Estimating Multi-Product Production Functions and Productivity using Control Functions

Emir Malikov

Department of Agricultural Economics & Rur. Soc.
Auburn University, Auburn, AL
emalikov@stlawu.edu

Selected Paper prepared for presentation at the 2016 Agricultural & Applied Economics Association Annual Meeting, Boston, Massachusetts, July 31 – August 2

Estimating Multi-Product Production Functions and Productivity using Control Functions

Emir Malikov*

First Draft: June 25, 2015
This Draft: November 30, 2015

Abstract

The existing control-function-based approaches to the identification of firm-level production functions are exclusively concerned with the estimation of single-output production functions despite that, in practice, most firms produce multiple outputs. While one can always opt to employ a single-product specification of the production process by a priori aggregating the firm's outputs, such a formulation is rarely an accurate portrayal of the firm's productive process. This paper extends the control-function-based approach to the structural identification and estimation of firm-level production functions and productivity to the *multi*-product setting. Specifically, I consider the nonparametric estimation of multi-product production functions. Among other advantages, explicit modeling of multiple outputs allows the identification of cross-output elasticities representing the technological trade-off between individual outputs along the firm's production possibilities frontier, which a traditional single-output production function approach is unable to deliver. To showcase the methodology, I apply it to study the multi-product production technology of Norwegian dairy farms during the 1998–2008 period.

Keywords: control function, dairy, endogeneity, multiple-output, production function, productivity, sieve estimation

JEL Classification: C14, D24, L10, Q12

*Department of Agricultural Economics & Rur. Soc., Auburn University, Auburn, AL; Department of Economics, St. Lawrence University, Canton, NY; *Email*: emalikov@stlawu.edu.

I am indebted to Subal Kumbhakar and Mike Tsionas for many thoughtful comments that helped improve this paper. I would also like to thank seminar participants at Auburn University for their insightful feedback.

1 Introduction

The identification and consistent estimation of the production function, which lies at the heart of the theory of the firm, is among the oldest empirical problems in economics dating back at least as early as the 19th century (Chambers, 1997). The presence of latent productivity is the primary obstacle to the identification of firm-level production functions. While usually unobserved by an econometrician, it is one of the key determinants of the firm's endogenous input allocation decisions resulting in its correlation with inputs entering the production function. If not accounted for explicitly, the presence of the firm's productivity leads to the classical endogeneity (omitted variable) problem as first discussed by Marschak and Andrews (1944). This endogeneity issue is also sometimes referred to as the "transmission bias" problem (Griliches and Mairesse, 1998).

Given a rather unsatisfactory performance of conventional approaches to tackling endogeneity in the production function context, such as fixed effects estimation or instrumenting for inputs using output and/or input prices [see Ackerberg, Caves and Frazer (2006), Gandhi, Navarro and Rivers (2013) for excellent discussions], the recently developed alternative control-function-based approach to the identification of firm-level production functions by Olley and Pakes (1996) and Levinsohn and Petrin (2003) has gained wide popularity among practitioners. Structural in nature, the approach makes use of the lagged inputs as the source of exogenous variation (under some assumptions about the firm's economic environment) and employs investment or intermediate inputs to proxy for latent productivity.

The existing control-function-based identification strategies are however all exclusively concerned with the estimation of *single*-output production functions despite that, in practice, most firms produce *multiple* outputs. This paper seeks to fill in this gap in the literature. Building on the framework of Gandhi et al. (2013), I extend the proxy approach to the structural identification and estimation of firm-level production functions and productivity to the *multi*-product setting.

Specifically, this paper considers the estimation of multi-product production functions where, motivated by Fernández, Koop and Steel (2000), I let outputs form an unspecified (aggregate) index function. Such a treatment of multiple outputs differs from an *a priori* aggregation, which most studies using a traditional single-output production function normally resort to, in that it *(i)* does not use *ad hoc* aggregation weights[1] and, more importantly, *(ii)* allows the identification of cross-output elasticities representing the *technological* trade-off between individual outputs along the firm's production possibilities frontier, which the traditional single-output formulation is unable to deliver. As in Levinsohn and Petrin (2003), unobserved firm-level productivity is proxied via inverting the conditional demand for non-dynamic intermediate inputs. In order to avoid a likely possibility of misspecifying the production function, I employ the nonparametric formulation not only for the control function but also for the production function itself, including the output index. Specifically, I estimate the multi-product production function along with the firm's productivity using cubic B-spline sieves. The identification strategy is robust to Ackerberg et al.'s (2006) and Gandhi et al.'s (2013) criticisms. The proposed two-stage estimation procedure is implemented by recasting it in a multiple-equation nonparametric GMM framework.

The approach to the identification of the firm's production technology pursued in this paper conceptually differs from other available alternatives in the literature, the bulk of which focuses on the estimation of "stochastic frontier" models of production. Regardless of whether the firm's production process is formulated as a single-output production function or a multiple-output ra-

[1] Single-output production functions are usually estimated with the output defined as total sales, whereby individual outputs are effectively aggregated using their relative prices.

dial/directional distance function,[2] most approaches to handling endogeneity in stochastic frontier models rely on output and/or input prices as the source of exogenous variation needed to achieve the identification of the frontier (e.g., Atkinson and Primont, 2002; Atkinson et al., 2003; Tsionas et al., 2015; Kumbhakar and Tsionas, 2015; Atkinson and Tsionas, 2015). In many instances, they also require additional assumptions such as the parameterization of the relationship between the endogenous regressors and the stochastic component in the model (e.g., Kutlu, 2010; Tran and Tsionas, 2013; Griffiths and Hajargasht, 2015). Some researchers may however have reservations about the practicality of using the price information as the source of exogenous variation in order to identify the firm's production technology. Not only are the data on prices often unavailable to practitioners or prone to measurement errors (Levinsohn and Petrin, 2003), but the use of prices may also be problematic on theoretical grounds (Griliches and Mairesse, 1998; Ackerberg et al., 2006, 2007). Specifically, the validity of output and/or input prices as exogenous instruments is normally justified by invoking the assumption of perfectly competitive markets in which firms operate. However, if firms were indeed price-takers, in theory, one should not observe the firm-level variation in prices[3] and, without such a variation, prices cannot be used as operable instruments. In contrast, if a researcher does observe the variation in output and/or input prices across individual firms, the latter variation likely reflects differences in the firm's market power or the quality of either inputs or output. In both instances, the variation in prices is unlikely to be exogenous to firms' decisions and hence cannot help the identification (see Gandhi et al., 2013, for more discussion).

In this paper, I therefore opt to pursue the structural identification strategy that does not rely on the firm-level variation in output and/or input prices. Along the lines of Gandhi et al. (2013), my approach to the identification of multi-product production functions relies on the information contained in the firm's first-order condition for an intermediate input, which does *not* require prices to vary across firms. On a related note, Amsler et al. (2015) discuss few available methods in the stochastic frontier literature that also do not require the price information for identification, which make use of either higher moment conditions or copulas (e.g., see Tran and Tsionas, 2015).

The proposed methodology is showcased by applying it to study the multi-product production technology of Norwegian dairy farms during the 1998–2008 period. To empirically assess the sensitivity of the production function and productivity estimates to the treatment of multiple outputs in the multi-product processes, I also estimate a conventional single-output production function where multiple outputs are *a priori* aggregated into a single measure (real total sales). I document dramatic distortions in farms' input elasticities, returns to scale, and productivity estimates obtained from a traditional single-output specification, which takes the multi-product nature of production technology for granted.

The rest of the paper proceeds as follows. Section 2 describes the model of multi-product production. I discuss the identification and estimation strategy in Section 3. Section 4 describes the data. The results are discussed in Section 5. Section 6 concludes.

2 Multi-Product Production Function

Consider a standard formulation of the production process of a firm i ($i = 1, \dots, n$) in the time period t ($t = 1, \dots, T$) in which physical capital K_{it}, land N_{it}, labor L_{it} and materials M_{it} (intermediate input) are being transformed into a *single* output Y_{it} via the time-varying production function $F_t(\cdot)$ given Hicks-neutral productivity. The literature that seeks to estimate the produc-

[2] For theoretical underpinnings of the multi-output multi-input distance function models, see Färe and Primont (1995).
[3] Except maybe across different regions.

tion function via the control function approach (e.g., Olley and Pakes, 1996; Levinsohn and Petrin, 2003; Ackerberg et al., 2006; Doraszelski and Jaumandreu, 2013) usually formalizes the stochastic production process as

$$Y_{it} = F_t(K_{it}, N_{it}, L_{it}, M_{it}) \exp\{w_{it} + \eta_{it}\}, \tag{2.1}$$

where $(w_{it} + \eta_{it})$ is the composite productivity term consisting of the persistent (first-order) Markovian productivity w_{it} and a random *i.i.d.* productivity shock η_{it}. The production function $F_t(\cdot)$ is often assumed to take the Cobb-Douglas form with varying assumptions about the timing of different input decisions.

While one can always justify a single-product formulation of the production process (2.1) by defining Y_{it} to be some aggregate measure of the firm's output, such as deflated total sales or revenue (in fact, as customarily done during the estimation), such a formulation is rarely an accurate portrayal of the firm's productive process in practice. Most firms produce multiple products which exhibit varying degrees of substitutability or complementarity between one another. These relationships between different outputs are of great interest to economists on their own, which one however cannot discern using the formulation in (2.1).

I generalize a single-output formulation of the production process in (2.1) to the case of *multiple* outputs along the lines of Fernández, Koop and Steel (2000). The multi-product (stochastic) production process is given by

$$H_t(Y_{1,it}, \ldots, Y_{S,it}) = F_t(K_{it}, N_{it}, L_{it}, M_{it}) \exp\{w_{it} + \eta_{it}\}, \tag{2.2}$$

where a firm i in the time period t is assumed to transform physical capital K_{it}, land N_{it}, labor L_{it} and materials M_{it} into S outputs $(Y_{1,it}, \ldots, Y_{S,it})$ via the time-varying production function $F_t(\cdot)$ given Hicks-neutral productivity. Following Fernández et al. (2000, 2002, 2005), outputs are combined into an aggregate (index) function $H_t(\cdot)$.[4]

An evident assumption embedded in (2.2) is the (homothetic) separability of functions $H_t(\cdot)$ and $F_t(\cdot)$. While at first this assumption may seem to be rather restrictive, it is however also imposed onto a standard single-output production function where Y_{it} is assumed to be separable from $F_t(\cdot)$. Clearly, (2.2) nests the single-product formulation in (2.1) as a special case, i.e., the case when $S = 1$ and function $H_t(\cdot)$ is normalized to an identity link function. Some recent specifications of similar separable multi-product production functions include Greene (2008), O'Donnell and Nguyen (2013) and O'Donnell (2014) among others.[5]

Further, if one is willing to parameterize production function $F_t(\cdot)$, say, by letting it take the Cobb-Douglas form, then some obvious choices for $H_t(\cdot)$ include the CES or Cobb-Douglas specifications. The only normalizing restriction required for the identification of $H_t(\cdot)$ is the (joint) linear homogeneity of the latter in outputs $(Y_{1,it}, \ldots, Y_{S,it})$.[6] Note that the same normalization is implied in the single-output production function (2.1), where the unknown coefficient in front of Y_{it} is being normalized to unity.

Following Ackerberg et al. (2006), I assume that, unlike a freely varying M_{it}, both K_{it} and L_{it} are subject to adjustment frictions (e.g., time-to-install, hiring costs) and thus are quasi-fixed.[7] It is perhaps even more natural to assume that land N_{it} is also a quasi-fixed input. Thus, M_{it} is

[4] Note that, unlike in Fernández et al. (2000, 2002, 2005), I do not parameterize neither function $F_t(\cdot)$ nor function $H_t(\cdot)$ leaving them to be unspecified nonparametric functions.

[5] However, note that in some applications, where different outputs require a different input mix, the separability may be quite unappealing. Should that be the case, the use of the proposed model is left up to a researcher's discretion.

[6] Which, if pursued, the CES and the Cobb-Douglas specifications of H_t can be easily required to maintain.

[7] Such a timing framework resolves the perfect collinearity problem pointed out by Ackerberg et al. (2006).

determined by the firm in period t, whereas K_{it}, N_{it} and L_{it} are determined in period $t-1$. Note that, despite all three K_{it}, N_{it} and L_{it} being quasi-fixed, K_{it} and N_{it} are the only two that are dynamic, i.e., state variables following deterministic (controlled) laws of motion.

Given the assumptions about the productivity above, the distributions of w_{it} and η_{it} can respectively be written as

$$\mathcal{P}_w\left(w_{it}|\ \Omega_{it-1}\right) = \mathcal{P}_w\left(w_{it}|\ w_{it-1}\right) \tag{2.3a}$$

$$\mathcal{P}_\eta\left(\eta_{it}|\ \Omega_{it}\right) = \mathcal{P}_\eta\left(\eta_{it}\right), \tag{2.3b}$$

where Ω_{it} denotes the information available to the firm for making period t decisions. From (2.3a), it follows that

$$w_{it} = \mathbb{E}[w_{it}|\ w_{it-1}] + \zeta_{it}, \quad \text{where} \quad \mathbb{E}[\zeta_{it}|\ \Omega_{it-1}] = \mathbb{E}[\zeta_{it}|\ w_{it-1}] = 0. \tag{2.4}$$

Here, ζ_{it} is the innovation in persistent productivity, unobservable to firms in period $t-1$. Firms do however observe $w_{it} \in \Omega_{it}$, which consists of $\mathbb{E}[w_{it}|\ w_{it-1}]$ and ζ_{it}, in period t when decisions concerning freely varying inputs are being made.

Similarly, from (2.3b), it follows that

$$\mathbb{E}[\eta_{it}|\ \Omega_{it}] = \mathbb{E}[\eta_{it}] = 0, \tag{2.5}$$

where I normalize the mean of η_{it} to zero. The above implies that the random shock η_{it} is observable to firms in period t only *ex post*, i.e., after all production decisions, including those about freely varying inputs, take place.

The described framework of the production process is quite similar to that considered by Olley and Pakes (1996), Levinsohn and Petrin (2003) and Doraszelski and Jaumandreu (2013). Putting the issue of multiple outputs aside, the primary differences between this paper's setup and theirs are as follows. First, in order to address Ackerberg et al.'s (2006) criticism, I assume M_{it} is the only freely varying input. Second, in order to minimize the chances of the misspecification bias, I do not restrict the production function $F_t(\cdot)$ and the output index $H_t(\cdot)$ to take any prespecified parametric forms but rather allow them to be unspecified (smooth) nonparametric functions.

3 Identification and Estimation Strategy

The estimation of the production function in (2.2) would have been trivial (given the assumptions), had the persistent productivity w_{it} been observable to an econometrician. Omitting it from the regression is also not an option, since the latter would give rise to the endogeneity problem given that w_{it} is correlated with the inputs (directly and/or through w_{it-1}). Conventional methods to tackle endogeneity in the production function context, such as fixed effects estimation or instrumenting for inputs using input prices, have been quite unsatisfactory in practice and problematic from the perspective of economic theory [see Ackerberg et al. (2006) and Gandhi et al. (2013) for excellent reviews]. Recent advances in the identification of firm-level production functions and productivity primarily include dynamic panel data methods (Arellano and Bond, 1991; Blundell and Bond, 1998) and control function approaches (Olley and Pakes, 1996; Levinsohn and Petrin, 2003; Ackerberg et al., 2006; Gandhi et al., 2013), which use the lagged inputs as the source of exogenous variation (under some assumptions about the firm's economic environment). In this paper, I follow the latter strand of the literature by considering the control function approach.

In what follows, I consider the identification and estimation of the gross multi-product production function. My approach to identification builds on Gandhi et al.'s (2013) framework (where I explicitly address their recent criticism), which I generalize to accommodate multiple outputs.

Making use of the definition of the (joint) linear homogeneity of $H_t(\cdot)$ in outputs — a normalizing restriction I require for the identification, as discussed in Section 2 — for some scalar $\kappa > 0$, I get

$$\begin{aligned} H_t(\kappa Y_{1,it},\ldots,\kappa Y_{S,it}) &= \kappa H_t(Y_{1,it},\ldots,Y_{S,it}) \\ &= \kappa F_t(K_{it},N_{it},L_{it},M_{it})\exp\{w_{it}+\eta_{it}\}, \end{aligned} \tag{3.1}$$

where I have made a substitution for $H_t(\cdot)$ in the second equality using (2.2). Setting κ equal to the *inverse* of one of the outputs, say $Y_{1,it}$, yields[8]

$$Y_{1,it} = \frac{F_t(K_{it},N_{it},L_{it},M_{it})}{H_t\left(1,\frac{Y_{2,it}}{Y_{1,it}},\ldots,\frac{Y_{S,it}}{Y_{1,it}}\right)}\exp\{w_{it}+\eta_{it}\}. \tag{3.2}$$

Taking logs of both sides of (3.2), I get

$$\begin{aligned} y_{1,it} &= f_t(K_{it},N_{it},L_{it},M_{it}) - h_t(1,Z_{2,it},\ldots,Z_{S,it}) + w_{it} + \eta_{it} \\ &\overset{\text{def}}{=} f_t(K_{it},N_{it},L_{it},M_{it}) + \widetilde{h}_t(Z_{2,it},\ldots,Z_{S,it}) + w_{it} + \eta_{it}, \end{aligned} \tag{3.3}$$

where the lower-case variables/functions denote the logs of the respective variables/functions (e.g., $y_{1,it} = \ln Y_{1,it}$) and, for convenience, I have defined $Z_{s,it} \overset{\text{def}}{=} \frac{Y_{s,it}}{Y_{1,it}}\ \forall\ s = 2,\ldots,S$ and $\widetilde{h}_t(\cdot) \overset{\text{def}}{=} -h_t(1,\cdot)$.

The production function in (3.3) however contains the unobservable (to an econometrician) productivity w_{it}. To control for it, I make use of the conditional intermediate input demand function $M_{it} = \mathbb{M}_t(K_{it},N_{it},L_{it},w_{it})$ generated by solving the firm's inter-temporal discounted lifetime profits maximization problem.[9] Given the profit-maximizing behavior by firms, $\mathbb{M}_t(\cdot)|\ M_{it} > 0$ must be strictly monotonic in w_{it} for any given (K_{it},N_{it},L_{it}) (see Levinsohn and Petrin, 2003). This monotonicity condition is identical to that for the investment function derived by Pakes (1994). Hence, $\mathbb{M}_t(\cdot)$ can be inverted to proxy for persistent productivity via $w_{it} = \mathbb{M}_t^{-1}(K_{it},N_{it},L_{it},M_{it})$. Specifically, from (3.3) I get

$$\begin{aligned} y_{1,it} &= f_t(K_{it},N_{it},L_{it},M_{it}) + \widetilde{h}_t(Z_{2,it},\ldots,Z_{S,it}) + \psi_t[w_{it-1}] + \zeta_{it} + \eta_{it} \\ &= f_t(K_{it},N_{it},L_{it},M_{it}) + \widetilde{h}_t(Z_{2,it},\ldots,Z_{S,it}) + \psi_t\left[\mathbb{M}_{t-1}^{-1}(K_{it-1},N_{it-1},L_{it-1},M_{it-1})\right] + \zeta_{it} + \eta_{it} \\ &\overset{\text{def}}{=} f_t(K_{it},N_{it},L_{it},M_{it}) + \widetilde{h}_t(Z_{2,it},\ldots,Z_{S,it}) + \varphi_t(K_{it-1},N_{it-1},L_{it-1},M_{it-1}) + \zeta_{it} + \eta_{it}, \end{aligned} \tag{3.4}$$

where I have also made use of the first-order Markovian nature of w_{it} from (2.4) in the first equality by letting $\mathbb{E}[w_{it}|\ w_{it-1}]$ be an unknown function $\psi_t[\cdot]$ and have defined $\varphi_t(\cdot) \overset{\text{def}}{=} \psi_t\left[\mathbb{M}_{t-1}^{-1}(\cdot)\right]$ in the last equality.

3.1 Identification

The production process framework described in Section 2 implies that *(i)* M_{it} appearing inside $f_t(\cdot)$ in (3.4) is correlated with the transitory productivity shock ζ_{it} and *(ii)* the output ratios

[8] It is easy to show that the normalization is invariant to the choice of output.

[9] Note that the demand function for M_{it} is not inter-temporal, i.e., dynamic, since M_{it} is assumed to be a freely varying input.

$(Z_{2,it}, \ldots, Z_{S,it})$ are correlated with both ζ_{it} and the random shock η_{it} as a result of simultaneity of all outputs.[10] One may think that the additive nonparametric model (3.4) can be seemingly identified on the basis of the following $(t-1)[(S-1)+4]+3$ moment conditions:

$$\mathbb{E}[\zeta_{it}+\eta_{it}|\ \Omega_{it-1}] = \mathbb{E}\left[\zeta_{it}+\eta_{it}\ \middle|\ \begin{array}{l} K_{it}, N_{it}, L_{it}, \\ K_{it-1}, N_{it-1}, L_{it-1}, M_{it-1}, Z_{2,it-1}, \ldots, Z_{S,it-1}, \ldots, \\ K_{i1}, N_{i1}, L_{i1}, M_{i1}, Z_{2,i1}, \ldots, Z_{S,i1} \end{array}\right] = 0, \quad (3.5)$$

where $(K_{it}, N_{it}, L_{it}, K_{it-1}, N_{it-1}, L_{it-1}, M_{it-1})$ instrument for themselves and one has the abundance of exogenous *lags* of $(K_{it}, N_{it}, L_{it}, M_{it}, Z_{2,it}, \ldots, Z_{S,it})$ to instrument for M_{it} and $(Z_{2,it}, \ldots, Z_{S,it})$.

However, despite the apparent abundance of valid instruments, equation (3.4) still is not identified due to the presence of freely varying M_{it} inside the production function $f_t(\cdot)$ – an issue recently pointed out by Gandhi et al. (2013). They show (in this paper's notation) that, when conditioned on the information set Ω_{it-1}, M_{it} entering $\mathbb{E}[f_t(\cdot)|\Omega_{it-1}]$ in the identifiable conditional expectation $\mathbb{E}[y_{1,it}|\ \Omega_{it-1}]$ from (3.4):

$$\begin{aligned} M_{it} &= \mathbb{M}_t(K_{it}, N_{it}, L_{it}, w_{it}) \\ &= \mathbb{M}_t(K_{it}, N_{it}, L_{it}, \psi_t(w_{it-1})+\zeta_{it}) \\ &= \mathbb{M}_t(K_{it}, N_{it}, L_{it}, \varphi_t(K_{it-1}, N_{it-1}, L_{it-1}, M_{it-1})+\zeta_{it}), \end{aligned} \quad (3.6)$$

is a function of the following observables $(K_{it}, N_{it}, L_{it}, K_{it-1}, N_{it-1}, L_{it-1}, M_{it-1})$. Comparing these variables with those entering $\mathbb{E}[f_t(\cdot)|\Omega_{it-1}]$ directly as well as the proxy for productivity $\varphi_t(\cdot)$, it is evident that the only extra source of variation for M_{it}, which has not already been included on the right-hand side of (3.4), is the *unobservable* ζ_{it}. In other words, M_{it} lacks an (excluded) instrument from outside of the production function. To tackle this under-identification problem, I augment the production function in (3.4) with an equation for the first-order condition (with respect to M_{it}) from the firm's profit-maximization problem, along the lines of Gandhi et al. (2013). Intuitively, the approach resembles the estimation of the system of simultaneous equations.

Since materials are a freely varying input, the firm's optimal choice of M_{it} can therefore be modeled as the *concentrated* expected profit-maximization problem[11] subject to the (already) optimal allocation of all dynamic and non-dynamic quasi-fixed inputs (K_{it}, N_{it}, L_{it}), i.e.,

$$\begin{aligned} \max_{M_{it}} \quad & \sum_{j=1}^{S} P_{j,t}\mathbb{E}[Y_{j,it}|\ \Omega_{it}] - P_{M,t}M_{it} = \\ & \sum_{j=1}^{S} P_{j,t}\frac{F_t(K_{it}, N_{it}, L_{it}, M_{it})\exp\{w_{it}\}\mathbb{E}[\exp\{\eta_{it}\}|\ \Omega_{it}]}{\widetilde{H}_t^j(Z_{1,it}^j, \ldots, Z_{S,it}^j)} - P_{M,t}M_{it}, \end{aligned} \quad (3.7)$$

where I have expressed each output $Y_{j,it}$ using (3.2) under the respective $\kappa = 1/Y_{j,it}$ normalization. I add the superscript j to the normalized output index $\widetilde{H}_t(\cdot)$ as well as the output ratios $Z_{s,it}$ to explicitly recognize that the index and the ratios change depending on the choice of the output used for the normalization. Further, $P_{j,t}$ and $P_{M,t}$ are the prices of output j and materials, respectively. Following the discussion in the Introduction, note that these prices do *not* need to vary across firms.

[10] As can easily be seen from (2.2).
[11] Under the risk neutrality of firms.

The first-order condition with respect to M_{it} is given by

$$\sum_{j=1}^{S} P_{j,t} \frac{\partial F_t(K_{it}, N_{it}, L_{it}, M_{it})}{\partial M_{it}} \frac{\exp\{w_{it}\} \times \theta}{\widetilde{H}_t^j(Z_{1,it}, \ldots, Z_{S,it})} = P_{M,t}, \tag{3.8}$$

where $\theta \stackrel{\text{def}}{=} \mathbb{E}[\exp\{\eta_{it}\}|\ \Omega_{it}]$ is a constant. Taking logs of both sides and recognizing that [due to the linear homogeneity of $H_t(\cdot)$]

$$\sum_{j=1}^{S} P_{j,t} \left[\widetilde{H}_t^j(Z_{1,it}^j, \ldots, Z_{S,it}^j)\right]^{-1} = [H_t(Y_{1,it}, \ldots, Y_{S,it})]^{-1} \sum_{j=1}^{S} P_{j,t} Y_{j,it},$$

from (3.8) I get

$$\ln\left[\frac{\partial F_t(K_{it}, N_{it}, L_{it}, M_{it})}{\partial M_{it}}\right] - h_t(Y_{1,it}, \ldots, Y_{S,it}) + w_{it} + \ln\theta = \ln\left[\frac{P_{M,t}}{\sum_{j=1}^{S} P_{j,t} Y_{j,it}}\right]. \tag{3.9}$$

Adding the log of M_{it} to both sides of (3.9) and then subtracting the latter from the normalized production function in logs (3.3) yields

$$\ln\left[\frac{\partial F_t(K_{it}, N_{it}, L_{it}, M_{it})}{\partial M_{it}} \frac{M_{it}}{F_t(K_{it}, N_{it}, L_{it}, M_{it})}\right] + \ln\theta - \eta_{it} = \ln\left[\frac{P_{M,t} M_{it}}{\sum_{j=1}^{S} P_{j,t} Y_{j,it}}\right]. \tag{3.10}$$

For convenience, I define function $G_t(K_{it}, N_{it}, L_{it}, M_{it}) \stackrel{\text{def}}{=} \frac{\partial F_t(K_{it}, N_{it}, L_{it}, M_{it})}{\partial M_{it}} \frac{M_{it}}{F_t(K_{it}, N_{it}, L_{it}, M_{it})}$, which equals the elasticity of the production function $F_t(\cdot)$ with respect to materials M_{it}. Further note that $V_{it} \stackrel{\text{def}}{=} \frac{P_{M,t} M_{it}}{\sum_{j=1}^{S} P_{j,t} Y_{j,it}}$ equals the nominal share of materials in the total nominal output and is usually observable from the data. Again, note that the construction of V_{it} does not require the data on firm-level output and/or input prices; knowledge of the firm's spending on materials and total nominal sales would suffice. Therefore, from (3.10) I can obtain a nonparametric equation that identifies both the function $G_t(\cdot)\theta$ and the random productivity residual η_{it}, i.e.,

$$\ln V_{it} = \ln\left[G_t(K_{it}, N_{it}, L_{it}, M_{it})\theta\right] - \eta_{it}, \tag{3.11}$$

on the basis of the following moment conditions[12]

$$\mathbb{E}[\eta_{it}|\ K_{it}, N_{it}, L_{it}, M_{it}] = 0. \tag{3.12}$$

I next identify (up to a constant) the production function $f_t(\cdot)$ from the already identified function $G_t(\cdot)$ by integrating the ratio of function $G_t(\cdot)$ and M_{it} over M_{it}. Specifically, note that

$$\frac{G_t(K_{it}, N_{it}, L_{it}, M_{it})}{M_{it}} = \frac{\partial \ln F_t(K_{it}, N_{it}, L_{it}, M_{it})}{\partial M_{it}} = \frac{\partial f_t(K_{it}, N_{it}, L_{it}, M_{it})}{\partial M_{it}}, \tag{3.13}$$

[12]Clearly, lags of $(K_{it}, N_{it}, L_{it}, M_{it})$ can also be added to the conditioning set. The entire set of valid moment conditions is

$$\mathbb{E}\left[\eta_{it} \left| \begin{array}{l} K_{it}, N_{it}, L_{it}, M_{it}, \\ K_{it-1}, N_{it-1}, L_{it-1}, M_{it-1}, Z_{2,it-1}, \ldots, Z_{S,it-1}, \ldots, \\ K_{i1}, N_{i1}, L_{i1}, M_{i1}, Z_{2,i1}, \ldots, Z_{S,i1} \end{array} \right.\right] = 0.$$

which, if integrated over M_{it}, yields

$$\int_0^{M_{it}} \frac{G_t(K_{it}, N_{it}, L_{it}, M_{it})}{M_{it}} \mathrm{d}M_{it} = f_t(K_{it}, N_{it}, L_{it}, M_{it}) + C_t(K_{it}, N_{it}, L_{it}), \tag{3.14}$$

where $C_t(\cdot)$ is the constant of integration that is a function of (K_{it}, N_{it}, L_{it}). In order to identify the production function $f_t(\cdot)$, I first need to identify the constant of integration $C_t(\cdot)$, since the integral $\int_0^{M_{it}} \frac{G_t(K_{it},N_{it},L_{it},M_{it})}{M_{it}} \mathrm{d}M_{it}$ can be easily obtained (identified), say via the numerical integration, using $G_t(\cdot)$ already identified from (3.11). For this, I first subtract (3.14) from the production function (3.3), i.e.,

$$y_{1,it} - \int_0^{M_{it}} \frac{G_t(K_{it}, N_{it}, L_{it}, M_{it})}{M_{it}} \mathrm{d}M_{it} - \eta_{it} = \widetilde{h}_t(Z_{2,it}, \ldots, Z_{S,it}) - C_t(K_{it}, N_{it}, L_{it}) + w_{it}. \tag{3.15}$$

Next, note that the left-hand side of (3.15) is fully identified from earlier and thus is observable. For convenience, I denote it as $y^*_{it} \stackrel{\text{def}}{=} y_{1,it} - \int_0^{M_{it}} \frac{G_t(K_{it},N_{it},L_{it},M_{it})}{M_{it}} \mathrm{d}M_{it} - \eta_{it}$. From (3.15), it then follows that productivity w_{it} is given by

$$w_{it} = y^*_{it} - \widetilde{h}_t(Z_{2,it}, \ldots, Z_{S,it}) + C_t(K_{it}, N_{it}, L_{it}). \tag{3.16}$$

Using the Markovian nature of the productivity, I substitute the right-hand side of (3.16) for w_{it} in (2.4):

$$\begin{aligned} y^*_{it} - &\widetilde{h}_t(Z_{2,it}, \ldots, Z_{S,it}) + C_t(K_{it}, N_{it}, L_{it}) = \\ &\psi_t\left[y^*_{it-1} - \widetilde{h}_t(Z_{2,it-1}, \ldots, Z_{S,it-1}) + C_t(K_{it-1}, N_{it-1}, L_{it-1})\right] + \zeta_{it}, \end{aligned} \tag{3.17}$$

which is a nonparametric equation that identifies both $\widetilde{h}_t(\cdot)$ and $C_t(\cdot)$ (up to a constant) on the basis of the following orthogonality conditions

$$\mathbb{E}\left[\zeta_{it} \middle| \begin{array}{l} K_{it}, N_{it}, L_{it}, \\ K_{it-1}, N_{it-1}, L_{it-1}, M_{it-1}, Z_{2,it-1}, \ldots, Z_{S,it-1}, \ldots, \\ K_{i1}, N_{i1}, L_{i1}, M_{i1}, Z_{2,i1}, \ldots, Z_{S,i1} \end{array}\right] = 0, \tag{3.18}$$

where $(K_{it}, N_{it}, L_{it}, K_{it-1}, N_{it-1}, L_{it-1}, Z_{2,it-1}, \ldots, Z_{S,it-1})$ instrument for themselves and the remaining lags are used to instrument for endogenous $(Z_{2,it}, \ldots, Z_{S,it})$. Thus, having identified $C_t(\cdot)$ I can now identify (up to a constant) the production function of interest $f_t(\cdot)$ from (3.14).

3.2 Estimation Procedure

I estimate the unknown production function $f_t(\cdot)$, the output index function $h_t(\cdot)$ and the (unobservable) persistent productivity w_{it} via the following two-stage procedure, which can also be combined into the multiple-equation nonparametric GMM system.

First Stage. I start by nonparametrically estimating the material share equation (3.11). In this paper, I employ the nonparametric sieve estimator.[13] Specifically, I approximate unknown $\ln\left[G_t(K_{it}, N_{it}, L_{it}, M_{it})\theta\right]$ in the logs of arguments using a series of basis functions $\mathscr{B}_r(\cdot)$, i.e.,

$$\ln\left[G_t(K_{it}, N_{it}, L_{it}, M_{it})\theta\right] \approx \sum_{r\geq 1} \boldsymbol{\alpha}_{t,r} \mathscr{B}_r(k_{it}, n_{it}, l_{it}, m_{it}),$$

[13] Feasible alternatives include the kernel-based Local Constant or Local Polynomial Least Squares estimators.

where $r = 1, \ldots, R_n$ and $R_n \to \infty$ slowly with n . I use cubic B-splines[14] for $\mathscr{B}_r(\cdot)$ due to their well-known robust finite-sample performance. Some other valid alternatives include Hermite or Laguerre polynomials as well as artificial neural network sieves. Given the orthogonality conditions in (3.12), the approximated function $\ln [G_t(K_{it}, N_{it}, L_{it}, M_{it})\theta]$ can be fitted via either the nonparametric least squares or nonparametric GMM method.

Recall that the residuals from fitting (3.11) are the (consistent) estimates of the random productivity shock η_{it}. Using these estimates $\widehat{\eta}_{it}$, I can consistently estimate $\theta \stackrel{\text{def}}{=} \mathbb{E}[\exp\{\eta_{it}\}]$ via $\widehat{\theta} = (nT)^{-1} \sum_{it} \exp\{\widehat{\eta}_{it}\}$. The latter allows us to obtain the estimate of $G_t(\cdot)$ free of constant θ:

$$\widehat{G}_t(K_{it}, N_{it}, L_{it}, M_{it}) = \exp \left\{ \sum_{r \geq 1} \widehat{\boldsymbol{\alpha}}'_{t,r} \mathscr{B}_r(k_{it}, n_{it}, l_{it}, m_{it}) - \ln \widehat{\theta} \right\},$$

which I then use to evaluate the integral in (3.14) in order to construct the estimate of $y^*_{1,it}$ to be used in the second stage:

$$\widehat{y}^*_{it} = y_{1,it} - \widehat{\int_0^{M_{it}} \frac{\widehat{G}_t(K_{it}, N_{it}, L_{it}, M_{it})}{M_{it}} \mathrm{d}M_{it}} - \widehat{\eta}_{it}.$$

Second Stage. Given the consistent estimates $\widehat{y}^*_{it}$ from the first stage, I next proceed with the nonparametric sieve estimation of (3.17), where I approximate (in logs) the unknown functions $\widetilde{h}_t(\cdot)$, $C_t(\cdot)$ and $\psi_t[\cdot]$ using cubic B-spline sieves:

$$\begin{aligned}
\widetilde{h}_t(Z_{2,it}, \ldots, Z_{S,it}) &\approx \sum_{r' \geq 1} \boldsymbol{\beta}'_{t,r'} \mathscr{B}_{r'}(z_{2,it}, \ldots, z_{S,it}) \\
C_t(K_{it}, N_{it}, L_{it}) &\approx \sum_{r'' \geq 1} \boldsymbol{\gamma}'_{t,r''} \mathscr{B}_{r''}(k_{it}, n_{it}, l_{it}) \\
\psi_t[\cdot] &\approx \sum_{r''' \geq 1} \boldsymbol{\delta}'_{t,r'''} \mathscr{B}_{r'''}(\cdot),
\end{aligned}$$

where $r' = 1, \ldots, R'_n$, $r'' = 1, \ldots, R''_n$ and $r''' = 1, \ldots, R'''_n$; R'_n, R''_n and R'''_n all increase with the sample size. The unknown functions above are estimated via GMM on the basis of moment conditions in (3.18).

The second-stage estimation readily provides the estimates of the aggregate index function $\widehat{\widetilde{h}}_t(\cdot)$ which informs us of the relationship between the multiple outputs that the firm produces. The production function $f_t(\cdot)$ is estimated from (3.14) using the second-stage estimates of $C_t(\cdot)$:

$$\widehat{f}_t(K_{it}, N_{it}, L_{it}, M_{it}) = \widehat{\int_0^{M_{it}} \frac{\widehat{G}_t(K_{it}, N_{it}, L_{it}, M_{it})}{M_{it}} \mathrm{d}M_{it}} - \widehat{C}_t(K_{it}, N_{it}, L_{it}) \ ,$$

and the unobservable productivity w_{it} is estimated using (3.3) as

$$\widehat{w}_{it} = y_{1,it} - \widehat{f}_t(K_{it}, N_{it}, L_{it}, M_{it}) - \widehat{\widetilde{h}}_t(Z_{2,it}, \ldots, Z_{S,it}) - \widehat{\eta}_{it} \ .$$

In all stages, the number of equispaced knots R_n, R'_n, R''_n and R'''_n are selected via the data-driven generalized cross-validation of Craven and Wahba (1979). Lastly, building on Wooldridge's

[14] More specifically, I use the tensor product of cubic B-splines for each element in $(k_{it}, n_{it}, l_{it}, m_{it})$.

(2009) suggestion, I can estimate both stages as a (sequential) system in the multiple-equation nonparametric GMM framework (where the instrument sets vary across equations), which not only improves efficiency but also permits the derivation of the robust variance-covariance matrix that accounts for the use of generated covariates in the second stage. A similar method-of-moments interpretation of multi-stage (parametric) estimators has also been proposed by Newey (1984). Specifically, I rewrite the two estimation stages in the form of the following orthogonality conditions:

$$\mathbb{E}\left[\begin{array}{c}\Big[\ln V_{it} - \sum_{r} \boldsymbol{\alpha}'_{t,r}\mathscr{B}_r(k_{it},n_{it},l_{it},m_{it})\Big]\begin{bmatrix}\mathscr{B}_1(k_{it},n_{it},l_{it},m_{it})\\ \vdots \\ \mathscr{B}_{R_n}(k_{it},n_{it},l_{it},m_{it})\end{bmatrix}\\ \left[\begin{array}{c} y^*_{it}(\boldsymbol{\alpha}_t) - \sum_{r'} \boldsymbol{\beta}'_{t,r'}\mathscr{B}_{r'}(z_{2,it},\ldots,z_{S,it}) + \sum_{r''} \boldsymbol{\gamma}'_{t,r''}\mathscr{B}_{r''}(k_{it},n_{it},l_{it}) - \\ \sum_{r'''} \boldsymbol{\delta}'_{t,r'''}\mathscr{B}_{r'''}\Big(y^*_{it-1}(\boldsymbol{\alpha}_t) - \sum_{r'} \boldsymbol{\beta}'_{t,r'}\mathscr{B}_{r'}(z_{2,it-1},\ldots,z_{S,it-1}) + \sum_{r''} \boldsymbol{\gamma}'_{t,r''}\mathscr{B}_{r''}(k_{it-1},n_{it-1},l_{it-1})\Big)\end{array}\right]\boldsymbol{\Xi}_{it}\end{array}\right] = \mathbf{0},$$

where $\boldsymbol{\alpha}_t = (\boldsymbol{\alpha}'_{t,1},\ldots,\boldsymbol{\alpha}'_{t,R_n})'$ and $\boldsymbol{\Xi}_{it}$ is a vector of basis functions formed using the instruments in (3.18). The vector of moment conditions above consists of two blocks, each corresponding to one of the stages in the estimation of the production function and productivity. The two blocks correspond to the moment conditions in (3.12) and (3.18), respectively. Under relatively mild regularity conditions, this two-stage nonparametric GMM sieve M-estimator is consistent and asymptotically normal (Chen, 2007; Chen et al., 2015).

4 Data

The micro-level data on dairy farms in Norway come from Tsionas et al. (2015) and are based on the Norwegian Farm Accountancy Survey administered by the Norwegian Agricultural Economics Research Institute. The survey includes the information on farm production and economic data collected annually from about 1,000 farms from different regions, farm size classes and types of farms. Participation in the survey is voluntary. There is no limit on the number of years a farm may be included in the survey. Approximately 10% of the farms surveyed are replaced every year. The farms are classified according to their main category of farming, defined in terms of the standard gross margins of the farm. Hence, the dairy farms in the considered sample are the farms, the largest share of the total standard gross margin of which is attributed to dairy production. The sample is an unbalanced panel with 4,667 observations on 902 farms observed during the 1998–2008 period.

Norwegian dairy farming is highly regulated (Jervell and Borgen, 2000). Throughout the past decades, various regulatory schemes have been set up to align aggregate milk production to domestic demand. A quota-based regulatory scheme was set up in 1983. From 1991, quotas of dairy farmers exiting the industry were used to reduce national milk supply, and there was no redistribution of quotas. The individual quotas have been reduced on several occasions in order to adjust total supply to domestic demand. The exit rate from the dairy sector was however slow for many years. Partly as a reaction to this outcome, a limited quota-trading scheme with quota prices defined by the government as well as the administrative reallocation of quotas was introduced in 1997. The objective of the change was to introduce greater flexibility to the quota system and to encourage structural changes in the sector. To maintain the regional distribution of production, the country was divided into milk-trade regions, and quota transfers were restricted to a given region only. The introduction of a quota-trading scheme in 1997 is likely to have led to structural changes in the production technologies by dairy farms in response to the deregulation. To mitigate the potential

Table 1. Summary Statistics for Norwegian Dairy Production, 1998–2008

Variable	Description and Units	Mean	Median	SD	2.5%	97.5%
Outputs						
Y_1	Milk, *Liters*	107,856.74	96,785.01	55,500.21	35,487.10	249,275.69
Y_2	Other Outputs, *Real EUR*	48,026.89	42,503.79	25,731.02	16,940.81	112,309.04
$Sales$	Total Sales, *Real EUR*	98,332.25	89,719.47	44,716.20	40,215.26	209,170.15
$Share_1$	Nominal Milk Output Share, *%*	51.73	51.99	7.56	34.63	65.71
Inputs						
N	Land, *Hectares (ha)*	24.66	22.00	12.78	8.60	56.83
L	Labor, *Hours (hr)*	3,858.57	3,700.00	1,226.59	1,985.95	6,756.34
M	Materials, *Real EUR*	29,623.93	26,027.21	15,943.81	9,828.12	67,710.67
K	Capital, *Real EUR*	28,992.50	25,726.11	15,969.56	9,865.54	69,873.07

Notes: SD is the sample standard deviation. 2.5% and 97.5% are the 2.5th and 97.5th percentiles of the empirical distribution, respectively.

presence of discontinuous shifts in the production function across the pre- and post-1997 periods, in this paper I opt to confine the analysis to the period from 1998 onwards.

Dairy farms are often involved not only in the production of milk but also in other farm production activities, such as the production of various types of meat, crop, etc. Following Sipiläinen et al. (2014) and Tsionas et al. (2015), I bundle all other non-milk products into an aggregate "other outputs" measure. Thus, the two outputs considered in this paper are: Y_1 – milk, measured in liters sold, and Y_2 – a single measure of all other outputs, which includes cattle and crop products. Since Y_2 includes several outputs, I measure it in monetary value terms, i.e., revenue from all these outputs. To convert the nominal value into the real terms, I first deflate Y_2 to real 2000 Norwegian Krones (NOK) using a weighted price index for cattle and crops, which I then convert from NOK to Euros (EUR) using the average exchange rate. Further, I specify the following four inputs: N – land, measured in hectares, L – own and hired farm labor, measured in hours, M – materials, including the cost of fertilizer, pesticides, preservatives, cost related to animal husbandry as well as purchased feed, and K – physical capital, which includes farm machinery. Similar to Y_2, materials and capital are measured in real 2000 EUR. All are deflated using a respective price index. Table 1 reports summary statistics for the variables. It also includes the real total sales ($Sales$), a measure of total output of a farm, as well as the share of milk products in the total nominal sales ($Share_1$).

5 Results

This section reports nonparametric sieve estimates of the firm-level production functions and productivity obtained using the proposed model, which explicitly formulates the multi-product nature of the production process by dairy farms in Norway. The results are obtained via the two-stage estimation procedure outlined in Section 3, which I estimate as a nonlinear system via multiple-equation nonparametric GMM.[15] In accordance with orthogonality conditions in (3.12) and (3.18), I use the following instruments (in logs). The spline basis functions of $(k_{it}, n_{it}, l_{it}, m_{it})$ are used for the first-stage material share equation (3.11). For the second-stage equation of the Markovian process of the productivity in (3.17), I use the spline basis functions of $(k_{it}, n_{it}, l_{it}, k_{it-1}, n_{it-1}, l_{it-1}, m_{it-1})$. The Hansen J-test of over-identifying restrictions produces a p-value of 0.99768 lending strong assurance to the model. As explained earlier, I select the order of approximation for sieves (in this

[15] I use the gradient-based BFGS algorithm to numerically locate the minimum of the GMM objective function.

Table 2. Production Function Estimates, 1999–2008

	Multi-Product				**Total Output**			
Elasticity	Mean	Q1	Median	Q3	Mean	Q1	Median	Q3
Land	0.386	0.310	0.367	0.454	0.276	0.249	0.276	0.303
	(0.002)				(0.001)			
Labor	0.165	0.139	0.182	0.218	0.149	0.122	0.155	0.174
	(0.002)				(0.002)			
Materials	0.290	0.267	0.295	0.317	0.290	0.267	0.295	0.317
	(0.001)				(0.001)			
Capital	0.274	0.224	0.267	0.324	0.252	0.194	0.237	0.282
	(0.002)				(0.002)			
RTS	1.118	1.041	1.121	1.201	0.969	0.891	0.961	1.053
	(0.003)				(0.004)			
Cross-Output	–0.759	–3.242	–1.358	–0.267				
	(0.132)							

Notes: Q1 and Q3 are the first and third quartiles of the empirical distribution, respectively. Bootstrap standard errors are in parentheses.

case, the number of knots) via data-driven generalized cross-validation. Also note that, since my identification strategy requires lags (of at least the first order) to be used in the second stage, in what follows I report the results for the 1999–2008 period, where the data for 1998 are used to instrument for endogenous variables in the year 1999.

In addition to the preferred model of the multi-product production (2.2) [hereby referred to as the "Multi-Product" model], I also estimate a more popular single-output production function given in (2.1). As is oftentimes done in the literature, here I use the deflated total sales ($Sales$) as an aggregate measure of the farm's total output. This single-output production function is also estimated in two stages via nonparametric GMM using cubic splines.[16] The estimation procedure closely follows Gandhi et al. (2013) and, essentially, is identical to the one developed in Section 3 except for the function $\widetilde{h}_t(\cdot)$ being suppressed in all equations. By estimating this single-output model [hereby referred to as the "Total Output" model], I am able to empirically assess the sensitivity of the production function and productivity estimates to the treatment of multiple outputs in the multi-product processes.

5.1 Production Function Estimates

Table 2 reports the summary of nonparametric estimates of the elasticities of the production function with respect to inputs, i.e., $\frac{\partial f_t(K_{it},N_{it},L_{it},M_{it})}{\partial a_{it}}$ $\forall\, a_{it} \in \{k_{it}, n_{it}, l_{it}, m_{it}\}$, for both models. Given that the elasticities do not have a closed-form expression owing to a nonparametric specification of the production function, I compute them using numerical derivatives of the estimated function $\widehat{f}_t(K_{it}, N_{it}, L_{it}, M_{it})$.

The empirical evidence suggests that the Total Output model, which aggregates all outputs into a single measure, tends to systematically *under*estimate the elasticity with respect to quasi-fixed land, labor and capital. The latter is also vividly demonstrated in Figure 1 which plots kernel densities[17] of these elasticities over all farm-years based on the two models. The distributions

[16] To ensure maximal comparability of the results from the two models, I employ the same instruments as the ones used for the Multi-Product model.

[17] The densities are estimated using the second-order Epanechnikov kernel with the optimal bandwidth parameters selected using the leave-one-out cross-validation.

Table 3. Rank Correlation Coefficients of RTS and Productivity Estimates across Models

	Returns to Scale		*Productivity*	
Year	Rank Corr.	SE	Rank Corr.	SE
1999	0.261	(0.047)	0.357	(0.047)
2000	0.291	(0.047)	0.331	(0.049)
2001	0.275	(0.045)	0.366	(0.049)
2002	0.284	(0.049)	0.387	(0.051)
2003	0.242	(0.049)	0.448	(0.045)
2004	0.268	(0.054)	0.437	(0.044)
2005	0.297	(0.050)	0.534	(0.035)
2006	0.368	(0.045)	0.467	(0.051)
2007	0.367	(0.050)	0.384	(0.055)
2008	0.396	(0.061)	0.467	(0.051)
1999–2008	0.286	(0.016)	0.405	(0.015)

Note: SE is the bootstrap standard error.

of estimates from the Total Output model are generally shifted leftward compared to those of estimates from the preferred Multi-Product model. The differences across the two models are the most distinct in the instance of the elasticity of land. The preferred Multi-Product model estimates the average elasticity of land at 0.386, whereas the conventional Total Output model produces a much lower average value of 0.276. Not only does the Multi-Product model predict the elasticity of production function with respect to land of higher magnitudes, but it also points to its significantly higher variability across individual farms as can be inferred based on the comparison of kernel densities in Figure 1.

I next examine the scale efficiency of dairy farms in Norway, where the former is defined as the firm's attainment of the efficient scale size associated with (unitary) constant returns to scale. The returns to scale measure is computed as the sum of estimated input elasticities, i.e., $\text{RTS}_{it} = \sum_{a\in\{k,n,l,m\}} \frac{\partial f_t(K_{it},K_{it},L_{it},M_{it})}{\partial a_{it}}$. The RTS measure of magnitudes less than/equal to/greater than one corresponds to decreasing/constant/increasing returns to scale. The results from the two models are summarized in Table 2.

While both models indicate the deviation from unitary returns to scale and hence point to some degree of scale inefficiency in the Norwegian dairy sector, the nature of this "inefficiency" is qualitatively different across the models. Specifically, while the Multi-Product model, on average, produces evidence in support of significant *increasing* returns to scale exhibited by dairy farms of magnitude at around 1.118, the competing Total Output model however fails to detect these scale economies and produces the average RTS estimate of 0.969 corresponding to *decreasing* returns to scale (i.e., diseconomies of scale). The dramatic differences in the returns to scale estimates across the two models are illustrated graphically in Figure 1. The results based on the proposed Multi-Product model are consistent with findings of increasing returns to scale enjoyed by dairy farms both in Norway and other European countries documented in the literature (e.g., Reinhard et al., 1999; Emvalomatis et al., 2011; Tsionas et al., 2015). Furthermore, not only does the Total Output model seems to grossly *under*estimate returns to scale across farms, but I also document very little correspondence between its estimates and those produced by the Multi-Product model. The left panel of Table 3 reports the Spearman rank correlation coefficients (along with bootstrap standard errors) for the returns to scale estimates from the two models for each year in the sample period. Here, I find a positive but rather weak association between estimates from the two models with the rank correlation coefficient ranging between 0.242 and 0.398.

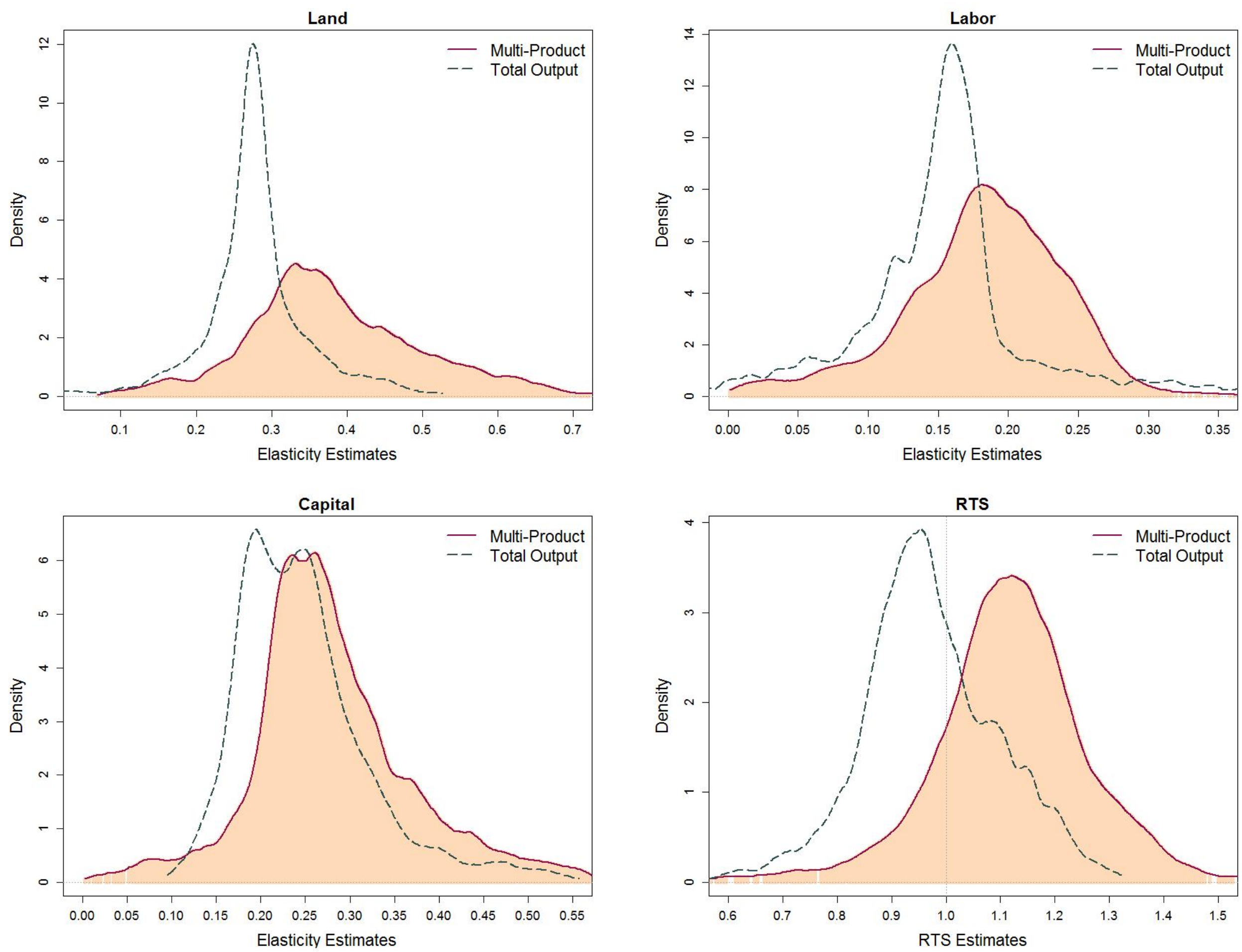

Figure 1. Kernel Densities of the Elasticity and RTS Estimates, 1999–2008

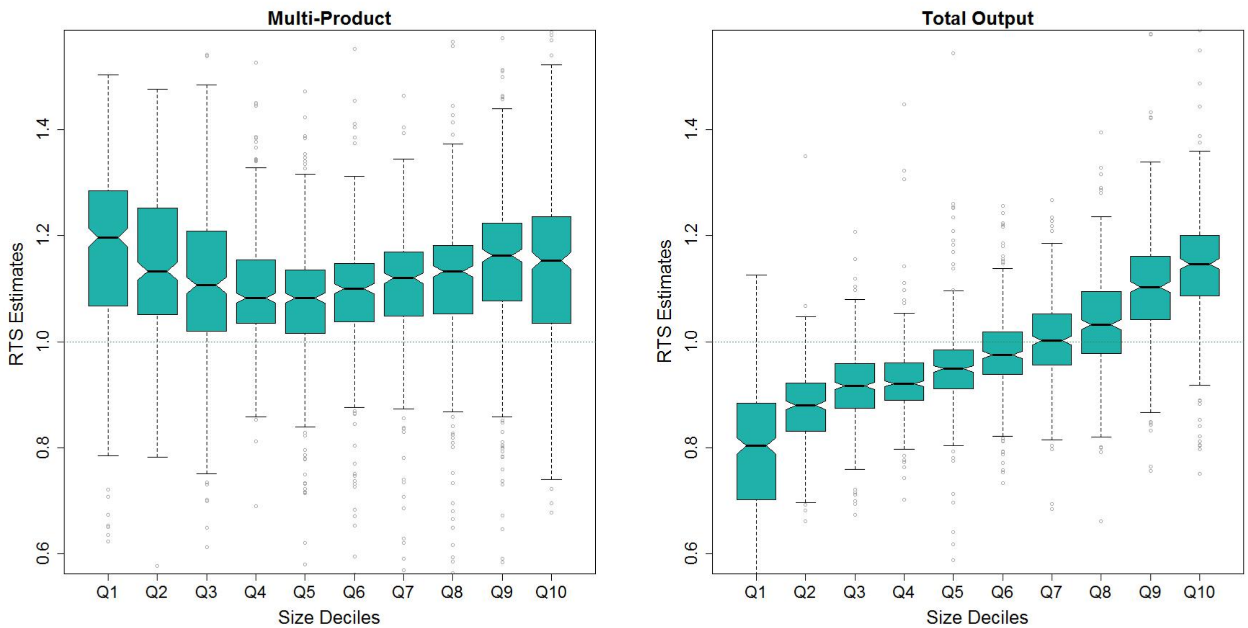

Figure 2. RTS vs. Farm Size, 1999–2008

It is also noteworthy that the two models point to rather different relations between returns to scale of a dairy farm and its size proxied by total sales/output. Normally, one would expect to see an inverse relationship between the two. Figure 2 shows distributions of the return to scale estimates across sample deciles of the farm size in the form of box-plots. The preferred Multi-Product model generally finds a weakly negative relation between scale economies and the size of a farm. There is an indication of an inverse relationship between the two amongst smaller farms in the bottom half of the size distribution (deciles 1 to 5). While the relationship seems to slightly reverse in the top half of the size distribution (deciles 6 to 9), the overall change in the median returns to scale from the first to tenth decile is still negative. In contrast, the estimates obtained from the Total Output model show a clear pattern suggesting that the magnitude of returns to scale is steadily increasing with the farm size, which many are likely to find somewhat hard to explain.

Before I proceed to the discussion of the productivity of dairy farms in the sample, I also consider the estimates of the cross-output elasticity $\frac{\partial y_{1,it}}{\partial y_{2,it}}$ from the preferred Multi-Output model. This elasticity can be intuitively interpreted as the technological "shadow price" capturing the technological trade-off between the production of two outputs *given* an input allocation and the level of productivity (along the firm's production possibilities frontier). Using implicit differentiation of the normalized multi-product production function in (3.3), I can compute the cross-output elasticity as follows:

$$\frac{\partial y_{1,it}}{\partial y_{2,it}} = \frac{\frac{\partial h_t(z_{2,it})}{\partial z_{2,it}}}{\frac{\partial h_t(z_{2,it})}{\partial z_{2,it}} - 1},$$

where elasticity $\frac{\partial h_t(z_{2,it})}{\partial z_{2,it}}$ is obtained using numerical derivatives of the estimated function $\widehat{h}_t(Z_{2,it})$. Table 2 reports the summary of these cross-output elasticity estimates. The empirical evidence confirms the trade-off between the two outputs at the statistically significant average "shadow price" of –0.759, suggesting that, on average, a 0.8% expansion in milk production can be attained by means of reducing other outputs by about 1% *ceteris paribus*. The estimation of the alternative single-output model however does not allow us to identify this technological metric for dairy production due to *a priori* bundling all outputs together.

5.2 Productivity Estimates

Following the literature (e.g., Olley and Pakes, 1996), I construct the estimates of farm-level productivity using the definition of the multi-product gross production function (2.2), i.e.,

$$\mathrm{P}_{it} = \exp\{\widehat{w}_{it} + \widehat{\eta}_{it}\},$$

where $\widehat{w}_{it}$ and $\widehat{\eta}_{it}$ are obtained in the first and second stages of the estimation procedure, respectively. Thus, in what follows I analyze the composite firm-level productivity defined as the exponentiated sum of both the persistent and random productivity. I first point to a rather weak association between productivity estimates obtained from the two models. The sample-wise rank correlation coefficient is estimated at 0.405 with year-specific coefficients falling in the range between 0.331 and 0.534 (see right panel of Table 3).

The second column of Table 4 reports the annual estimates of the average productivity P_t for all dairy farms in the sample. The average aggregate productivity is computed as the farm-total-output-weighted average of farm-level productivity measures for each year, i.e.,

$$\mathrm{P}_t = \sum_i \varpi_{it}\mathrm{P}_{it}, \quad \text{where} \quad \varpi_{it} = \frac{Sales_{it}}{\sum_j Sales_{jt}} \quad \forall \quad t.$$

Since comparison of absolute magnitudes of productivity estimates across the two models is somewhat meaningless,[18] I normalize P_t to unity in 1999. Along with the weighted-average productivity P_t, Table 4 also reports its corresponding annual productivity growth rates (see the first column) computed as the log difference, i.e., $d\log(\mathrm{P}_t) = \log(\mathrm{P}_t) - \log(\mathrm{P}_{t-1})$.

According to the estimates from the preferred Multi-Product model, the average annual productivity growth rate in the Norwegian dairy sector from 2000 to 2008 was around –0.12% per annum with some rather sharp, both up- and downward, fluctuations over the course of years, resulting in a cumulative nine-year decline of 1.1%. While the Total Output model also detects fluctuations in the productivity growth among dairy farms during the sample period, the 1999–2008 average annual productivity growth rate is however positive and is estimated to be at around 0.44% per year. While the finding of negative average productivity growth in dairy farming as suggested by the preferred model may seem somewhat perplexing, similar results have also been reported for other European countries, such as Poland (see Brümmer et al., 2002). Furthermore, a positive productivity growth in dairy sectors found for other European countries like Germany, Iceland and the Netherlands (Brümmer et al., 2002; Muluwork Atsbeha et al., 2012), which share an institutional environment similar to that of Norway, or even Norway itself (Sipiläinen et al., 2014) are likely to be biased and thus misleading. The cited studies model the dairy production process in the form of either input or output distance function while taking the endogeneity of inputs and outputs (due to simultaneity) for granted. Thus, those earlier estimates of productivity growth are likely to suffer from simultaneity biases. In contrast, the proposed model takes careful account of endogeneity of both inputs and outputs. Lastly, the finding of a negative trend in productivity among dairy farms in Norway may be (at least) partly attributed to the regulatory environment in the sector dominated by the Norwegian Agricultural Marketing Board that exercises a high degree of control over milk pricing via price discrimination, pooling arrangements and price supports

[18] Because the normalized multi-product and single-output production functions have different left-hand-side variables.

Table 4. Productivity Estimates

		Decomposition			*Percentile Ratios*		
Year	$d\log(\mathrm{P}_t)$	P_t	$\overline{\mathrm{P}}_t$	$\mathrm{P}_t - \overline{\mathrm{P}}_t$	75/25	90/10	95/5
			Multi-Product				
1999		1.000	0.997	0.002	1.147	1.339	1.465
2000	1.27%	1.012	1.010	0.002	1.124	1.296	1.457
2001	–2.67%	0.986	0.983	0.003	1.127	1.334	1.465
2002	–1.29%	0.973	0.970	0.003	1.117	1.310	1.467
2003	0.33%	0.976	0.971	0.004	1.165	1.366	1.510
2004	0.56%	0.982	0.978	0.003	1.154	1.358	1.523
2005	0.15%	0.983	0.975	0.007	1.179	1.383	1.568
2006	–0.95%	0.974	0.965	0.008	1.187	1.417	1.567
2007	–2.47%	0.950	0.941	0.008	1.175	1.404	1.582
2008	3.96%	0.989	0.979	0.009	1.198	1.406	1.627
1999–2008	–0.12%†				1.158	1.365	1.534
			Total Output				
1999		1.000	0.986	0.013	1.067	1.176	1.248
2000	–1.25%	0.987	0.972	0.014	1.069	1.158	1.241
2001	–1.26%	0.975	0.960	0.015	1.068	1.173	1.264
2002	0.10%	0.976	0.964	0.011	1.068	1.168	1.242
2003	–0.21%	0.973	0.963	0.010	1.067	1.166	1.238
2004	2.00%	0.993	0.983	0.010	1.077	1.158	1.228
2005	–0.39%	0.989	0.977	0.012	1.082	1.160	1.248
2006	0.59%	0.995	0.981	0.014	1.085	1.175	1.230
2007	0.26%	0.998	0.982	0.016	1.084	1.203	1.258
2008	4.17%	1.040	1.028	0.011	1.081	1.217	1.316
1999–2008	0.44%†				1.079	1.178	1.256

Notes: P_t is the normalized firm-output-share-weighted average estimate of firm-level productivity; $d\log(\mathrm{P}_t)$ is the first log difference in P_t; $\overline{\mathrm{P}}_t$ is the normalized simple average of firm-level productivity; $\mathrm{P}_t - \overline{\mathrm{P}}_t$ equals the sample covariance between $\overline{\mathrm{P}}_t$ and firm output.
† Annualized rate of growth.

(Brunstad et al., 2005). These market distortions may well discourage individual dairy farms from improving their efficiency and productivity. Furthermore, dairy farms in Norway may have even less incentive to improve productivity of the milk production in the face of the government-subsidized "multi-functionality" policies, where the latter refer to policies aimed to support non-agricultural aspects of dairy farming such as environmental conservation, cultural heritage tourism and the maintenance of countryside scenery.

Changes in the (weighted) aggregate productivity indices can be attributed to two primary sources: *(i)* a secular increase or decrease in productivity across firms in the industry and *(ii)* the reallocation of fixed factors towards more productive farms which would enable the latter to produce more output (Olley and Pakes, 1996). To differentiate between these two sources, I decompose the weighted aggregate productivity P_t into two components:

$$\begin{aligned}
\mathrm{P}_t &= \sum_i \varpi_{it}\mathrm{P}_{it} \\
&= \sum_i \left[\overline{\varpi}_t + (\varpi_{it} - \overline{\varpi}_t)\right]\left[\overline{\mathrm{P}}_t + \left(\mathrm{P}_{it} - \overline{\mathrm{P}}_t\right)\right] \\
&= \overline{\mathrm{P}}_t + \sum_i (\varpi_{it} - \overline{\varpi}_t)\left(\mathrm{P}_{it} - \overline{\mathrm{P}}_t\right) \quad \forall \quad t,
\end{aligned}$$

where $\overline{\mathrm{P}}_t = 1/n_t \sum_i \mathrm{P}_{it}$ and $\overline{\varpi}_t = 1/n_t$ are the unweighted average productivity and unweighted output share (a uniform weight), respectively. According to the above decomposition, the (aggregate) weighted average productivity P_t is a sum of the *un*weighted average of farm-level productivity $\overline{\mathrm{P}}_t$ and a sample covariance between the farm-level (total) output and productivity $\mathrm{P}_t - \overline{\mathrm{P}}_t = \sum_i (\varpi_{it} - \overline{\varpi}_t)\left(\mathrm{P}_{it} - \overline{\mathrm{P}}_t\right)$. While changes in the first component represent a secular trend in productivity, yearly changes in the covariance term capture the reallocation of economic activity from less productive to more productive farms. The larger the covariance term, the larger the (total) output share of more productive farms in the dairy sector.

The decomposition results are presented in Table 4 (columns 2 through 4). The Multi-Product model suggests that there is a steady increase in the covariance between farms' total output and productivity during the sample period, which indicates that the decrease in the weighted aggregate productivity P_t can be attributed entirely to a secular decrease in average productivity $\overline{\mathrm{P}}_t$ as opposed to the reallocation of resources towards less productive farms. The Total Output model draws a starkly different picture, whereby no reallocation of resources towards more productive dairy farms has largely taken place and the sector has rather enjoyed, with some fluctuations, an overall increase in secular productivity. The differences in the dynamics of secular (unweighted) average productivity across the two models are vividly shown in Figure 3, which yet again highlights the importance of explicit modeling of the multi-product nature of production processes.

I next study the distribution of productivity across farms. I begin by looking at the variation of productivity across farms, which I measure using the ratios of the 75th to 25th, 90th to 10th and 95th to 5th percentiles of the empirical productivity distribution. The last three columns of Table 4 report such productivity ratios for all years. According to the preferred Multi-Product model, I find that the dispersion of the productivity distribution has been increasing over the course of years. On average, a dairy farm at the 75th/90th/95th percentile was 16%/37%/53% more productive than a firm at the 25th/10th/5th percentile. In contrast, not only does the Total Output model provide a much weaker evidence of a widening productivity gap between farms on both ends of the distribution, but it also finds this productivity gap to be of a significantly smaller magnitude.

Lastly, I examine the relationship (if any) between the farm's productivity and its size proxied by total sales/output. I seek to assess this relationship both distribution-wise and on average. To

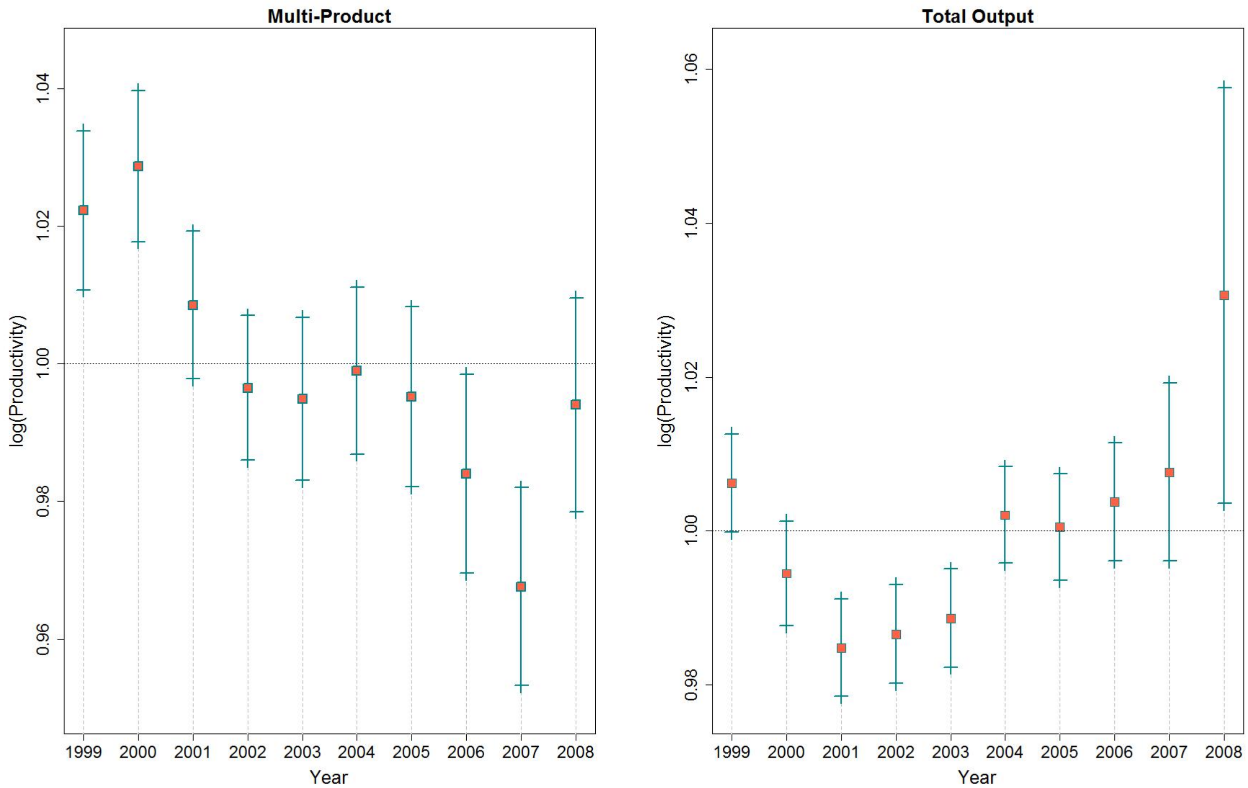

Figure 3. (Unweighted) Average Productivity with the 95% Confidence Bounds

accomplish the former, I estimate bivariate kernel densities[19] of farm-level productivity and total output (in logs) for both models. Figure 4 depicts a contour plot for these densities. The sub-figures suggest that, when the multi-product nature of dairy production is modeled explicitly (as in the Multi-Product model), the productivity appears to be distributed over farms of different size with *no* particular pattern. However, aggregating both outputs into one (as in the Total Output model) produces the results that are suggestive of a generally positive relationship between productivity and size. I also reaffirm this finding at the mean. I do so by estimating a nonparametric local-linear mean regression[20] of (logged) productivity on (logged) total output, which produces farm-year-specific estimates of gradients (marginal effects). I find that, in the instance of the Multi-Product model, the marginal effect of the farm size on its (mean) productivity is statistically insignificant for 82% of farm-years,[21] whereas in the case of the Total Output model, the effect is significantly positive for 99% of the sample.

To conclude, I document dramatic distortions in farms' input elasticities, returns to scale, and productivity estimates obtained from a traditional single-output specification of the production function, which *a priori* aggregates multiple outputs.

[19] I employ an axis-aligned bivariate Gaussian kernel, evaluated on a square grid using the normal reference bandwidth.

[20] I use the second-order Epanechnikov kernel with the optimal bandwidth parameters selected using the leave-one-out least-squares cross-validation.

[21] The remaining 18% of observations are split as follows: 15% are significantly greater than zero and 3% are significantly less than zero.

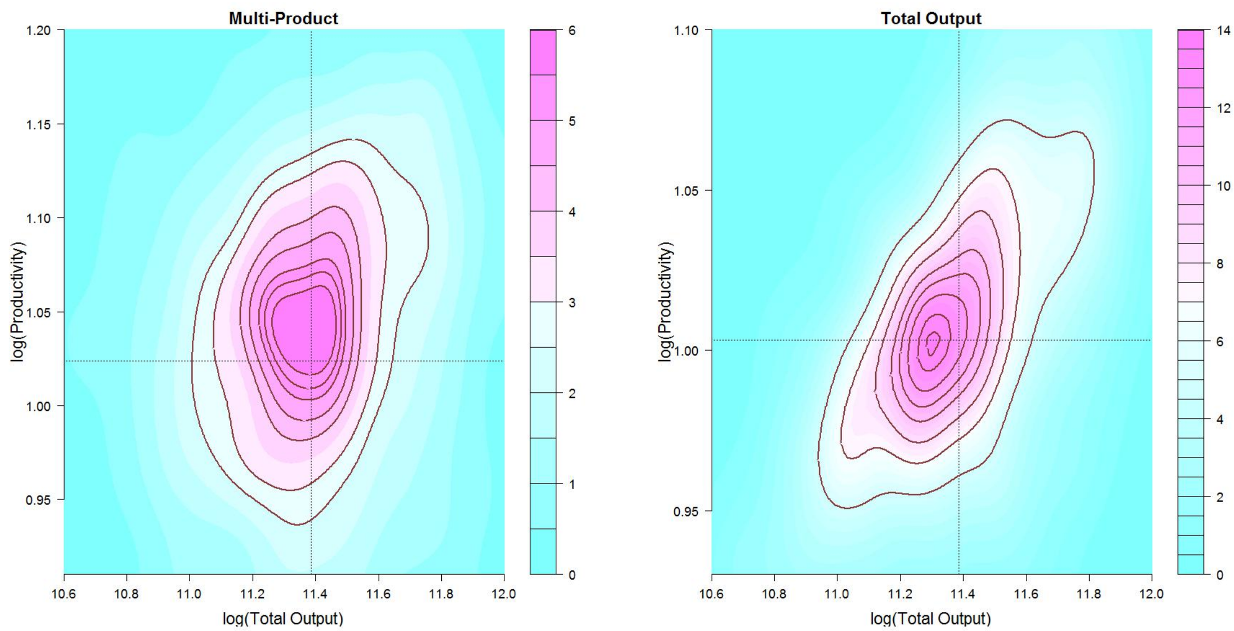

Figure 4. (Log) Productivity vs. Farm Size, 1999–2008

6 Conclusion

The existing control-function-based approaches to the identification of firm-level production functions are exclusively concerned with the estimation of *single*-output production functions despite that, in practice, most firms produce *multiple* outputs. While one can always opt to employ a single-product specification of the production process by *a priori* aggregating the firm's outputs, such a formulation is rarely an accurate portrayal of the firm's productive process.

This paper contributes to the literature by extending the proxy approach to the structural identification and estimation of firm-level production functions and productivity to the *multi*-product setting. Specifically, I consider the nonparametric estimation of multi-product stochastic production functions, where outputs form an unspecified (aggregate) index function. Such a treatment of multiple outputs differs from an *a priori* aggregation, which most studies using a traditional single-output production function normally resort to, in that it does not use *ad hoc* aggregation weights as well as allows the identification of cross-output elasticities representing the technological trade-off between individual outputs along the firm's production possibilities frontier. To avoid misspecification, I employ the nonparametric formulation not only for the control function but also for the production function itself, including the output index. I showcase the proposed methodology by applying it to study the multi-product production technology of Norwegian dairy farms during the 1998–2008 period.

References

Ackerberg, D. A., Benkard, C. L., Berry, S., and Pakes, A. (2007). Econometric tools for analyzing market outcomes. In Heckman, J. J. and Leamer, E. E., editors, *Handbook of Econometrics*, volume 6A. North Holland.

Ackerberg, D. A., Caves, K., and Frazer, G. (2006). Structural identification of production functions. Mimeo, University of California Los Angeles.

Amsler, C., Prokhorov, A., and Schmidt, P. (2015). Endogeneity in stochastic frontier models. *Journal of Econometrics*. forthcoming.

Arellano, M. and Bond, S. (1991). Some tests of specification for panel data: Monte Carlo evidence and an application to employment equations. *Review of Economic Studies*, 58:277–297.

Atkinson, S. E., Cornwell, C., and Honerkamp, O. (2003). Measuring and decomposing productivity change: Stochastic distance function estimation versus data envelopment analysis. *Journal of Business and Economics Statistics*, 21(2):284–294.

Atkinson, S. E. and Primont, D. (2002). Stochastic estimation of firm technology, inefficiency, and productivity growth using shadow cost and distance functions. *Journal of Econometrics*, 108:203–225.

Atkinson, S. E. and Tsionas, E. G. (2015). Directional distance functions: Optimal endogenous instruments. *Journal of Econometrics*. forthcoming.

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

Brümmer, B., Glauben, T., and Thijssen, G. (2002). Decomposition of productivity growth using distance functions: The case of dairy farms in three European countries. *American Journal of Agricultural Economics*, 2002:628–644.

Brunstad, R. J., Gaasland, I., and Vårdal, E. (2005). Efficiency losses in milk marketing boards – the importance of exports. *Nordic Journal of Political Economy*, 31:77–97.

Chambers, R. G. (1997). *Applied Production Analysis: A Dual Approach*. Cambridge University Press.

Chen, X. (2007). Large sample sieve estimation of semi-nonparametric models. In Heckman, J. J. and Leamer, E. E., editors, *Handbook of Econometrics*, volume 6B. North Holland.

Chen, X., Hahn, J., Liao, Z., and Ridder, G. (2015). Asymptotic properties of nonparametric two-step sieve estimates. Working Paper, University of California Los Angeles.

Craven, P. and Wahba, G. (1979). Smoothing noisy data with spline functions. *Numerische Mathematik*, 13:377–403.

Doraszelski, U. and Jaumandreu, J. (2013). R&D and productivity: Estimating endogenous productivity. *Review of Economic Studies*, 80:1338–1383.

Emvalomatis, G., Stefanou, S. E., and Oude Lansink, A. (2011). A reduced-form model for dynamic efficiency measurement: Application to dairy farms in Germany and the Netherlands. *American Journal of Agricultural Economics*, 93:161–174.

Färe, R. and Primont, D. (1995). *Multi-Output Production and Duality: Theory and Applications*. Kluwer Academic Publishers, Boston, MA.

Fernández, C., Koop, G., and Steel, M. F. J. (2000). A Bayesian analysis of multiple-output production frontiers. *Journal of Econometrics*, 98:47–79.

Fernández, C., Koop, G., and Steel, M. F. J. (2002). Multiple-output production with undesirable outputs: An application to nitrogen surplus in agriculture. *Journal of the American Statistical Association*, 97(458):432–442.

Fernández, C., Koop, G., and Steel, M. F. J. (2005). Alternative efficiency measures for multiple-output production. *Journal of Econometrics*, 126:411–444.

Gandhi, A., Navarro, S., and Rivers, D. (2013). On the identification of production functions: How heterogeneous is productivity? Working Paper, University of Wisconsin – Madison.

Greene, W. H. (2008). The econometric approach to efficiency analysis. In Fried, H. O., Lovell, C. A. K., and Schmidt, S. S., editors, *The Measurement of Productive Efficiency and Productivity Change*. Oxford University Press.

Griffiths, W. E. and Hajargasht, G. (2015). Some models of stochastic frontiers with endogeneity. *Journal of Econometrics*. forthcoming.

Griliches, Z. and Mairesse, J. (1998). Production functions: The search for identification. In *Econometrics and Economic Theory in the Twentieth Century: The Ragnar Frisch Centennial Symposium*, pages 169–203. Cambridge University Press.

Jervell, A. M. and Borgen, S. O. (2000). Distribution of dairy production rights through quotas: The Norwegian case. In Schwarzweller, H. K. and Davidson, A. P., editors, *Dairy Industry Restructuring*. Elsevier Science Inc., New York.

Kumbhakar, S. C. and Tsionas, E. G. (2015). The good, the bad and the technology: Endogeneity in environmental production models. *Journal of Econometrics*. forthcoming.

Kutlu (2010). Battese-Coelli estimator with endogenous regressors. *Economics Letters*, 2010:79–81.

Levinsohn, J. and Petrin, A. (2003). Estimating production functions using inputs to control for unobservables. *Review of Economic Studies*, 70:317–341.

Marschak, J. and Andrews, W. H. (1944). Random simultaneous equations and the theory of production. *Econometrica*, 12:143–205.

Muluwork Atsbeha, D., Kristofersson, D., and Rickertsen, K. (2012). Animal breeding and productivity growth of dairy farms. *American Journal of Agricultural Economics*, 94:996–1012.

Newey, W. K. (1984). A method of moments interpretation of sequential estimators. *Economics Letters*, 14(2):201–206.

O'Donnell, C. J. (2014). Econometric estimation of distance functions and associated measures of productivity and efficiency change. *Journal of Productivity Analysis*, 41:187–200.

O'Donnell, C. J. and Nguyen, K. (2013). An econometric approach to estimating support prices and measures of productivity change in public hospitals. *Journal of Productivity Analysis*, 40:323–335.

Olley, G. S. and Pakes, A. (1996). The dynamics of productivity in the telecommunications equipment industry. *Econometrica*, 64:1263–1297.

Pakes, A. (1994). Dynamic structural models, problems and prospects Part ii: Mixed continuous-discrete control problems, and market interactions. In Sims, C., editor, *Advances in Econometrics: Sixth World Congress*, volume 2. Cambridge University Press.

Reinhard, S., Lovell, C. A. K., and Thijssen, G. (1999). Econometric estimation of technical and environmental efficiency: An application to Dutch dairy farms. *American Journal of Agricultural Economics*, 81:44–60.

Sipiläinen, T., Kumbhakar, S. C., and Lien, G. (2014). Performance of dairy farms in Findland and Norway from 1991 to 2008. *European Review of Agricultural Economics*, 41:63–86.

Tran, K. C. and Tsionas, E. G. (2013). GMM estimation of stochastic frontier model with endogenous regressors. *Economics Letters*, 2013:233–236.

Tran, K. C. and Tsionas, E. G. (2015). Endogeneity in stochastic frontier models: Copula approach without external instruments. *Economics Letters*, 2013:233–236.

Tsionas, E. G., Kumbhakar, S. C., and Malikov, E. (2015). Estimation of input distance functions: A system approach. *American Journal of Agricultural Economics*. forthcoming.

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