

**An Instrumental Variables Panel Data Approach to
Farm Specific Efficiency Estimation**

Robert Gardner
Department of Agricultural Economics
Michigan State University

1998 American Agricultural Economics Association Meeting
Salt Lake City, Utah

An Instrumental Variables Panel Data Approach to Farm Specific Efficiency Estimation

A benefit of combining time series and cross sectional data is the ability to account for unobservable individual effects. These effects are inherently part of the error term and can bias a statistical model should a relationship exist with the explanatory variables. For example, land quality differences would be an unobservable individual effect that could bias a model where land was an explanatory variable.

The fixed effects panel data model is a common method of handling such a bias (Mundlack). This model, in essence, demeans the data on a firm by firm basis which removes all time invariant information and expresses it as part of individual firm constants. Since individual firm effects are time invariant, this demeaning removes them from the error term, effectively eliminating any related biasedness.

This fixed effects or within-groups estimation procedure is not fully efficient due to lost information contained in the eliminated means. Variation across individuals is ignored and parameter estimation of time invariant variables is impossible as these variables are eliminated in the demeaning process. Furthermore, Hausman and Taylor state that errors in measured variables are exacerbated by this data transformation.

A solution to these problems is provided by an estimator that is, in essence, a hybrid of the fixed and random effects panel data models (Gardner). Here, an instrumental variables model is specified with an instrument set containing the demeaned data, as in a fixed effects model, plus the weighted means of only the data found not to cause a biased regression. The strength of this technique is the ability to obtain parameter estimates for relevant time invariant variables using information otherwise lost in the firm constants of a fixed effects model. This model allows efficiency to be estimated more accurately from the resulting firm constants as only information related to individual effects will remain (Schmidt and Sickles).

Statistical Model and Estimation Procedure:

A log-linearized Cobb Douglas production function gives estimates of input elasticities (the parameters β) which, assuming constant returns to scale, should sum to unity in a correctly specified model. To obtain an instrumental variables specification of this, the Gauss-Markov estimator, θ , must be identified (Gardner). In the balanced panel data case, this is the minimum variance matrix-weighted average of the within (σ_v) and between (σ_μ) group estimators as shown in (1).

$$(1) \quad \mathbf{2} = \sqrt{\mathbf{F}_v^2 / (\mathbf{F}_v^2 + T\mathbf{F}_\mu^2)}$$

$\text{Cov}(u_{it}|X_{it}, Z_i) = \sigma_v^2 \mathbf{I}_{TN} + T\sigma_\mu^2 \mathbf{P}_v$ is the familiar block-diagonal matrix elicited from the standard panel data model where Q_v transforms a vector of observations into a vector of deviations from group means and P_v produces a vector of group means. With a little manipulation,

$$(2) \quad = \mathbf{F}_v^2 (Q_v + \frac{1}{2^2} P_v) \text{ becomes: } \mathbf{S}^{-1/2} = \frac{1}{\mathbf{F}_v} (Q_v + \mathbf{2}I$$

As Hausman points out, transforming the traditional panel data model by $\mathbf{S}^{-1/2}$ is equivalent to simple differencing of the observations as shown in (3) and (4).

$$(3) \quad \mathbf{S}^{-1/2}\mathbf{Y} = \mathbf{S}^{-1/2}\mathbf{X}_{it}\boldsymbol{\beta} + \mathbf{S}^{-1/2}\mathbf{Z}_i(\boldsymbol{\theta} + \mathbf{S}^{-1/2}\mathbf{u}_{it})$$

which, by substitution with (2), becomes:

$$(4) \quad (Q_v + \mathbf{2}P_v)Y_{it} = (Y_{it} - \bar{Y}) + \mathbf{2}\bar{Y} = Y_{it} - (1 - \mathbf{2})\bar{Y}$$

with X_{it} , the time varying explanatory variable, and u_{it} , the error term, being transformed similarly and Z_i , the time invariant explanatory variable, becoming θZ_i because of its time invariance.

An instrumental variables model is then specified with the appropriate set of instruments $\{Q_v, \theta X^*, \theta Z^*\}$; where X^* and Z^* are the θ -weighted means not related to individual firm effects (as identified below) and used for instruments to reproduce the θ -weighted means of the regressors. In the balanced panel case, where θ is constant, omission of this weighting is inconsequential.

In the unbalanced case, the model will be transformed on a firm by firm basis:

$$(5) \quad \mathbf{S}^{-1/2}Y_{it} = Y_{it} - (1 - \mathbf{2}_i)\bar{Y}_i = (Y_{it} - \bar{Y}) + \mathbf{2}_i\bar{Y}_i$$

where the Gauss-Markov estimator, θ , from (1) will now be expressed as θ_i , varying between firms depending on the individual number of observations.

$$(9) \quad \mathbf{2}_i = \sqrt{\mathbf{F}_v^2 / (\mathbf{F}_v^2 + T_i \mathbf{F}_u^2)}$$

Again, an instrumental variables estimator is sought with the appropriate set of instruments being $\{Q_v, \theta_i X^*, \theta_i Z^*\}$. In the unbalanced panel case, where θ_i is not constant across firms, this firm specific weighting is necessary to elicit an accurate model.

Hausman developed a test to identify the appropriate set of instruments (especially X^* and Z^* , identified above). Here, parameter estimates of the instrumental variables model are compared against the corresponding fixed effects model estimates which are, by definition, unbiased. Using this test, only the weighted means of the variables not causing the instrumental variables model to deviate statistically from the unbiased fixed effects model will be added as instruments.

Once this model has been formulated, parameter estimates can be obtained for time invariant variables and firms can be ranked by efficiency (Schmidt and Sickles). Efficiency ranking is done by identifying the corresponding individual firm constants, α_i , and then subtracting each of these constants from the highest value firm constant as shown in (7).

$$(7) \quad u_i = \alpha^* - \alpha_i \text{ where: } \alpha^* = \max(\alpha_i)$$

The most efficient firm is assumed to be 100% efficient.

Data and Model Specification

Institution of the North American Free Trade Act (N.A.F.T.A.) and the Generalized Agreement on Tariffs and Trade (G.A.T.T.) has spawned much discussion on the competitiveness

of American and Canadian dairy farms in light of increased trade. This is because considerable differences exist between the United States and Canadian dairy industries as a result of differences in policy and environment.

In the United States, the dairy sector has been facing falling prices, forcing farmers to be even more competitive by increasing herd size and productivity. On the other hand, Canadian dairy farmers produce milk under a quota system that is costly to obtain but guarantees milk prices to quota holders. This milk price is considerably higher than those received by American farms, even taking into account the exchange rate. Consequently, Canadian dairy markets are sought after feverishly by United States milk marketers.

Unfortunately for these American marketers, the quota system also restricts growth in the Canadian dairy sector and does little to promote increased efficiency of milk production. In fact, most Canadian expansion is in nondairy enterprises resulting in a lower degree of specialization and increased land bases. Consequently, Canadian dairy farmers are less competitive than their American counterparts and therefore unwilling to allow American milk across their borders.

To study this situation, a panel data base of dairy farm financial and production records was constructed from various sources from New York and Pennsylvania (Ford, et.al.) and the bordering Canadian province of Quebec (Gardner and Oehmke). Variables of interest were precisely defined to ensure uniformity across regions and to allow accurate comparisons of each area's dairy sectors.

A log-linearized Cobb-Douglas production function was then specified for each region with output expressed as logarithm of gross revenue and inputs expressed as logarithms of capital, materials and labor. The Capital variables were average total assets (less values of quota in the Quebec sample, cows and land), number cows, and acres land. The Materials variable was total

variable cost minus interest, insurance, tax and hired labor costs while labor was expressed as total hours each of operator, family and hired labor.

Given the unique nature of dairy operations, such factors as managerial ability and land and cow quality are unobservable firm specific effects that can cause biasedness in a Ordinary Least Squares or a Generalized Least Squares (Random Effects) specification. A Fixed Effects model can alleviate this problem but excludes variables such as tillable land and cow numbers which are time invariant or nearly so. Thus, the instrumental variables specification is warranted.

RESULTS:

In table 1, results of the instrumental variables model, as applied to the different regions, are compared. Individual effects caused no biasedness in the New York farms and consequently all data means were added (in effect, a Generalized Least Squares or Random Effects model was used). In the Pennsylvania farms, only operator and family labor means could be added to the data set without causing a biasedness while analysis of the Quebec data set showed individual effects were related only to the capital and materials variables.

Overall, in the United States' sample, operator labor and materials contributed more (had a higher elasticity of production) to the production of gross revenues than in the Quebec sample where these parameter estimates were not significantly different than zero. Family labor, hired labor, capital and land contributed more in the Quebec sample than in the American farms.

Table 2 looks at the correlation of efficiency rankings generated from each model with profit as well as herd size, degree of specialization (as determined by percent milk sales of total sales), and observed growth in equity and herd size. This allowed assessment of the viability of this type of

efficiency estimation and its relation to dairy structure.

As shown, the United States' samples had a higher correlation between efficiency and profit than the Quebec sample. Furthermore, only the United States farms showed a strong positive correlation between efficiency and herd size. However, New York and Quebec models showed little correlation between degree of efficiency and specialization while efficiency was negatively correlated with specialization in the Pennsylvania sample.

Growth in equity and herd size were positively correlated with efficiency in all cases. In the American sample, there was a stronger relationship between efficiency and herd growth than equity growth which was the opposite of that seen in the Quebec sample. The hypothesis that each correlation was equal to zero was rejected universally with 95% confidence using a t-test.

Conclusions

Overall, the instrumental variables panel data model presented here allowed efficiency estimation without the drawbacks of an Ordinary Least Squares, Generalized Least Squares or Fixed Effects specification. Parameter estimates of time invariant variables, normally unobtainable in a fixed effects specification, could be obtained without incurring a biased model. As a result, only information related to unobservable firm effects were used to elicit efficiency ratings of farms in each data set.

The statistical equivalence of the New York fixed effects and random effects models may have been reflective of the homogeneity of this particular data set. Conversely, the Pennsylvania and Quebec samples contained farms that were more diverse in size and degree of specialization in dairy. This increased diversity would be more apt to impact such variables as capital and material in an

unobservable manner and prevent the inclusion of their means in the instrument set.

In regards to the actual models, the United States' models were distinctively similar despite coming from different data sets and especially when compared to the Canadian model. The American samples showed a decreased reliance on capital and an increased reliance on materials in the production of total revenues. This may be reflective of increases in herd sizes and productivity resulting in a more efficient use of capital and an increase in material use. The Quebec model reflected the more diverse operations where total revenues were produced more capital intensively; especially through cropping enterprises where the land parameter estimate became significant.

Lastly, efficiency estimated this way seemed to be related to dairy structure considerations. The more efficient farms were more profitable and exhibited the most growth. The United States' samples had a stronger relationship between efficiency and herd growth and a weaker relationship with equity growth than the Canadian sample. Unconstrained by a quota system, efficient United States' farmers focused on herd expansion to maintain revenues during falling prices. This expansion probably was debt financed, possibly explaining the lower correlation between efficiency and equity growth. Efficient Canadian farmers concentrated expansion on nondairy enterprises, which may have explained the increased emphasis on equity growth over herd growth.

References

Ford, S.A., R.W. Gardner, S. Gripp, S. Harsh, W. Knoblauch,

A. Novakovic, L. Putnam, M. Stephenson, A. Weersink, and

R. Yonkers. *A Descriptive Analysis of the Characteristics and Financial Performance of Dairy Farms in Michigan, New York, Ontario, Pennsylvania and Wisconsin*. Cornell

University Research Bulletin 96-08. Gardner, Robert W. "Unobservable Individual Effects in Unbalanced Panel Data." *Econ. Letters* 68(1998):24-8.

Gardner, R.W. and J.E. Oehmke. *Dairy Farm Financial/Production Data Base from Lake and North-Central States and Provinces: A Description*. Michigan State University Staff Paper 97-29.

Hausman, J.A. "Specification Tests in Econometrics." *Econometrica* 46(1978): 1251-1272.

Hausman, J.A. and W.B. Taylor. "Panel Data and Unobservable Individual Effects." *Econometrica* 49(1981): 1377-1399.

Mundlak, Y. "Empirical Production Function Free of Management Bias." *J. Farm Econ.* 43(1961): 44-56.

Schmidt, P. and R. Sickles. "Production Frontiers and Panel Data." *J. Bus. and Econ. Statist.* (1984):367-370.

Table 1. Parameter Estimates for Great Lakes Dairy Farm**Panel Data Instrumental Variables Models**

Variable	MODEL		
	New York	Pennsylvania	Quebec
Constant ^a	3.88 (19.6)	4.95 (23.4)	2.29 (12.25)
Operator Lbr	.028 (3.59)	.014 (1.99)	.012 (1.12)
Family Lbr	-.0001 (.16)	.0001 (0.1)	.003 (2.15)
Hired Lbr	-.001 (1.1)	.003 (1.8)	.014 (4.18)
Capital	.024 (2.1)	.042 (2.96)	.350 (10.5)
Cows	.34 (18.2)	.36 (15.1)	.39 (9.39)
Land	.0006 (.12)	-.005 (0.8)	.056 (2.63)
Materials	.53 (35.8)	.43 (31.6)	.28 (9.73)
Hausman Test	.77198 ^b	.56552 ^b	9.11 ^b
β^c	.92	.85	1.09
F Test	1,005	25,192	16,748

^a The "constant" in this model is θ_i .

^b Failed to rejected hypothesis $\beta_{\text{within}} = \beta_{\text{O(G)LS}}$ with 95% confidence.

^c This should sum to unity with constant returns to scale.

Table 2. Correlational Analysis of Great Lake Dairy Farms
Efficiency versus Dairy Farm Characteristics

	EFFICIENCY		
	New York	Pennsylvania	Quebec
PROFIT	.69	.67	.31
% SPECIALIZED ^a	-.03	-.19	-.07
HERD SIZE	.56	.55	.05
EQUITY GROWTH	.24	.08	.26
HERD GROWTH	.33	.13	.11

^a % SPECIALIZED is measured as percent milk sales of total sales.