a switching regression analysis of farm households in Germany. In: Applied Economics 41 (1): 45-54.

GRILICHES, Z. and J. HAUSMAN (1986): Errors in Variables in Panel Data. In: Journal of Econometrics 31 (1): 93-118.

GRILICHES, Z. and J. MAIRESSE (1998): Production Functions: The Search for Identification. In: Strom, S. (ed.): Econometrics and economic theory in the 20th century. Cambridge University Press, Cambridge: 169-203.

HASTIE, T., R. TIBSHIRANI and J. FRIEDMAN (2009): The Elements of Statistical Learning: Data Mining, Inference, and Prediction. 2nd edition. Springer, New York.

HEMPELL, T. (2005): What's spurious, what's real? Measuring the productivity impacts of ICT at the firm-level. In: Empirical Economics 30 (2): 427-464.

KAZUKAUSKAS, A., C. NEWMAN and F. THORNE (2010): Analysing the Effect of Decoupling on Agricultural Production: Evidence from Irish Dairy Farms using the Olley and Pakes Approach. In: German Journal of Agricultural Economics 59 (3): 144-157.

KLEIBERGEN, F. and R. PAAP (2006): Generalized reduced rank tests using the singular value decomposition. In: Journal of Econometrics 133 (1): 97-126. doi:10.1016/j.jeconom.2005.02.011.

KLOSS, M. (2017): Factor productivity in EU agriculture: A microeconometric perspective. Studies on the Agricultural and Food Sector in Transition Economies 89. IAMO, Halle (Saale).

LATRUFFE, L. and C. NAUGES, (2013): Technical efficiency and conversion to organic farming: the case of France. In: European Review of Agricultural Economics 41 (2): 227-253.

LATRUFFE, L., B.E. BRAVO-URETA, A. CARPENTIER, Y. DESJEUX and V.H. MOREIRA (2017): Subsidies and Technical Efficiency in Agriculture: Evidence from European Dairy Farms. In American Journal of Agricultural Economics 99 (3): 783-799. DOI: 10.1093/ajae/aaw077.

LEVINSOHN, J. and A. PETRIN (2003): Estimating Production Functions Using Inputs to Control for Unobservables. In: Review of Economic Studies 70 (2): 317-341.

MARSCHAK, J. and W.H. ANDREWS (1944): Random Simultaneous Equations and the Theory of Production. In: Econometrica 12 (3/4): 143-205.

MUNDLAK, Y. (1961): Empirical Production Function Free of Management Bias. In: Journal of Farm Economics 43 (1): 44-56.

– (2001): Production and supply. In: Gardner, B.L. and G.C. Rausser (eds.): Handbook of Agricultural Economics. Volume 1A. Agricultural Production, Elsevier, Amsterdam: 3-85.

MUNDLAK, Y., R. BUTZER and D.F. LARSON (2012): Heterogeneous technology and panel data: The case of the agricultural production function. In: Journal of Development Economics 99 (1): 139-149.

OLLEY, G.S. and A. PAKES (1996): The dynamics of productivity in the telecommunications equipment industry. In: Econometrica 64 (6): 1263-1297.

PETRICK, M. and M. KLOSS (2018): Identifying Factor Productivity from micro-data: The case of EU agriculture. IAMO Discussion Papers. IAMO, Halle (Saale). In: https://www.iamo.de/publikationen/iamo-discussion-papers/, Abruf: 22.04.2018.

PETRICK, M. and L. LATRUFFE (2006): Contractual relations in agricultural credit markets: a hedonic pricing approach with application to Poland. In: Journal of Agricultural Economics 57 (1): 49-63.

PETRIN, A. and J. LEVINSOHN (2012): Measuring aggregate productivity growth using plant-level data. In: The RAND Journal of Economics 43 (4): 705-725.

POLLAK, R.A. (1985): A Transaction Cost Approach to Families and Households. In: Journal of Economic Literature 23 (2): 581-608.

RIZOV, M., J. POKRIVCAK and P. CIAIAN (2013). CAP Subsidies and Productivity of the EU Farms. In: Journal of Agricultural Economics 64 (3): 537-557. In: http://doi.org/10.1111/1477-9552.12030.

ROODMAN, D. (2009): How to do xtabond2: An introduction to difference and system GMM in Stata. In: Stata Journal 9 (1): 86-136.

SHUMWAY, C.R. (1995): Recent Duality Contributions in Production Economics. In: Journal of Agricultural and Resource Economics 20 (1): 178-194.

WOOLDRIDGE, J.M. (2009): On estimating firm-level production functions using proxy variables to control for unobservables. In: Economics Letters 104 (3): 112-114.

– (2010): Econometric analysis of cross section and panel data. 2nd edition. MIT Press, Cambridge, MA.

ZHENGFEI, G., A. LANSINK, M. VAN ITTERSUM and A. WOSSINK (2006): Integrating Agronomic Principles into Production Function Specification: A Dichotomy of Growth Inputs and Facilitating Inputs. In: American Journal of Agricultural Economics 88 (1): 203-214.

Contact author:
DR. MATHIAS KLOSS
University Hospital Halle
Magdeburger Str. 24, 06112 Halle (Saale)
e-mail: mkloss@posteo.de