Staff Papers Series

STAFF PAPER P81-17 July 6, 1981

THE MARKET AS AN EXPERIMENT: PROBLEMS OF DATA GENERATION

Willis L. Peterson

Department of Agricultural and Applied Economics

Staff Papers are published without formal review in the Department of Agricultural and Applied Economics.

University of Minnesota
Institute of Agriculture, Forestry and Home Economics
St. Paul, Minnesota 55108

June 1981

THE MARKET AS AN EXPERIMENT: PROBLEMS OF DATA GENERATION

Willis L. Peterson

I. INTRODUCTION

Unlike our colleagues in the physical and biological sciences who in large part utilize data from controlled experiments to test hypotheses, we in the social sciences with the possible exception of psychology obtain much of our data by observing human behavior in uncontrolled settings. In economics this includes the observation of price and quantity variables in markets and input-output relationships in production activities. While much progress has been made during the past half century in developing and applying new measurement techniques, there appears to have been relatively little concern over whether the data that are utilized in the models are capable of providing the measurements sought. The question here is not one of data accuracy. This is not to depreciate the importance of accurate data but the "garbage-in-garbage-out" syndrome is well known and methods to assess specification bias are widely used. Rather the question at hand is one of data capability; are the observations obtained from markets or firms capable of testing the hypotheses in question? Obviously the answer is not always yes.

The main purpose of this essay is to call attention to cases where data that are normally used to estimate functional relationships may not be up to the task. Much of the discussion to follow may give the appearance of being negative or pessimistic in nature; its intended purpose is to be realistic regarding what can be expected form the data. The discussion will focus on three areas: 1. supply and demand estimation in product markets, 2. the same functions in labor markets, and 3. the estimation of production and cost functions from data obtained by observing firm behavior. While the main emphasis is on agriculture, the problems discussed are not unique to research in this industry.

II. PRODUCT MARKETS

A. Supply functions. It is well known that in order to estimate either a supply or demand function it is necessary to assume that price and quantity observations correspond to points of equilibrium. Such points cannot be observed if there are effective ceiling or support prices in existence.

Consider first support prices, which have been common in agriculture. If quantities sold are taken, the price-quantity observations will at best trace out a demand curve. If quantities produced are used, the observations may or may not trace out a supply curve. A valid supply curve cannot be measured if the price supports are accompanied by input restrictions in an effort to hold down surpluses. For example, land has been often restricted during periods of agricultural price supports. In such cases the observations may trace out a psuedo-supply curve, less elastic than would prevail in an unfettered market. As long as there are at least imperfect substitutes for a restricted input, such as fertilizer for land, output will increase with higher prices but not as cheaply and not as much as when all resources can be utilized. The resulting elasticities will understate the ability of producers to respond to higher prices in situations where such a policy is not in effect. In the case of ceiling prices, the observations may trace out a supply curve only if quantities produced are used rather than quantities sold.

In the absence of policies which maintain disequilibria in markets, probably the most troublesome aspect of supply estimation is obtaining the right kind of variation in price and quantity. In order to estimate a supply curve it is of course necessary to have variation in demand. If all the supply shifters are held constant (statistically) the various demand curves will trace out points along a given supply function. However, for products where demand is relatively stable, most of the variation in price and quantity will be due to shifts in supply, which in turn make it very difficult to obtain accurate estimates of supply elasticities. Such is the case for most agricultural products. The problem is greatest where per capita consumption of the product has declined, such as butter and fluid milk, resulting in a relatively stable total market demand over time in spite of population and income growth. In this situation it will not be possible to observe much range along a given supply curve which in turn reduces the reliability of the estimated supply elasticity.

Assuming that there is sufficient variation in price, the next question is whether the duration of the variation is long enough to bring forth a supply response. There are two aspects to this question: 1. the length of time required to convince producers that the price change is worthwhile responding to, and 2. the time it takes for producers to follow through on a change in their price expectations. For the purpose of changing production, a pattern of short term, year-to-year price fluctuations may be close to that of zero price variation, at least in the minds of producers. Increased output generally calls for increased investment. Producers will not invest unless they can be reasonably sure that price will remain at a higher level long enough to make the investment profitable. Then after expectations are changed, additional time is required to follow through on investment decisions. One might question whether the practice of using actual price in the preceding year as a proxy for expected price in the current year will capture the proper lags between price changes and the resulting changes in output in most production activities.

The use of lagged price also creates an identification problem. If current year values of the shift variables are used in the supply function, there is no way of accounting for the source of the price change in period t-1.

And the source of a price change should influence the response to it. For example, an increase in lagged price can be due to a decrease in supply, say because of bad weather in period t-1, or due to an increase in demand. Understandably the supply response should be larger in the latter situation than in the former.

Changes in the state-of-the-art or technology is another troublesome supply shifter. First there is the problem of how to measure such changes. A commonly used proxy for technology is a time trend variable. Aside from the fact that "time" does not explain anything, a problem with this procedure is that time is likely to be highly correlated with the two major demand shifters for most goods and services: population and per capita income. In order to estimate a supply function, the supply shifters must not be highly correlated with the demand shifts. If they are we end up measuring the elasticity of a dot. In this case the chances of getting a positive elasticity are not much greater than that of obtaining a zero or negative estimate.

One might question whether technology should be a supply shifter. Investment in research and development (R & D) which produces new technology is in principle no different than investment in new plant and equipment: both increase the output of the industry in question. Because changes in investment represent an important means of adjusting to product price changes it is incorrect to hold constant conventional investment when measuring supply response. If profitability considerations influence the level and direction of R & D, then technology also should be looked upon as a factor which facilitates a price response rather than as an exogenous shift variable. Granted the lag between product price changes and output changes coming from investment in R & D may be longer than for conventional investment. But it is an important source of supply response nevertheless.

The same argument can be made for public sector research. If the allocation of resources to such investment is governed by considerations of social profitability, the technology generated also can be viewed as a response to price changes. For example, the increase in both private and public sector research on energy is not unrelated to the realtive increase in the price of energy. By holding technology constant in the supply function, we, therefore, understate the response to price changes, particularly in the long run. In order for price to show an effect, those things influenced by price must be allowed to vary, otherwise the estimated supply elasticity will be constrained to zero.

Economic theory suggests that product markets should generate prices that are highly correlated between products. Consider two products Y_1 and Y_2 that are substitutes in production. In order to estimate the supply of Y_1, there must be shifts in the demand for Y_1. An increase in demand for Y_1, say, will increase its price and cause resources to be drawn away from Y_2 thereby raising its price provided it is a specific factor of production to the industry. If the markets for Y_1 and Y_2 are in equilibrium, as assumed, there will be perfect correlation between the prices of these products which in turn rules out separate estimates of their own price and cross elasticities of supply. Granted exogenous changes in variables which affect one product but not the other, or affect the two differently, may occur and therefore prevent perfect correlation between the two prices. But there is no theoretical reason why this must occur.

A similar multicollinearity problem exists between product and input prices. For example, an increase in real input prices (quality adjusted) leads to an increase in the price of the product. In this case there need not be perfect correlation between the related output prices if the products are produced with different input combinations. But there still should be perfect correlation between input and output prices providing the factor and product markets are in equilibrium, unless those exogenous changes mentioned above happen to occur.

Of course we have no way of knowing whether the absence of perfect correlation between prices of related products or inputs is due to exogenous factors or to disequilibria in product and/or factor markets. Because markets are not likely to adjust instantaneously to new equilibrium points, it behooves us to pay more attention to possible disequilibrium bias.

Disequilibria in factor markets are particularly difficult to assess because of input quality changes. What may appear to be an increase in an input price, for example, may in fact be a decrease one the observed price is adjusted for quality. Input prices are meaningless unless they are adjusted for quality.

It is well known that the existence of imperfectly competitive sellers in an industry rules out the existence of a supply curve for the industry, at least as conventionally defined. However, with shifts in demand for the product in question one can still estimate the percent change in quantity supplied resulting from each one percent change in price, i.e., a supply elasticity, providing the effort is not stymied by one or more of the problems mentioned above. The simultaneous determination of price and quantity by imperfectly competitive firms is not particularly troublesome for empirical estimation of supply elasticities because we are not interested in the level of price but in the effect of a change in price. Even though we cannot talk about supply curves of oil, steel, or automobiles, for example, we can still talk about and try to measure their supply elasticities.

B. Demand functions. Much of the preceding discussion on supply estimation also applies to demand. Ever since E. J. Working point out Henry Moore's classic mistake, it has become well known that one must hold constant the demand shifters and rely on shifts in supply to trace out points of equilibrium along a demand function (Working).

In the case of effective price controls it is not possible to observe points along a demand curve because quantity is determined by supply only. An attempt to estimate a demand curve under these conditions will produce neither a demand nor a supply function because the demand function should contain the demand shifters.

In estimating a demand function ideally one would like to have large and prolonged shifts in supply. If the supply shifts are relatively small and short lived in nature, only a small portion of the demand function can be observed and the results may not be reliable. Larger shifts in supply occur in expanding or contracting industries. Unfortunately few products exhibit both expansion and contraction in the same set of observations. If supply is increasing the observations will lie along points relatively low on the demand curve in comparison to the case where supply is decreasing. As a result the estimated demand elasticity in arithmatically linear demand functions could be affected.

Similar to supply the intercorrelation of the demand and supply shifters also presents a problem for demand estimation. In a growing economy the major demand shifts of population and per capita income changes are likely to be correlated with one or more supply shifters such as changes in number of firms (or "fixed" resources) and technology. If demand for a product is increasing, the growth in its profitability will attract new firms and/or resources and stimulate private R & D as well as public research. As a result the demand shifters may serve as proxies for the supply shifters again leaving us with the impossible task of estimating the elasticity of a dot.

As in the case of supply the duration of price changes is likely to affect the demand response to these changes. Short term changes will not likely bring forth as large as a response as long term changes of the same magnitude, particularly if the short term changes are random in nature. If price stays put for a relatively long period consumers have a better opportunity of learning of the price change and adapting their consumption patterns to it. Time for adjustment is particularly important for infrequently purchased items such as consumer durables. Therefore the estimated demand elasticity will depend on the nature of the supply shifts, in the same way that the estimated supply elasticity depends on the nature of demand shifts.

How price is measured also can affect the duration of the price change. If price is measured at a point in time such as at the beginning, in the middle, or at the end of the year, the duration of the price change can vary between one and 364 days for the same price-quantity observation. Obviously the price response should be greater in the latter situation than in the former. The averaging of monthly, weekly, or daily price quotations will mitigate this problem but may create other distortions. For example the same average price may be obtained for a period regardless of whether price is rising, falling, or constant over the period. One would expect the price response to vary across the three situations.

Whether one is estimating the demand for a specific product or a general category of products produced by imperfectly competitive firms, it is not possible to observe points of intersection between the demand curve in question and various marginal cost curves as in a perfectly competitive market. In itself this is not a problem. Abstracting from the intercorrelation problem, changes in quantity demanded resulting from changes in price (price elasticity of demand) still can be measured. It is interesting to note that it is no more or no less correct to estimate supply elasticities for goods produced by imperfectly competitive firms than it is to estimate demand elasticities for products purchased from these firms.

III. LABOR MARKETS

The existence of unemployment above the frictional or natural level implies prolonged disequilibrium in the labor market. The sources of this unemployment will determine whether or not it is possible to accurately estimate demand or supply functions for labor. Consider first the supply function.

In cases where the wage is maintained above the market equilibrium by a strong industrial union or by minimum wage laws it will not be possible to observe points along a supply curve because quantity will be determined by demand.

Even in the absence of market distortions one pecularity of the labor market may rule out accurate estimates of supply. Wage rates are notoriously sticky on the downside. Both management and labor appear to prefer a lay-off to a wage cut, at least for short-run reductions in labor demand. Consequently the price-quantity coordinate will be read off at the demand rather than supply function again making it impossible to estimate supply.

On the demand side of the labor market the existence of wage ceilings precludes the estimation of labor demand for the same reason that price ceilings rule out the estimation of product demand; points along the demand function cannot be observed.

In estimating the demand for labor, or any other inputs, there is the question of what should be held constant: the prices of other inputs or their quantities? While both are theoretically correct, the method used will influence the estimated demand elasticity. In general the elasticity will be larger if prices of other inputs are held constant (allowing their quantities to vary) than vice versa. The former would seem to be the more realistic measure of demand elasticity in the factor markets, since employers have no reason to hold constant quantities of substitutes or complementary inputs when the price of the input in question changes. Input demand elasticities derived indirectly from production functions therefore are not very realistic.

There is also the correlation problem between the labor supply and demand shifters. Probably most troublesome is the long run correlation between population growth (a major labor supply shifter) and the increase in capital and improvements in technology (major demand shifters). Consequently, this supply shifter may serve as a proxy for these two demand shifters, and vice versa, again preventing the estimation of true labor supply or demand functions.

IV. PRODUCTION RELATIONSHIPS

Most production function studies in economics are of the Cobb-Douglas tradition whereby data are obtained from observations on firm behavior in a market as opposed to data generated by controlled experiments. While there seems to be an awareness of problems stemming from simultaneous equation bias (Marschak and Andrews) and specification bias (Theil, Griliches), problems stemming from the data generating process itself appear to command less attention.

In order to obtain separate estimates of production elasticities it is necessary that firms utilize different input combinations. This means that firms must face different relative input prices. If the firms included in a cross-section sample are located in a relatively small geographic area there is little chance that the firms will in fact face different relative input prices. In this case we would expect all firms to use about the same input mix which means that only a single point is observable on a unit isoquant.

A common result of attempts to estimate production functions from such data usually is a high intercorrelation of the dependent variables with one picking up the major share of the variation.

If a person is successful in estimating a production function from such data, one should ask, why? If relative input prices have changed a short time before the cross section observations are taken, it is possible that some firms have adjusted faster than others. This may be due to differences in expectations of future prices or to differences in managerial ability.

In the case of macro production function studies where data are obtained across an entire country the common input price problem may be less troublesome. For example, in the U.S. wage rates have been lower relative to the price of capital in the South than in other sections of the country, although this difference is becoming smaller particularly when labor quality (education) is taken into account. For other intermediate inputs, the situation is not much better at the macro than at the micro level.

In spite of the problems discussed above in relation to cross section data, most production function estimates utilize such data. This is to be expected because time series data present even greater problems. In a growing economy labor tends to be inversely correlated with output and capital while all other variables tend to be highly correlated in a positive manner. The statistical results of such estimates commonly yield a negative or at least a statistically insignificant coefficient on labor with one or two of the other inputs or a time trend variable picking up the bulk at the remaining variation . Again this result is to be expected since factor markets should make input prices highly correlated the same way as product markets make product prices highly correlated.

In recent years the use of more sophisticated functional forms such as the CES and translog production functions has become more common. The well known advantage of such functions is that they allow the substitution elasticities to take on any value instead of assuming a common value of one as in the Cobb-Douglas function. The question is, are the data capable of providing measures of substitution elasticity? Probably not; at least not always.

The estimation of substitution elasticities where firms face common input prices presents a dilemma. In order to estimate this parameter it is necessary to observe points of tangency between a unit isoquant and various iso-cost lines. If firms face common input prices there will be only one such point of tangency which rules out an estimate of the substitution elasticity. If firms are observed utilizing different input mixes under such conditions, it implies disequilibria which in turn violates the necessary conditions. Hence the dilemma: if there is equilibrium the substitution elasticity cannot be estimated, but if it can be estimated due to the existence of disequilibria, it shouldn't be. The same argument holds true for the use of cost functions to estimate substitution elasticities.

In cases where there appear to be differences in relative input prices, it is crucial in measuring substitution elasticities to take account of input quality differences. Otherwise what may appear to be points of tangency along a single isoquant may in fact be points along two or more different isoquants.

To the extent that relative input price differences exist, in cross section data they are most likely to be found between capital and labor. Although capital prices should not vary between states at a point in time, as mentioned above the South has experienced somewhat lower wages than the rest of the country, at least in the past.

Of course, the mere existence of relative factor price differences does not guarantee unbiased estimates of substitution elasticities; cost minimizing equilibria must also exist. Such conditions may not exist in dynamic factor market situations. As argued in respect to supply, producers should not be expected to adjust instantaneously to price changes. The degree of adjustment to a relative price change that has taken place will likely influence the magnitude of the estimated substitution elasticity. A complete adjustment to a price change will yield a larger substitution elasticity than a partial adjustment. The degree of adjustment is likely to be a function of how long producers expect the price change to remain in effect which in turn is likely to be a function of how long the change has endured. For example, if one cuts in at a point in time five years after a relative price change has occurred, a greater adjustment (larger substitution elasticity) should be observed than if the price change is more recent. Once and for all relative factor price changes should yield larger substitution elasticities than temporary, year-to-year changes. Therefore the nature of demand and supply shifts in the factor markets can be expected to influence the size of the estimated substitution elasticities.

It should be understood, however, that an old fashioned production function, such as the Cobb-Douglass, does not require cost minimizing (or profit maximizing) conditions. As long as point of disequilibria lie along an isoquant, the production function can be estimated, as a purely physical relationship. However accurate estimates of conventional production functions do require that all firms in the sample are equal in technical efficiency, or inefficiency as the case may be. This may be one reason why estimates of aggregate production functions (where firms in a state, or county are averaged together) tend to be more "resonable" and stable than estimates from firm level observations. While technical efficiency between individual firms is likely to vary, it is less likely that on the average firms in one state will be more or less technically efficient than firms in another state. Also aggregate functions are easier to specify; variables such as the skills, motivation and ambition of the manager and other labor, which are important for individual firms but virtually impossible to measure, should average out over many firms.

Concluding Remarks

Data obtained by observing market behavior rarely meet the stringent conditions necessary to obtain unbiased estimates of demand, supply, production, and substitution elasticities. Lack of significant variation of relative prices, particularly long term variation, along with high intercorrelation of prices, and the existence of disequilibria are likely to bias empirical estimates of the above parameters. A good understanding of the industry under consideration including firm and market behavior should help prevent major errors caused by attempts to obtain estimates of parameters from data incapable of providing these estimates.