Comment on "Simulation and Estimation of Hedonic Models" by Heckman, Matzkin and Nesheim

Keane, Michael

May 2003

Comment on "Simulation and Estimation of Hedonic Models" by Heckman, Matzkin and Nesheim

by

Michael P. Keane
Yale University

May 28, 2003

We always need *a priori* identifying assumptions in order to learn anything of interest from data, beyond perhaps the simplest of descriptive statistics. Data cannot simply "speak" and reveal interesting aspects of economic behavior to an investigator whose mind is a blank slate, devoid of economic priors.[1] The chapter by Heckman, Matzkin and Nesheim (HMN) provides an excellent illustration of this point, by clarifying the types of identifying assumptions that are necessary to learn anything about the parameters that characterize supply and demand in hedonic models. Work like this provides a useful palliative to the overly sanguine – and currently prevalent - view that, if we can just find "natural experiments" or "clever instruments," we can learn interesting things about behavior without making strong *a priori* assumptions, and without using "too much" economic theory.[2]

The HMN chapter illustrates the point that the validity of instruments hinges on the assumed economic structure. That is, we need to think carefully about theory in order to decide whether (or under what conditions) a particular "instrument" will enable us to learn about some aspect of behavior. When the economic assumptions that underlie the validity of instruments are left implicit, the proper interpretation of inferences is obscured. Of course, this is a point that Heckman has stressed not just here but in a number of recent papers.[3]

The HMN chapter is representative of an important research program in econometrics that seeks to clarify, in the context of various different economic models, the kinds of identifying assumptions that are necessary in order to draw inferences about behavioral parameters of interest. The identification issues in the hedonic models that HMN examine are especially challenging, because these models are highly nonlinear, and because equilibrium considerations are critical. Progress in this area has been slow, but HMN's contribution substantially advances our knowledge.

While identification of behavioral parameters always requires some *a priori* assumptions, the estimation of complex nonlinear models, such as hedonic models, raises additional problems. In such cases, additional assumptions, not necessary for identification per se, are typically needed in order to make estimation practical. Practical computational considerations may dictate that specific distributions be chosen for stochastic terms, functional forms chosen for utility

[1] By "data" I mean the joint distribution of observed variables. To use the language of the Cowles Commission, "Suppose … B is faced with the problem of identifying … the structural equations that alone reflect specified laws of economic behavior ... Statistical observation will in favorable circumstances permit him to estimate … the probability distribution of the variables. Under no circumstances whatever will passive statistical observation permit him to distinguish between different mathematically equivalent ways of writing down that distribution … The only way in which he can hope to identify and measure individual structural equations … is with the help of a priori specifications of the form of each structural equation" - see Koopmans, Rubin and Leipnik (1950).

[2] For instance, Angrist and Krueger (1999) state that "An alternative to structural modeling, … the "experimentalist" approach, … puts front and center the problem of identifying causal effects from specific events or situations," where, by "events or situations," they are referring to "natural experiments" that generate exogenous variation in certain variables that would otherwise be endogenous in the behavioral relationship of interest. They go on to state that: "the difference is primarily one of emphasis, because structural modeling generally incorporates some assumptions about exogenous variability in certain variables and quasi-experimental analyses require some theoretical assumptions." This statement draws a distinction between "assumptions about exogenous variability" vs. "theoretical assumptions," and the implication is that the theoretical assumptions are less central (or weaker) in the "experimentalist" approach. But I would take issue with this distinction and this assertion. An assumption that a variable is exogenous always rests on a set of theoretical assumptions, even in the unusual instance that the variable is subject to random assignment. As a concrete illustration, Rosenzweig and Wolpin (2000) present a plausible example where even draft lottery numbers are endogenous. In the "experimentalist" approach, the theoretical (or behavioral) assumptions that underlie exogeneity assumptions are often quite strong, and they are often left implicit.

[3] See for instance Heckman (1997). Another good discussion is Rosenzweig and Wolpin (2000).

functions, and so on. When we seek to draw inferences about a particular behavioral parameter, such as a price elasticity of demand, we would like to distinguish between those assumptions that we make for computational reasons vs. those that are essential for identification of the parameter.

Why is this distinction important? The argument runs as follows: It is desirable that identification should not rest entirely on functional form assumptions that lack a basis in economic theory. If a particular functional form assumption is not necessary for identification, then we may hope that as we relax that assumption – say as computational advances or larger sample sizes make it feasible to do so – then our fundamental inferences will remain intact.[4] On the other hand, we know that we can never dispense with the assumptions that are fundamental for identification. In the interest of truth in advertising we should try to be clear about what those fundamental assumptions are. A consumer of our results can then decide how much weight he/she wants to place on them, based on his/her priors on the validity of our identifying assumptions. However, in the context of a model as complex as the hedonic equilibrium model, the distinction between assumptions necessary for identification and those made for computational practicality is not immediately obvious.

Let me now turn to the specifics of HMN. It will be helpful to first lay out some notation. Let z denote the quantity of an amenity that a consumer decides to consume, and let the equilibrium price schedule be $P(z)$. Utility, which is linear in the consumption of other goods, is given by $U^* = [Y-P(z)] + U(z, \tau)$, where Y is income. The utility that the consumer gets from consumption of the amenity depends on the taste shifter τ. Restrict the class of models so that $P(z)$ is twice continuously differentiable.

Now, the fundamental problem in identifying the parameters that characterize demand and/or supply in a hedonic model using data from a single market is that there is no exogenous variation in price. Let's focus on estimating the demand function. If the consumer chooses amenity level z^* then he/she faces marginal price $P'(z^*)$. Consumers with different tastes will choose different z's and therefore face different marginal prices. Thus, the marginal price that each consumer faces is endogenous. In fact, the price is, in effect, a choice variable.

Furthermore, variables that shift the price schedule $P(z)$ by altering the supply side of the market (i.e., cost shifters) are not valid instruments. When the $P(z)$ schedule shifts it induces a change in the marginal price that a consumer faces which stems, at least in part, from the change in the consumer's choice of z. Generally this change depends on the consumer's tastes, so that the change in the marginal price confronting a consumer is endogenous. Thus, the desired experiment of presenting a consumer of given tastes with an exogenous shift in marginal price, is not implemented. This means further that identification of demand function parameters off of variation of the price schedule across multiple markets will not work, even if the price schedule differences arise exclusively from different cost conditions.

So how do HMN propose to get around this problem? They make two key types of identifying assumptions. First, they assume that the vector of consumer taste shifters τ can be decomposed into (x, ε), where x is observed by the analyst and ε is unobserved. Furthermore, the observed taste shifter x is assumed to be independent of unobserved tastes ε.

[4] Of course, this is only a hope that may not be born out. For example, Heckman and Sedlacek (1986) noted that the multinomial probit (MNP) model is identified without exclusion restrictions. But in Keane (1992) I found that, even in very large samples, the MNP likelihood is extremely flat along certain ranges of the parameter space unless exclusion restrictions are invoked. Thus, even in very large samples, inference in this model will be very sensitive to the assumed exclusions, despite the fact that the model is formally identified without them.

Second, they need to place some structure on the marginal utility function $U_z(z, x, \varepsilon)$. In the "additive" model they assume that:

$$U_z(z, x, \varepsilon) = m(z) + n(x) + \varepsilon$$

while in the "non-additive" model they assume that:

$$U_z(z, x, \varepsilon) = M(q(z, x), \varepsilon).$$

In the additive case $m(z)$ is an unknown function (decreasing in z if we have diminishing marginal utility). In the "non-additive" model, $M(\cdot,\cdot)$ is an unknown function that is strictly decreasing in q and monotonic in ε, while q is a known function that is increasing in z and x.[5]

I wish to focus initially on the role played by the first assumption: that there exists an observed taste shifter variable x that is independent of unobserved tastes ε. In order to do this, it will be useful to consider the simple case where marginal utility is linear in z and x:

$$U_z(z, x, \varepsilon) = -Az + \theta_0 + \theta_1 x + \varepsilon \qquad A>0.$$

This gives the first order condition:

$$(1) \qquad P'(z^*) = -Az^* + \theta_0 + \theta_1 x + \varepsilon$$

As HMN note, the marginal price function $P'(z)$ will be a nonlinear function of z except in the very special case of Tinbergen's (1956) linear-quadratic-normal model.[6] HMN assume that one has an estimate of $P'(z)$ in hand, and treat this function as known in their identification analysis.

Now, to understand the role of the taste shifter variable x, it is instructive to rewrite the first order condition like a demand function:

$$(2) \qquad z^* = -\frac{1}{A}[P'(z^*) - (\theta_0 + \theta_1 x)] + \frac{\varepsilon}{A}$$

Note that a shift in x has an effect on consumer demand that is exactly like the effect of an exogenous shift in the marginal price schedule. Thus, intuitively, the exogenous variation in x will substitute for the exogenous price variation that we lack, enabling us to learn about the demand function parameters. Since $P'(z)$ is generally a nonlinear function of z, both the z^* that a consumer chooses and the resultant $P'(z^*)$ will, in general, be nonlinear functions of x. Hence, we can use nonlinear functions of x (such as powers of x), along with $(1, x)$, as instruments to identify θ_0, θ_1 and A in (2). This is the really key point of the HMN identification analysis.

It is worth stressing that this IV procedure only fails in the linear-quadratic-normal special case, where the marginal price function is linear. If $P'(z) = \pi_0 + \pi_1 z$, then we have that z^*

---

[5] In the non-additive case these assumptions only enable HMN to identify the utility function parameters up to scale and location normalizations, so they need an additional assumption to pin down these parameters. For instance, this can take the form of normalizing M at some value of x and ε.

[6] That is, to get a linear $P'(z)$ function, ε would have to be normal, and, furthermore, we would need a linear marginal cost function with normal errors on the firm side as well.

$= (A + \pi_1)^{-1}[-\pi_0 + \theta_0 + \theta_1 x + \varepsilon]$. Since z^* and $P'(z^*)$ are both linear functions of x, powers of x beyond the first are not useful as instruments, and the model is under-identified.[7,8]

Now, even granting that is reasonable to assume that $P'(z)$ is a nonlinear function of z, it might appear that the IV method still hinges on an "arbitrary" assumption that the $n(x)$ function does not involve higher order terms in x. The IV method would indeed fail if $P'(z^*)$ and z^* happened to lie in a space spanned by polynomials in x of order less than or equal to the order of the $n(x)$ function. But in a forthcoming paper, Ecklund, Heckman and Nesheim (2003) show that, except for very special cases (like, e,g., the linear-quadratic-normal case), this will not happen. The structure of the hedonic model implies that $P'(z^*)$ will not generally lie in such a space.

Once we understand the argument for why powers of x will generally provide valid instruments for estimation of the model, we see that the very existence of a variable like x is really the critical assumption. The difficulty is that, in many empirical applications, it will hard to find observables like x that shift the marginal utility of consumption but that are independent of unobserved determinants of tastes ε. For instance, having kids might shift one's taste for risky work, but it is also likely to be correlated with one's "baseline" tastes for risk. Thus, in any empirical application, the challenge will be to find convincing instruments x. The HMN analysis is very useful because it clarifies what properties a good instrument should have. But, as is typically the case, the economic theory does not deliver the instrument.

This brings me to a consideration of the second type of assumption used by HMN, their functional form assumptions. On way to interpret these assumptions is that they place additional requirements on x. In the additive model, in addition to requiring that x be independent of ε and act like a shifter of $P'(z^*)$, we further require that x acts *as if* it induces a parallel shift in the marginal price schedule (i.e., a shift that is constant at all levels of z).

The non-additive model relaxes this assumption, but the relaxation is not strict because it comes at a price. The assumption that q is a known function is quite restrictive. To give a very simple illustration, suppose we let $M(q,\varepsilon) = -Aq + \varepsilon$, and $q(z,x) = z - (\theta_0 + \theta_1 x + \theta_2 xz)$. The we obtain the first order condition $P'(z^*) = -Az^* + A[\theta_0 + \theta_1 x + \theta_2 xz] + \varepsilon$. This allows x to shift the marginal utility from consumption of z by different amounts at different levels of z. But, since θ_0, θ_1 and θ_2 are assumed known, the analyst is assumed to know *a priori* precisely how the degree of the shift varies with z.[9]

The analysis here reminded me of some work in the transportation literature that has dealt with an analogous problem in a very different context. Basically, the problem is that, in many instances, not only is there no exogenous variation in transport mode prices – there is no variation *period*. How does one estimate a price elasticity of demand for the NYC subway when everyone faces the same fare? One idea is to find a variable that acts just like a price shifter. For instance, people live different distances from the nearest subway station. The "full price" of

---

[7] Another way to see this is to note that estimation of the reduced form for z would give us estimates of $(A + \pi_1)^{-1}(\theta_0 - \pi_0)$ and of $(A + \pi_1)^{-1}\theta_1$, from the intercept and slope coefficient respectively. Since π_0 and π_1 are known, this gives us only two equations in the three unknowns A, θ_0 and θ_1. But, if $P'(z)$ were a quadratic function of z, it would add a reduced from coefficient without changing the number of unknown structural parameters, thus giving exact identification.

[8] An obvious point is that, since we can estimate $P'(z)$ in a first stage, we can always check its properties before proceeding. If we don't find evidence of significant non-linearity, we would not proceed with the HMN approach. But, it turns out, that is exactly the case where it would be appropriate to estimate the linear-quadratic-normal model using FIML, because we know the unobserved taste shocks must be approximately normal.

[9] Another way to understand the assumption is to say that the marginal utility of consumption of z varies with a known index of z and x. One case where this is natural is if x represents a known non-tradable endowment of z itself.

taking the subway includes the fare plus the monetized value of the time it takes to get to the station (which could be operationalized using wage rates).[10] The point is that there are really two avenues for obtaining variables that act like shifts in the marginal price schedule: we can find variables like x that shift the marginal utility of consumption, or we can search for variables that enter symmetrically with price in the consumer's budget constraint. This provides another possible avenue for finding good instruments, and it may be especially appealing because then we know that an additivity assumption should hold when we write the first order condition.

HMN have considered a number of possible estimators based on their identification analysis. The construction of these estimators illustrates my earlier point that estimation often involves a mixture of fundamental identifying assumptions and assumptions made for tractability. The specification in their equation (23) involves the choice of particular orders for the $m(z)$ and $n(x)$ polynomials. The density of ε is estimated using a kernel with a particular bandwidth. The function $P_z(z)$ and its derivatives need to be estimated in a first stage (perhaps using polynomials in z). In their Monte-Carlo analysis the authors have the advantage of knowing the proper orders for $m(z)$ and $n(x)$. In any real application these will not be known. On the one hand one might want to err on the side of specifying too high an order for $n(x)$ in order to avoid misspecification. But increasing the order of $n(x)$ probably will entail a high cost in terms of lost efficiency, since identification will then rely on the dependence of $E[z|x]$ on successively higher order terms in x. Future Monte Carlo work should explore this issue.

Finally, let me conclude by noting that, while I am very impressed by HMN's effort to clarify identification issues in hedonic models, I have some reservations about the general course that that broader econometric research program on "semi-parametric" or "non-parametric" identification has taken. The problem, in my view, is that this program has typically adopted an asymmetric perspective towards the legitimacy or desirability of using certain types of identifying assumptions, and I am unclear about what underlies these judgments.

In seeking to identify behavioral parameters, one can typically choose from amongst several types of *a priori* identifying assumptions. Loosely speaking, from reading the recent identification literature, one gets the sense that parametric assumptions on error distributions or functions of interest are generally viewed as quite undesirable. Assumptions like additive separability of errors, exclusion restrictions, and independence or mean independence of errors and covariates are viewed as far preferable.[11] The usual argument is that economic theory doesn't deliver the former type of assumption. But it must be admitted that it doesn't typically deliver the latter type either. For instance, nothing in the theory of hedonic models will tell us whether a particular variable is a good candidate for x. This will always involve an *a priori* judgment.[12]

Furthermore, while a normality assumption for stochastic terms or a linear or quadratic functional form assumption may seem arbitrary, do we really want to go the opposite extreme of

---

[10] This is meant by way of illustration. I would not necessarily argue that distance from the subway is exogenous with respect to ε.

[11] To quote Marschak (1950), "There are many competing sets of *a priori* restrictions that can be imposed upon the structural parameters without contradicting what we know of human behavior … Regarding the great variety of functions equally appropriate, on *a priori* grounds, to describe structural economic relations, one may expect some help from the statisticians' recent attempts at nonparametric estimation of distribution functions … Certain weak *a priori* restrictions on the structural relations, such as the sign of certain partial derivatives, the independence of successive shocks, etc., the economist can assert with better conscience than the restrictions upon, say, the degree of polynomials chosen to describe the structural relations."

[12] To quote Koopmans, Rubin and Marsckak (1950), "… the distinction between exogenous and endogenous variables is a theoretical, *a prior* distinction …"

letting error distributions and/or functional forms be non-parametric? In most contexts, I would have a strong prior that error distributions can be very well approximated by low dimensional mixtures of normals, and that functions of interest can be well approximated by flexible polynomial functions. In many contexts such distributional or functional form assumptions seem at least as plausible as any exclusion or mean independence restriction one is likely to propose.

In an ongoing research program, John Geweke and I have been developing and implementing computational methods to do Bayesian inference in models with very flexible mixture-of-normals error distributions and flexible polynomial functional forms (see Geweke and Keane (1999, 2000, 2001)). These flexible parametric methods can handle multivariate conditioning much more easily than can semi-parametric methods. Other related work in this genre has used priors that functions of interest are smooth, monotonic, concave, etc.[13] In principle such methods could also be used to achieve identification in various contexts. We view this research program as a viable alternative to the classical semi-parametric estimation agenda.

As I noted at the outset, all inference about structure relies on *a priori* identifying assumptions. The Bayesian framework has the advantage that one can use it to learn about the extent to which priors drive inferences. It is possible to write down a structure that is not completely identified from the data. Nevertheless placing proper priors on the non-identified parameters generates a well-defined posterior. One can then examine the sensitivity of inferences regarding the structural parameters to the specification of the priors.[14] This has the potential to avoid the asymmetry of the classical semi-parametric approach, because one doesn't have to dogmatically adhere to a subset of the possible *a priori* identifying assumptions. As a concrete example, in the hedonic model, one could allow x to be correlated with ε, and use priors to control the strength of this correlation, thus gauging the sensitivity of inferences to variation along this dimension.

In conclusion, the HMN analysis substantially clarifies issues of identification of hedonic models. These issues have vexed and confused economists for many years, and this work points the way to renewed progress in this important area.

References

Angrist, J. and A. Krueger (1999). Empirical Strategies in Labor Economics. Handbook of Labor Economics, Ashenfelter and Card (eds.), Elsevier Science B.V., 3463-3568.1277-1366.

Barnett, W.A., Geweke, J. and M. Wolfe (1991). Semi-nonparametric Bayesian Estimation of the Asymptotically Ideal Production Model. Journal of Econometrics, 49(1-2): 5-50.

Eckland, I. J. Heckman and L. Nesheim (2003). Identification and Estimation of Hedonic Models. Journal of Political Economy, forthcoming.

Geweke, J. and M. Keane (1999). Mixture of Normals Probit Models. Analysis of Panels and Limited Dependent Variable Models, Hsiao, Lahiri, Lee and Pesaran (eds.), Cambridge University Press, 49-78.

[13] See, for instance, Barnett, Geweke and Wolfe (1991), Gordon (1996), Romeo (1999).
[14] Lancaster's (1997) analysis of the job search model is an example of this type of approach.

Geweke, J. and M. Keane (2000). An Empirical Analysis of Male Income Dynamics in the PSID: 1968-1989. Journal of Econometrics, 96(2): 293-356.

Geweke, J. and M. Keane (2001). Computationally Intensive Methods for Integration in Econometrics. Handbook of Econometrics: Vol. 5, J.J. Heckman and E.E. Leamer (eds.), Elsevier Science B.V., 3463-3568.

Gordon, S. (1996). Using Mixtures of Flexible Functional Forms to Estimate Factor Demand Elasticities. Canadian Journal of Economics, 29(3): 717-736.

Heckman, J. (1997). Instrumental Variables: A Study of Implicit Behavioral Assumptions Used in Making Program Evaluations. Journal of Human Resources, 32(3): 441-462.

Heckman, J. and G. Sedlacek (1985). Heterogeneity, Aggregation and Market Wage Functions: An Empirical Model of Self-Selection in the Labor Market. Journal of Political Economy, 93(6): 1077-1125.

Keane, M. (1992). A Note on Identification in the Multinomial Probit Model. Journal of Business and Economic Statistics, 10(2): 193-200.

Koopmans, T.C., H. Rubin and R.B. Leipnik (1950). Measuring the Equation Systems of Dynamic Economics. Cowles Commission Monograph No. 10: Statistical Inference in Dynamic Economic Models, T.C. Koopmans (ed.), John Wiley & Sons, New York.

Lancaster, T. (1997). Exact Structural Inference in Optimal Job Search Models. Journal of Business and Economic Statistics, 15(2): 165-179.

Marschak, J., (1950). Statistical Inference in Economics: An Introduction. Cowles Commission Monograph No. 10: Statistical Inference in Dynamic Economic Models, T.C. Koopmans (ed.), John Wiley & Sons, New York.

Romeo, C.J. (1999). Conducting Inference in Semiparametric Duration Models Under Inequality Restrictions on the Shape of the Hazard Implied by Job Search Theory. Journal of Applied Econometrics, 14(6): 587-605.

Rosenzweig, M. and K. Wolpin (2000). Natural "Natural Experiments" in Economics. Journal of Economic Literature, 38(4): 827-874.

Tinbergen, J. (1956). On the Theory of Income Distribution. Weltwirtschaftliches Archiv, 77: 155-173.