ON THE ESTIMATION OF STRUCTURAL HEDONIC PRICE MODELS*

James N. Brown

and

Harvey S. Rosen

Econometric Research Program
Research Memorandum No. 284

June 1981

*Part of Rosen's work on this paper was supported by NSF. We are grateful to John Ham and Richard Quandt for useful comments.

Econometric Research Program
Princeton University
207 Dickinson Hall
Princeton, New Jersey

I. Introduction

Many commodities can be viewed as bundles of individual attributes for which no explicit markets exist. It is often of interest to estimate structural demand and supply functions for these attributes, but the absence of directly observable attribute prices poses a problem for such estimation. In an influential paper published several years ago, Rosen [1974] proposed an estimation procedure to surmount this problem. This procedure has since been used in a number of applications (see, for example, Harrison and Rubinfeld [1978] or Witte, et. al [1979]). The purpose of this note is to point out certain pitfalls in Rosen's procedure, which, if ignored, could lead to major identification problems.

In Section II we summarize briefly the key aspects of Rosen's method as it has been applied in the literature. Section III discusses the potential problems inherent in this procedure. These problems are illustrated with a specific example in Section IV, and Section V concludes with a few suggestions for future research.

II. Rosen's [1974] Procedure

For convenience, we retain Rosen's notation. Suppose that some good is composed of n attributes, $Z_1, \ldots, Z_n$. In general, the price of this good will depend on the quantities of the various attributes of which it is composed; hence one can express this price as some function $P(Z)$. Further, by differentiating $P(Z)$ with respect to its i^{th} argument, Z_i, one can derive the market equilibrium price function for Z_i, $P_i(Z)$, that is implicit in $P(Z)$.

Now, assume that consumers' marginal willingness to pay for Z_i is some function $F_i(\cdot)$ of the quantities of all attributes as well as a vector of exogenous shift variables, Y_1. Similarly, assume the the marginal supply price of Z_i can be written as $G_i(Z,Y_2)$, where Y_2 is a vector of exogenous supply shift variables.

Interpreting $P_i(Z)$ as the (implicit) market price for attribute Z_i, the usual equilibrium assumptions lead to the following model to be estimated:

(1) $P_i(Z) = F_i(Z_1,\ldots,Z_n,Y_1)$ (demand)

(2) $P_i(Z) = G_i(Z_1,\ldots,Z_n,Y_2)$ (supply)

for $i = 1,\ldots n$. In this model, there are $2n$ equations to determine the $2n$ exogenous variables $P_1,\ldots,P_n, Z_1,\ldots,Z_n$.

Rosen suggests the following two-stage estimation procedure:

> "First, estimate $P(Z)$ by the usual hedonic method, without regard to Y_1 and Y_2. That is, regress observed differentiated products' prices, P, on all their characteristics, Z, using the best fitting functional form. This econometrically duplicates the information acquired by agents in the market, on the basis of which they make their decisions. Denote the resulting estimate of the function $P(Z)$ by $\hat{P}(Z)$. Next, compute a set of implicit marginal prices, $\partial P(Z)/\partial Z_i = \hat{P}_i(Z)$ for each buyer and seller evaluated at the amounts of characteristics (numerical values of Z) actually bought or sold, as the case may be. Finally, use estimated marginal prices $\hat{P}_i(Z)$ as endogenous variables in the second-stage simultaneous estimation of equations [1] and [2]. Estimation of marginal prices plays the same role here as do direct observations on prices in the standard theory and converts the second stage estimation into a garden variety identification problem." (Our emphasis.)

III. Some Pitfalls

Contrary to Rosen's original statement, we claim that marginal attribute prices constructed as above will not necessarily play the same role in estimation that direct observations on prices would play if they were available. Because such constructed prices are created only from observed sample quantities, any *new* information that they may provide (i.e., any information beyond that already provided directly by observed sample quantities) can only come from *a priori* restrictions placed on the functional form of the price function $P(Z)$. In the absence of such additional restrictions, second stage "structural" estimation of the sort suggested by Rosen may only reproduce the information already provided by the first-stage estimation of the $P(Z)$ function.

To see this point most simply, consider the following example. Suppose that the price function $P(Z)$ is second order in the Z's :

$$(3)\quad P(Z)=\delta_o+\sum_{i=1}^{n}\delta_i Z_i+\sum_{i=1}^{n}\sum_{j=i}^{n}\delta_{ij}Z_iZ_j+u\ ,$$

where the δ's are parameters and u is an error term (observation subscripts have been suppressed for the sake of clarity). In this case, the marginal implicit price for Z_i will be estimated by

$$(4)\quad \hat{P}_i=\frac{\partial\hat{P}}{\partial Z_i}=\hat{\delta}_i+\sum_{j\neq i}\hat{\delta}_{ij}Z_j+2\hat{\delta}_{ii}Z_i\ ,$$

where $\hat{}$'s denote estimated values.

Suppose further that the (inverse) demand and supply functions are linear in the Z's . Then, following Rosen's procedure, the equations to be estimated will be

$$(5) \qquad \hat{P}_k = \gamma_{ko} + \sum_{i=1}^{n} \gamma_{ki} Z_i + \gamma_k Y_1 + \omega_k \qquad \text{(demand)}$$

$$(6) \qquad \hat{P}_k = \phi_{ko} + \sum_{i=1}^{n} \phi_{ki} Z_i + \phi_k Y_2 + \varepsilon_k \qquad \text{(supply)}$$

for $k = 1, \ldots n$; where ω_k and ε_k are random errors.[1] and the γ's and ϕ's are structural parameters to be estimated.

In this particular example, Rosen's procedure will fail to identify the structural demand and supply functions that are of key interest. Indeed, the $\hat{\gamma}$'s and $\hat{\phi}$'s that emerge will be simple functions of the first stage $\hat{\delta}$'s, and can be derived exactly without doing any second stage estimation at all. This fact is easily seen by looking at the ordinary least squares estimator for γ_{ki} in (5) .

$$(7) \qquad \hat{\gamma}_{ki} \text{ (OLS)} \equiv \frac{S_{\hat{P}_k \cdot Z_i}}{S_{Z_i \cdot Z_i}}$$

where $S_{\hat{P}_k \cdot Z_i}$ denotes the sample partial covariance of $\hat{P}_k$ and Z_i (i.e., the sample covariance of those elements of $\hat{P}_k$ and Z_i that are uncorrelated with all other Z's and Y_1) and where $S_{Z_i \cdot Z_i}$ denotes analogously the sample partial variance of Z_i . Noting from (3) that $S_{\hat{P}_k \cdot Z_i} = \hat{\delta}_{ki} S_{Z_i \cdot Z_i}$, it can be seen immediately that applying ordinary least squares to equation (5) will not provide estimates of the γ_{ki}, but rather will simply yield various functions of the previously estimated parameters of equation (1). Specifically, the coefficient on Z_i will be $2\hat{\delta}_{ii}$, and that on Z_j will be $\hat{\delta}_{ij}$.

---

1. In this paper, we do not investigate the implications of alternative error structures in (5) and (6). For such a discussion, see Epple [1980].

This argument applies identically to the estimation of equation (6). Given that $\hat{P}_k$ is an exact function of the Z's, the presence of Y_2 rather than Y_1 in equation (6) will have no effect on the ratio in (7), and thus OLS estimation of equation (6) will lead to the same estimated coefficients for the Z's as produced by OLS estimation of equation (5), none of which reflects the structural parameters γ_{ki} and ϕ_{ki}. These results remain unchanged when (5) and (6) are estimated by two-stage least squares or, more generally, any instrumental variables method.[2]

The fundamental point illustrated by this example is that marginal "prices" constructed only from quantities do not in themselves add any information to that already provided by observations on quantities. The source of any additional information provided by such prices must lie in the a priori restrictions placed on the manner in which observed quantities are translated into marginal prices. Thus, even abstracting from standard structural identification issues, implementation of Rosen's procedure

---

2. Proof: In this case, the estimator for γ_{ki} is

$$(8) \quad \hat{\gamma}_{ki}(2SLS) = \frac{S_{\hat{P}_k \cdot \hat{Z}_i}}{S_{\hat{Z}_i \cdot \hat{Z}_i}}$$

where $\hat{Z}_i$ denotes the first stage constructed instrument for Z_i, and where the variance and covariance notation remains as before. In this case, noting from (3) that $S_{\hat{P}_k \cdot \hat{Z}_i} \equiv \hat{\delta}_{ki} S_{Z_i \cdot \hat{Z}_i}$, and recalling that, by construction $S_{Z_i \cdot \hat{Z}_i} = S_{\hat{Z}_i \cdot \hat{Z}_i}$, it is once again seen that because of the particular identity relation linking $\hat{P}_k$ and the Z's, estimation of (5) and (6) using the constructed $\hat{P}_k$ will only reproduce the previously estimated parameters of (3). Finally, because the ratios in expressions (7) and (8) would be unchanged if the observed sample values of P_k and $\hat{Z}_i$ were first transformed by the same weighting scheme, using three-stage least squares rather than two-stage least squares or, more generally, using any generalized least squares procedure to estimate equations (5) and (6) would lead to the same results.

requires prior restrictions on functional form.

In general, problems of the sort that arise in the preceding example can be avoided only if the marginal price function for Z_i cannot be expressed as some exact combination of the arguments of the structural demand or supply functions. If, for example, the market locus were assumed to be m^{th} order in the Z's, then the marginal price function for Z_i would be of order $m-1$ in the Z's, and so long as the demand and supply functions for Z_i were assumed to be of order $m-2$ or less in the Z's, second-stage estimation of (5) and (6) would not merely duplicate information from the market locus.[3] The structural interpretation of the resulting estimates, of course, would depend upon the validity of the underlying restrictions.

Alternatively, if marginal prices were first estimated from equations fit separately for spatially distinct markets and then used to estimate a common underlying structure for all markets, simple reproduction of first-stage information again could be avoided. To elaborate, if observations were drawn from r separate markets, then r versions of the market locus (3) could be estimated, and for each observation, P_i would depend not only upon the Z vector associated with the observation, but also upon other market-specific factors reflected in the various coefficient estimates. If one were then willing to impose the condition that the structural demand and supply parameters be identical across markets, even though the hedonic price loci are not, then identification could be achieved. Of course, if there were not sufficient variation in the hedonic locus

---

3. Because any non-linear function can be represented arbitrarily closely by a polynomial of some order, we consider only a polynomial market locus.

parameters across markets, nonsense estimates of (5) and (6) would nevertheless be obtained. In any case, unlike the garden variety simultaneous estimation problem, selection of appropriate shift variables is not the only key to identification of the structural parameters.[4]

IV. An Example

Witte, Sumka, and Erekson [1979] (hereinafter WSE) have applied Rosen's [1974] technique to the study of housing markets. They specify a second order hedonic price equation like (3), and linear demand and supply equations like (5) and (6). Their hedonic price equation is estimated separately for four different cities, but the parameters of (5) and (6) are constrained to be identical across the cities. Although they do not note it, the fact that they use different cities is the only thing that saves WSE from estimating a series of identities in the second stage. Let us investigate, therefore, whether or not there is enough "in-between" variation in their sample of cities for WSE to obtain meaningful structural estimates.

The arguments of Section II suggest that if WSE had assumed one overall hedonic price equation, their structural estimates of (5) and (6) would have displayed the following characteristics:

(a) In the bid and offer functions for any characteristic Z_k the estimated coefficients on the i^{th} characteristics would be equal;

(b) The estimated coefficients on Z_i in the bid and offer functions for the k^{th} characteristic would be equal to the estimated coefficients on Z_k in the bid and offer functions for the i^{th} characteristic;

---

4. Of course, the rank as well as order restrictions must be satisfied.

(c) The estimated coefficient on Z_i in the bid and offer functions for the k^{th} characteristic would be equal to $2\hat{\delta}_{ii}$ for $k = i$, and equal to $\hat{\delta}_{ki}$ for $k \neq i$;

(d) The estimated coefficients for the vector Y_1 in all demand equations and for the vector Y_2 in all supply equations should be equal to zero.

To facilitate discussion of point (a), we have reproduced in Table I the relevant coefficients from WSE's bid and offer equations, using our notation. For each set of bid and offer equations there are three pairs of coefficients to compare, or nine comparisons altogether. Inspection of the table indicates that in seven out of nine cases, the coefficients in each cell are similar in magnitude (the two exceptions are the own coefficients in the demand and supply equations for Z_1 and Z_2). As for point (b), the symmetry it implies is also clearly evident in Table I.

For evidence regarding point (c), consider Table II below, which reproduces WSE's estimated coefficients from the market locus equations that they fit separately for their four cities. Because there are four values for each $\hat{\delta}_{ij}$ (one for each city), the γ_{ij} and ϕ_{ij} that WSE estimate from a pooled sample will be a combination of these $\hat{\delta}_{ij}$'s. WSE present insufficient data for us to determine these weights, so we are unable to predict the exact relation that must exist between the estimates in Table II and those in Table I. Nevertheless, the numbers in the two tables lend at least qualitative support for point (c). For example, note from Table II the tendency $|\delta_{11}| > |\delta_{22}| > |\delta_{33}|$. This is mirrored in the decline of the coefficients of Table I in absolute value as one reads along the main diagonal from northwest to southeast. Similarly, note that the the estimated values of δ_{13} are small in absolute value,

TABLE I

Partial Listing of WSE's Demand and Supply Equations*

Dependent Variables	Independent Variable** Z_1	Z_2	Z_3
Demand Price for Z_1	-8.65	5.00	0.41
	(4.78)	(4.63)	(0.36)
Supply Price for Z_1	11.08	7.83	-0.74
	(2.87)	(2.76)	(0.49)
Demand Price for Z_2	8.12	-6.97	0.41
	(2.49)	(2.41)	(0.19)
Supply Price for Z_2	6.41	-0.71	0.28
	(1.16)	(1.12)	(0.20)
Demand Price for Z_3	-0.28	0.38	-0.03
	(0.19)	(0.19)	(0.01)
Supply Price for Z_3	0.12	-0.02	0.01
	(0.09)	(0.09)	(0.02)

* Source: Witte, et.al.,[1979, p. 1161]. Numbers in parentheses are standard errors.

** Z_1 = dwelling quality; Z_2 = dwelling size; Z_3 = lot size.

TABLE II

Partial Listing of WSE's Market Locus Parameters*

Parameter / City	δ_{11}	δ_{22}	δ_{33}	δ_{12}	δ_{13}	δ_{23}
Greenville	-7.40	-3.23	**	**	**	0.65
	(3.05)	(1.21)				(0.27)
Kinston	8.53	-0.78	-0.001	-2.00	0.75	-0.17
	(6.75)	(1.25)	(0.02)	(4.95)	(0.83)	(0.31)
Lexington	9.29	-0.40	-0.0110	6.13	0.19	-0.05
	(2.77)	(1.41)	(0.01)	(2.29)	(0.24)	(0.10)
Statesville	**	-2.47	**	14.18	**	**
		(1.02)		(3.94)		

* Source: Witte, et. al.,[1979, p. 1158]. Numbers in parentheses are standard errors.

** WSE excluded these variables because they did not add significantly to the explanatory power of the regression. See WSE [1979, p.1157, fn. 13].

as are the estimated coefficients for Z_3 in the demand and supply equations for Z_1 , and for Z_1 in the demand and supply equations for Z_3 .[5]

The evidence seems to suggest that in the WSE data, there was insufficient variation across cities to allow the computation of meaningful structural coefficients.

V. Conclusions

We have shown the inappropriateness of assuming that equations like (5) and (6) form a conventional simultaneous equation model without consideration of how the marginal attribute prices were generated from the sample data. When it is assumed that the prices are generated by a single hedonic price equation, the only way to identify the structural coefficients is by (possibly arbitrary) restrictions on functional form. When it is assumed that the prices are generated by several different equations, there may be sufficient "in-between" variation to generate meaningful structural

---

5, Consistent with point (d), in only eight out of twenty-four cases are WSE's estimated coefficients on demand and supply shift variables statistically non-zero at less than a .10 level of significance. (To conserve space, these coefficients are not reported in Table I.) It is tempting to apply conventional statistical tests for the conformity of WSE's results with our predictions. However, WSE provide insufficient information for us to state a null hypothesis that should apply when prices are constructed individually for each city but demand and supply functions are estimated for all cities. Further, if we had such information, our predictions should hold exactly, so the issue of conformity is really not statistical in nature. Anything less than perfect conformity would contradict our predictions. At this stage, the most we can say is that our predictions would hold exactly if: (a) a full set of interactive dummy variables were included in the structural equations estimated by WSE; (b) the structural equations estimated by WSE were estimated separately for each city as were the market locus equations; or (c) the market locus equations were estimated from the pooled sample, as were the structural equations estimated by WSE.

parameters even without such constraints on functional forms. However, one must then impose the restriction that the structural parameters are identical across markets even though the hedonic locus parameters are not.

Although we have pointed out some pitfalls in the usual methods for estimating structural hedonic price models, we have not stated a general set of conditions under which identification is possible. Given the importance of hedonic models in applied research, the search for such conditions deserves serious attention.

References

Epple, Dennis, "Hedonic Prices and Implicit Markets: Estimating Demand and Supply Functions for Differentiated Products," mimeo, Carnegie-Mellon University, August, 1980.

Harrison, David and Daniel L. Rubinfeld, "Hedonic Housing Prices and the Demand for Clean Air," Journal of Environmental Economics and Management 5, March, 1978, pp. 81-102.

Rosen, Sherwin, "Hedonic Prices and Implicit Markets: Product Differentiation in Pure Competition," Journal of Political Economy 82, 1974, pp. 34-55.

Witte, Ann D., Howard J. Sumka, and Homer Erekson, "An Estimate of a Structural Hedonic Price Model of the Housing Market: An Application of Rosen's Theory of Implicit Markets," Econometrica 47, 1979, pp. 1151-73.