Productivity, Farm Size and Competitive Advantages of Small Farms: Evidence from the U.S. Dairy Industry

Motivation: Over the past few decades, milk production has shifted from traditional producing regions including the Lake States and Corn Belt towards large dairy operations in Western regions. Evolving from the significant transition, production structure and technology of the U.S. dairy industry vary greatly across regions. For instance, a majority of dairy farms in Wisconsin have less than 200 cows compared to 1,000 cows of the average California dairy farm. Some studies indicate important scale economy employed by large farms and a declining trend of small and family-oriented operations. In traditional dairy regions, small- and medium-sized dairy farms have also been found to remain profitable and active in production. Small dairy farms benefit from being more diversified and vertically integrated Summer and Wolf (2002). They tend to be more involved in producing diversified outputs besides milk. These products directly contribute to farm income or are used as intermediate inputs, which benefits farmers by mitigating risk and improving factor productivity of inputs. Although there is an extensive literature on productivity analysis of the U.S. dairy industry, some important issues are still unresolved and require careful investigation. For example, due to data limitation, two practices are commonly applied in production function estimation and productivity analysis: (i) home grown feed and family labor are ignored, and (ii) price deflated revenue is used as proxy for output quantity, which treats milk as homogenous product and thus ignores component and quality differences. These issues have important implications for estimation of production technology and productivity, especially for small dairy farms, in which homegrown feed, family labor and milk quality differential provide important competitive advantages for their survival.

The objective of the paper is two-fold. The first is to quantify the contribution of homegrown feed and family labor to farm's production and productivity. This is made possible by the newly constructed panel dataset of the dairy sector using the Agricultural Resource Management Surveys (ARMS) in 2000, 2005 and 2010. The data is constructed by following the method in Weber et al. (2016). To our knowledge, this is the first paper using detailed input and output data of ARMS to investigate dairy farm productivity in a panel data setting. Information on farm inputs such as family labor and homegrown feed allows us to investigate dynamic changes in factor productivity across dairy farms and to quantify competitive advantages of small dairy farms compared with large operations. A novel structural model is adopted to control for endogeneity in production function estimation. Secondly, employing both milk quantity and revenue measures reported in ARMS, we will estimate revenue production functions, which enables us to obtain quality-adjusted productivity measures of dairy farms in various sizes.

Empirical Models: *(1) Productivity estimation.* A farmer is assumed to produce Q units of milk with five inputs: labor (L), capital (K), materials and energy (M), feed (H) and cows (C). Consider a Cobb-Douglas production function: $q_{it} = \beta_0 + \beta_l l_{it} + \beta_k k_{it} + \beta_m m_{it} + \beta_h h_{it} + \beta_c c_{it} + \omega_{it} + \epsilon_{it}$, where q_{it} is the log of output and l_{it}, k_{it}, m_{it}, h_{it} and c_{it} are the logs of inputs. $\beta = \{\beta_0, \beta_l, \beta_k, \beta_m, \beta_h, \beta_c\}$ is a vector of parameters to estimate. ϵ_{it} represents idiosyncratic shocks. ω_{it} denotes productivity shocks, which are observed by farmers, but not by econometrician, and therefore is potentially correlated with input choices. To deal with the endogeneity problem, we adopt the methods of Levinsohn and Petrin (2003) and Ackerberg et al. (2015) to obtain consistent estimation of productivity ω_{it}.

LP uses the intermediate material demand (m_{it}) as a proxy for productivity, the optimal amount of which is determined by $m_{it} = f_t(k_{it}, c_{it}, \omega_{it})$. Assuming f_t is strictly increasing in ω_{it}, we have $\omega_{it} = f_t^{-1}(k_{it}, c_{it}, \omega_{it})$. Plugging into the production function and collecting terms, we have: $q_{it} = \beta_l l_{it} + \beta_h h_{it} + \Phi_t(k_{it}, c_{it}, m_{it}) + \epsilon_{it}$. Productivity is assumed to follow the law of motion: $\omega_{it} = g_t(\omega_{it-1}) + v_{it}$. After obtaining the productivity innovation v_{it}, the GMM method is applied to obtain estimates of β_k, β_c and β_m. To solve the identification problem that a variable input such as feed and labor may be correlated with unobserved productivity, Ackerberg et al. (2015) and Wooldridge (2009) propose to estimate all coefficients in one step with additional moment conditions.

Production functions and corresponding productivity measures (ω_{it}^{lf}, ω_{it}^{l}, ω_{it}^{f}, ω_{it}) will be estimated by using different measures of labor and feed inputs, where ω_{it}^{lf} corresponds to the productivity measure accounting for both homegrown feed and family labor, and ω_{it}^{l}, ω_{it}^{f} and ω_{it} denote the measures estimated by excluding one or both factors in the production function estimation. Comparison across the estimates will quantify the contribution of homegrown feed and family labor to farm productivity. The difference between ω_{it}^{lf} and ω_{it} also measures the magnitude of distortion of the productivity estimates reported in the existing literature, which using incomplete measures of feed and labor inputs.

(2) Revenue production function estimation with quality-adjusted productivity. Milk is priced based on its components in the US market. Following the literature, e.g., De Loecker (2011), we estimate a revenue production function, which incorporates a demand system into the production function estimation. Dairy farmers are assumed to be price-taker and therefore price of milk with the same quality is identical within a state s, which is denoted by P_{st}. Milk demand for farm i in state s is $Q_{it} = Q_{st}exp(\xi_{it})$, where Q_{st} denotes an aggregate demand shifter and ξ_{it} the unobserved shocks. Adding demand to the revenue function $R_{it} = Q_{it}P_{it}$, we have the log revenue function: $\widetilde{r}_{it} = \beta_l l_{it} + \beta_k k + \beta_m m_{it} + \beta_h h_{it} + \beta_c c_{it} + \beta_s q_{st} + \omega_{it} + \xi_{it} + \epsilon_{it}$, where $\widetilde{r}_{it} = r_{ir} - p_{st}$ is log deflated revenue. Under the state-specified demand system, unobserved prices are captured by input variations and aggregate demand q_{st}.

Farm-level prices may be affected by other factors such as milk components and regional price difference. Therefore, unobserved demand shock ξ_{it} can be decomposed into: $\xi_{it} = \xi_s + \tau\varphi_i + \widetilde{\xi}_{it}$, where ξ_s and $\widetilde{\xi}_{it}$ are unobserved state- and farm-specific demand shocks, respectively. φ_i represents milk quality-index for farm i that affects milk price received by each farm under the component-based milk pricing. To recover φ_i, we regress computed farm-specific milk price $\left(\frac{R_i}{Q_i}\right)$ on percentages of fat and protein (I_{Fati} and $I_{proteini}$) and the somatic cell counts (SCC_i), a commonly applied milk quality measure: $\frac{R_i}{Q_i} = \alpha_0 + \alpha_f I_{Fati} + \alpha_p I_{Proteini} + \alpha_{SCC1} SCC_i + \alpha_{SCC2} SCC_i^2$. With estimated coefficients, farm-specific quality index φ_i is obtained as predicted value of milk price: $\varphi_i = \hat{\left(\frac{R_i}{Q_i}\right)}$. Therefore, the revenue production function can be specified as: $\widetilde{r}_{it} = \beta_l l_{it} + \beta_k k + \beta_m m_{it} + \beta_h h_{it} + \beta_c c_{it} + \beta_s q_{st} + \sum_{s \in S(i)} \delta_s D_{si} + \delta_i \varphi_i + \omega_{it} + \epsilon_{it}$. This can be estimated following the procedure similar to LP and ACF methods described above. Quality-adjusted productivity measures will allow us to compare the productivity differences across farms and disclose competitive advantage of small farms. The hypothesis is that small farms produce milk with relative high quality in order to compete with large farms with quantity advantage.

Future Work: The original weights of the ARMS data cannot be used for the new panel dataset, which is generated by matching a farm identifier across sample years. We will figure out how to appropriately assign weights to incorporate the selection of both the survey and the panel construction. This matters as relative large dairy farms are more likely to be interviewed more than once, which may generate upward biases.

Reference: The detailed references can be found here.