University of Wisconsin-Madison
Department of Agricultural & Applied Economics

Staff Paper No. 545 March 2010

A Semi-Parametric Analysis of Technology, with an Application to U.S. Dairy Farms

By

Gustavos de los Campos, Jeremy Foltz, Jean-Paul Chavas
and Daniel Gianola

AGRICULTURAL &

APPLIED ECONOMICS

STAFF PAPER SERIES

A Semi-Parametric Analysis of Technology, with an Application to U.S. Dairy Farms

by

Gustavo de los Campos[1,2,3]

Jeremy Foltz[1],

Jean-Paul Chavas[1]

and

Daniel Gianola[2]

Abstract: This article proposes a semi-parametric stochastic frontier model (SPSF) in which components of the technology and of technical efficiency are represented using semi-parametric methods and estimated in a Bayesian framework. The approach is illustrated in an application to US farm data. The analysis shows important scale economies for small and medium herds and constant return to scale for larger herds. With the exception of labor, estimates of marginal products were close to the value expected under profit maximization. Finally, the results suggest important opportunities to increase productivity through reductions in technical inefficiencies.

[1] Department of Agricultural and Applied Economics, University of Wisconsin-Madison.

[2] Department of Animal Science, University of Wisconsin-Madison.

[3] Gustavo de los Campos is now at University of Alabama at Birmingham. Corresponding author: gdeloscampos@gmail.com 443 Ryals public Health Building, 1665 University Boulevard, Birmingham, AL 35294.

The study of economic properties of a technology and the measurement of technical efficiency are central themes in production economics. The production possibility set (PPS; Varian 1992) is the most general way of characterizing a technology. However, the determination of a PPS from a finite sample of production plans requires making assumptions about properties of the PPS (e.g., free-disposal or convexity) and of the nature of the production process (e.g., deterministic versus stochastic).

The method of stochastic frontiers (SF, e.g., Aigner Lovell and Schmith 1977) and data envelopment analysis (DEA, e.g., Charnes Cooper and Rhodes 1978; Varian 1984) provide two distinctive approaches for estimation of a PPS. DEA aims at estimation of a PPS based on minimal assumptions about it, but regards the production process as deterministic. This neglects any role played by measurement errors[1]. On the other hand, the method of SF's considers the possibility that measurement errors can affect production. However, standard applications require parametric assumptions about the boundary of the PPS and about the joint distribution of firm efficiencies and measurement errors.

This article contributes to the literature on production economics by presenting a multiple-output/multiple-input Bayesian semi-parametric stochastic frontier (SPSF) in which objects describing the technology and firm inefficiencies are modeled using semi-parametric methods. The SPSF model presented in this article extends the single-output

[1] For our purposes, in addition to truly measurement error, the term includes all factors not accounted for in the model.

SPSF presented by de los Campos (2009) to a multiple-output setting. The methodology is illustrated in an application to the technology of U.S. dairy farms.

The US Dairy industry have seen extensive structural and technological change in recent decades (e.g., Short 2004; Blayney et al. 2006; Miller and Blayney 2006; MacDonald *et al.* 2007). Since 1980, the number of dairy farms declined by 75% and the number of dairy cows dropped by 17%. As a result, herd size has increased steadily. At the same time, per-cow production increased substantially (about 50% in the last 20 years), and the number of farms and cows contracted whereas total milk production increased by a third. The changes in milk production have varied across regions. There has been an increased importance of Western states in which large and highly specialized dairy firms are commonly found. These structural changes are also seen in more traditional regions such as the Midwest or Northeast regions.

While specialization and concentration has been the dominant trend, pasture-based production systems (predominantly in the the Midwest and the Northeast, as discussed by Foltz and Lang 2005) and organic dairy farms (e.g., Barham Brock and Foltz 2006) have remained present as alternative production systems.

The increasing farm size may possibly be indicative of scale economies. Farm-level data (e.g., Short 2004; MacDonald *et al*. 2007) show that milk yield per cow increases with herd size. Although this may reflect different input usage, recent USDA's estimates indicate that average expenses per unit of milk decline with farm size (Short 2004; MacDonald *et al.* 2007). However, these reports are based on observed expenses

and not on estimates of a cost function, which means that they may be due to technology, efficiency or price effects, or a combination of all of these.

The most recent studies of the technology of the U.S. Dairy industry (Tauer and Misra 2005, and Mosheim and Lovell 2006) are based on the USDA Costs-Returns survey of 2000. The studies by Tauer and Mishra (2005) focus on milk only and did control for the effect of prices on the cost function. McDonald *et al.* (2007) argued that milk and beef are jointly produced, and that separation of costs associated to each activity is based on arbitrary assumptions. Using the same dataset, Mosheim and Lovell (2006) modeled the cost function of the whole farm using a parametric model in which the cost function was indexed by several prices and had an aggregate output index that considered both milk and beef production.

Our article complements the above literature by providing a characterization of the technology of U.S. dairy farms based on the most up-to-date available data. It considers milk and other outputs and uses semi-parametric Bayesian methods to model the technology and farm production efficiency of the U.S. dairy sector.

A Semi-Parametric Stochastic Frontier model for multiple Inputs/Outputs

The semi-parametric stochastic frontier (SPSF) discussed in this section is an extension of the single-output SPSF presented in de los Campos et al. (2009)[2]. Before describing the model a brief review of parametric stochastic frontier (SF) is given.

Standard Stochastic Frontier Model

The SF framework was proposed simultaneously by Aigner *et al.* (1977), Meeusen and van den Broeck (1977) and Battese and Corra (1977). When used to describe a single-output process, the output equation is $y_i = f(\mathbf{x}_i) + \delta_i + \varepsilon_i$, where; y_i is the output of the ith firm; $f(\mathbf{x}_i)$ is a production function, representing the expected maximum level of output given inputs $\mathbf{x}_i$; $\delta_i \leq 0$ is an inefficiency term, which models departures from the frontier due to firm-specific factors; and ε_i is a zero-mean random shock representing measurement errors.

Without additional assumptions about the shape of $f(\mathbf{x}_i)$ or about the joint distribution of $\{\delta_i, \varepsilon_i\}$, it is not possible to separate the effects of δ_i from ε_i, and the model suffers from an identification problem. Typically, identification is attained by making parametric assumptions about $f(\mathbf{x}_i)$ and about the joint distribution of $\{\delta_i, \varepsilon_i\}$

[2] de los Campos et al. (2009) also used simulations to evaluate the robustness of a SPSF with respect to changes in the assumptions about the data generating process.

(e.g., Aigner Lovel and Schmidt 1977; Meeusen and van den Broeck 1977; Stevenson 1980; Greene 1990).

The need for parametric assumptions about $f(\mathbf{x}_i)$ and about the joint distribution of $\{\delta_i, \varepsilon_i\}$ has been a source of criticism of the SF. Several authors have proposed to extend the SF model by relaxing these assumptions. For example, in Griffin and Steel (2004) a Dirichlet process (DP, e.g., Ferguson 1973; Antoniak 1974) was used to model, non-parametrically, the distribution of the δ_i's in a SF where $f(\mathbf{x}_i)$ was parametric. Alternatively, Fan Li and Weersink (1996) presented a model where the distribution of $\{\delta_i, \varepsilon_i\}$ is parametric, but $f(\mathbf{x}_i)$ is non-parametric. The model described next combines these ideas in a unified framework that can accommodate multiple outputs.

A multiple-output semi-parametric stochastic frontier (SPSF)

For simplicity, we describe a model for two outputs. This matches the application presented below where dairy farms produce two outputs: milk and other outputs. Extensions to more than two outputs are straightforward. The equations for the two outputs are:

$$\begin{aligned} y_{1i} &= f_1(\mathbf{x}_i) + \delta_{1i} + \varepsilon_{1i} \\ y_{2i} &= f_2(\mathbf{x}_i) + f_{21}(y_{1i}) + \delta_{2i} + \varepsilon_{2i} \qquad i = 1,\ldots,n. \end{aligned} \tag{1}$$

Above: $f_1(\mathbf{x}_i)$ is a semi-parametric function relating input variables, $\mathbf{x}_i$, to the first output; $\delta_{1i} + \varepsilon_{1i}$ is a two-term model residual, where the first term is non-positive,

$\delta_{1i} \leq 0$, and the second one satisfies $E(\varepsilon_{1i}) = 0$. The second equation is a multi-output production function, which is normalized on the second output. The first term $f_2(\mathbf{x}_i)$ captures the effects of inputs on the second output. The second term $f_{21}(y_{1i})$ captures the technology tradeoff between the two outputs. The functions $f_1(\mathbf{x}_i) = f_{1i}$, $f_2(\mathbf{x}_i) = f_{2i}$ and $f_{21}(y_{1i}) = f_{21i}$ are taken to be semi-parametric objects capturing patterns in the input-output and output-output relationships associated with the upper-bound of the feasible set representing the underlying technology.

The ε_i's are assumed to be Gaussian, independent and identically distributed (IID) across firms, $[\varepsilon_{i1}, \varepsilon_{i2}] \overset{IID}{\sim} N\left[\mathbf{0}, \begin{pmatrix} \sigma_1^2 & 0 \\ 0 & \sigma_2^2 \end{pmatrix}\right]$. The residual covariance matrix is assumed to be diagonal because the association between outputs that is not accounted for by input use is modeled by the recursion, $f_{21}(y_{1i})$. The above assumptions give the following likelihood function:

$$p(\mathbf{y}_1, \mathbf{y}_2 | \mathbf{f}_1, \mathbf{f}_2, \mathbf{f}_{21}, \boldsymbol{\delta}_1, \boldsymbol{\delta}_2, \sigma_1^2, \sigma_2^2) = \prod_{i=1}^{n} N(y_{1i} | f_{1i} + \delta_{1i}, \sigma_1^2) N(y_{2i} | f_{2i} + f_{21i} + \delta_{2i}, \sigma_2^2), \quad (2)$$

where: $\mathbf{y}_1 = \{y_{1i}\}$; $\mathbf{y}_2 = \{y_{2i}\}$; $\mathbf{f}_1 = \{f_{1i}\}$; $\mathbf{f}_2 = \{f_{2i}\}$; $\mathbf{f}_{21} = \{f_{21i}\}$; $\boldsymbol{\delta}_1 = \{\delta_{1i}\}$; and, $\boldsymbol{\delta}_2 = \{\delta_{2i}\}$. The likelihood function in (2) depends on several unknowns $\{\mathbf{f}_1, \mathbf{f}_2, \mathbf{f}_{21}, \boldsymbol{\delta}_1, \boldsymbol{\delta}_1, \sigma_1^2, \sigma_2^2\}$. In a Bayesian setting, a prior distribution is assigned to these unknowns. We structure the prior distribution in the following manner,

$$p(\mathbf{f}_1, \mathbf{f}_2, \mathbf{f}_{21}, \boldsymbol{\delta}_1, \boldsymbol{\delta}_1, \sigma_1^2, \sigma_2^2) = p(\mathbf{f}_1, \mathbf{f}_2, \mathbf{f}_{21}) p(\boldsymbol{\delta}_1, \boldsymbol{\delta}_1) p(\sigma_1^2) p(\sigma_2^2). \quad (3)$$

A standard choice of prior for the residual variances is the scaled inverse chi-square distribution, $p(\sigma_{.}^2) = \chi^{-2}(\sigma_{.}^2 | df_{.}, S_{.})$, with $df_{.}$ degrees of freedom and prior scale $S_{.}$[3]. Choosing small prior degrees of freedom reduces the influence of prior on inferences. The remaining components of the prior distribution, $p(\mathbf{f}_1, \mathbf{f}_2, \mathbf{f}_{21}) p(\boldsymbol{\delta}_1, \boldsymbol{\delta}_1)$, are described next.

Semi-parametric representation of functions using Gaussian processes. A long line of literature has showed how Gaussian processes can be used to describe functions semi-parametrically (e.g., Wahba 1990; Ruppert Wand and Carroll 2003; Shawe-Taylor and Cristianini 2004; Rasmussen and Williams 2006). In this approach, the vector containing the evaluations of a function $\mathbf{f}_{.} = [f_{.}(\mathbf{x}_1),..., f_{.}(\mathbf{x}_n)]'$ at points in the input space are viewed as a multivariate normal vector $\mathbf{f}_{.} \sim N(\mathbf{f}_{.}^0, \mathbf{K}\sigma_{f.}^2)$, where $\mathbf{f}_{.}^0 = [f_{.}^0(\mathbf{x}_1),..., f_{.}^0(\mathbf{x}_n)]'$, the mean vector, is used to represent components of the function that are not to be penalized (i.e., its estimation is based on the likelihood function only) and the (co)variance matrix $\mathbf{K}\sigma_{f.}^2 = \{K(\mathbf{x}_i, \mathbf{x}_j)\}\sigma_{f.}^2$ controls how much the evaluations of the function depart from the mean vector.

If $f_{.}^0(\mathbf{x}_i)$ is structured such that it represents a parametric model, e.g., $f_{.}^0(\mathbf{x}_1) = \mu. + \sum_j x_{ij}\beta_{.j}$, the parametric specification will appear as a special case with

[3] Here, and throughout this work, distributions are presented using the parameterizations described in Gelman Carlin Stern and Rubin (2004).

$\sigma^2_{f.} = 0$. Therefore, fitting the model with unknown σ^2_f versus $\sigma^2_f = 0$ provides a method for testing the parametric model. To keep the formulation general, one can use a set of q basis functions and let $f_.^0(\mathbf{x}_1) = \mu. + \sum_{k=`}^{q} \phi_k(\mathbf{x}_i)\beta_{.k}$, or in matrix notation, $\mathbf{f}_.^0 = \mathbf{1}\mu_. + \mathbf{\Phi}\boldsymbol{\beta}_.$. Here, $\{\phi_k(\mathbf{x}_i)\}_{k-1}^q$ are basis functions introduced to model the relationship between inputs and $f_.^0(\mathbf{x}_1)$. One can use standard basis functions of parametric models (e.g., polynomials, logarithm, exponential), or basis functions such as splines (e.g., Hastie 1992) to approximate the mean vector locally.

The evaluations of the unknown function are $f_.(\mathbf{x}_i) = f_.^0(\mathbf{x}_i) + \xi_.(\mathbf{x}_i)$, where $\xi_.(\mathbf{x}_i) = f_.(\mathbf{x}_i) - f_.^0(\mathbf{x}_i)$ represents deviates from the mean, $f_.^0(\mathbf{x}_i)$. The distribution of this deviates is $\boldsymbol{\xi}_. \sim N(\mathbf{0}, \mathbf{K}\sigma^2_{f.})$. Therefore, the (co)variance function $K(\mathbf{x}_i, \mathbf{x}_j)$, also known as the reproducing kenrel, defines a notion of smoothness of these deviations with respect to input space. The prior correlation for these deviates is, $Cor[\xi_.(\mathbf{x}_i), \xi_.(\mathbf{x}_j)] = \frac{K(\mathbf{x}_i, \mathbf{x}_j)}{\sqrt{K(\mathbf{x}_i, \mathbf{x}_i)K(\mathbf{x}_j, \mathbf{x}_j)}}$. Shawe-Taylor and Cristianini (2004) discuss several ways in which $K(\mathbf{x}_i, \mathbf{x}_j)$ can be structured. For example, given a distance function, $d(\mathbf{x}_i, \mathbf{x}_j)$ one can choose the kernel to be $K(\mathbf{x}_i, \mathbf{x}_j) = \exp\{-d(\mathbf{x}_i, \mathbf{x}_j)\}$, which gives higher prior correlation for observations that are close to each other, in the sense of $d(\mathbf{x}_i, \mathbf{x}_j)$.

Our model contains three functions, $f_1(\mathbf{x}_i)$, $f_2(\mathbf{x}_i)$ and $f_{21}(y_{1i})$. We consider

$$\begin{pmatrix} \mathbf{f}_1 \\ \mathbf{f}_2 \\ \mathbf{f}_{21} \end{pmatrix} \sim N\left[\begin{pmatrix} \mathbf{1}\mu_1 + \mathbf{\Phi}_x\boldsymbol{\beta}_1 \\ \mathbf{1}\mu_2 + \mathbf{\Phi}_x\boldsymbol{\beta}_2 \\ \mathbf{\Phi}_y\boldsymbol{\beta}_{21} \end{pmatrix}, \begin{pmatrix} \mathbf{G}\otimes\mathbf{K} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} \end{pmatrix}\right], \tag{4}$$

where $\mathbf{\Phi}_x$ and $\mathbf{\Phi}_y$ are incidence matrix containing basis functions evaluated in inputs vectors and in $\{y_{1i}\}$, respectively; $\mathbf{G} = \begin{pmatrix} \sigma^2_{f1} & \sigma^2_{f1f2} \\ \sigma^2_{f1f2} & \sigma^2_{f2} \end{pmatrix}$; and $\mathbf{K} = \{K(\mathbf{x}_i, \mathbf{x}_j)\}$.

The above distribution is indexed by several unknowns: $\{\mu_1, \mu_2, \boldsymbol{\beta}_1, \boldsymbol{\beta}_2, \boldsymbol{\beta}_{21}, \mathbf{G}\}$. In a Bayesian setting, a prior distribution needs to be assigned to these unknowns. We choose the prior of these unknowns to be proportional to a scaled inverted Wishart distribution with prior degree of belief df_G and prior scale $\mathbf{S}_G$. That is,

$$p(\mu_1\mu_2, \boldsymbol{\beta}_1, \boldsymbol{\beta}_2, \boldsymbol{\beta}_{21}, \mathbf{G}) \propto IW(\mathbf{G}|df_G, \mathbf{S}_G). \tag{5}$$

With this prior, there is no shrinkage of the parameters of the mean vector, $\{\mu_1, \mu_2, \boldsymbol{\beta}_1, \boldsymbol{\beta}_2, \boldsymbol{\beta}_{21}\}$. The Inverse Wishart is conjugate to the Gaussian prior where $\mathbf{G}$ enters, and one can choose df_G to be small (relative to sample size) in order to reduce the effect of the prior on inferences.

Modeling firm inefficiencies non-parametrically. Following Griffin and Steel (2004), we used a Dirichelet process, DP (e.g., Ferguson 1973; Antoniak 1974) to model the joint distribution of δ_{1i} and δ_{2i}, non-parametrically. For a discussion of DP, see Neal (1998). Briefly, DP's can be described as the limit (as the number of components goes to infinity) of a finite-mixture model. In finite-mixture models, the density function of a random variable (or vector), z_i, is represented as the weighted average of K components,

that is, $p\left(z_i \middle| \pi_1,...,\pi_K, \mathbf{\Psi}_1,...,\mathbf{\Psi}_K\right) = \sum_{j=1}^{K} \pi_j F_j\left(z_i \middle| \mathbf{\Psi}_j\right)$. Here, $\left(\pi_1,...,\pi_K\right)$ are mixing proportions satisfying $\sum_{j=1}^{K} \pi_j = 1$, and $F_j\left(z_i \middle| \mathbf{\Psi}_j\right)$ is the density of the jth mixture component. In many applications, these components are members of a parametric family indexed by some parameter vector ($\mathbf{\Psi}_j$). In this setting the parametric model appears as a special case of the finite mixture model, with K=1. This formulation gives flexibility to the probability model and allows approximation of the densities that may not be approximated well by standard parametric models.

In our case, using (1) and (2), and following Neal (1998), the conditional distribution of $\left\{y_{1i}, y_{2i}, \delta_{1i}, \delta_{2i}\right\}$ can be described as the limit, as K→∞, of the following hierarchy,

$$\begin{cases} p\left(\mathbf{y}_1, \mathbf{y}_2 \mid \mathbf{f}_1, \mathbf{f}_2, \mathbf{f}_{21}, \boldsymbol{\theta}_1, \boldsymbol{\theta}_2, \mathbf{c}, \mathbf{R}\right) = \prod_{i=1}^{n} N\left(y_{1i} \middle| f_{1i} + \theta_{1ci}, \sigma_1^2\right) N\left(y_{2i} \middle| f_{2i} + f_{21i} + \theta_{2ci}, \sigma_2^2\right) \\ p\left(\boldsymbol{\theta}_1, \boldsymbol{\theta}_2\right) = \prod_{j=1}^{K} F_0\left(\theta_{1j}, \theta_{2j}\right) \\ p\left(c_i \middle| p_1,..., p_K\right) = \prod_{i=1}^{n} \prod_{j=1}^{K} p_j^{1(c_i = j)} \\ p\left(p_1,..., p_K \middle| \alpha\right) = \prod_{j=1}^{K} p_j^{\frac{\alpha}{K}} \end{cases} \tag{6}$$

The first level of the above hierarchy is as in (2), with $\left\{\delta_{1i} = \theta_{1ci}\,;\, \delta_{2i} = \theta_{2ci}\right\}$. That is, conditional on a set of indicator variables $c_i \in \left(1,...,K\right)$ that link observations in the sample $\left\{i\right\}$ to components of a mixture $\left\{j\right\}$ and on the means of each of the mixture

components $\{\theta_{1j}, \theta_{2j}\}$, the δ's are replaced by the corresponding cluster means: $\{\delta_{1i} = \theta_{1ci}; \delta_{2i} = \theta_{2ci}\}$.

The second level gives the prior probability of the cluster means, $\{\theta_{1j}, \theta_{2j}\}$, which are independent draws from a base distribution, F_0. In our model F_0 is a half-bivariate-normal distribution with support in $\Re_- \times \Re_-$ and dispersion parameter $\mathbf{\Omega}$, denoted as, $N^-\left[\begin{pmatrix}\theta_{1j}\\ \theta_{2j}\end{pmatrix}\middle|\, \mathbf{0}, \mathbf{\Omega}\right]$. The last two levels give the probability model of the indicator variables that link observations to components of the mixture. In the last level, α is a concentration parameter, with α/K controlling how much the distribution of the inefficiencies depart from F_0. In (6), the influence of the prior on the posterior distribution of firm inefficiencies can be controlled by choosing $\mathbf{\Omega}$ and α. Choosing $\mathbf{\Omega}$ with large diagonal values (relative to the sample variance of the outputs) and small α reduces the influence of the prior.

Inferences using Bayesian Methods

In a Bayesian setting, inferences about model unknowns (and functions thereof) are made based on the distribution of the parameters given the data, a posterior distribution. Following Bayes' rule, the posterior distribution is, $p(\boldsymbol{\omega}|\mathbf{y}) = \frac{p(\mathbf{y}|\boldsymbol{\omega})p(\boldsymbol{\omega})}{p(\mathbf{y})}$, where: $p(\boldsymbol{\omega}|\mathbf{y})$ is the conditional distribution of model unknowns, $\boldsymbol{\omega}$, given data $\mathbf{y}$; $p(\mathbf{y}|\boldsymbol{\omega})$ is the conditional distribution of the data given the parameters (a likelihood function when viewed as a function of $\boldsymbol{\omega}$ for fixed $\mathbf{y}$); $p(\boldsymbol{\omega})$ is the joint prior distribution of model unknowns, and $p(\mathbf{y}) = \int p(\mathbf{y}, \boldsymbol{\omega})\partial\boldsymbol{\omega} = \int p(\mathbf{y}|\boldsymbol{\omega})p(\boldsymbol{\omega})\partial\boldsymbol{\omega}$ is the marginal distribution of the data. This last integral is typically difficult to compute. However, from the point of view of the posterior distribution $p(\mathbf{y})$ is just a constant of integration. Then, the posterior distribution is proportional to the product of the likelihood and of the prior distribution, $p(\boldsymbol{\omega}|\mathbf{y}) \propto p(\mathbf{y}|\boldsymbol{\omega})p(\boldsymbol{\omega})$.

Collecting the elements of the model previously described, given by equations (2)-(6), the posterior distribution of all unknowns in the SPSF model becomes,

$$
\begin{aligned}
p\left(\mu_1,\mu_2,\boldsymbol{\beta}_1,\boldsymbol{\beta}_2,\boldsymbol{\beta}_{21},\mathbf{f}_1,\mathbf{f}_2,\mathbf{f}_{21},\boldsymbol{\theta}_1,\boldsymbol{\theta}_2,\mathbf{c},\sigma_1^2,\sigma_2^2,\mathbf{G}\middle|\mathbf{y}_1,\mathbf{y}_2\right)\propto \\
&\prod_{i=1}^{n} N\left(y_{1i}\middle|f_{1i}+\theta_{1ci},\sigma_1^2\right)N\left(y_{2i}\middle|f_{2i}+f_{21i}+\theta_{2ci},\sigma_2^2\right) \\
&\times N\left[\begin{pmatrix}\mathbf{f}_1\\ \mathbf{f}_2\\ \mathbf{f}_{21}\end{pmatrix}\middle|\begin{pmatrix}\mathbf{1}\mu_1+\boldsymbol{\Phi}_x\boldsymbol{\beta}_1\\ \mathbf{1}\mu_2+\boldsymbol{\Phi}_x\boldsymbol{\beta}_2\\ \boldsymbol{\Phi}_y\boldsymbol{\beta}_{21}\end{pmatrix},\begin{pmatrix}\mathbf{G}\otimes\mathbf{K}_{\mathbf{X}} & \mathbf{0}\\ \mathbf{0} & \mathbf{0}\end{pmatrix}\right] \\
&\times\left\{\prod_{j=1}^{K} N^{-}\left[\begin{pmatrix}\theta_{1j}\\ \theta_{2j}\end{pmatrix}\middle|\mathbf{0},\boldsymbol{\Omega}\right]\right\}\left\{\prod_{i=1}^{n}\prod_{j=1}^{K} p_j^{1(c_i=j)}\right\}\left\{\prod_{j=1}^{K} p_j^{\frac{\alpha}{K}}\right\} \\
&\times IW\left(\mathbf{G}\middle|df_G,\mathbf{S}_G\right)\chi^{-2}\left(\sigma_1^2\middle|df_1,S_1\right)\chi^{-2}\left(\sigma_2^2\middle|df_2,S_2\right).
\end{aligned}
\tag{7}
$$

Although this distribution does not have closed form, a Gibbs sampler (with a Metropolis-Hastings step employed to implement the DP) can be used to draw samples from the above posterior distribution. Equation (7) is the SPSF we use below to estimate the production technology and its features.

Inferences about features of the production set and of firm inefficiencies. Posterior samples of $\left(f_{1i}, f_{2i}, f_{12i}, \delta_{1i}, \delta_{2i}\right)$ can be used to describe properties of the technology and to measure technical efficiency. For example, the second equation in (1), $y_{2i} = f_2\left(\mathbf{x}_i\right) + f_{21}\left(y_{1i}\right) + \delta_{2i} + \varepsilon_{2i}$ is a multi-output production function, and the posterior distribution of the δ_{2i}'s can be used to arrive at point estimates (e.g., posterior means) and measures of uncertainty (e.g., posterior standard deviations) of firm inefficiencies.

When y_1 and y_2 can be meaningfully added, the marginal product of aggregate output can be obtained from (1) as

$$
\frac{\partial\left(y_1+y_2\right)}{\partial x_j}=\frac{\partial f_1}{\partial x_j}+\frac{\partial f_2}{\partial x_j}+\frac{\partial f_{21}}{\partial x_j}=\frac{\partial f_1}{\partial x_j}+\frac{\partial f_2}{\partial x_j}+\frac{\partial f_{21}}{\partial y_1}\frac{\partial f_1}{\partial x_j}. \tag{8}
$$

In a non-parametric context, the above derivatives do not have closed form. However, the expression in (8) can be approximated using

$$\frac{\partial f_{.}}{\partial x_j} \approx \frac{\Delta f_{.}}{\Delta x_j} = \frac{f_{.}\big|_{\mathbf{x}_{-j}, x_j + k_2} - f_{.}\big|_{\mathbf{x}_{-j}, x_j - k_1}}{k_1 + k_2}, \tag{9}$$

and $$\frac{\partial f_{21}}{\partial y} \approx \frac{\Delta f_{21}}{\Delta f_1} = \frac{f_{21}\big|_{y_1 = E(y_1 | \mathbf{x}_{-j}, x_j + k_2)} - f_{21}\big|_{y_1 = E(y_1 | \mathbf{x}_{-j}, x_j - k_1)}}{E\left(y_1 \middle| \mathbf{x}_{-j}, x_j + k_2\right) - E\left(y_1 \middle| \mathbf{x}_{-j}, x_j - k_1\right)}, \tag{10}$$

for some small value of $k_1, k_2 > 0$. Here, $\mathbf{x}_{-j}$ denotes some fixed value for all inputs other than x_j, and $f_{.}\big|_{\mathbf{x}_{-j}, x_j + k_2}$ and $f_{.}\big|_{\mathbf{x}_{-j}, x_j - k_1}$ denote the evaluation of function $f_{.}$ at $\{\mathbf{x}_{-j}, x_j + k_2\}$ and $\{\mathbf{x}_{-j}, x_j - k_1\}$, respectively. In a Bayesian setting, (9) and (10) are stochastic because so are $\{\mu_1, \mu_2, \boldsymbol{\beta}_1, \boldsymbol{\beta}_2, \boldsymbol{\beta}_3, \mathbf{G}\}$. Draws from the posterior distributions of expressions (9) and (10) can be obtained by evaluating the expression from the posterior distributions of $\{\mu_1, \mu_2, \boldsymbol{\beta}_1, \boldsymbol{\beta}_2, \boldsymbol{\beta}_3, \mathbf{G}\}$. This allows us to obtain point-estimates and measures of uncertainty of marginal products.

Semi-parametric assessment of technology of US conventional dairy farming

In this section, the SPSF model just described, (7), is used to study the technology and assess technical the efficiency of US Dairy farms.

Data and models

The data come from the 2006 USDA-ARMS Dairy Costs and Returns Survey. A detailed description of sampling scheme and questionnaire used is available at the USDA-ARMS website (U.S. Department of Agriculture 2008). This survey covered 1,814 dairy farms (1,462 conventional and 352 organic). Our analysis focuses on the conventional dairy farms. The sample contained 6 firms with a very large number of COWS (more than 5,000). Due to the semi-parametric nature of the model, very little can be said about the technology at scales of operations that are beyond the sample information. Because of this, we focus the analysis on farms with up to 5,000 cows. Other reasons for removing firms from the data were the presence of missing values and of suspiciously values for some of the inputs/outputs variables. Only 3% of the data was discarded because of editing procedures. After editing, the data include 1,408 conventional dairy farms.

Variables. For ease of interpretation, variables are expressed relative to the number of COWS. Our model includes two outputs: Milk Yield (MY, U$S/COWS/year) and "other products" (OP, U$S/COWS/year), the latter including livestock products, manure and other outputs. Livestock products (primarily culled cows) were by far the most important item in OP (85%). Inputs included: purchased feed (PF, U$S/COWS/year), home-grown feed (HF, the opportunity cost of the home-grown feed, including grain, pasture, silage, etc. expressed in U$S/COWS/year); paid labor (PL, hours/week/COWS); unpaid labor and managerial time (UL, hours/week/COWS), and other costs (OC, including veterinary costs, farm overhead, etc., U$S/COWS/year). The

estimated value of building and equipments (B&E, U$S/COWS) was used as proxy for capital stock.

In the original data, outputs and inputs were valued at the prices faced by the farm. Expenses on each of the inputs or revenue from MY or OP may be different across farms because of technology, efficiency or price effects. Due to the spatial pattern in prices and the association between technology, firm size and regions (e.g., McDonald *et al.* 2007), it is difficult to separate price from technology and efficiency effects when netputs are valued at the prices faced by the firm. In order to avoid this problem, when separation of quantity and price was possible, netputs were valued at a single national average price. The price used to value outputs and inputs was the average price for the good of in question in the dataset.

Milk was priced at a single national price (no information about milk quality was available). Livestock was priced within category (heifers, replacement cows, replacement bulls, culled cows, culled bulls, calves), and the netput for this item was defined as sales - purchases + change in inventory. The remaining items of OP were as estimated by the Economic Research Service of the U.S. Department of Agriculture, i.e., valued at prices faced by the firm. This should not have a strong influence on results since MY was about 90% of total output, and of the remaining 10%, livestock products represented an average of 85%. Paid labor (PL) and unpaid labor (UL) were expressed in hours per milking COWS per week.

The original survey has information about quantities and expenses in purchased and home-grown feed. An attempt was made to express PF and HF as quantity indexes.

However, we did not pursue this idea as the quantity variables were of poor quality. On the other hand, OC and B&E are aggregates computed from a large number of expenses, for most of which prices were not available. Because of these reasons, PF, HF, OC and B&E were expressed as expenses rather than quantity indexes. Regional dummy variables were included in the model to control for systematic effects due to region.

Table 1 shows summary statistics of the variables included in the model. The average herd size was 393 COWS. The sample included a fairly wide range of herd sizes. In general all variables had large variability across firms. Milk production (MY) represented more than 90% of the total value of outputs of the dairy enterprise. All farms had some expenses on purchased feed (PF); however, some firms had no expenses on home-grown feed (HF). On average, paid labor (PL) was the most important source of labor in the farms; however, some farms used unpaid labor (UL) only. Note that for some farms other products (OP) was negative, these are farms acquire most of the replacement form the market and therefore OP becomes an input to the farm.

Table 1, about here

Table 2 shows summary statistics by region. The body of the table presents within-region means. And the last column gives the percentage of the sample variance of each of the variables that can be described as between-region sample variability from a standard analysis of variance. The Heartland and Northern-Crescent regions include some of the more traditional dairy production areas with a relatively large proportion of small

farms that depend on UL and HF heavily. The Fruitful Rim includes many Western counties where milk production has expanded recently, and has a relatively large concentration of large operations that depend on PF heavily, and make a relatively low use of labor, especially, very low use of UL.

Table 2, about here

Econometric Models

The analysis uses the model described by (7) with $\mathbf{\Phi}_x$ containing regional dummy variables and the basis function of an additive natural spline (e.g., Hastie 1992) with 3 *df* per input. Specifically, the mean vectors were represented as,

$$f_{01i} = \mu_1 + \sum_{j=1}^{5} \phi_{ij} \beta_{1j} + \sum_{j=6}^{12} \sum_{k=1}^{3} \phi_{jk}\left(x_{i,j-5}\right) \beta_{1,5+3(j-6)+k} = \mu_1 + \boldsymbol{\varphi}'_{x,i} \boldsymbol{\beta}_1$$

$$f_{02i} = \mu_2 + \sum_{j=1}^{5} \phi_{ij} \beta_{2j} + \sum_{j=6}^{12} \sum_{k=1}^{3} \phi_{jk}\left(x_{i,j-5}\right) \beta_{2,5+3(j-6)+k} = \mu_2 + \boldsymbol{\varphi}'_{x,i} \boldsymbol{\beta}_2$$

where: ϕ_{ij}'s ($j=1,\ldots,5$) are dummy variables for five regions as defined by the Economic Research Service of the U.S. Department of Agriculture (2008); and $\phi_{jk}(.)$ ($j=6,\ldots 13;$ $k=1,2,3$) is the kth basis function of the natural cubic spline associated to the ($j-5)^{th}$ input variable. The function relating the two outputs, was also structured the basis function of a natural spline with 3 df, that is, $f_{021i} = \sum_{k=1}^{3} \phi_k\left(y_{1i}\right) \beta_{21k} = \boldsymbol{\varphi}'_{y,i} \boldsymbol{\beta}_{21}$,

where, $\phi_k(y_{1i})$ is the corresponding basis function associated to y_{1i}.

To evaluate how well the mean vector describes the patterns relating inputs and outputs two models were fitted: in AM, standing for "additive model", with **G=0** in (7), therefore,

$$\begin{pmatrix} \mathbf{f}_1 \\ \mathbf{f}_2 \\ \mathbf{f}_{21} \end{pmatrix} = \begin{pmatrix} \mathbf{1}\mu_1 + \mathbf{\Phi}_x \boldsymbol{\beta}_1 \\ \mathbf{1}\mu_2 + \mathbf{\Phi}_x \boldsymbol{\beta}_2 \\ \mathbf{\Phi}_y \boldsymbol{\beta}_{21} \end{pmatrix}.$$

The second model (KM, standing for "kernel model") was fitted with unknown $\mathbf{G}$ and using a Gaussian kernel based on a standardized Euclidean distance, $K_X(i,j) = \exp\left\{ -\sum_{k=1}^{7} \frac{1}{V_k} \left\| x_{ik} - x_{jk} \right\|^2 \right\}$ where V_k is the sample variance of the kth input. With the Gaussian kernel, the prior correlation drops as points get further apart in input spaced, as measured by the standardized Euclidean distance.

Inference. Samples from the posterior distribution of the SPSF, (7), were obtained using a Gibbs sampler, with a Metropolis-Hastings step (e.g., Gelman Carlin Stern and Rubin 2004) used to draw samples of the components of the DP. The algorithm used to obtain the samples is as described in de los Campos (2009). Convergence to the posterior distribution was checked by inspecting trace plots of variance components. Inferences were based on posterior means, posterior standard deviations and highest-posterior density credibility regions (HPD, i.e., the minimum interval $[k_l, k_u]$ containing at least 95% of the samples from the posterior distribution) computed using 30,000 samples

obtained after discarding the 5,000 samples as burn-in (e.g., Gelman Carlin Stern and Rubin 2004).

Results and discussion

Table 3 shows the posterior means (standard deviations) of the variance components by model and output. The estimated residual variances $\left(\sigma_1^2, \sigma_2^2\right)$ were slightly smaller in the kernel model (KM) than in the additive model (AM), and the estimated diagonal elements of $\mathbf{G}$ $\left(\sigma_{f1}^2, \sigma_{f2}^2\right)$ were small relative to the sample variance of the outputs. These results suggest that the Gaussian process $\left(\mathbf{f}_1, \mathbf{f}_2, \mathbf{f}_{21}\right)$ was dominated by its mean $\left(\mathbf{f}_1^0, \mathbf{f}_2^0, \mathbf{f}_{21}^0\right)$, and that dispersion around this mean was very small. In agreement with these, we observed that the empirical correlations between the estimated posterior means of several items from AM and KM were extremely large. For example: $Cor\left(\hat{f}_{1i}\,;\,\tilde{f}_{1i}\right) = .991$; $Cor\left(\hat{f}_{1i} + \hat{\delta}_{1i}\,;\,\tilde{f}_{1i} + \tilde{\delta}_{1i}\right) = .995$; $Cor\left(\hat{f}_{2i} + \hat{f}_{21i}\,;\,\tilde{f}_{2i} + \tilde{f}_{21i}\right) = .960$; and, $Cor\left(\hat{f}_{2i} + \hat{f}_{21i} + \hat{\delta}_{2i}\,;\,\tilde{f}_{2i} + \tilde{f}_{21i} + \tilde{\delta}_{2i}\right) = .997$. Here, $\hat{\theta}$ and $\tilde{\theta}$ denote the estimated posterior means of θ under AM and KM, respectively.

All the above results suggest that an AM provides a good (local) approximation to the type of patterns observed in the data. In what remains of the article, we based the discussion on results from the AM.

Table 3, about here

As stated, the equation of OP is interpretable as a production function, and δ_{2i} can be used to assess technical efficiency. Figure 1 shows a histogram of the estimated posterior means of δ_{2i}. Only a few firms had a posterior mean of δ_{2i} close to zero (i.e., they are found to be efficient); most of the firms had a posterior mean of about -$550, this value is about 18,5% of the average revenue per COW ($2,971) observed in the firms included in the sample. Thus, we conclude there is much prospect for increasing productivity by reducing technical inefficiency. Given that our measure of technical efficiency is standardized on the first output (milk) and that most of OP correspond to livestock products, δ_{2i} may be reflecting differences across firms in management, reproductive performance and/or health status of the herd that are not associated to input use. Indeed, poor reproductive performance or high incidence of diseases such as mastitis or metabolic disease lead to high culling rates and increased needs of replacements (e.g., Lehenbauer and Oltjen 1998).

Figure 1, about here.

We now turn onto the analysis of response to inputs, labor and capital. Figure 2 shows the estimated expected total output, $\hat{f}_{1i} + \hat{f}_{2i} + \hat{f}_{21i}$ versus purchased feed (PF) and home-grown feed (HF). This was evaluated holding the remaining predictors at their mean values and under full efficiency (i.e., $\delta_{1i} + \delta_{2i} = 0$). The dashed vertical lines are

the .15, .50 and .85 sample quantiles of the predictor whose values are displayed in the x-axis. Expected total output increased almost linearly with PF. Response to expenses on HF was considerably smaller, and showed a concave-shape, with an upward trend up to 2,500U$S/COW and a downward pattern thereafter. However, within the range defined by the sample quantiles .15-.85 of HF, the response to expenses in HF was close to linear. The differences in response to expenses in PF and HF may reflect several factors, such as better quality per dollar of PF or a relatively low utilization of HF (which in large proportion involve pastures and silage). However, the response to HF may have been under-estimated if the actual cost of producing the food is smaller than the opportunity cost used to estimate its value.

Figure 2, about here

Figure 3 provides the counterparts of figure 2 for paid labor (PL), unpaid labor (UL), other costs (OC) and building and equipment (B&E). Labor, both paid and unpaid, had a relatively small effect on total output, with PL showing a concave relationship with output, with a maximum at around 1.2 hours/week/COW. Other things being equal, output showed a small decreasing trend as UL increased. The estimated response curve for OC was close-to-linear up to 500U$S/COW and showed decreasing marginal returns (MR) thereafter. A similar pattern was observed for B&E.

Figure 3, about here.

In our SPSF, capital, inputs and output variables were expressed in a per-cow basis, which means that, other things equal, the estimated response in total output to changes in COWS can be used as a measure of scale economies. A positive slope of the curve relating estimated expected outputs to levels of COWS indicates increasing returns to scale; absence of response to changes in COWS are indicative of constant returns to scale and a negative slope of the aforementioned curve is indicative of decreasing returns to scale. The estimated expected output per COW for different herd sizes (at the mean value of the remaining predictors and under efficiency) is depicted in figure 4. Other things being equal, output per cow increased as herd size did up to about 2,500 COWS. After this level the curve becomes almost flat, with a slightly decreasing trend. These results suggest increasing returns to scale over a fairly wide range of herd size and constant (or slightly decreasing) returns to scale for large herds. More than 90% of the firms in the sample had a herd size for which, locally, the technology exhibit increasing returns to scale.

The above results are in agreement with previous reports that either provides evidence suggesting scale economies (e.g., Jones 1999; Short 2004; MacDonald *et al.* 2007) or estimated it using econometric models (e.g., Mosheim and Lovell 2006). Overall, our results provide evidence in favor of variable-returns to scale. In this respect, our results are in agreement with Chavas (2001) who reported that the average cost function in agriculture in developed countries tends to be L shaped, i.e., increasing returns to scale for small-medium size operations, and constant returns to scale for large firm-size. Jones (1999), using data from Wisconsin, based on expenses and not an

estimate of the cost function, provides evidence suggesting that the region of increasing returns to scale covers only a small range of herd size (0-300 COWS). Our results indicate that scale economies are strong in this range of herd size; however, scale economies do not seem to dissipate over 300 cows. This difference may be due to the fact that our data, unlike the one used by Jones, has a national coverage and includes many large operations in Western states.

Tauer and Mishra (2005) regressed variable, fixed and total cost pre unit of milk on herd size using a SF model in which COW and dummy variables for region entered on a cost frontier and COW was also entered as an effect in the distribution of firm inefficiencies. Using this model, they found that COW was a significant factor determining firm inefficiencies but did not significantly affect the frontier.[4] They concluded that higher cost of production on many small farms is due to inefficiency rather than technology. However, these results, and their interpretation, are not comparable with ours in several respects. First, the model by Tauer and Mishra is not strictly a cost frontier since prices were not included in the model. Second the article by Tauer and Mishra focuses on milk only and ours account for other outputs as well. Most importantly, they include firm size as predictor in the objects that describe the technology and also as a covariate in the distribution of firm inefficiencies. In contrast, we take the point of view that input use and firm size are technological factors.

[4] They reported a significant effect of COW on the frontier for fixed cost, but not so for variable and total costs.

Marginal Products. Under the assumption of profit maximization, the marginal product (MP) should be equal to the marginal cost (MC) for each input. For the input variables expressed in U$S/COW/year (PF, HF, OC), this condition is achieved when the local slope of the revenue function is equal to one. For other variables, the evaluation of the condition requires considering the price of the unit in which the predictor is expressed. Our semi-parametric estimation of the production function allows the estimated marginal products to vary with respect to input use. In this context; we evaluated expressions (9) and (10) for each of the predictor variables at different quantiles at each iteration of the Gibbs sampler. These evaluations were used to estimate the posterior means and posterior standard deviation of marginal products (MP) that are presented in table 4. We also indicate with "stars" the cases where 1 (0) was not included in a 95% highest-posterior density (HPD) interval of MP_{PF}, MP_{HF} and MP_{OC} (MP_{PL}, MP_{UL}, $MP_{B\&E}$, MP_{COWS}).

The estimated MP_{PF} was close to 1 at the median and 0.9 quantile, and slightly smaller, but always positive at $q_{.10}$. A 95% HPD interval for MP_{PF} included 1 at $q_{.50}$, $q_{.90}$. For MP_{HF} the picture was reversed: MP_{HF} was higher at $q_{.10}$ (here the 95% HPD included one) and lower (and the HPD intervals did not include 1) at quantiles $q_{.50}$ and $q_{.90}$. Estimates of MP_{OC} were close to 1 at the .10 quantile, and below 1 for other quantiles.

The average hourly wage in the data was U$S 9. At this rate, an increase in one hour/week/cow implies an additional annual cost of 468 U$S/COW (9×52). The posterior means of MP_{PL} and of MP_{UL} are clearly smaller than this value; however the most

important feature of the posterior distribution of MP_{PL} and of MP_{UL} is the large posterior SD. In almost all cases HPD intervals included zero.

The posterior means of the $MP_{B\&E}$ were positive at all quantiles, and results suggest diminishing $MP_{B\&E}$. The estimated posterior means at different quantiles ranged from .08 to .02. The condition for profit maximization for this item is that MP equals depreciation rate; these values are reasonably close to depreciation rates for fixed capital.

Finally, the marginal response to an increase in herd size was positive at the three quantiles, but estimates are suggestive of diminishing marginal returns. This suggests important economies of scale for small and medium size herds, and constant returns to scale for large herds. Other things being equal, i.e., keeping fixed expenses and capital (in a per-COW basis), increasing herd size by ten COWS at the median (170 COWS) gives an expected increase in net benefit for the total herd of $2,110/year.

Table 4, about here.

The estimated posterior means of the marginal products (MP) obtained with the kernel model (KM, not presented) were similar to those in table 4. However, this model yielded considerably higher levels of posterior uncertainty. This happens because the introduction of random deviates from the mean vector increased the local variability of the production surface. Due to this fact, the quality of the approximation to MPs may be poorer.

Concluding Remarks

We proposed a multiple-output semi-parametric stochastic frontier (SPSF) model in which components of the technology and of technical efficiency are described. The methodology includes parametric models as special cases. As such, our proposed framework provides a more flexible method than parametric models, and can be used to test the validity of parametric assumptions.

The analysis was presented in a multi-output context, and applied to the assessment of technology on U.S. dairy farms. Using the USDA-ARMS 2006 Dairy survey, we investigated how the patterns of production practices (including input use, herd size and output levels) vary across farms and regions and documented how small farms tend to depend on unpaid labor and home-grown feed. In contrast, large farms are more specialized on the input side (i.e., they rely more heavily on purchased feed), use less labor, and make little use of unpaid labor. The data also confirm the association between region, herd size and the use of some inputs (especially paid labor and home-grown feed), as reported in previous studies.

With the exception of labor, all input and capital variables showed positive and diminishing marginal returns, with estimates of marginal products that are close to what is expected under profit maximization. Also, our results suggest the existence of strong scale economies for small and medium size herds, and approximately constant returns to scale for large herds. This is in agreement with previous studies. It is also consistent with the increased size of dairy farms observed over the last decade. However, in contrast to

most previous literature, our results indicate that the region of increasing returns to scale spans over a fairly wide range of herd size, including what one may now consider medium size herds.

Finally, our results detected significant technical inefficiency. This means that there are important opportunities for increasing productivity by improving management and the efficiency of production. Moreover, our results are in agreement with previous studies (e.g., Lehenbauer and Oltjen 1998) reporting that high culling rates (due to reproductive problems and high incidence of diseases) is one of the main challenges faced by conventional dairy farms in the U.S.

ACKNOWLEDGMENTS

The authors wants to thank Robert Battaglia of the US Department of agricluture for his collaboration in the prearation of the agreement that gave access to data form the 2006 Cost and Returns Survey, and for providing facilities to process the data. Financial support by the Willard F. Mueller Fellowship, the Wisconsin Agriculture Experiment Station and grant DMS-NSF DMS-044371 is acknowledged.

Literature Cited

Aigner D., C.A.C, Lovell and P. Schmidt. 1977. "Formulation and estimation of stochastic frontier production function models." *Journal of Econometrics* 6: 21-37.

Antoniak C.E. 1974. "Mixture of Dirichlet processes with applications to Bayesian nonparametric problems." *The Annals of Statistics* 2:1152-1174.

Batesse G. E. and G.S. Corra. 1977. "Estimation of a production frontier model: with application to the pastoral zone of eastern Australia." *Australian Journal of Agricultural Economics*, 21: 169-179.

Blayney D., M. Gehlhar, C.H. Bolling, K. Jones, S. Langley, M.A. Normile and A. Somwaru. U.S. 2006. "Dairy at global crossroads". U.S. Department of Agriculture, Economic Research Service, research report No 28. Available at: http://www.ers.usda.gov/publications/err28/err28.pdf

Braham B.L., C. Brock, and J. Foltz. 2006. "Organic dairy farms in Wisconsin: prosperous, modern and expansive." PATS, research report No. 16. Available at: http://www.pats.wisc.edu/pdf%20documents/Organic_Dairy_Farms_in_Wisconsin_6_16_06%20with%20executive%20summary%20and%20front%20cover2.pdf

Charnes A., W.W. Cooper, and E. Rhodes. 1978. "Measuring the inefficiency of decision making units." *European Journal of Operational Research* 2: 429-444.

Chavas J.P. 2001. "Structural change in agricutlural production." In: Bruce L. Gardner and Gordon C. Rausser, ets., *Handbook of agricultural economics*, V1A, Agricultural production, Amsterdam: Elsevier, pp. 263-285.

de los Campos, G. 2009. "Semi-parametric methods with applications to quantitative genetics and production economics". PhD Dissertation, University of Wisconsin-Madison.

Economic Research Service, U.S. Department of Agriculture. 2008. "Farm resource regions". Available at: http://www.ers.usda.gov/Publications/aib760/aib-760.pdf

Fan Y., Q. Li and A. Weersink. 1996. "Semiparametric estimation of stochastic production frontier models." *Journal of Business & Economic Statistics*, American Statistical Association, 14:460-68, October.

Ferguson T.S. 1973. "A Bayesian analysis of some nonparametric problems." *The Annals of Statistics* 1:209-230.

Foltz, Jeremy D. and G. Lang. 2005. "The Adoption and Impact of Management Intensive Rotational Grazing (MIRG) on Connecticut Dairy Farms." *Renewable Agriculture and Food Systems*. 20: 261-266.

Gelman, A., J.B. Carlin, H.S. Stern and D.B. Rubin. 2004. *Bayesian data analysis*. Chapman & Hall, London.

Greene, W. 1990. "A Gamma distributed stochastic frontier model." *Journal of Econometrics*, 1990, 46: 141-164.

Griffin J.E. and M.F.J. Steel. 2004. "Semiparametric Bayesian inference for stochastic frontier models." Journal of Econometrics 123:121-152.

Hastie T. J. 1992. *Generalized additive models*. Chapter VII of Statistical models in S eds J. M. Chambers and T. J. Hastie, Wadsworth & Brooks/Cole

Jones B. 1999. “Growth in dairy farms: the consequences of taking big steps or small ones when expanding.” In: Four state dairy extension conference, Wi and St. Paul, MN.

Lehenbauer T.W. and J.W. Oltjen. 1998. “Dairy cow culling strategies: making economic culling decisions.” *J Dairy Sci*.: 81:264-271.

MacDonald J.M., E.J. O’ Donoghue, W.D. McBride, R.F., Nehring, C.L. Sandretto, and R. Mosheim. 2007. “Profits, costs and the changing structure of dairy Farming”. U.S. Department of Agriculture, Economic Research Service, research report No. (ERR-47). Available at: http://www.ers.usda.gov/Publications/ERR47/

Meeusen W. and J. van den Broeck. 1977. “Efficiency estimation from Cobb-Douglas production function with composed error.” *International Economic Review* 8:435-444.

Miller J.J. and D. Blayney. 2006. “Dairy Backgrounder”. U.S. Department of Agriculture, Economic Research Service. Available at: http://www.ers.usda.gov/Publications/LDP/2006/07Jul/LDPM14501/

Mosheim, R. and C.A.K. Lovell. 2006. “Economic efficiency, structure and scale economies in the U.S. dairy sector.” In: Annual meetings of the American Economic Association, Long Beach, CA.

Neal R. 1998. “Markov chain sampling methods for Dirichlet Process mixture models.” *Journal of Computational and Graphical Statistics*, 9:249-265.

Rasmussen C. E. and C.K.I. Williams. 2006. *Gaussian processes for machine learning (adaptive computation and machine learning)* MIT Press, Cambridge, MA.

Ruppert D., M.P. Wand, and R.J. Carroll. 2003. *Semiparametric regression*. Cambridge Series in Statistical and Probabilistic Mathematics.

Shawe-Taylor J. and N. Cristianini. 2004. *Kernel methods for pattern analysis*. University Press Cambridge, UK.

Short S.D. 2004. "Characteristics and production costs of U.S. dairy operations". U.S. Department of Agriculture, Economic Research Service, Statistical Bulletin No SB974-6. Available at: http://www.ers.usda.gov/publications/sb974-6/sb974-6.pdf

Stevenson R. 1980. "Likelihood functions for generalized stochastic frontier estimation." *Journal of Econometrics*, 1980, 13, pp. 57-66.

Tauer, L.W. and A.K. Mishra. 2005. "Can small dairy farm remain competitive in the U.S. agriculture? " *Food Policy* 31:458-468.

U.S. Department of Agriculture. 2008. "Agricultural resource management survey, Phase III". Available at: http://www.ers.usda.gov/Data/ARMS/FarmsDocumentation.htm

Varian H.R. 1984. "The nonparametric approach to production analysis." *Econometrica* 52: 579-97.

Varian H.R. 1992. *Microeconomic analysis* 3rd Edn. W.W. Norton & Company, New York.

Wahba G. 1990. *Spline models for observational data*. SIAM, Philadelphia, PA.

Tables and Figures

Table 1. Descriptive Summary of Input/Output Variables.

Name	Description	Units	Mean	SD[a]	$q_{.05}$ [a]	$q_{.95}$ [b]
Milking COWS	Stock	Heads	393	603	35	1,559
Milk Yield (MY)	Sales	U$S/COWS/year	2,721	793	1,387	3,968
Other products (OP)	Livestock and other	U$S/COWS/year	246	243	-171	637
Purchased Feed (PF)	All purchased feed	U$S/COWS/year	851	431	164	1,606
Home-Grown Feed (HF)	Grain, silage, grass	U$S/COWS/year	640	567	0	1,634
Paid Perm. Labor (PL)	---	Hr/COWS/week	0.35	0.41	0.00	1.08
Unpaid Labor[b] (UL)	---	Hr/COWS/week	0.67	0.77	0.03	2.23
Other costs (OC)	---	U$S/COWS	719	326	274	1,310
Build&Equip. (B&E)	Value of capital stock	U$S/COWS	2,780	2,433	296	7,141

a: SD=standard deviation; $q_{.05}$ and $q_{.95}$ denote the .05 and .95 sample quantiles, respectively; b: Includes manager's time.

Table 2. Within-Region Means and Between-Regions Variability.

	East. Uplands (n=176)	Fruitful Rim (n=276)	Heart-land (n=255)	North. Crescent (n=463)	South. Seaboard (n=146)	Other (n=92)	Variance associated to Region (%)[a]
	Regional Average						
Number of Milking COWS	145	698	227	297	312	1028	16.8
Milk Yield [b]	2229	2877	2689	2896	2528	2702	7.8
Other Products [b]	238	230	273	274	177	205	1.8
Purchase Feed [b]	718	1183	653	745	947	1033	20.3
Homegrown Feed [b]	646	426	690	857	476	305	10.9
Paid Labor [c]	0.386	0.396	0.306	0.520	0.461	0.330	3.5
Unpaid Labor [c]	0.895	0.348	0.925	0.770	0.491	0.289	9.3
Other Costs [b]	603	642	751	836	687	538	9.2
Bulding & Eq. [b]	3282	1615	3405	3415	2186	1332	11.4

a: Between-region variance as percentage of the overall sample variance of the corresponding variable computed from standard analysis of variance (ANOVA); b: in U$S/COWS/year; c: in hours/week/COWS.

Table 3. Estimates of Posterior Means (Standard Deviations) of Variance Components by Output and Model.

Model:	Milk Yield (U$S/COW)		Other Products (U$S/COW)	
	σ_1^2	σ_{f1}^2	σ_2^2	σ_{f2}^2
(In the metric of the output)				
AM	241,680 (24,710)	---	17,050 (2,254)	---
KM	230,838 (22,085)	32,947 (11,606)	15,525 (2,233)	2,190 (772)
(Relative to sample variance of the output)[a]				
AM	38.39%	---	28.90%	---
KM	36.67%	5.23%	26.31%	3.71%

Note: σ_1^2 and σ_2^2 are residual variances of the corresponding models; AM is a 'fixed effects' model, KM adds, relative to AM, random deviations from the mean vector of AM that follow a Gaussian process whose variances are σ_{f1}^2 and σ_{f2}^2, for outputs one and two, respectively.

a: The sample variance of Milk Yield and Other products were 629,564 and 59,002, respectively (both in U$S/COW^2).

Table 4. Posterior Mean (SD) of Marginal Product (MP) of Each of the Predictor Variables, at Selected Quantiles (Additive-Model).

Predictor	Quantiles[a]		
	.10	.50	.90
Purchase Feed (U$S/COW)	.656* (.129)	.885 (.106)	.926 (.078)
Home-Grown Feed (U$S/COW)	.860 (.180)	.338* (.062)	.182* (.050)
Paid Labor (Hours/week/COW)[b]	183.2* (68.3)	24.3 (61.3)	-172.8 (352.0)
Unpaid Labor (Hours/week/COW)[b]	---	-97.6 (93.7)	-84.6 (42.9)
Other Costs (U$S/COW)	1.118 (.190)	.700* (.119)	.471* (.074)
Building & Eq. (U$S/COW)	.081 (.056)	.040* (.015)	.019 (.012)
Number of Milking Cows	2.874* (.792)	1.242* (.270)	.297* (.086)

a: The MPs are evaluated at the .10, .50 and .90 sample quantiles of the distribution of the corresponding predictors, with all other predictors set at the mean-value; b: Unpaid labor also includes manager's time. The average weekly wage of paid labor in the data was U$S9 per hour; *: Indicates that 1 (for Purchased Feed, Home-Grown Feed and Other Costs) or 0 (for the remaining variables) does not belong to a 95% HPD interval.

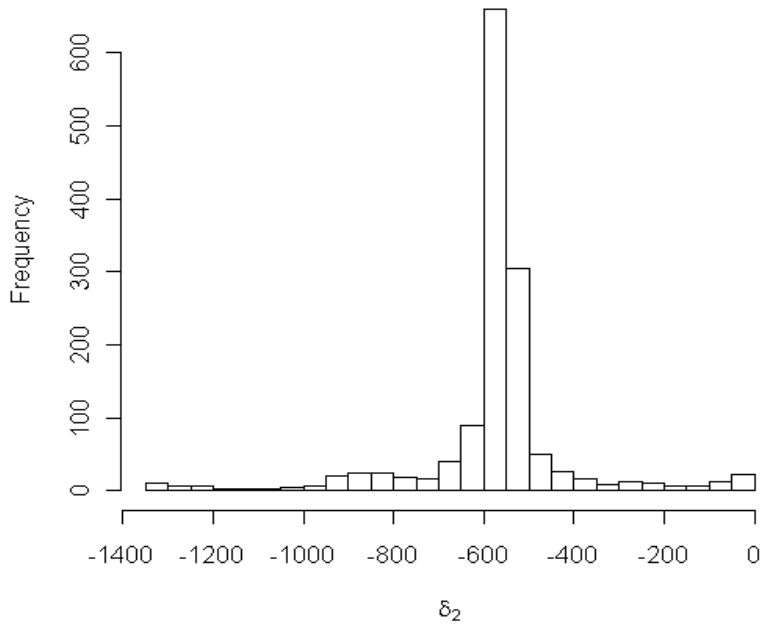

Figure 1. Histogram of posterior means of a measure of technical efficiency (in U$S/COW, standardized with respect to milk production and measured in the direction of products other than milk).

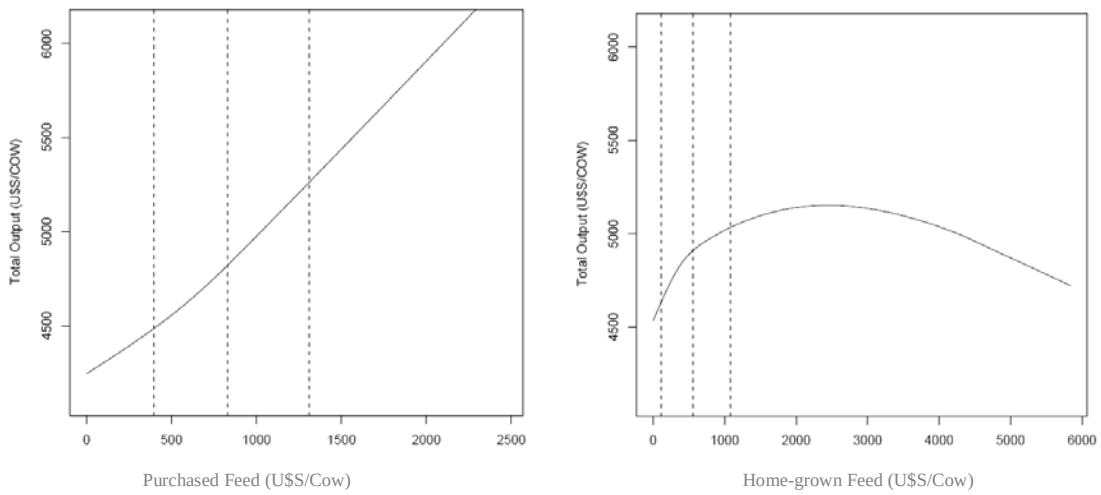

Figure 2. Estimated expected total output under efficiency, $\hat{f}_{1i} + \hat{f}_{2i} + \hat{f}_{21i}$, versus expenses in purchased (PF, left) and home-grown feed (HF, right). The curves display the estimated posterior means of $\hat{f}_{1i} + \hat{f}_{2i} + \hat{f}_{21i}$ evaluated at different values of PF and HF and at the mean-value of the remaining predictors. The dashed lines are sample quantiles (.15, .50, .85) of the variable in the horizontal axis.

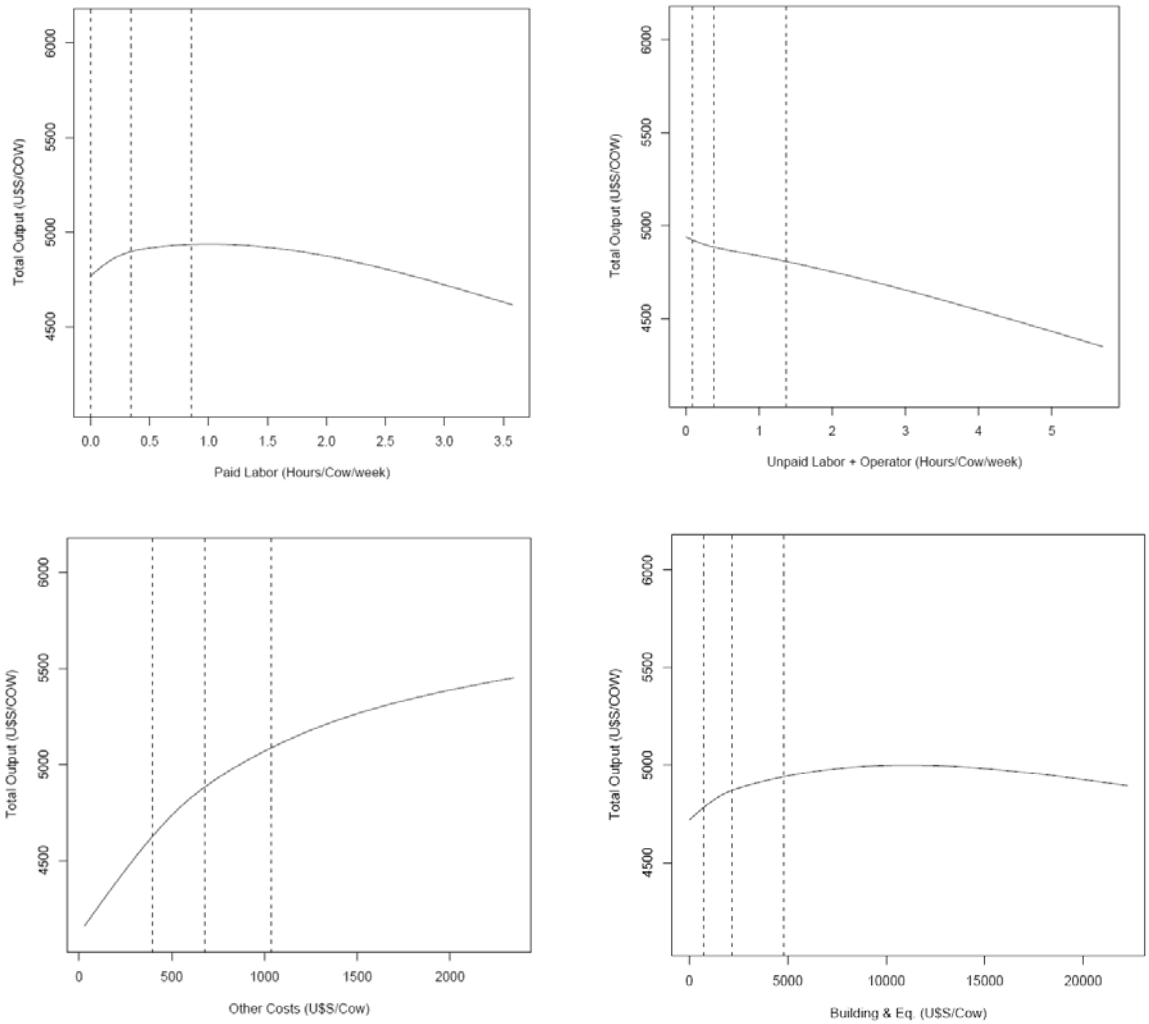

Figure 3. Estimated expected total output under efficiency, $\hat{f}_{1i} + \hat{f}_{2i} + \hat{f}_{21i}$, versus paid labor (PL, top-left panel) , unpaid labor (UL, top-right panel), expenses in other cost (OC, lower-left panel) and the value of the capital stock (B&E, lower-right panel). The curve displays the estimated posterior means of $\hat{f}_{1i} + \hat{f}_{2i} + \hat{f}_{21i}$ evaluated at different values of the predictor whose values are displayed in the x-axis, and at the mean-value of the remaining predictors. The dashed lines are sample quantiles (.15, .50, .85) of the variable in the horizontal axis.

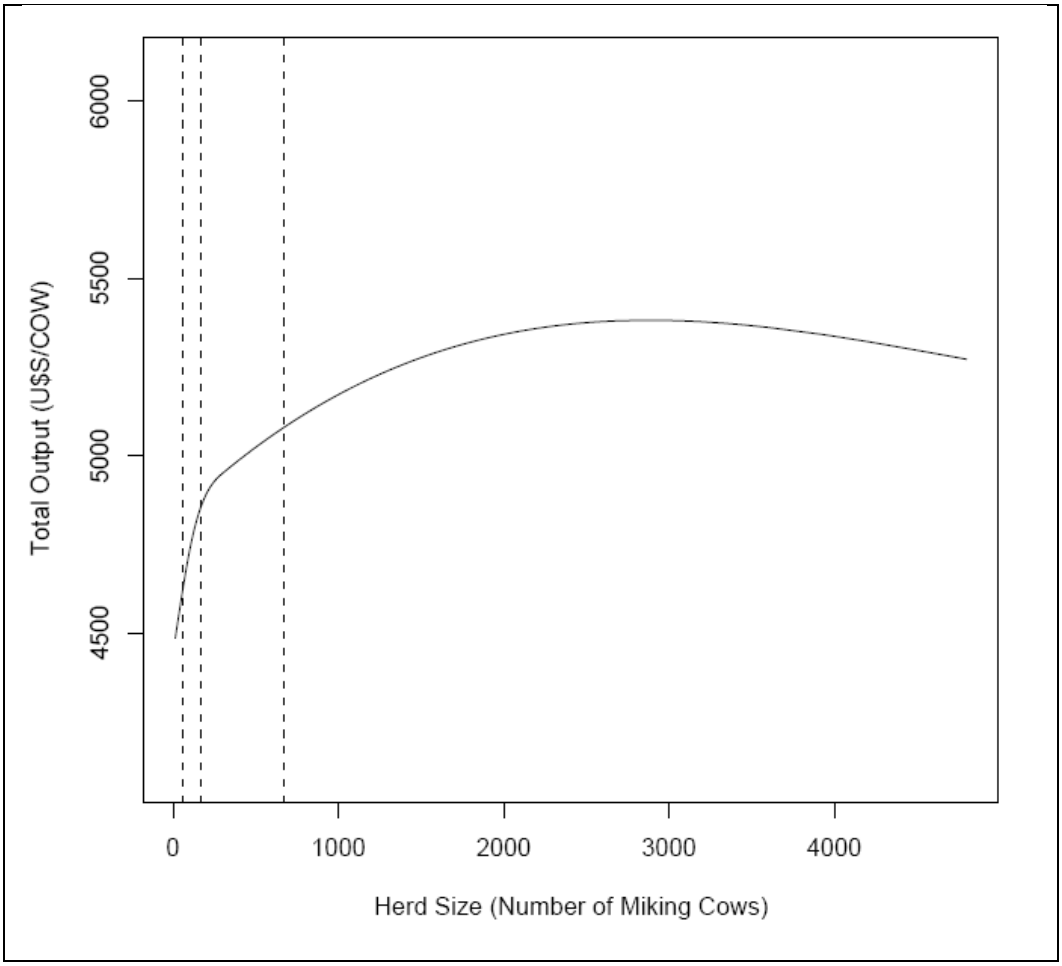

Figure 4. Estimated expected total output per-cow under efficiency, $\hat{f}_{1i} + \hat{f}_{2i} + \hat{f}_{21i}$, versus herd size (COWS). The curve displays the estimated posterior mean of $\hat{f}_{1i} + \hat{f}_{2i} + \hat{f}_{21i}$ evaluated at different values of COWS and at the mean-value of the remaining predictors. The dashed lines are sample quantiles (.15, .50, .85) of COW.