ELSEVIER

Journal of Econometrics 126 (2005) 525–548

JOURNAL OF Econometrics

www.elsevier.com/locate/econbase

Product diversification, production systems, and economic performance in U.S. agricultural production

Catherine J. Morrison Paul[a,*], Richard Nehring[b]

[a]*Department of Agricultural and Resource Economics, University of California, One Shields Avenue, Davis, CA 95616, USA*

[b]*Natural Resource Economics Division, Economic Research Service, U.S. Department of Agriculture, Washington DC, USA*

Available online 2 July 2004

Abstract

The U.S. agricultural sector is experiencing significant structural change. Farm size is rising and activities are broadening, including more off-farm employment, implying economic incentives for larger and more diversified farms, and complementarities among agricultural netputs. We quantify such patterns for farms in the corn belt, by measuring scale economies, and output and input contributions and jointness. We estimate the multi-output and -input production technology by stochastic frontier techniques applied to output and input distance functions. We find that both scope and scale economies have important economic performance implications, and that an input-oriented framework including off-farm income best characterizes agricultural production.
© 2004 Elsevier B.V. All rights reserved.

JEL Classification: O3; O13

Keywords: Scale economies; Scope economies; Jointness; Stochastic frontier; Distance function

*Corresponding author. Tel: +1-530-752-0469; fax: +1-530-752-5614.
E-mail address: cjmpaul@primal.ucdavis.edu (C.J.M. Paul).

0304-4076/$ - see front matter © 2004 Elsevier B.V. All rights reserved.
doi:10.1016/j.jeconom.2004.05.012

1. Introduction

Average farm size in the U.S. Heartland[1] rose by 18 percent between 1980 and 2000. Similarly strong growth in farm size occurred in the Lake and Northern Plains states, although slower growth is evident in other areas (4 percent in the Prairie Gateway).[2] Agricultural production is also highly concentrated in large farms, with "large and very large"[3] family farms making up only 8 percent of all farms in 1998, but accounting for 53 percent of agricultural production. These large farms were "viable economic businesses" in 1998, in the sense that they generated positive profits, whereas "[m]ost farm typology groups did not report adequate income to cover expenses." (USDA, 2001a). Such patterns suggest that significant scale economies exist in modern agriculture, and that this technological reality is putting critical pressure on the small family farm.

In addition to the apparent importance of scale economies, product diversity or scope economies seems to contribute considerably to farms' economic performance. The USDA/ERS Family Farm Report (2001a) states that: "…diversification is a significant factor explaining differences in the level and variability of income between higher and lower performing small farms. Financially successful small farms tend to be more diversified." The Report also notes that production of multiple outputs is most prevalent for high-sales farms, and that diversification affects input demand as well as economic performance. Alexander et al. (2001) also document the importance of product diversity, based on a survey of Iowa farmers' production practices. They find that 94 percent of the respondents grew both soybeans and corn, with the balance differing by year. More than half also grew other crops, and 60 percent raised livestock.

Another type of farm "output"—or contribution to revenue—is also notably affecting the economic health (or even viability) of family farms; off-farm income has become increasingly important in many agricultural areas. USDA (2001a) finds that "farm households rel[y] heavily on off-farm jobs," with 55 percent of farm households reporting that the operator, spouse, or both worked off-farm to increase farm household income, and USDA (2001b) documents that off-farm income now dominates net farm business income in the U.S.[4] Farmers in more rural states are, however, less able to enhance their revenues through off-farm earnings than those in more urban environments (Gardner, 2001).[5]

[1]As recently defined by the USDA; states for these regions are listed in Appendix A.

[2]USDA *Agricultural Statistics*, selected issues, National Agricultural Statistics Service, Washington, DC.

[3]The USDA classifies large farms as those with $250,000 –$500,000 farm revenue, and very large farms as those with more than $500,000 revenue; see Appendix A, Table 7.

[4]Income from farming in the U.S., measured by net-farm cash income, was $55.7 billion in 1999, as compared to income from off-farm sources of $124 billion (USDA, 2001b).

[5]Gardner (2001) shows that the growth of farmers' income is significantly negatively related to the rural proportion of the state's population. As he notes, this supports Schultz's (1950) hypothesis that a more urban environment increases farmers' incomes through enhanced off-farm earnings opportunities, as well as the demand for farm products and services.

The economic performance of U.S. agricultural producers seems also to be increasingly influenced by input jointness or complementarity, and embodied technological change. "Production systems" have been recognized as a progressively greater presence in agricultural markets. Alexander and Goodhue (2002) show that accounting for the mounting complementarity of agricultural inputs (such as seeds, pesticides, and labor or machinery devoted to tilling), and substitutability across production systems (rather than individual inputs), is crucial for analysis of transgenic (or genetically modified, GM) seed demand. Increased effectiveness of the inputs used for production, through adoption of new technology such as that embodied in GM crops, may also have enhanced farms' performance and competitiveness.[6]

These output and input (netput) relationships affect the shape and shifts of the production technology for U.S. agricultural producers, and thus how efficiently farms of different sizes and with different netput composition mixes might be producing. However, farm/farmer characteristics may also affect observed productivity. USDA (2001a) documents key dissimilarities in hours worked, age, education, debt, and management methods, that affect both overall agricultural productive performance and the benefits obtained from innovative adoption. If some types of farmers are producing in a technically inefficient manner, this will affect the observed economic performance.

In this paper, we attempt to quantify these types of scale, scope, system, and efficiency effects determining the economic performance of farms in the U.S. Corn Belt. We focus on output and input jointness and implied complementarities, by measuring scale economies (overall relationships between inputs and outputs), scope economies (output relationships), and input substitutability/complementarity (input relationships).[7]

The farms in our data sample generate a variety of outputs; they produce corn, soybeans, other crops and animal products (dairy, livestock), and earn off-farm income. These production processes rely on a wide variety of inputs, some of which may be linked together in production systems, including labor, fuel, fertilizer, seed, feed, machinery, land, other livestock-specific materials, other crop-specific materials, and other general expenses.

The data for these farms, from a U.S. Department of Agriculture (USDA) farm survey, comprise 15,218 observations across 5 years (1996–2000). These data are summarized in terms of cohorts, by averaging similar farms in like areas for each observation, resulting in a balanced pseudo-data panel of 650 observations (13 cohorts for 10 states by year). Our estimates based on these data therefore represent cross-farm or -cohort as well as spatial and temporal variations in production processes.

[6]This has been documented, for example, by Fernandez-Cornejo and McBride (2000).

[7]Risk from yield and price uncertainties or market power may also affect jointness and diversification, as discussed further below. This behavioral motivation for observed output and input composition is not a direct focus of this primal model, although it should be recognized for interpretation.

We represent the underlying multi-output, multi-input technology of these farms allowing for deviations from the production frontier (technical inefficiency, TE), with alternatively an output or input orientation and with and without off-farm income, by estimating output and input distance functions by stochastic production frontier methods. Our estimates allow us to compute and evaluate a range of measures reflecting the output and input relationships that contribute to economic performance.

The alternative perspectives of the output and input distance function frameworks provide useful comparisons for our analysis. Homogeneity requirements imply that production expansion is at least implicitly evaluated holding output composition (for the output distance function) or input composition (for the input distance function) fixed. This implies different "takes" on which relationships are the most crucial for appropriate representation of production processes; input (output) contributions and substitution may better be captured by the output (input) distance function. By contrast, the output and input distance functions, respectively, reflect output and input contributions or shadow values in a relative (ratio) form. These perspectives could therefore provide quite different implications, although the primary results of our analyses are quite consistent.

Our findings indicate more complementarity or jointness, and yet less consistent composition, for outputs as compared to inputs. That is, diversification is clearly productive but output composition varies more than input mix across type of farm, so scope as well as scale economies are important economic performance "drivers". This suggests that an input distance function implicitly based on constant input composition but directly allowing for a full range of output relationships is advantageous for representing U.S. agricultural production processes (although characterizing scale economies holding input composition constant may overestimate their impact).

Further, off-farm income appears empirically as well as anecdotally to be a key aspect of economic performance and economic viability, especially for small farms. Allowing for this component of farm "output" suggests slightly more efficiency and less scale economies than when the focus is solely on farm business. It also somewhat improves the representation of production processes for the input-oriented model, but exacerbates problems with the output-oriented model. This seems due to the widely varying role of off-farm income across farm types, that augments the variability of output composition when it is included as part of total farm activity or revenue.

2. The models

To explore the roles of scale economies, product diversification, and production systems on farms' economic performance, we require a multi-output, multi-input specification of the technology that allows us to represent interactions among these netputs. Such a specification may be characterized from the output or input perspective, via the output or input sets $P(\mathbf{X}, \mathbf{R})$ or $L(\mathbf{Y}, \mathbf{R})$. $P(\mathbf{X}, \mathbf{R})$ is the set of

output vectors **Y** which can be produced using the input vector **X**, given the levels of external or shift factors in the vector **R**, and $L(\mathbf{Y}, \mathbf{R})$ is the inverse—the set of all **X** vectors that can produce **Y** given **R**.

These relationships can be used to develop estimable distance functions, again with either an output or input orientation. The output (O) distance function $D^{\mathrm{O}}(\mathbf{X}, \mathbf{Y}, \mathbf{R})$ identifies the most **Y** possible to produce given **X**, and the input (I) distance function $D^{\mathrm{I}}(\mathbf{X}, \mathbf{Y}, \mathbf{R})$ the least **X** necessary to produce **Y**, defined according to $P(\mathbf{X}, \mathbf{R})$ or $L(\mathbf{Y}, \mathbf{R})$, respectively. More formally, as developed by Färe (1988), Färe et al. (1994), and Färe and Primont (1995):

$$D^{\mathrm{O}}(\mathbf{X}, \mathbf{Y}, \mathbf{R}) = \min\{\Theta : (\mathbf{Y}/\Theta) \in P(\mathbf{X}, \mathbf{R})\} \tag{1a}$$

and

$$D^{\mathrm{I}}(\mathbf{X}, \mathbf{Y}, \mathbf{R}) = \max\{\rho : (\mathbf{X}/\rho) \in L(\mathbf{Y}, \mathbf{R})\}. \tag{1b}$$

$D^{\mathrm{O}}(\mathbf{X}, \mathbf{Y}, \mathbf{R})$ can thus be interpreted as a multi-output production function, and $D^{\mathrm{I}}(\mathbf{X}, \mathbf{Y}, \mathbf{R})$ as a multi-input input-requirement function, with both allowing for deviations (distance) from the frontier. These primal functions represent technical (substitution) relationships amng and across the inputs and outputs—not economic optimization.[8] Thus the deviations from the frontier are interpreted in terms of technical efficiency, TE.

To empirically implement these functions, linear homogeneity with respect to outputs (for D^{O}) or inputs (for D^{I}) must be imposed. As described by Lovell et al. (1994), this can be accomplished for D^{O} by normalizing by one of the outputs; homogeneity implies $D^{\mathrm{O}}(\mathbf{X}, \omega\mathbf{Y}, \mathbf{R}) = \omega D^{\mathrm{O}}(\mathbf{X}, \mathbf{Y}, \mathbf{R})$ for any $\omega > 0$, so if ω is set arbitrarily at $1/Y_1$, $D^{\mathrm{O}}(\mathbf{X}, \mathbf{Y}/Y_1, \mathbf{R}) = D^{\mathrm{O}}(\mathbf{X}, \mathbf{Y}^*, \mathbf{R})$, where $\mathbf{Y}^* = \mathbf{Y}/Y_1$. The input distance function is analogously normalized by one input; $D^{\mathrm{I}}(\omega\mathbf{X}, \mathbf{Y}, \mathbf{R}) = \omega D^{\mathrm{I}}(\mathbf{X}, \mathbf{Y}, \mathbf{R})$ for any $\omega > 0$; so for $\omega = 1/X_1$, $D^{\mathrm{I}}(\mathbf{X}, \mathbf{Y}, \mathbf{R})/X_1 = D^{\mathrm{I}}(\mathbf{X}/X_1, \mathbf{Y}, \mathbf{R}) = D^{\mathrm{I}}(\mathbf{X}^*, \mathbf{Y}, \mathbf{R})$, where $\mathbf{X}^* = \mathbf{X}/X_1$.

We define **Y**, **X**, and **R** based on USDA annual farm survey data that allow us to distinguish a broad range of both outputs and inputs, and thus to evaluate the impacts of product diversification in terms of output jointness, and of production systems in terms of input complementarity. The data are for farms in states for which corn is a major component of agricultural output, that produce any combination of crops and animal products. The farm-level data were used to construct a pseudo panel data set based on cohorts, to deal with the problem of linking annual cross-section data intertemporally (more detail is provided in the data appendix).

For outputs, **Y**, we distinguish three types of crops—corn, Y_{C}, soybean, Y_{S}, and "other", Y_{O}—each of which comprises a large percentage of farm income (see Table 5 of Appendix A) and have been impacted by biotech adoption during this time period.[9] We also separate out animal output production (meat or dairy), Y_{A},

[8] Mundlak (1996) has shown that when information is incomplete, panel data fixed effects estimation based on a primal model may be preferable to a dual estimator that could generate statistically inefficient estimates due to price variability, as pointed out by an anonymous referee.

[9] The planting of insect resistant corn and herbicide tolerant soybeans increased dramatically over our sample period.

and off-farm income, Y_I. Although Y_I is not really a farm "output", it represents revenue generated from the effort of the farm family, so for our off-farm model with Y_I included **Y** can be interpreted as a farm activity rather than output vector.

We also separate out a variety of inputs, **X**: labor, X_L, fuel (energy), X_E, fertilizer, X_F, seed, X_{SD}, feed, X_{FD}, livestock (animal) expenses, X_A, custom crop expenses, X_C, other expenses, X_O, machinery, X_K, and (quality-adjusted) land, X_{LD}.[10] The material categories in particular are broken down more than is usually possible, permitting us to explicitly link complementary inputs that might comprise a production system. Note that the labor input for our off-farm specification is augmented to include effort devoted to off-farm activities, based on the opportunity cost of the associated operator or spouse.

The farm-level data also allow us to identify farm- and farmer-specific characteristics that we treat as shift factors, or components of the **R** vector, R_j. Farmer characteristics are age, AGE, and education, ED. Farm characteristics are a debt–asset ratio, DA, and the proportions of land that are rented, RNT, and planted in GM corn and soybean crops, GM_C and GM_S. We also include dummy variables for each year, T_{1996}–T_{2000} (T_{1996} left out for estimation) and for four size classes or typologies (loosely from small to large), residential farms, RES, small family farms, SM, large family farms, LG, and very large family and corporate operations, CORP (with RES left out for estimation).

The R_j factors are treated as fixed effects, except for GM_C and GM_S, for which cross-effects are included because one would expect these factors to interact with specific inputs or outputs. For the output-oriented specifications that focus on input use, GM_C and GM_S are "interacted" with X_{SD} and X_C (which capture seed and pesticide purchases), because enhanced embodied technology might be expected to increase the productive contribution of seeds, and reduce the contribution of pesticides. For the input-oriented specifications, we interact GM_C and GM_S with Y_C and Y_S to directly capture yield and output substitution impacts of genetically modified seeds.

Summary statistics for these **Y**, **X**, and **R** variables are provided in Table 5 of Appendix A for the year 2000, in total and distinguished by cohort type. The reported values are averages across all the farms in the sample. Both the levels and composition of inputs and outputs vary by type of farm, with the average farm in the survey generating slightly more than half its farm revenues from crop production, and making about a third of its total revenue from off-farm income.[11]

[10] X_A includes not only the cost of purchased livestock, but also bedding and litter, and medical expenditures (medical supplies, veterinary and custom services for livestock). X_C includes hauling and machine hire, irrigation, and pesticide expenses. X_O includes general business expenses that cannot be ascribed specifically to crops or animals, such as interest and insurance.

[11] Note that capital and land services, as well as wages/prices of the farm operator/livestock, are imputed, so overall input payments in terms of opportunity costs exceed revenues for the average farm.

One issue that arises for implementing the distance function models is which of the outputs or inputs might be used as normalizing factors. Glossing over the econometric issues associated with a numeraire netput,[12] and recognizing that the final results are invariant to this choice (Coelli and Perelman, 2000), there could still be economic reasons for choosing Y_1 or X_1. For our output distance function specifications, we specify $Y_1 = Y_C$, because we are focusing on farms where corn is the primary commodity crop. For our input specifications, we specify $X_1 = X_{LD}$, so all other inputs are represented relative to land. This is consistent with the typical agricultural economics approach to production modeling in terms of yields, or inputs (and outputs) per acre.

Another issue to underscore before moving to empirical implementation is differences embedded in the output versus input perspective, due to both the dual nature of the output- and input-oriented frameworks and fixities when one evaluates farm growth or expansion. That is, enhanced economic performance is represented by the output distance function through output expansion given input use, and conversely by the input function through input contraction given output production. This expansion or contraction is, however, based on observed output or input composition, respectively; output or input ratios are held constant when measuring output or input elasticities (and thus scale economies). So in the output distance function model outputs are not as much "choice variables" as are inputs, and vice versa.

The choice of specification therefore depends on whether one believes production jointness or systems are more fundamental on the output or input side. If inputs are essentially fixed for a farmer, then output composition is the primary economic performance determinant and an input-oriented function is more appropriate. If, however, the balance of inputs used is more flexible than outputs produced, an output-oriented function is preferable.

3. Empirical implementation

3.1. The model

For empirical implementation of our models, we assume the distance functions can be approximated by translog functional forms, which limit a priori restrictions on the relationships among the outputs and inputs that we wish to explore. However, since we treat most of the **R** factors as fixed effects, the functions are not fully flexible.[13] The resulting functions are of

[12]As Coelli and Perelman argue, this should not present econometric problems because " only ratios of the outputs appear as regressors and these ratios may be assumed to be exogenous, since the distance function is defined for radial (proportional) expansion of all outputs, given the input levels, and hence by definition the output ratios are held constant for each firm."

[13]This reduces the number of cross-terms with our specification, which is so disaggregated across outputs and inputs that with full flexibility the cross-R_j terms have negligible significance.

the form:

$$\begin{aligned}\ln D_{it}^{\mathrm{O}}/Y_{1,it} &= \alpha_0 + \sum_j \alpha_j \ln R_{jit} + \sum_m \alpha_{m*} \ln Y_{mit}^* \\ &\quad + 0.5 \sum_m \sum_n \beta_{m*n*} \ln Y_{mit}^* \ln Y_{nit}^* \\ &\quad + \sum_k \alpha_k \ln X_{kit} + \sum_{k(=\mathrm{SD,C})s} \beta_{ks} \ln X_{kit} \mathrm{GM}_s \\ &\quad + 0.5 \sum_k \sum_1 \beta_{kl} \ln X_{kit} \ln X_{lit} \\ &\quad + \sum_k \sum_m \beta_{m*k} \ln Y_{mit}^* \ln X_{kit} \\ &= \mathrm{TL}^{\mathrm{O}}(\mathbf{R}, \mathbf{Y}^*, \mathbf{X}) \end{aligned} \tag{2a}$$

and

$$\begin{aligned}\ln D_{it}^{\mathrm{I}}/X_{1,it} &= \alpha_0 + \sum_j \alpha_j \ln R_{jit} + \sum_k \alpha_{k*} \ln X_{kit}^* \\ &\quad + 0.5 \sum_k \sum_l \beta_{k*l*} \ln X_{kit}^* \ln X_{lit}^* \\ &\quad + \sum_m \alpha_m \ln Y_{mit} + \sum_{mj} \beta_{m(=\mathrm{C,S})s} \ln Y_{mit} \mathrm{GM}_s \\ &\quad + 0.5 \sum_m \sum_n \beta_{mn} \ln Y_{mit} \ln Y_{nit} \\ &\quad + \sum_k \sum_m \beta_{k*m} \ln X_{kit}^* \ln Y_{mit} \\ &= \mathrm{TL}^{\mathrm{I}}(\mathbf{R}, \mathbf{X}^*, \mathbf{Y}), \end{aligned} \tag{2b}$$

where i denotes farm, t time period, m, n the outputs, k, l the inputs, j the external effects, and s the types of GM seeds (C,S). Rewriting these functions with $-\ln D_{it}^{\mathrm{O}} = -u_{it}^{\mathrm{O}}$ and $-\ln D_{it}^{\mathrm{I}} = -u_{it}^{\mathrm{I}}$ as one-sided error terms, and including "white noise" error terms ${v_{it}}^{\mathrm{O}}$ and ${v_{it}}^{\mathrm{I}}$ representing random factors such as measurement error or unobserved inputs, results in the estimating equations:

$$-\ln Y_{1,it} = \mathrm{TL}^{\mathrm{O}}(\mathbf{R}, \mathbf{Y}^*, \mathbf{X}) - u_{it}^{\mathrm{O}} + v_{it}^{\mathrm{O}} \tag{3a}$$

and

$$-\ln X_{1,it} = \mathrm{TL}^{\mathrm{I}}(\mathbf{R}, \mathbf{X}^*, \mathbf{Y}) - u_{it}^{\mathrm{I}} + v_{it}^{\mathrm{I}}. \tag{3b}$$

Coefficient estimates for these equations have the opposite signs from those for a standard production or input requirement function. $\partial \mathrm{TL}^{\mathrm{O}}/\partial X_k$, for example, represents the overall change in outputs (the change in Y_1 with all output ratios, and thus output composition, constant) with a change in X_k. However, from the distance function perspective this estimated "marginal product" is negative instead of positive. Similarly, $\partial \mathrm{TL}^{\mathrm{I}}/\partial X_1$ represents the overall change in inputs (given input composition) with a change in Y_1, but this "marginal input requirement"

(or, loosely, "marginal cost") of Y_1 is again negative. To interpret the measures from (3) more similarly to those from these more familiar functions, we thus reverse their sign:

$$\ln\ Y_{1,it} = -\mathrm{TL}^{\mathrm{O}}(\mathbf{R}, \mathbf{Y}^*, \mathbf{X}) + u_{it}^{\mathrm{O}} - v_{it}^{\mathrm{O}}, \tag{4a}$$

$$\ln\ X_{1,it} = -\mathrm{TL}^{\mathrm{I}}(\mathbf{R}, \mathbf{X}^*, \mathbf{Y}) - u_{it}^{\mathrm{I}} - v_{it}^{\mathrm{I}}. \tag{4b}$$

These equations are written in a standard stochastic production frontier form (with a two-part error term representing deviations from the frontier and random error). They can therefore be estimated econometrically using maximum likelihood techniques,[14] assuming that v_{it}^{O} or v_{it}^{I} are independently and identically distributed random variables, $\mathrm{N}(0, \sigma_v^2)$, and u_{it}^{O} or u_{it}^{I} are nonpositive random variables independently distributed as truncations at zero of $\mathrm{N}(0, \sigma_u^2)$. For our estimation, we have applied the error components model of Battese and Coelli (1992), using Tim Coelli's FRONTIER program, as in Coelli and Perelman (2000) and Paul et al. (2000).

3.2. The performance measures

Various measures that summarize production processes and act as performance indicators can be constructed as derivatives or elasticities from our estimated model. For example, combined first-order netput elasticities represent scale economies, and thus capture the extent to which productivity increases with growth. Such measures provide insights about the competitive disadvantages faced by small farms, and thus farmers' incentives to expand their scale of production to enhance their competitiveness. Individual first-order elasticities characterize input- or output-specific contributions to these economies, and identify the productive contributions of farm/farmer characteristics or unmeasured effects specific to a given year or cohort. Second-order elasticities reflect production complementarities (or biases) that reflect economic performance impacts from output or input jointness.

More specifically, to develop these performance measures we first focus on the overall **Y**–**X** relationship. From the output distance function, this relationship, measured as the sum of the output elasticities for each input, shows how much overall output would increase from a 1 percent increase in each input, analogous to a returns to scale estimate from a production function. That is, the "output elasticity" for input X_k, $-\varepsilon_{D^{\mathrm{O}},X_k} = -\partial \ln D^{\mathrm{O}}/\partial \ln X_k = \partial \ln Y_1/\partial \ln X_k = \varepsilon_{Y,X_k}$, represents the percent change in Y_1 from a 1 percent change in X_k, holding all output ratios Y^* (and thus output composition) constant. Such elasticities thus represent the returns to or output contributions from X_k changes, like output elasticities from production function estimation. Also, if $\partial Y_1/\partial X_k$ is interpreted as the marginal product of X_k (an increase in *overall* output from an increase in X_k, MP_k), ε_{Y,X_k} can be thought of as the "output share" of X_k (relative to Y_1); $\varepsilon_{Y,X_k} = MP_k X_k / Y_1$. Summing these

[14]This method, initially developed by Aigner et al. (1977) and Meeusen and van den Broeck (1977), is discussed in depth in Coelli et al. (1998).

measures results in the output-oriented distance function-based scale economy measure (Färe and Primont, 1995) $-\varepsilon_{D^{\mathrm{O}},X} = -\sum_k \partial \ln D^{\mathrm{O}}/\partial \ln X_k = \sum_k \partial \ln Y_1/\partial \ln X_k = \sum_k \varepsilon_{Y,X_k} = \varepsilon_{Y,X}$. $\varepsilon_{Y,X} > 1$ implies increasing returns to scale; input increases generate a more than proportionate output expansion (with proportional changes in all outputs).

For the input distance function, the **X**–**Y** scale economy relationship, represented as a sum of the individual input elasticities conversely to the output-oriented measure, reflects how much overall input use must increase to support a 1 percent increase in all outputs. This is similar to a cost function-based scale economy measure that captures input use changes required for output growth, although this primal measure represents just the technical relationship—not input choice.

Formally, the individual input elasticity summarizing the input expansion required for a 1 percent increase in Y_m is $-\varepsilon_{D^{\mathrm{I}},Y_m} = -\partial \ln D^{\mathrm{I}}/\partial \ln Y_m = \partial \ln X_1/\partial \ln Y_m = \varepsilon_{X,Y_m}$. Dual to the "output share" notion above, such a measure can be thought of as an "input share" of Y_m (relative to X_1): $\partial \ln X_1/\partial \ln Y_m = (\partial X_1/\partial Y_m)\, Y_m/X_1$. In combination, these elasticities represent scale economies: $-\varepsilon_{D^{\mathrm{I}},Y} = -\sum_m \partial \ln D^{\mathrm{I}}/\partial \ln Y_m = \sum_m \partial \ln X_1/\partial \ln Y_m = \sum_m \varepsilon_{X,Y_m} = \varepsilon_{X,Y}$ consistent with Baumol et al. (1982) for a multiple-output cost model and Färe and Primont (1995) for an output distance function. The extent of scale economies (for proportional changes in all inputs) is implied by the short-fall of $\varepsilon_{X,Y}$ from 1.

The first-order elasticities ε_{YX_k}, ε_{YX}, ε_{XY_m}, and ε_{XY} can also be decomposed into second-order effects reflecting input or output composition changes as scale expands. This information is implied by technological bias measures, indicating for the output distance function how the X_k output elasticity or share (ε_{YX_k}) adapts to a change in another input, and the reverse for the input distance function. Such measures thus provide insights about input and output jointness, or production systems. The performance impact of such netput complementarity is represented by a combination of the biases. If overall output (input) relationships are complementary, an increase in one output (input) enhances the contributions of other outputs (inputs) and thus performance.

Specifically, for the output distance function $\varepsilon_{YX_k,X_1} = \partial \varepsilon_{Y,X_k}/\partial \ln X_1$ represents the impact on the contribution or share of input X_k from an increase in X_1. If X_k and X_1 in some sense "move together" (are complementary or act as a system), an increase in X_1 shifts up the share and thus marginal product of $X_k : \varepsilon_{YX_k,X_1} > 0$. This measure in the translog context collapses to the cross-input β_{kl} coefficient estimate; with symmetry, $\varepsilon_{YX_k,X_1} = \beta_{kl} = \varepsilon_{YX_l,X_k}$. Similarly, $\varepsilon_{XY_m,Y_n} = \partial \varepsilon_{X,Y_m}/\partial \ln Y_n$ from the input distance function represents the increase in the Y_m input share if Y_n also increases. If $\varepsilon_{XY_m,Y_n} < 0$, output jointness or complementarity is implied; input use does not have to increase as much to expand Y_m if the Y_n level is higher. This elasticity is represented by the cross-output coefficient estimate $\beta_{mn} : \varepsilon_{XY_m,Y_n} = \beta_{mn} = \varepsilon_{XY_n,Y_m}$.

In addition to information about input (output) patterns, some insights about output (input) contributions may be distilled from the output (input) distance function, but they are relative due to the ratio form of the arguments of the function. That is, from the output perspective, $-\partial D^{\mathrm{O}}/\partial Y_m = \partial Y_1/\partial Y_m^* = r_m^*$ reflects the

(negative) shadow value of Y_m relative to Y_1, or, loosely, the slope of the production possibility frontier.[15] $\varepsilon_{Y,Y_m} = \partial \ln Y_1/\partial \ln Y_m^* = r_m^* Y_m^*/Y_1$ therefore represents the "shadow share" or contribution of Y_m relative to Y_1, and the coefficient $\beta_{m^*n^*} = \partial\varepsilon_{Y,Y_m}/\partial \ln Y_n^*$ reflects the change in this share from a change in output (ratio) Y_n^*. Analogous input relationships may be characterized from the input distance function. However, these relative measures are less readily interpretable than the "absolute" measures overviewed above that directly identify the productive contributions of each output (input).

The performance impacts of the **R** vector components also can be estimated as the distance function elasticities $-\varepsilon_{D^{\mathrm{O}},R_j} = -\partial \ln D^{\mathrm{O}}/\partial R_j = \partial \ln Y_1/\partial R_j = \varepsilon_{Y,R_j}$ and $-\varepsilon_{D^{\mathrm{I}},R_j} = -\partial \ln D^{\mathrm{I}}/\partial R_j = \partial \ln X_1/\partial R_j = \varepsilon_{X,R_j}$; $\varepsilon_{Y,R_j} > 0$ (more output is produced for a given input vector if R_j is greater), or $\varepsilon_{X,R_j} < 0$ (less input is required to produce a given output vector if R_j is greater), implies enhanced productivity or performance from R_j. Since most of our R_j variables are included only as fixed effects or overall shift factors, these impacts are implied simply from their associated estimated parameters (they have only a first-order effect and no biases). For $\mathrm{GM_C}$ and $\mathrm{GM_S}$, however, cross-effects with X_{SD} and X_{S}, and Y_{C} and Y_{S}, are embedded in ε_{Y,R_j} and ε_{X,R_j}, respectively.

Finally, from the one-sided error terms, u_{it}^{O} or u_{it}^{I}, we can quantify the levels of (residual) technical efficiency, $\mathrm{TE^O} = \exp(u_{it}{}^{\mathrm{O}})$, and $\mathrm{TE^I} = \exp(u_{it}^{\mathrm{I}})$ (see Coelli et al., 1998, for elaboration). That is, the deviation of a particular observation from the estimated frontier identifies the remaining apparent technical inefficiency after the impacts from all our measured factors are taken into account. The deviations of the TE measures from 1 indicate the percent by which production would have to increase, or input use would have to decrease, to reach the production frontier.

4. Empirical results

We estimated four alternative models, (4a) and (4b) with and without off-farm income, which we denote our output and input "base" and "off-farm" specifications. Due to our single estimating equation and large number of parameters, one might expect multicollinearity and thus statistical insignificance to be evident. We found most squared-input terms to be insignificant, and the cross-Y–X terms to be virtually always insignificant (consistent with separability of **Y** and **X**). However, the cross- and squared-output terms were almost invariably significant across specifications, and many cross-input terms were also significant. Thus, we set only the β_{k^*m} or β_{m^*k} terms, and the insignificant β_{kk} or $\beta_{k^*k^*}$ terms, to zero.

[15] Färe (1988) and Färe and Grosskopf (1990) showed that the distance function duality with the revenue function can be used to define the revenue-deflated shadow price of Y_{m} via a distance-function oriented Shephard's lemma. The interpretability of these relative measures and their corresponding second-order derivatives is more limited than the measures focused on above, which have a more direct linkage to standard economic notions of, e.g., marginal products or output shares. However, these relationships can be manipulated to obtain information about substitutability, as illustrated in Paul et al. (2000).

Table 1
Scale economy and output/input elasticities, all specifications (evaluated at the average values of the data)

	Output					Input			
	Base		Off-farm			Base		Off-farm	
$\varepsilon_{Y,X}$	1.176	0.03	1.091	0.03	$\varepsilon_{X,Y}$	0.654	0.01	0.732	0.02
TE^{O}	0.883		0.927		TE^{I}	0.929		0.937	
ε_{Y,X_L}	0.090	0.04	0.334	0.03	ε_{X,X^*_L}	−0.260	0.02	−0.304	0.02
ε_{Y,X_E}	0.050	0.04	−0.008	0.03	ε_{X,X^*_E}	0.003	0.02	−0.010	0.02
ε_{Y,X_F}	0.121	0.03	0.115	0.03	ε_{X,X^*_F}	−0.059	0.02	−0.104	0.02
$\varepsilon_{Y,X_{SD}}$	0.261	0.03	0.185	0.03	$\varepsilon_{X,X^*_{SD}}$	−0.149	0.02	−0.136	0.02
$\varepsilon_{Y,X_{FD}}$	0.190	0.02	0.144	0.01	$\varepsilon_{X,X^*_{FD}}$	−0.101	0.01	−0.105	0.01
ε_{Y,X_A}	0.066	0.01	0.062	0.01	ε_{X,X^*_S}	−0.040	0.01	−0.038	0.01
ε_{Y,X_C}	0.195	0.03	0.141	0.03	ε_{X,X^*_C}	−0.058	0.02	−0.049	0.02
ε_{Y,X_O}	0.106	0.04	0.001	0.04	ε_{X,X^*_O}	−0.102	0.03	−0.071	0.03
ε_{Y,X_K}	−0.021	0.04	−0.003	0.03	ε_{X,X^*_K}	−0.020	0.02	−0.012	0.02
$\varepsilon_{Y,X_{LD}}$	0.116	0.03	0.120	0.02					
					ε_{X,Y_C}	0.162	0.01	0.150	0.01
ε_{Y,Y^*_S}	−0.228	0.01	−0.178	0.01	ε_{X,Y_S}	0.138	0.01	0.136	0.01
ε_{Y,Y^*_O}	−0.163	0.01	−0.113	0.01	ε_{X,Y_O}	0.095	0.01	0.093	0.01
ε_{Y,Y^*_A}	−0.400	0.01	−0.342	0.01	ε_{X,Y_A}	0.258	0.01	0.240	0.01
ε_{Y,Y^*_I}			−0.198	0.01	ε_{X,Y_I}			0.085	0.01
ε_{Y,GM_C}	0.003	0.001	0.003	0.001	ε_{X,GM_C}	−0.001	0.001	−0.001	0.001
ε_{Y,GM_S}	−0.0004	0.0006	0.0003	0.0005	ε_{X,GM_S}	0.001	0.0004	0.0005	0.0004

The first-order elasticity and TE estimates for the four models on average across the entire sample are presented in Table 1, along with their standard errors.[16] The primary overall measures, representing output/input patterns and performance incentives for increasing the scale and diversity of farm operations, are the scale elasticities $\varepsilon_{Y,X}$ and $\varepsilon_{X,Y}$. The presented measures suggest significant scale economies,[17] especially for the input-oriented specifications (recall that $\varepsilon_{Y,X} > 1$ and $\varepsilon_{X,Y} < 1$ indicate scale economies). The high estimated returns for the input specification may be attributable to output jointness or scope economies, which are embodied in this measure but not captured in the output specification. In fact, since

[16]These estimates are computed using the delta method to evaluate the elasticity formulas at the average values of the variables in the data. This method is based on linearizing the elasticity functions around the estimated parameter values, and then using standard formulas for the variances and covariances of linear functions of random variables. These elasticity estimates are very consistent in magnitude with average elasticity estimates computed instead by estimating the elasticities for each data point, and then averaging the elasticity estimates.

[17]Note that one might expect some scale economies to be captured by the farm-type dummies for SM, LG, and CORP combined cohorts, because these types are largely defined according to size. However, preliminary empirical investigation showed that any implied bias is not substantive. Estimated $\varepsilon_{Y,X}$ for the base output specification without the cohort dummies is, for example, 1.22 rather than 1.18.

output but not input mix is directly represented in the input function (and vice versa), scope economies could be over-stated for the input specification as well as under-stated for the output specification for which output composition variations are not accommodated.

The estimated scale economies are somewhat lower when off-farm income is included. This implies that the increasing prevalence of off-farm income for small farmers combats their scale disadvantages from only farm business activities. It could alternatively suggest that a model implicitly based on constant output composition is not representative when the extent of off-farm income is cohort-specific.

The individual input and output contributions underlying the scale elasticities are also presented in Table 1.[18] Note first, for the output specification that $\varepsilon_{Y,X_{\mathrm{K}}}$ is negative for the base model and both $\varepsilon_{X,Y_{\mathrm{E}}}$ and $\varepsilon_{Y,K}$ for the off-farm model, although they are both small in magnitude and statistically insignificant. This could suggest that X_{E} and X_{K} comprise a "production system" or are complementary; they jointly contribute to output, so their individual contributions are not well identified.[19] The contribution of "other" inputs is also much smaller, and of labor is much greater, when off-farm income is recognized. The remaining measures are roughly similar across the base and off-farm models, with only labor and land generating larger returns (shares) in the off-farm model. Also note that seed, feed, and crop expenses (largely pesticides) seem the most important drivers of overall farm output, perhaps because they are the variable inputs that determine the productivity of the other, more fixed, inputs. The $\varepsilon_{Y,Y_{m*}}$ elasticities are more similar across the base and off-farm models, with animal outputs comprising by far the greatest output share, but off-farm income contributing nearly 20 percent.

The individual estimated netput shares from the input-oriented models are more consistent across the base and off-farm specifications (although the contribution of X_{E} is again the wrong sign but insignificant for the base model), with off-farm income contributing significantly to input use (but not as much as to output: $\varepsilon_{X,Y_{\mathrm{I}}} = 0.085$). This is in part because the labor of the farm operator who works off the farm is captured in our labor measure (so $\varepsilon_{X,X_{\mathrm{L}*}}$ rises in absolute value to -0.304 from -0.260 on average), but also because recognizing off-farm income increases the role of X_{E} (more fuel is likely to be required to support off-farm activities) and X_{F} (fertilizer may, for example, substitute for labor-intensive tilling practices). By contrast, the contribution of capital declines for the off-farm model. The $\varepsilon_{X,Y_{\mathrm{m}}}$ elasticities also confirm that animal outputs require the greatest input share, with corn second, and soybeans third.

These estimated shares, weighted implicitly by their estimated marginal products or shadow values, are broadly comparable to the actual shares, based on market

[18]Risk may have an impact on the " marginal product" or contribution of inputs as well as outputs, although the direction of impact is not definitive. Ramaswami (1992) found that whether yield uncertainty increases or reduces the marginal products of inputs, and thus input use, depends on whether input use is risk-increasing or decreasing.

[19]This makes conceptual sense, and is consistent with the overall insignificance of the cross-X_{E} coefficients (evident from appendix Table 6), with a strongly positive (although not statistically significant) $\beta_{X_E X_K}$ coefficient.

prices, implied by the output and input values presented in Table 5 of Appendix A. It is worth noting that soybeans appear to have larger (output *and* input) shares than "other" crops, whereas this is reversed in the primary data; soybeans thus seem to have greater "value", but also to be more input-intensive. For inputs, labor's share is implicitly larger in terms of shadow value than the numeraire input, land (except for the output base specification), contrary to their market values; opportunity costs of land (labor) appear larger (smaller) than its true marginal product. The shadow value of seed is also higher, but the values of capital and livestock inputs lower, than their market values (shares).

Note also that the $\varepsilon_{Y,\mathrm{GM_S}}$ and $\varepsilon_{X,\mathrm{GM_S}}$ measures suggest that planting GM corn has a (statistically) significant but very small contribution to performance. Planting GM soybeans seems to reduce productivity, although negligibly, which is consistent with the notion that associated benefits such as managerial effort savings are not well measured.[20]

Finally, measured efficiency (TE) is quite high, especially for the off-farm specifications, with average levels of approximately 0.93–0.94 for the input models, and 0.88–0.93 for the output specifications. These estimates suggest that a large amount of farm diversity is explained by the broad characterization of output and input relationships in these models, even with the restricting implications of constant output (input) composition in the input (output) models.

Overall, the first-order estimates reported in Table 1 suggest that recognizing off-farm income provides additional insights over the base models. We will thus focus for our remaining discussion of the empirical results on the off-farm specifications, for which parameter estimates and *t*-statistics are presented in Table 6 of Appendix A.[21] The many significant coefficients suggest quite strong explanatory power of these detailed models.

To further evaluate the implications from our estimates about netput complementarities and their contribution to scale economies, we can focus on the (second order) cross-effects or biases. These estimates, as shown above, are represented by the cross-parameters of the estimated functions, reproduced in Table 2 in matrix form. The X_{SD}–X_{F} "cell" for the output specification, for example, represents $\beta_{\mathrm{SD_F}} - \beta_{F_{\mathrm{SD}}}$.

Recall that the signs and magnitudes of these cross-effects indicate the extent of input or output jointness, with a positive (negative) value implying complementarity for the inputs (outputs). Table 2 thus provides two interesting results. First, the cross-output relationships (from the input specification) are predominantly negative, and half are significant, whereas the cross-input relationships (from the output specification) are quite evenly balanced in terms of negative and positives, and statistical significance. Second, the cross-terms tend to be much smaller for the outputs than the inputs.

In combination, these results suggest that scope economies, or diversification, contribute significantly to economic performance, but that the linkages between the

[20]See Marra (2001).

[21]We do not present the coefficients for all the specifications to reduce the volume of tables.

Table 2
Cross-terms representing input and output jointness, off-farm model

Output	X_L	X_E	X_F	X_{SD}	X_{FD}	X_A	X_C	X_O	X_K
X_E	0.007								
X_F	−0.034	−0.024							
X_{SD}	−0.181	0.034	−0.032						
X_{FD}	−0.036	−0.044	0.003	−0.025					
X_A	0.015	0.033	−0.005	0.009	0.010				
X_C	0.169	0.072	0.030	−0.009	−0.009	−0.016			
X_O	−0.063	−0.103	0.294	−0.040	0.022	−0.003	−0.266		
X_K	−0.011	0.062	−0.171	0.194	0.048	−0.021	0.032	0.185	
X_{LD}	0.130	−0.048	−0.063	−0.079	0.044	−0.006	0.118	−0.056	−0.006
GM_C				−0.003			0.003		
GM_S				0.001			0.000		
Input									
	Y_C	Y_S	Y_O	Y_A					
Y_S	−0.006								
Y_O	−0.009	−0.013							
Y_A	−0.001	−0.011	−0.004						
Y_I	−0.002	0.003	−0.008	−0.012					
GM_C	0.001	−0.00004							
GM_S	−0.001	0.0004							

outputs are not as strong as for inputs. Input composition thus seems to be more consistent, or fixed across farm type or time, than output mix, particularly when off-farm income is included (as implied from Table 1). This implies that an input-oriented specification, that explicitly recognizes varying output composition, is more appropriate for representing farm production processes.[22]

Note also that, although few clear insights emerge in the literature about how uncertainty and risk affect farm production processes and productivity (Pope, 1987), additional unmeasured jointness may arise from risk rather than technical complementarities. As suggested by Just and Pope (1978), even if two outputs are not correlated the production of one output is reduced if uncertainty over the other input rises, and if they *are* correlated a positive correlation exacerbates and a negative correlation reduces this effect. This jointness is also related to input use, because uncertainty causes variations in the marginal products or contributions of inputs across products that depend on risk factors.[23]

[22]This result is somewhat surprising since the inputs are so much more disaggregated than the outputs that one might expect a specification targeting input use patterns would provide more insights than one focusing on output production patterns. It is, however, consistent with Williams and Shumway's (1998) finding that much more aggregation of input than output categories is supported by empirical tests.

[23]The important roles of uncertainty and risk in explaining production, and particularly diversification, behavior, emphasized by Newberry and Stiglitz (1981), was noted by an anonymous referee. However, more explicit consideration of these behavioral motivations is beyond the scope of our primal model.

Table 3
Primary measures, different farm types, off-farm

	Output					Input			
	Res	SM	LG	CORP		Res	SM	LG	CORP
$\varepsilon_{Y,X}$	1.114	1.086	1.083	1.085	$\varepsilon_{X,Y}$	0.593	0.650	0.805	0.855
TE^{O}	0.928	0.910	0.940	0.927	TE^{I}	0.935	0.932	0.938	0.942
$\varepsilon_{Y,X_{\mathrm{L}}}$	0.419	0.387	0.306	0.234	$\varepsilon_{X,X_{\mathrm{L}}}$	−0.379	−0.360	−0.281	−0.202
$\varepsilon_{Y,X_{\mathrm{E}}}$	−0.064	−0.021	0.025	0.018	$\varepsilon_{X,X_{\mathrm{E}}}$	0.004	−0.009	−0.024	−0.008
$\varepsilon_{Y,X_{\mathrm{F}}}$	0.169	0.098	0.083	0.121	$\varepsilon_{X,X_{\mathrm{F}}}$	−0.091	−0.094	−0.113	−0.115
$\varepsilon_{Y,X_{\mathrm{SD}}}$	0.178	0.244	0.197	0.116	$\varepsilon_{X,X_{\mathrm{SD}}}$	−0.093	−0.132	−0.164	−0.144
$\varepsilon_{Y,X_{\mathrm{FD}}}$	0.123	0.120	0.158	0.172	$\varepsilon_{X,X_{\mathrm{FD}}}$	−0.095	−0.083	−0.113	−0.126
$\varepsilon_{Y,X_{\mathrm{A}}}$	0.048	0.053	0.062	0.085	$\varepsilon_{X,X_{\mathrm{A}}}$	−0.011	−0.029	−0.045	−0.063
$\varepsilon_{Y,X_{\mathrm{C}}}$	0.105	0.150	0.160	0.142	$\varepsilon_{X,X_{\mathrm{C}}}$	−0.066	−0.067	−0.043	−0.022
$\varepsilon_{Y,X_{\mathrm{O}}}$	−0.030	−0.019	0.031	0.011	$\varepsilon_{X,X_{\mathrm{O}}}$	−0.071	−0.084	−0.076	−0.050
$\varepsilon_{Y,X_{\mathrm{K}}}$	−0.051	−0.055	0.001	0.092	$\varepsilon_{X,X_{\mathrm{K}}}$	0.042	0.052	−0.024	−0.112
$\varepsilon_{Y,X_{\mathrm{LD}}}$	0.217	0.129	0.061	0.093					
					$\varepsilon_{X,Y_{\mathrm{C}}}$	0.088	0.104	0.194	0.200
$\varepsilon_{Y,Y^*_{\mathrm{S}}}$	−0.124	−0.188	−0.212	−0.179	$\varepsilon_{X,Y_{\mathrm{S}}}$	0.097	0.125	0.166	0.148
$\varepsilon_{Y,Y^*_{\mathrm{O}}}$	−0.123	−0.110	−0.096	−0.129	$\varepsilon_{X,Y_{\mathrm{O}}}$	0.074	0.081	0.091	0.127
$\varepsilon_{Y,Y^*_{\mathrm{A}}}$	−0.284	−0.321	−0.350	−0.411	$\varepsilon_{X,Y_{\mathrm{A}}}$	0.170	0.206	0.258	0.319
$\varepsilon_{Y,Y^*_{\mathrm{I}}}$	−0.350	−0.249	−0.139	−0.072	$\varepsilon_{X,Y_{\mathrm{I}}}$	0.142	0.107	0.067	0.031
$\varepsilon_{Y,\mathrm{GM}_{\mathrm{C}}}$	0.005	0.004	0.002	0.001	$\varepsilon_{X,\mathrm{GM}_{\mathrm{C}}}$	−0.002	−0.002	0.000	0.000
$\varepsilon_{Y,\mathrm{GM}_{\mathrm{S}}}$	−0.001	0.000	0.001	0.001	$\varepsilon_{X,\mathrm{GM}_{\mathrm{S}}}$	0.001	0.001	0.000	0.000

It is also useful to consider differences in the first-order measures across cohort and time, from the elasticities presented in Tables 3 and 4.[24] The scale elasticities from the output model indicate that returns to scale vary little by cohort, which is not intuitively very plausible. Those for the input model, by contrast, show larger potential scale economies for the smaller farms.[25] This discrepancy supports the suggestion that the input model better captures output composition differences that are an important component of economic performance for these farms. Both models also suggest that the efficiency of small family farms is lower than for other farm

[24]The statistical significance of these measures varies little from the average over all the data, reported in Table 1, because the parameters and standard errors are the same; the measures are simply evaluated at different data points. We therefore omit the standard errors to simplify the tables. Note also that the time trends are fairly smooth, so we only present estimates for every other year for brevity.

[25]Estimations were also carried out separately by cohort rather than pooled, to evaluate the suggestion by an anonymous referee that the pooling across very different types of farms may be affecting our results. The resulting scale economy estimates were very comparable for the input-oriented model, although the RES farms had less scale economies than estimated in the pooled model, and significantly less than SM. The individual input and output elasticities were also very similar. The output-oriented measures differed more, with the most scale economies appearing for LG farms and the least for SM; the individual netput elasticities also were somewhat perverse (including some large negative values) for the RES and SM farms.

Table 4
Primary measures, first and last years, off-farm

	Output				Input		
	1996	1998	2000		1996	1998	2000
$\varepsilon_{Y,X}$	1.120	1.083	1.051	$\varepsilon_{X,Y}$	0.686	0.713	0.754
TE	0.863	0.937	0.972	TE^{I}	0.899	0.941	0.966
$\varepsilon_{Y,X_{\mathrm{L}}}$	0.403	0.339	0.288	$\varepsilon_{X,X_{\mathrm{L}}}$	−0.342	−0.310	−0.288
$\varepsilon_{Y,X_{\mathrm{E}}}$	−0.043	0.013	0.002	$\varepsilon_{X,X_{\mathrm{E}}}$	0.008	0.002	−0.026
$\varepsilon_{Y,X_{\mathrm{F}}}$	0.062	0.125	0.134	$\varepsilon_{X,X_{\mathrm{F}}}$	−0.075	−0.091	−0.132
$\varepsilon_{Y,X_{\mathrm{SD}}}$	0.205	0.202	0.157	$\varepsilon_{X,X_{\mathrm{SD}}}$	−0.143	−0.152	−0.123
$\varepsilon_{Y,X_{\mathrm{FD}}}$	0.155	0.137	0.135	$\varepsilon_{X,X_{\mathrm{FD}}}$	−0.105	−0.087	−0.105
$\varepsilon_{Y,X_{\mathrm{A}}}$	0.063	0.063	0.064	$\varepsilon_{X,X_{\mathrm{A}}}$	−0.043	−0.054	−0.028
$\varepsilon_{Y,X_{\mathrm{C}}}$	0.197	0.112	0.140	$\varepsilon_{X,X_{\mathrm{C}}}$	−0.069	−0.051	−0.028
$\varepsilon_{Y,X_{\mathrm{O}}}$	−0.011	−0.111	0.008	$\varepsilon_{X,X_{\mathrm{O}}}$	−0.082	−0.020	−0.065
$\varepsilon_{Y,X_{\mathrm{K}}}$	−0.075	0.070	0.045	$\varepsilon_{X,X_{\mathrm{K}}}$	0.021	−0.049	−0.044
$\varepsilon_{Y,X_{\mathrm{LD}}}$	0.165	0.132	0.079				
				$\varepsilon_{X,Y_{\mathrm{C}}}$	0.151	0.164	0.136
$\varepsilon_{Y,Y^*_{\mathrm{S}}}$	−0.173	−0.180	−0.186	$\varepsilon_{X,Y_{\mathrm{S}}}$	0.115	0.131	0.148
$\varepsilon_{Y,Y^*_{\mathrm{O}}}$	−0.108	−0.111	−0.119	$\varepsilon_{X,Y_{\mathrm{O}}}$	0.089	0.090	0.097
$\varepsilon_{Y,Y^*_{\mathrm{A}}}$	−0.346	−0.339	−0.336	$\varepsilon_{X,Y_{\mathrm{A}}}$	0.241	0.236	0.239
$\varepsilon_{Y,Y^*_{\mathrm{I}}}$	−0.203	−0.197	−0.205	$\varepsilon_{X,Y_{\mathrm{I}}}$	0.088	0.087	0.085
$\varepsilon_{Y,\mathrm{GM_C}}$	0.003	0.003	0.002	$\varepsilon_{X,\mathrm{GM_C}}$	−0.001	−0.001	−0.001
$\varepsilon_{Y,\mathrm{GM_S}}$	0.00002	0.0003	0.0004	$\varepsilon_{X,\mathrm{GM_S}}$	0.0005	0.0005	0.001

types, as one would expect, although the efficiency scores vary little (especially for the input model) and both models indicate that returns to scale are slightly decreasing, and efficiency appreciably increasing, over time.[26]

Other interesting implications about input and output contributions, such as a lower share or contribution of X_{L} for farms in the larger cohorts, but higher contributions of X_{F}, X_{FD}, X_{A} and X_{K}, are also apparent. The productive contribution of labor seems also to be decreasing over time, along with X_{SD} and X_{C}, and that of X_{E}, X_{F}, and X_{K} to be increasing. The output patterns are less definitive; the output contributions of Y_{C}, Y_{O} and Y_{A} are larger for CORP than any other cohorts, but since $\varepsilon_{X,Y}$ is also smaller the individual output elasticities from the input model are not directly comparable. If deflated by overall scale economies, there is no clear cohort-specific pattern of ε_{X,Y_m} elasticities, as is also true over time (although the share of soybeans is rising). Note also that the (positive) productive

[26]Separate estimation by year resulted in primarily insignificant and more variable parameter estimates (especially for the output specification), such as decreasing and then increasing scale economies over time (driven largely by X_{SD}, X_{FD}, and X_{O} for the output model and Y_{S} for the input model), although the overall result of significant scale economies remained in all years. The returns to $\mathrm{GM_C}$ and $\mathrm{GM_S}$ also were larger in magnitude and varied more across years and specification.

impacts of planting GM corn are even less substantive for farms in the larger as compared to the smaller cohorts.

Some further insights may be obtained from the coefficients on the R_j factors, from Table 6 of Appendix A, although some are difficult to interpret. First, the coefficients on the time dummies for the output model suggest that overall productivity has been declining ($\varepsilon_{Y,t} = \alpha_t < 0$, $t = 1997$–2000), which if interpreted in terms of technical change implies (statistically significant) technical regress. The input model, however, suggests some (statistically insignificant) increase in productivity, at least in the first half of the sample. Although one might expect little evidence of technical progress over such a short time frame, particularly when the time dummies are picking up all year-specific external factors (such as weather) that affect production, this again suggests that the output-oriented model may have limitations for representing production processes when output composition differences are an important performance driver. Estimates for both models also imply that the larger cohorts produce less output per unit of input overall. This result is somewhat perverse, although it could reflect very different input and output mixes (more physical and livestock capital requirements, for example, or higher-valued outputs) for the larger farms.

The remaining farm/farmer characteristics have little significant (or consistent) impact. For example, older farmers appear to be more (but not statistically significantly) productive. Additional debt (relative to equity) also seems counter-productive, although only significantly for the input specification and a greater proportion of rented land seems from the output—but not the input—specification to contribute to performance.

5. Concluding remarks

In this paper, we estimate and evaluate measures of economic performance, focusing on scale economies and their underlying output and input composition patterns, for farmers in the U.S. Corn Belt. The alternative output and input distance function specifications used for analysis have somewhat different perspectives and provide slightly different implications, but generate consistent messages. This consistency is somewhat surprising, given the asymmetric treatment of output and input relationships, but is encouraging for the econometric implementation of distance function models.

Overall, we find strong scale economies and output jointness, high and increasing efficiency levels, and similar input and output contributions across both specifications. Scale and scope (output diversification) economies thus seem to have central roles in explaining productivity and motivating growth patterns in the U.S. agricultural sector.

Some implications about preferable specifications for analysis of agricultural production processes also emerge from our estimates. In particular, our input-oriented distance function model, by contrast to the output-based specification, provides insights about the extent and distribution of scope as well as scale

economies—which is crucial with substantive output composition variability. Input mix seems relatively constant, at least within cohorts, perhaps due to both input fixities and production systems. The output-oriented framework is also less able to effectively measure the key productive contribution of off-farm income (likely because this exacerbates the problem of variable output composition, since it is particularly important for smaller farms), and suggests erroneous temporal shifts. Thus, an input-oriented model is advantageous for the characterization of agricultural production and performance, although the primal productivity literature tends to focus on output-oriented (usually production function) models that limit consideration of output composition differences.

Appendix A. Data

We use U.S. farm level data from the 1996, 1997, 1998, 1999, and 2000 Agricultural Resources Management Study (ARMS) Phase III surveys. ARMS is an annual survey covering farms in the 48 contiguous states, conducted by the National Agricultural Statistics Service, USDA, in cooperation with the Economic Research Service. The relatively homogeneous corn/soybean region in the Corn Belt was selected because it represents major corn and soybean cropping patterns where GMO use is prevalent, and where off-farm employment opportunities and urbanization trends are important. Ten corn-states are distinguished in the data: IL, IN, IA, KS, MN, MO, NE, SD, OH, and WI (Table 5).

Our four farm outputs—corn, soybeans, other crops, and livestock—are measured as total value of production. Off-farm income is measured as off-farm pay (before taxes and other withholdings, including cash wages, salaries, tips, commissions, piece rate payments, bonuses, military pay, etc.).[27]

For the inputs, labor is annual per-farm expenditures on labor; energy is expenditures on gasoline, diesel fuel and other fuels; fertilizer is expenditures on fertilizer, lime and other chemicals; seed is expenditures on seeds, livestock expenses are expenses incurred in feeding and other operating expenses in raising livestock; crop expenses are pesticides and custom services; and "other expenses" include miscellaneous operating expenses. Labor is augmented for the off-farm models by adding a wage bill for operator and spouse hours worked off-farm, valued at the hire wage rate to approximate the use of farm and off-farm labor in a multi-activity enterprise.[28] Capital machinery is measured as the annualized flow of capital services from assets (excluding land). Land is measured as an annuity based on a 20-year life and 10 percent rate of interest, and an annualized flow of services from land, valued at the quality-adjusted price of land.[29]

[27]Off-farm income used in the analysis does not include net income from operating another business or other sources of income, although such sources of off-farm income are likely to become increasingly important.

[28]The ARMS survey does not collect information on other input expenses for time spent off-farm.

[29]In efficiency analysis, spatial differences in land quality prevent the direct comparison of observed prices. Land in agricultural production is typically quite heterogeneous in terms of soil type, associated soil

Table 5
Summary Statistics, 2000, averages (values, total and each cohort)

	Full sample	RES	SM	LG	CORP
Farms	2,714	593	526	879	716
Y_C	17,890	3,078	8,684	53,175	99,765
Y_S	14,839	3,613	41,285	74,038	74,038
Y_O	17,154	1,800	6,460	40,616	143,872
Y_A	42,058	5,128	15,210	65,746	440,049
Y_I	43,854	60,930	23,225	24,216	20,534
$Y_{TOTAL,\ farm}$	94,655	14,212	72,165	234,454	758,440
Y_{TOTAL}	138,509	75,142	95,390	258,670	778,974
X_L	20,462	9,976	21,896	36,267	71,122
X_E	3,611	855	2,624	9,258	18,399
X_F	10,776	2,337	6,613	29,045	57,067
X_{SD}	5,771	1,155	3,293	14,717	35,435
X_{FD}	10,721	1,449	2,773	14,672	122,329
X_A	12,756	1,020	1,868	13,455	169,643
X_C	6,978	1,563	4,303	18,266	38,566
X_O	12,996	4,240	9,030	29,665	67,072
X_K	14,622	4,430	10,642	36,054	69,157
X_{LD}	48,373	18,648	39,653	113,671	187,499
$X_{TOTAL,var}$	84,071	22,595	52,400	165,345	579,633
X_{TOTAL}	147,066	45,673	102,695	315,070	836,289
AGE	54	54	58	49	49
ED*	3	3	2	3	3
DA	13	8	9	16	21
RNT	49	38	36	59	56
GM_C	30	20	29	30	33
GM_S	57	57	57	55	60

*ED: 1 = no high school, 2 = high school/equivalent, 3 = some college, 4 = 4 year degree, 5 = graduate school. DA is the debt-equity ratio, and RNT is the proportion of rented land.

To translate the nominal values of outputs and inputs into real terms, all variables are deflated by the estimated increase or decrease in cost of production in 1997–2000 compared to 1996 (in terms of agricultural prices) (Table 6).

For empirical production studies using panel data, the temporal pattern of a given farm's production behavior must be established. In the absence of genuine panel data, repeated cross-sections of data across farm typologies may instead be used to construct pseudo panel data (see Deaton, 1985; Verbeek and Nijman, 1992; Heshmati and Kumbhakar, 1997). Pseudo panels are created by grouping the individual observations into homogeneous cohorts, demarcated on the basis of

(*footnote continued*)
characteristics, and other productivity-related factors across states and agricultural statistics districts and counties within states. Failing to account for these differences would lead to a biased measure of the land input, and of economic measures (see, for example, Alvarez and Gonzalez, 1999).

Table 6
Coefficient estimates, off-farm specifications

Output			Input			Output			Input		
Coeff	Est	t-stat	Coeff	Est	t-stat	Coeff	Est	t-stat	Coeff	Est	t-stat
α_0	1.628	1.31	α_0	5.007	8.49	$\beta_{X_A X_O}$	−0.003	−0.16	$\beta_{X_{A^*} X_{O^*}}$	0.021	1.37
						$\beta_{X_A X_K}$	−0.021	−1.02	$\beta_{X_{A^*} X_{K^*}}$	−0.006	−0.41
α_{X_L}	0.137	0.62	$\alpha_{X_L^*}$	−0.056	−0.90	$\beta_{X_A X_{LD}}$	−0.006	−0.40			
α_{X_E}	0.247	0.31	$\alpha_{X_E^*}$	−0.011	−0.24	$\beta_{X_C GM_C}$	0.003	2.33			
α_{X_F}	0.321	0.43	$\alpha_{X_F^*}$	−0.124	−2.01	$\beta_{X_C GM_S}$	0.000	−0.40			
$\alpha_{X_{SD}}$	1.716	3.12	$\alpha_{X_{SD}^*}$	−0.159	−3.24	$\beta_{X_C X_O}$	−0.266	−6.60	$\beta_{X_{C^*} X_{O^*}}$	0.132	4.55
$\alpha_{X_{FD}}$	−0.089	−0.42	$\alpha_{X_{FD}^*}$	−0.005	−0.13	$\beta_{X_C X_K}$	0.032	1.10	$\beta_{X_{C^*} X_{K^*}}$	−0.061	−2.73
α_{X_A}	−0.050	−0.32	$\alpha_{X_A^*}$	−0.138	−4.72	$\beta_{X_C X_{LD}}$	0.118	3.02			
α_{X_C}	−1.246	−2.97	$\alpha_{X_C^*}$	−0.030	−0.47	$\beta_{X_O X_K}$	0.185	4.26	$\beta_{X_{O^*} X_{K^*}}$	−0.149	−4.85
α_{X_O}	0.194	0.24	$\alpha_{X_O^*}$	0.036	0.46	$\beta_{X_O X_{LD}}$	−0.056	−0.83			
α_{X_K}	0.117	0.15	$\alpha_{X_K^*}$	−0.183	−2.58	$\beta_{X_K X_K}$	−0.148	−5.02			
$\alpha_{X_{LD}}$	−0.427	−1.45				$\beta_{X_O X_{LD}}$	−0.006	−0.12	$\beta_{X_{K^*} X_{K^*}}$	0.055	2.67
$\beta_{X_L X_E}$	0.007	0.09	$\beta_{X_L^* X_E^*}$	0.018	0.68				β_{Y_C}	−0.034	−0.57
$\beta_{X_L X_F}$	−0.034	−0.56	$\beta_{X_L^* X_F^*}$	0.027	1.02	$\beta_{Y_S^*}$	−0.198	−13.24	β_{Y_S}	−0.009	−0.16
$\beta_{X_L X_{SD}}$	−0.181	−3.06	$\beta_{X_L^* X_{SD}^*}$	0.044	1.95	$\beta_{Y_O^*}$	−0.123	−12.85	β_{Y_O}	0.074	1.55
$\beta_{X_L X_{FD}}$	−0.036	−1.45	$\beta_{X_L^* X_{FD}^*}$	0.032	2.62	$\beta_{Y_A^*}$	−0.303	−21.74	β_{Y_A}	0.062	0.79
$\beta_{X_L X_A}$	0.015	0.75	$\beta_{X_L^* X_A^*}$	0.008	0.80	$\beta_{Y_I^*}$	−0.203	−14.57	β_{Y_I}	0.115	1.49
$\beta_{X_L X_C}$	0.169	3.43	$\beta_{X_L^* X_C^*}$	−0.054	−2.35						
$\beta_{X_L X_O}$	−0.063	−1.03	$\beta_{X_L^* X_O^*}$	0.019	1.06				$\beta_{Y_C Y_C}$	0.019	8.61
$\beta_{X_L X_K}$	−0.011	−0.16	$\beta_{X_L^* X_K^*}$	0.046	2.18	$\beta_{Y_S^* Y_S^*}$	−0.043	−12.72	$\beta_{Y_S Y_S}$	0.021	6.93
$\beta_{X_L X_{LD}}$	0.130	2.67				$\beta_{Y_O^* Y_O^*}$	−0.022	−9.67	$\beta_{Y_O Y_O}$	0.019	10.54
$\beta_{X_E X_F}$	−0.024	−0.58	$\beta_{X_E^* X_F^*}$	−0.029	−0.98	$\beta_{Y_A^* Y_A^*}$	−0.024	−7.74	$\beta_{Y_A Y_A}$	0.022	7.76
$\beta_{X_E X_{SD}}$	0.034	0.89	$\beta_{X_E^* X_{SD}^*}$	−0.007	−0.30	$\beta_{Y_I^* Y_I^*}$	−0.024	−15.29	$\beta_{Y_I Y_I}$	0.008	2.54
$\beta_{X_E X_{FD}}$	−0.044	−1.62	$\beta_{X_E^* X_{FD}^*}$	0.041	1.99						
$\beta_{X_E X_A}$	0.033	1.60	$\beta_{X_E^* X_A^*}$	−0.032	−1.93				$\beta_{Y_C Y_S}$	−0.006	−1.66
$\beta_{X_E X_C}$	0.072	2.08	$\beta_{X_E^* X_C^*}$	0.019	0.62				$\beta_{Y_C Y_O}$	−0.009	−3.20
$\beta_{X_E X_O}$	−0.103	−2.01	$\beta_{X_E^* X_O^*}$	0.020	0.50				$\beta_{Y_C Y_A}$	−0.001	−0.25
$\beta_{X_E X_K}$	0.062	1.26	$\beta_{X_E^* X_K^*}$	0.014	0.39				$\beta_{Y_C Y_I}$	−0.002	−0.33
$\beta_{X_E X_{LD}}$	−0.048	−1.18							$\beta_{Y_C GM_C}$	0.001	1.40
$\beta_{X_F X_F}$	−0.010	−0.44	$\beta_{X_F^* X_F^*}$	−0.033	−2.24				$\beta_{Y_C GM_S}$	−0.001	−2.84
$\beta_{X_F X_{SD}}$	−0.032	−1.01	$\beta_{X_F^* X_{SD}^*}$	0.042	2.73	$\beta_{Y_S^* Y_O^*}$	0.026	6.21	$\beta_{Y_S Y_O}$	−0.013	−4.15
$\beta_{X_F X_{FD}}$	0.003	0.12	$\beta_{X_F^* X_{FD}^*}$	−0.024	−1.68	$\beta_{Y_S^* Y_A^*}$	0.018	4.02	$\beta_{Y_S Y_A}$	−0.011	−2.82
$\beta_{X_F X_A}$	−0.005	−0.30	$\beta_{X_E^* X_A^*}$	0.034	2.60	$\beta_{Y_S^* Y_I^*}$	0.012	3.72	$\beta_{Y_S Y_I}$	0.003	0.54
$\beta_{X_F X_C}$	0.030	1.31	$\beta_{X_F^* X_C^*}$	0.001	0.06				$\beta_{Y_S GM_C}$	−0.00001	−0.09
$\beta_{X_F X_O}$	0.294	6.33	$\beta_{X_F^* X_O^*}$	−0.137	−3.43				$\beta_{Y_S GM_S}$	0.00045	1.88
$\beta_{X_F X_K}$	−0.171	−5.31	$\beta_{X_F^* X_K^*}$	0.135	5.03	$\beta_{Y_O^* Y_A^*}$	0.008	2.04	$\beta_{Y_O Y_A}$	−0.004	−1.42
$\beta_{X_F X_{LD}}$	−0.063	−1.40				$\beta_{Y_O^* Y_I^*}$	0.002	0.72	$\beta_{Y_O Y_I}$	−0.008	−1.91
$\beta_{X_{SD} GM_C}$	−0.003	−3.26				$\beta_{Y_A^* Y_I^*}$	0.020	4.97	$\beta_{Y_A Y_I}$	−0.012	−1.93
$\beta_{X_{SD} GM_S}$	0.001	1.01									
$\beta_{X_{SD} X_{FD}}$	−0.025	−1.10	$\beta_{X_{SD}^* X_{FD}^*}$	0.049	3.44	α_{1997}	−0.179	−4.12	α_{1997}	−0.054	−1.56
$\beta_{X_{SD} X_A}$	0.009	0.65	$\beta_{X_{SD}^* X_A^*}$	−0.026	−2.41	α_{1998}	−0.214	−4.49	α_{1998}	−0.057	−1.32
$\beta_{X_{SD} X_C}$	−0.009	−0.44	$\beta_{X_{SD}^* X_C^*}$	−0.033	−2.98	α_{1999}	−0.444	−6.37	α_{1999}	0.026	0.51
$\beta_{X_{SD} X_O}$	−0.040	−1.21	$\beta_{X_{SD}^* X_O^*}$	0.060	2.23	α_{2000}	−0.468	−6.70	α_{2000}	0.002	0.04

Table 6 (*continued*)

Output			Input			Output			Input		
Coeff	Est	*t*-stat	Coeff	Est	*t*-stat	Coeff	Est	*t*-stat	Coeff	Est	*t*-stat
$\beta_{X_{SD}X_K}$	0.194	6.60	$\beta_{X^*_{SD}X^*_K}$	−0.093	−4.96	α_{SM}	−0.185	−4.35	α_{SM}	0.111	3.46
$\beta_{X_{SD}X_{LD}}$	−0.079	−2.09				α_{LG}	−0.091	−2.05	α_{LG}	0.131	3.89
$\beta_{X_{FD}X_A}$	0.010	2.00	$\beta_{X^*_{FD}X^*_K}$	−0.010	−1.69	α_{CORP}	−0.105	−1.74	α_{CORP}	0.183	3.77
$\beta_{X_{FD}X_C}$	−0.009	−0.52	$\beta_{X^*_{FD}X^*_C}$	0.000	0.03	α_{AGE}	0.000	−0.15	α_{AGE}	−0.002	−1.62
$\beta_{X_{FD}X_O}$	0.022	0.87	$\beta_{X^*_{FD}X^*_O}$	−0.004	−0.18	α_{ED}	−0.027	−1.18	α_{ED}	0.010	1.31
$\beta_{X_{FD}X_K}$	0.048	2.06	$\beta_{X^*_{FD}X^*_K}$	−0.025	−1.44	α_{DA}	−0.001	−0.72	α_{DA}	0.003	2.82
$\beta_{X_{FD}X_{LD}}$	0.044	2.14	$\beta_{X^*_{FD}X^*_{FD}}$	−0.010	−1.61	α_{GM_C}	0.002	2.46	α_{RNT}	0.000	0.67
						α_{GM_C}	0.008	2.56	α_{GM_C}	−0.006	−2.34
$\beta_{X_AX_C}$	−0.016	−0.85	$\beta_{X^*_AX^*_C}$	−0.003	−0.26	α_{GM_S}	−0.003	−1.70	α_{GM_S}	0.001	0.93

Table 7
The farm typology groups

Small Family Farms (sales less than $250,000)
1. Limited resource: Any small farm with: gross sales less than $100,000, total farm assets less than $150,000, and total operator household income less than $20,000. Limited-resource farmers may report farming, a nonfarm occupation, or retirement as their major occupation.
2. Retirement: Small farms whose operators report they are retired (excludes limited-resource farms operated by retired farmers).
3. Residential/lifestyle: Small farms whose operators report a major occupation other than farming (excludes limited-resource farms with operators reporting a nonfarm major occupation).
4. Farming occupation/lower-sales: Small farms with sales less than $100,000 whose operators report farming as their major occupation (excludes limited-resource farms whose operators report farming as their major occupation).
5. Farming occupation/higher sales: Small farms with sales between $100,000 and $249,999 whose operators report farming as their major occupation.

Other farms
6. Large family farms: Sales between $250,000 and $499,999.
7. Very large family farms: Sales of $500,000 or more.
8. Nonfamily farms: Farms organized as nonfamily corporations or cooperatives, as well as farms operated by hired managers.

Source: U.S. Department of Agriculture, Economic Research Service (USDA/ERS). "Farm Resource Regions." Agricultural Information Bulletin #760, September 2000. http://www.ers.usda.gov/publications/aib760/aib-760.pdf.

common observable time-invariant characteristics such as location, quality or size of land, or scope of agricultural activities relative to off-farm activities. The subsequent economic analysis then uses the cohort means rather than the individual farm-level observations.

The farm typology groups recently developed at the ERS and described in Table 7 of Appendix A allow us to assign our farm-level data to cohorts by typology and sub typology, by state, by year. The data in typologies 1–3 (limited resource, retirement,

Table 8
Cohort definitions

Small farms			Large farms		
Cohort	Typology	GV Sales	Cohort	Typology	GV Sales
COH1	1–3	<2,499	COH9	6	250,000–330,000
COH2	1–3	2,500–29,999	COH10	6	330,000–410,000
COH3	1–3	>30,000	COH11	6	>410,000
COH4	4	<10,000	COH12	7–8	<1,000,000
COH5	4	10,000–29,999	COH13	7–8	>1,000,000
COH6	4	30,000–100,000			
COH7	5	100,000–174,999			
COH8	5	175,000–249,999			

For our fixed effects: RES = COH1-3, SM = COH4–6, LG = COH7-10, and CORP = COH11–13.

and residential) are relatively limited compared to the traditional farm data in typologies 4–8. Typologies 1–3 were therefore grouped into three cohorts by level of agricultural sales.[30] Similarly, the data in typologies 4 and 6 were used to form three cohorts, while data in typologies 5 and 7 were grouped into two cohorts each. These categories are summarized in Table 8 in Appendix A. The resulting panel data set consists of 13 cohorts by state, for 1996–2000, measured as the weighted mean values of the variables to be analyzed. In total we have 650 annual observations (130 per year, a balanced panel), summarizing the activities of 2127 farms in 1996, 4305 in 1997, 2479 in 1998, 3593 in 1999, and 2714 in 2000.

A summary of our data for 2000 is presented in Table 5 of Appendix A.[31] The table categorizes the output and input data and the characteristics by the four cohort groups we have distinguished for presentation of the results—residential farms (RES, cohorts 1–3), small family farms (SM, cohorts 4–6), larger family farms (LG, cohorts 7–10), and very large and nonfamily farms (CORP, cohorts 11–13).

References

Aigner, D.J., Lovell, C.A.K., Schmidt, P., 1977. Formulation and estimation of stochastic frontier production function models. Journal of Econometrics 6, 21–37.

Alexander, C., Goodhue, R.E., 2002. The pricing of innovations, an application to specialized corn traits. Agribusiness, Summer, pp. 333–349.

Alexander, C., Fernandez-Cornejo, J., Goodhue, R.E., 2001. Effects of the GMO controversy on corn-soybean farmers acreage allocation decisions. Manuscript. University of California, Davis.

[30]If assets instead of sales are used to distinguish cohorts, to use a more clearly time-invariant characteristic, the results for scale economies and input and output contributions are negligibly affected. However, the estimated impacts of farm/farmer characteristics differ. This makes sense since the cohorts to some extent summarize differing farm characteristics, but is not a concern for our purposes since we are not focused on these factors and do not find them at all definitive empirically.

[31]Summaries for other years are available by request from the authors.

Alvarez, A.M., Gonzalez, E., 1999. Using cross-section data to adjust technical efficiency indexes estimated with panel data. American Journal of Agricultural Economics 81, 894–901.

Battese, G.E., Coelli, T.J., 1992. Frontier production functions, technical efficiency and panel data, with application to paddy farmers in India. The Journal of Productivity Analysis 3, 153–169.

Baumol, W.J., Panzar, J.C., Willig, R.D., 1982. Contestable Markets and the Theory of Industry Structure. Harcourt Brace Jovanovich, New York.

Coelli, T.J., Perelman, S., 2000. Technical efficiency of European railways. A distance function approach. Applied Economics 32, 1967–1976.

Coelli, T.J., Rao, D.S.P., Battese, G.E., 1998. An Introduction to Efficiency and Productivity Analysis. Kluwer Academic Publishers, Boston.

Deaton, A., 1985. Panel data from time series cross-sections. Journal of Econometrics 30, 109–126.

Färe, R.S., 1988. Fundamentals of Production Theory. Springer, Berlin.

Färe, R.S., Grosskopf, S., 1990. A distance function approach to price efficiency. Journal of Public Economics 43, 123–126.

Färe, R.S., Primont, D., 1995. Multi-Output Production and Duality, Theory and Applications. Kluwer Academic Publishers, Boston.

Färe, R.S., Grosskopf, S., Lovell, C.A.K., 1994. Production Frontiers. Cambridge University Press, Cambridge.

Fernandez-Cornejo, J., McBride, W.D., 2000. Genetically engineered crops for pest management in U.S. agriculture, farm-level effects. Agricultural Economic Report No. 786, USDA/ERS.

Gardner, B.L., 2001. How U.S. agriculture learned to grow, causes and consequences, Manuscript, University of Maryland.

Heshmati, A., Kumbhakar, S.C., 1997. Estimation of technical efficiency in Swedish crop farms, a pseudo panel data approach. Journal of Agricultural Economics 48 (1), 22–37.

Just, R.E., Pope, R.D., 1978. Stochastic specification of production functions and economic implications. Journal of Econometrics 7, 67–86.

Lovell, C.A.K., Richardson, S., Travers, P., Wood, L.L., 1994. Resources and functionings. In: Eichhorn, W. (Ed.), A New View of Inequality in Australia, In Models and Measurement of Welfare and Inequality. Springer, Berlin.

Marra, M.C., 2001. The Farm Level Impacts of Transgenic Crops, A Critical Review of the Evidence, Chapter 8 in Agricultural Biotechnology, Markets and Policies in an International Setting, Johns Hopkins Press and International Food Policy Research Institute, Baltimore, pp. 155–184.

Meeusen, W., van den Broeck, J., 1977. Efficiency Estimation from Cobb–Douglas Production Functions with Composed Error. International Economic Review 18, 435–444.

Mundlak, Y., 1996. Production Function Estimation, Reviving the Primal. Econometrica 64 (2), 431–438.

Newberry, D.M., Stiglitz, J.E., 1981. The Theory of Commodity Price Stabilization, A Study in the Economics of Risk. Clarendon Press, Oxford.

Paul, C.J.M., Johnson, W., Frengley, G., 2000. Efficiency in New Zealand sheep and cattle farming. The Impacts of Regulatory Reform, Review of Economics and Statistics 82 (2), 325–337.

Pope, R.D., 1987. A discussion. In: Kilmer, R.L., Armbruster, W.J. (Eds.), Economic Efficiency in Agricultural and Food Marketing. Iowa State University Press, Iowa, pp. 63–66.

Ramaswami, B., 1992. Production risk and optimal input decisions. American Journal of Agricultural Economics 74, 860–869.

Schultz, T.W., 1950. Reflections on poverty within agriculture. Journal of Political Economy 63, 1–15.

USDA/ERS, 2001a. Structural and financial characteristics of U.S. farms. In: Hoppe, R.A. (Ed.), 2001 Family Farm Report, Resource Economics Division, Agriculture Information Bulletin No. 768.

USDA/ERS, 2001b. Food and agricultural policy, taking stock for the new century, Report, September.

Verbeek, M., Nijman, T., 1992. Can cohort data be treated as genuine panel data?''. Empirical Economics 17, 9–23.

Williams, S.P., Shumway, C.R., 1998. Testing for behavioral objective and aggregation opportunities in U.S. agricultural data. American Journal of Agricultural Economics 80, 195–207.