Appendix 1—Economies of Scale and Scope and Technical Efficiency

This discussion uses two different but related methodologies and data sets and follows the analysis described in Nehring et al., 2005. First, using 1996-2000 survey data, we use an input distance function to estimate scale economies and technical efficiency, and compare these performance measures at the farm and household levels. Second, using 2000 survey data, we set up a multi-activity cost function to analyze labor allocation decisions within the farm operator household and estimate scope economies. We interpret off-farm income-generating activities as outputs, along with corn, soybeans, livestock, and other crops. For both estimations, we use detailed survey information of the farm operator household from USDA's Agricultural Resource Management Survey (ARMS).

Economies of Scale and Technical Efficiency

The analysis of production structure and performance requires representing the underlying multi-dimensional (input and output) production technology. This may be formalized by specifying a transformation function, $T(\mathbf{X},\mathbf{Y},\mathbf{R})=0$, which summarizes the production frontier in terms of an input vector $\mathbf{X}$, an output vector $\mathbf{Y}$, and a vector of external production determinants $\mathbf{R}$. This information on the production technology can also be characterized via an input set, $L(\mathbf{Y},\mathbf{R})$, representing the set of all X vectors that can produce Y, given the exogenous factors $\mathbf{R}$.

An input distance function (denoted by superscript I) identifies the least input use possible for producing the given output vector, defined according to $L(\mathbf{Y},\mathbf{R})$:

$$(1) \quad D^{1}(\mathbf{X},\mathbf{Y},\mathbf{R}) = Max\{\rho : (x/\rho) \in L(\mathbf{Y},\mathbf{R})\}.$$

This multi-input, input-requirement function allows for deviations from the frontier. It is also conceptually similar to a cost function, if allocative efficiency is assumed, in the sense that it implies minimum input or resource use for production of a given output vector (and thus, implicitly, costs). However, it does so in a primal/technical optimization or efficiency context, with no economic optimization implied.

For the farm-level model, the $\mathbf{Y}$ vector contains Y_1 = crops (corn, soybeans, and other crops), Y_2 = livestock, and, for the household-level model, $Y_1{}^*$ = crops and livestock, and $Y_2{}^*$ = off-farm income-generating activities, as farm "outputs." With $Y_2{}^*$ included, one might think of $\mathbf{Y}$ as a multi-activity rather than a multi-output vector. The components of X are defined as X_1 = land (LD), X_2 = hired labor (L), X_3 = operator labor (including hours worked off-farm)(K), X_4 = spouse labor (including hours worked off-farm) (E), X_5 = capital (F), and X_6 = materials (M).

The scale economies measure may be computed from the estimated model via derivatives or scale elasticities: $-\varepsilon_{DIY} = -\Sigma_m \partial ln\ D^I(\mathbf{X},\mathbf{Y},t)/\partial ln\ Y_m = \varepsilon_{X1Y}$ for M outputs Y_m (similar to the treatment in Baumol et al. (1982) for a multiple-output cost model, and consistent with the output distance function formula in Färe and Primont (1995). However, the inverse measure is more comparable to the cost literature, where the extent of increasing returns or scale economies is implied by the shortfall of the measure from 1. Again, this measure is based on evaluation of (scale) expansion from a given input composition base.

The distance function can be approximated by a translog functional as follows:

$$(2)\quad \ln D^I_{it}/X_{1,it} = \alpha_0 + \alpha_t t + \Sigma_m \alpha_m \ln X^*_{mit} + 0.5\, \Sigma_m \Sigma_n \beta_{mn} \ln X^*_{mit} \ln X^*_{nit} + \Sigma_k \alpha_k \ln Y_{kit} + 0.5\, \Sigma_k \Sigma_l \alpha_{kl} \ln Y_{kit} \ln Y_{lit} + \Sigma_k \Sigma_m \delta_{km} \ln Y_{kit} \ln X^*_{mit},$$

or

$$(3)\quad -\ln X_{1,it} = \alpha_0 + \alpha_t t + \Sigma_m \alpha_m \ln X^*_{mit} + 0.5\, \Sigma_m \Sigma_n \beta_{mn} \ln X^*_{mit} \ln X^*_{nit} + \Sigma_k \alpha_k \ln Y_{kit} + 0.5\, \Sigma_k \Sigma_l \alpha_{kl} \ln Y_{kit} \ln Y_{lit} + \Sigma_k \Sigma_m \delta_{km} \ln Y_{kit} \ln X^*_{mit} - \ln D^I_{it},$$

where i denotes farm and t time period. This functional relationship, which embodies a full set of interactions among the ***X, Y*** and *t* arguments of the distance function, can more succinctly be written as: $-\ln X_{1,it} = TL(\mathbf{X}/X_1, \mathbf{Y}, t) = TL(\mathbf{X}^*, \mathbf{Y}, t)$.

The input distance function is well-suited to measure technical efficiency. For empirical estimation of technical efficiency, we append a symmetric error term, v, to equation (3) and change the notation "$- \ln D^I_{it}$" to "u." The resulting function (with the subscripts it suppressed for notational simplicity) is: $-\ln X_1 = TL(\mathbf{X}^*, \mathbf{Y}, t) + v - u$, where the term (- *u*) may be interpreted as inefficiency (as technical efficiency measures the distance from the frontier). This method is known as a stochastic frontier production function, where output of a firm is a function of a set of inputs, inefficiency (- u) and a random error v (Aigner et al., 1977; Greene, 1995, 1997, 2000).

To estimate the function, we used Coelli's FRONTIER program (Coelli, 1996), based on the error components model of Battese and Coelli (1992). Since -u represents inefficiency, the technical efficiency scores are given by $\exp(-u) = D^I(\mathbf{X}^*, \mathbf{Y}, t)$. If a firm is not technically inefficient (the firm is on the frontier), u is equal to 0 and its technical efficiency score is 1.

In the absence of genuine panel data, repeated cross-sections of data across farm typologies are used to construct a pseudo-panel data set (see Deaton, 1985; Heshmati and Kumbhakar, 1992; Verbeek and Nijman, 1993). The pseudo-panels are created by grouping the individual observations into a number of homogeneous cohorts, demarcated on the basis of their common observable time-invariant characteristics, such as location and ERS farm typology. The subsequent economic analysis then uses the cohort means rather than the individual farm-level observations. ERS farm typology categories are summarized in Nehring et al. (2005). The resulting pseudo panel data set consists of 13 cohorts by State, for 1996-2000, measured as the weighted mean values of the variables to be analyzed. There are a total of 650 annual observations (130 per year), summarizing the activities of 1,934 farms in 1996, 3,890 in 1997, 2,311 in 1998, 3,201 in 1999, and 2,394 in 2000.

Economies of Scope

When a firm produces more than one output, there is a qualitative change in the production structure that makes the concept of economies of scale developed for a single output insufficient. For multiproduct firms, production economies may arise not only because the size of the firm is increased but also due to advantages derived from producing several outputs together rather than separately. Thus, more than one measure is necessary to capture the economies (or diseconomies) related to the scale of operation (volume of output) and the economies related to the scope of the operation (composition of output or product mix). The concepts of economies of scale and scope for multiproduct firms have been developed by Panzar and Willig (1977, 1981) and Baumol et al. (1982). They have been used in agriculture by Akridge and Hertel (1986) and Fernandez-Cornejo et al. (1992).

Economies of scope measure the cost savings due to simultaneous production relative to the cost of separate production. In general, scope economies occur when the cost of producing all products together is lower than producing them separately.

Formally, consider a partition of the output set N into two (disjoint) groups T and $N\text{-}T$. Let Y_T, Y_{N-T} be the output quantity (subvector) of each of the two groups and Y_N (or simply Y) the output vector, which consists of all the outputs. The respective cost functions $C(Y_T)$, $C(Y_{N-T})$ give the minimum cost of producing the two output groups separately, and $C(Y_N)$ denotes the minimum cost of producing them together (Nehring et al., 2005).

The degree of economies of scope *(SC)* relative to the (output) set T is defined as:

(1) $$\mathrm{SC} = [\,C(Y_T)+C(Y_{N-T}) - C(Y_N)\,] \,/\, C(Y_N)$$

where SC will be positive if there are economies of scope and negative if there are diseconomies of scope. In our case, we consider the first subset of the partition to include the four conventional outputs (corn, soybeans, other crops, and livestock), $N=\{1,2,3,4\}$ and the second subset the non-conventional off-farm income-generating activities, $N\text{-}T=\{5\}$.

Farms that produce the two output groups separately are those that either produce conventional outputs with no off-farm activities or else those with off-farm work but no conventional outputs. While the sample includes farm households that produce conventional outputs and no off-farm activities, it technically does not include household with zero traditional outputs. However, the sample does include many farm households with very small revenues from traditional outputs because, for statistical purposes, a U.S. farm is currently defined as "any place from which $1,000 or more of agricultural products were sold or normally would have been sold during the year under consideration." (USDA, 2005).

The well-developed restricted cost function is used to estimate the scope economies. Consider n outputs, m variable inputs, and s fixed inputs and other exogenous factors such as location or weather proxies; $\boldsymbol{Y} = (Y1,...Yn)'$

denotes the vector of outputs, $\boldsymbol{X} = (X1,...Xm)'$ denotes the vector of variable inputs, $\boldsymbol{Z} = (Z_1,...Zs)'$ is the vector of non-negative quasi-fixed inputs and other (exogenous) factors, and $\boldsymbol{W} = (W1,...Wm)'$ denotes the price vector of variable inputs. The restricted profit function is defined by:

$$(3) \qquad C(W,Y,Z) = Min\ [W'X : \in\ T].$$

Under the usual assumptions on the technology (production possibilities set *T*), the restricted cost function is well defined and satisfies the usual regularity conditions.

Using a normalized quadratic variable cost function, which can be viewed as a second-order Taylor series approximation to the true cost function, we obtain:

$$(4) \quad C(W,Y,Z) = a_0 + (a'b'c')\begin{bmatrix} W \\ Y \\ Z \end{bmatrix} = 1/2(W'Y'Z')\begin{bmatrix} B & E & F \\ E' & C & G \\ F' & G' & D \end{bmatrix}\begin{bmatrix} W \\ Y \\ Z \end{bmatrix}$$

where ***W*** is a vector of normalized variable input prices, a_0 is a scalar parameter, and ***a, b,*** and ***c*** are vectors of constants of the same dimension as ***W, Y,*** and ***Z***. The parameter matrices ***B, C,*** and ***H*** are symmetric and of the appropriate dimensions. Similarly, ***E, F,*** and ***G*** are matrices of unknown parameters.

Using Shephard's lemma, we obtain the demand functions for variable inputs which is estimated together with the cost function. We consider five outputs *Y* (corn, soybeans, other crops, livestock, and operator and spouse off-farm labor), five inputs *X* (hired labor, operator labor, spouse labor, miscellaneous inputs, and pesticides), and use the pesticides price as the *numeraire*. In addition the cost function is specified with two exogenous factors (Nehring et al., 2005).

The normalized quadratic variable cost function and the four cost-share equations are estimated in an iterated seemingly unrelated regression (ITSUR) framework using data for year 2000. The adjusted R^2's were 0.99 for the quadratic cost function, 0.26 for the hired labor input, 0.21 for the operator labor equation, 0.30 for the spouse labor equation, and 0.60 for the miscellaneous inputs equation. However, 48 percent of coefficients for the joint estimates are significant at the 10 percent level.

The own-price effects for the inputs exhibit the expected negative signs. The own-price effect for hired labor is significant at the 10-percent level, while the own-price effects for operator labor and spouse labor are not significant in this cross-section. The own price elasticity of demand for hired labor is highly elastic, with a value of -2.62. In contrast, the own- price elasticities of demand for operator and spouse labor are highly inelastic, with values of -0.105 and -0.283. These results, however, are not directly comparable with cost function studies in the literature that do not include off-farm income-generating activities as an output.