Comparing Hedonic and Random Utility Models of Demand with an Application to PC's.[1]

Patrick Bajari and C. Lanier Benkard

Duke University and NBER
Stanford GSB and NBER

August 13, 2004

Abstract

Random utility and hedonic models are both commonly used to recover consumer preferences in differentiated product markets. In this paper, we compare these two methods. We begin by describing a method, proposed by Bajari and Benkard (2004), for estimating a structural hedonic model. Second, we discuss some criticisms of both models made in the literature. Finally, we compare the we compare the hedonic estimator to discrete choice models by studying retail demand for personal computers. The models generate very different demand elasticities and implications for consumer welfare.

1 Introduction.

In this paper, we compare random utility models (henceforth RUM) and hedonic models of demand. Both methods are commonly used to recover consumer preferences in a differentiated product market. Hedonics are commonly used in Public or Natural Resource Economics to estimate, for example, consumer willingness to pay for high quality schools or clean air. RUMs are often applied in Industrial Organization and Marketing. One application is to recover product level demand curves for measuring market power or merger analysis.

RUM and hedonics models use different assumptions. In RUMs, consumer utility depends on stochastic preference shocks while in hedonic models these shocks are omitted. These shocks might introduce too much "taste for product" and lead to counter-intuitive implications for welfare analysis. Hedonic models, on the other hand, have been criticized for making competition between products too localized. Many authors are discussed these issues (see Ackerberg and Rysman (forthcoming), Berry and Pakes (2000), Caplin and Nalebuff (1991), Petrin (2002) and Andersen, de Palma and Thisse (1992)). However, to the best of our knowledge, there are no empirical papers that quantify the importance of these issues in applications.

In this paper, we attempt to address this gap in the literature by comparing estimates from these two commonly used methods. We begin by reviewing a particular method proposed by Bajari and Benkard (2004, henceforth BB) for estimating hedonic models. (Earlier examples of structural hedonic models

include models of horizontal product differentiation such as Gorman (1980) and Lancaster (1966), models of vertical product differentiation, such as Shaked and Sutton (1987) and Bresnahan (1987), as well as Rosen (1974)'s model.) The method we discuss here slightly modifies the BB estimator in order to make it easier to compute. Next, we theoretically compare hedonic models to random utility models. We establish that when the number of products is large RUM models, compared to a hedonic alternative, will tend to lead to more market power and high measures of consumer surplus. Finally, we study a unique data set of retail personal computer sales. This data set allows us to quantify the importance the theoretical properties we established in an application.

This paper makes three contributions to the literature. First, we propose a simpler, more "user friendly" version of BB (2004). BB consider a hedonic model that generalizes the standard hedonic model in three ways. First, they allow for consumers to be heterogeneous in their willingness to pay for product attributes. In a linear hedonic regression, the implicit price for a product attribute is commonly interpreted as the marginal willingness to pay and does not differ across consumers. BB consider a nonparametric framework that allows the marginal willingness to pay to differ across consumers. Their approach is similar to allowing a nonparametric distribution of random coefficients in the framework of Berry, Levinsohn and Pakes (1995, henceforth BLP). Second, BB derive consumer preferences for "omitted" product characteristics that are observed by consumers but not by the economist. In standard models, such as BLP, accounting for these attributes in important in order to generate reasonable price elasticities. Finally, BB discuss a framework that allows the economist to recover preferences in a hedonic model with discrete attributes. The standard analysis of the hedonic model exploits a first order condition to uncover structural willingness to pay. To the best of our knowledge, BB is the only paper that suggests a nonparametric hedonic model that allows

3

for all three of the properties discussed above. The approach we discuss in this paper allows for the three features of the BB approach discussed above: nonparametric random coefficients, omitted product attributes and discrete characteristics. However, the approach described in this paper is much simpler to estimate and can be easily programmed using standard statistical packages.

The second contribution is that we formalize some potential limitations of standard RUMs. While some of these results have antecedents in the literature, we believe that it is useful to summarize criticisms of the models mentioned in the literature and analytically establish exactly which modeling assumptions generate these properties. Also, some of the results that we prove are new to the literature (we will discuss the relation of the results to the literature in the text).

RUMs and hedonic models use different assumptions. In RUMs, consumer utility depends on stochastic preference shocks while in hedonic models these shocks are omitted. These shocks might introduce too much "taste for product" and lead to counter-intuitive implications for welfare analysis. Hedonic models, on the other hand, may be criticized for going to the opposite extreme and generating a demand system where competition is too localized. Many authors are discussed these issues (see Ackerberg and Rysman (forthcoming), Berry and Pakes (2000), Caplin and Nalebuff (1991), Petrin (2002) and Andersen, de Palma and Thisse (1992)). However, to the best of our knowledge, there are no empirical papers that quantify the importance of these issues in applications.

Finally, we compare estimates from hedonic and RUM models. In December 1999, in a data set with seven hundred products, we find that own price elasticities for the top five products in the market are on average -42 from the hedonic model compared to -1.4 from a version of the logit model with omitted product attributes discussed in Berry (1994). The welfare loss to consumers from removing their most

4

desired product and giving them their second most desired product is \$160 million in the logit model and only \$8 million in the hedonic model. These results are consistent with our analytic results which establish that random utility models generate more taste for product than alternative specifications.

Researchers such as Berry and Pakes (2001) and Ackerberg and Rysman (forthcoming) have suggested that "tapering off" the error terms in standard RUM models may lead to improved estimates in some applications. Hedonics are one method for removing the influence of stochastic preference shocks from the demand system. Both hedonics and RUM are routinely used in applied work. Our results suggest that welfare analysis may be sensitive to the choice of method and that economists may want to check the sensitivity of his conclusions.

2 A Hedonic Model of Consumer Demand

In this section, we describe a hedonic model of consumer demand.[2] In the model, the economist observes the choices of $i = 1, ..., I$ consumers during $t = 1, ..., T$ time periods. A product $j = 1, ..., J$ is a bundle of two types of attributes. The first is a $1 \times K$ vector x_j of the product characteristics that are observed by both the economist and the consumer. The second is a scalar, ξ_j, which is an omitted product attribute which is observed by the consumer, but not the economist. The price of product j is p_j.

Prices in the market are determined in equilibrium by the interaction of buyers and sellers. At time t, the function $\mathbf{p}_t$ denotes the equilibrium map from product characteristics to prices:

$$p_j = \mathbf{p}_t(x_j, \xi_j). \tag{1}$$

In BB (2004), we establish under fairly weak assumptions that the function $\mathbf{p}_t$ exists and is strictly increas-

---

[2] The exposition in this section closely follows Bajari and Benkard (2004) and Bajari and Kahn (2004).

5

ing in ξ_j.[3] In most of what follows, we shall drop the subscript t in equation (1) since we will typically not pool observations across markets in the estimator.

Households are assumed to be static utility maximizers. Utility is a function of the product characteristics (x_j, ξ_j) and consumption of a composite commodity c. We normalize the price of the composite commodity to one. The utility that consumer i receives for product j, u_{ij}, can therefore be written as

$$u_{ij} = u_i(x_j, \xi_j, c). \tag{2}$$

Suppose that household i has income of y_i. Product $j^*(i)$ is utility maximizing for household i if

$$j^*(i) = \arg\max_j \ u_i(x_j, \xi_j, c). \tag{3}$$
$$\mathbf{p}(x_j, \xi_j) + c \leq y_i$$

Note that in equation (3) we have substituted the budget constraint directly into the utility function.

Suppose that product j^* with characteristics $\left(x_{j^*,1}, ..., x_{j^*,K}, \xi_{j^*}\right)$ is utility maximizing for household i. If characteristic k is continuous, then the following first order condition must hold:

---

[3] The intuition behind the argument is simple. We suppose, as in BLP, that ξ_j is interpreted as a vertical characteristic that is always positively valued at the margin by the consumers. Then prices must be an increasing function of ξ_j. Suppose, by contradiction, that it is possible to get a higher value of ξ_j for a lower price, holding all else fixed. Then some product would be strictly dominated and would not be observed in the market. Note that this assumption does not depend on assumptions about the supply side.

$$\frac{\partial u_i(x_{j^*}, \xi_{j^*}, y_i - p_{j^*})}{\partial x_{j,k}} - \frac{\partial u_i(x_{j^*}, \xi_{j^*}, y_i - p_{j^*})}{\partial c} \frac{\partial \mathbf{p}(x_{j^*}, \xi_{j^*})}{\partial x_{j,k}} = 0 \tag{4}$$

$$\frac{\frac{\partial u_i(x_{j^*}, \xi_{j^*}, y_i - p_{j^*})}{\partial x_{j,k}}}{\frac{\partial u_i(x_{j^*}, \xi_{j^*}, y_i - p_{j^*})}{\partial c}} = \frac{\partial \mathbf{p}(x_{j^*}, \xi_{j^*})}{\partial x_{j,k}}. \tag{5}$$

Equation (5) is the familiar condition that the marginal rate of substitution between a continuous character-istic and the composite commodity is equal to the partial derivative of the hedonic. That is, the willingness to pay for a marginal unit of characteristic k is equal to the implicit price of k at the margin.

In many data sets, the choices of a given household may only be observed once, or a handful of times. Obviously, it is not possible to learn household i's entire utility function $u_i(x_j, \xi_j, c)$ from just a single or handful of choices. If there is no violation of the strong axiom, many possible utility functions could possibly rationalize such a small number of observations.

In order to identify preference, it is therefore necessary to make some restrictions on the function $u_i(x_j, \xi_j, c)$. In what follows, we will try to make restrictions that are minimal in the sense that the model will be just identified, that is, it exhausts all of the degrees of freedom in the data. It is most useful to illustrate our approach to modeling $u_i(x_j, \xi_j, c)$ by considering the example of computers which we will use in our applied section. Suppose that the relevant characteristics of the computer are:

- CPU-A measure of CPU performance (defined in the data section)
- RAM- The amount of Random Access Memory
- HD-Hard drive capacity of the system
- $CDROM$-A dummy variable equal to one if the system includes a CD-ROM drive.
- MMX-A dummy variable equal to one if the processor is equipped with Matrix Math eXtensions (MMX)
- $SCSI$-A dummy variable equal to one if the system is equipped with SCSI (Small Computer System Interface).

- DVD-A dummy variable equal to one if the system includes a zip drive
- NIC-A dummy variable equal to one if the system has a Network Interface Card (NIC)
- MON-A dummy variable equal to one if the system has a monitor.
- ZIP-A dummy variable equal to one if the system has a zip drive
- DT-A dummy variable equal to one if the system is a desktop (as opposed to a tower).
- $REFURB$-A dummy variable equal to one if the system is refurbished.
- $MODEM$-A dummy variable equal to one if the system includes a modem.
- $WINNT4$-A dummy variable equal to one if the operation system is Windows NT 4.0.
- $WINNT$-A dummy variable equal to one if the operating system is Windows NT.
- $WIN98$-A dummy variable equal to one if the operating system is Windows 98.
- $WIN95$-A dummy variable equal to one if the operating system is Windows 95.
- ξ_j-An omitted product attribute.

One specification for consumer preference we might consider is:

$$
\begin{aligned}
u_{ij} \;=\; & \beta_{i,1}\log(CPU_j) + \beta_{i,2}\log(RAM_j) + \beta_{i,3}\log(HD_j) + \beta_{i,4}CDROM_j + \beta_{i,5}MMX_j + \beta_{i,6}SCSI_j \quad (6) \\
& \beta_{i,7}DVD_j + \beta_{i,8}NIC_j + \beta_{i,9}MON_j + \beta_{i,10}ZIP_j + \beta_{i,11}DT_j + \beta_{i,12}REFURB_j + \beta_{i,13}MODEM_j \\
& \beta_{i,14}WINNT4_j + \beta_{i,15}WINNT_j + \beta_{i,16}\log WIN98_j + \beta_{i,17}WIN95 + \beta_{i,18}\log(\xi_j) + c
\end{aligned}
$$

In (6), household i's utility is a log-linear function of the product characteristics that we model as continuous and a linear function of the discrete characteristics. The coefficients $\beta_{i,1} - \beta_{i,18}$ are allowed to vary by household i. Following the literature, we will refer to these as random coefficients. In commonly used models, such as the multinomial logit or probit, it is commonly assumed that $\beta_i = \beta$, that is, the marginal valuation of a product characteristic is identical for all households. Our specification is considerably more general in the sense that the marginal valuations are allowed to differ by households. A second difference is

8

that there is no random normal or extreme value preference shock in the utility function. As we will discuss later, this is because the model above will be just identified if we do not impose functional form assumptions on the distribution of the β_i. Therefore, it is not possible to both identify a nonparametric distribution of random coefficients and a random preference shock. We opt to not include a random preference shock in our model since this would only be identified if we made parametric assumptions about the distribution of the β_i.

It is also possible to model the joint distribution of taste parameters and household demographics. Let $d_i = (d_1, ..., d_D)$ be a $1 \times D$ vector of household i's demographic characteristics. Let

$$\beta_{i,k} = f_k(d_i; \theta) + \eta_{i,k} \tag{7}$$

$$E\left(\eta_i | d_i\right) = 0. \tag{8}$$

Modeling household preferences as a function of demographics is common in the literature. Equations (7)-(8) allow the economist to learn about the joint distribution of tastes and demographics. The term $f_k(d_i; \theta)$ is the projection of tastes onto a parametric model of household demographics. The residual, $\eta_{i,k}$, can be interpreted as a household specific preference shock for characteristic k. Including equations similar to (7)-(8) is useful for instance in studies of residential sorting (see Bayer, McMillan and Rueben (2004)) where the goal is to understand the distribution of demographics by community. In some demand systems, such as Petrin 2002, demographic characteristics are natural taste shifter (e.g. families with children are more likely minivan buyers). Finally, knowledge of demographics may be useful in marketing exercises involving couponing or target marketing.

If we make the functional form assumption (6), then the first order conditions for the continuous product

characteristics can be written as:

$$\beta_{i,1} = \frac{\partial \mathbf{p}}{\partial CPU} CPU_j \tag{9}$$

$$\beta_{i,2} = \frac{\partial \mathbf{p}}{\partial RAM} RAM_j \tag{10}$$

$$\beta_{i,3} = \frac{\partial \mathbf{p}}{\partial RAM} HD_j \tag{11}$$

$$\beta_{i,18} = \frac{\partial \mathbf{p}}{\partial \xi} \xi \tag{12}$$

The equations above are a simple application of (5) to the model (7). In the following section, we will demonstrate how the empirical analogue of (9)-(12) can be used to recover preference parameters for households that choose to purchase a computer during a given period.[4]

There is no first order condition associated with the discrete attributes such as CDROM. A necessary condition for optimality, however, is that if the user chooses CDROM, she receives higher utility than if she does not. For instance, suppose that household i chooses product j^*. Define $\widehat{x}_i$ as a vector of observed characteristics with $CDROM = 1$ and all other elements set equal to their corresponding values in x_{j^*}. Define $\overline{x}_i$ similarly, except $CDROM = 0$. The implicit price faced by household i for $CDROM$ in the market is then $\frac{\Delta p}{\Delta CDROM} = \mathbf{p}(\widehat{x}_i, \xi_j) - \mathbf{p}(\overline{x}_i, \xi_j)$. Utility maximization implies that

$$[CDROM_{j^*} = 1] \implies \left[\beta_{i,4} > \frac{\Delta p}{\Delta CDROM}\right], \tag{13}$$

$$[CDROM_{j^*} = 0] \implies \left[\beta_{i,4} < \frac{\Delta p}{\Delta CDROM}\right]. \tag{14}$$

---

[4] In principal, it is possible to augment the model to include an outside good, as discussed in Bajari and Benkard (2004). However, we do not consider that possibility here.

That is, if household i chooses SCSI, we can infer that i's preference parameter exceeds the implicit price for this characteristic.

3 ESTIMATION

Our estimation approach involves three steps. For simplicity, we focus on the case where a household is only observed choosing a single time. While long panels of choices by individual households may be available in some applications, such as scanner data, more commonly the economist observes a given household's choice at most once. In the first step, the economist estimates a flexible hedonic price function. In the second step, the economist applies the empirical analogue of the first order conditions (9)-(12). Finally, the economist regresses estimated preference parameters on household demographics to recover an estimate of (7). We also discuss how to estimate preferences for dichotomous characteristics by maximum likelihood.

3.1 First Step: Estimating the Hedonic Price Function

The are a number of methods that can be used for flexibly estimating the hedonic. The method that we will use is based on local linear modeling. Here, we will describe the main steps in the estimator. Readers interested in a more detailed discussion can consult Fan and Gijbels (1996). This is because the method is particularly simple to program and because we have found that it performed well compared to alternative methods.

Fix a particular product j^*. Then then our estimate of the hedonic price function will be:

$$p_{j^*} = \widehat{\alpha}_{0,j^*} + \widehat{\alpha}_{1,j^*} x_{1,j} + ... + \widehat{\alpha}_{K,j^*} x_{K,j} + \xi_j \tag{15}$$

In our notation, we denote the estimated hedonic coefficients as $\widehat{\alpha}_{0,j^*} - \widehat{\alpha}_{K,j^*}$ to emphasize the fact that our

estimates will be local in the regression coefficient with differ for each value of j^*. In equation (15), the error term ξ_j will be interpreted as an omitted product attribute. While there are certainly other interpretations of the hedonic residual, we believe that this interpretation is the most important in our data. Given the lack of appropriate instruments, we maintain the standard hedonic assumption that the unobserved product characteristics are independent of the observed product characteristics.

Weighted least squares is the used to estimate $\widehat{\alpha}_{0,j^*} - \widehat{\alpha}_{K,j^*}$ for every product in the market. That is,

$$\alpha_{j^*} \;=\; \arg\min_{\alpha} \left(\mathbf{p} - \mathbf{X}\alpha\right)' \mathbf{W}(\mathbf{p} - \mathbf{X}\alpha) \tag{16}$$

$$\mathbf{p} \;=\; [p_j], \; \mathbf{X} = [x_j], \mathbf{W} = diag\{K_h(x_j - x_{j*})\} \tag{17}$$

In equation (16) and (17), $\mathbf{p}$ is a $J \times 1$ vector of all prices for all products j in a given metropolitan area, $\mathbf{X}$ is a $J \times (K+1)$ matrix of regressors, which for each product includes an intercept and seven characteristics and $\mathbf{W}$ is a $J \times J$ matrix of kernel weights. Notice that in our analysis, we do not pool observations across markets t. The hedonic is meant to capture the budget constraint faced by consumer i. Since different markets may be in different equilibria, pooling data would not be appropriate. Also, notice that the kernel weights are a function of the distance between the characteristics of product j^* and product j. Thus, the local linear regression assigns greater importance to observations with characteristics close to j^*.

Our estimates of equation (16) and (17) allow us to recover an estimate of the unobserved product characteristic

$$\xi_{j^*} = p_{j^*} - (\widehat{\alpha}_{0,j^*} + \widehat{\alpha}_{1,j^*}x_{1,j} + ... + \widehat{\alpha}_{K,j^*}x_{K,j}) \tag{18}$$

We note that the local linear regression imposes stronger assumptions than the framework used in BB. The

omitted product characteristic, ξ_{j^*}, is assumed to be additively separable in the local linear framework. In BB, this assumption is relaxed.

In local linear regressions, the choice of kernel and bandwidth is extremely important. In our application, we will specify the weights as follows:

$$K_h(z) = \frac{\Pi_k N\left(\frac{z_k}{1.5 \cdot \widehat{\sigma}_k^2}\right)}{\frac{x_k}{1.5 \cdot \widehat{\sigma}_k^2}} \tag{19}$$

In equation (18), the function K is a product of standard normal distributions which we denote using N. In equation (18), for the kth characteristic, we evaluate the normal distribution at $\frac{z_k}{1.5 \cdot \widehat{\sigma}_k^2}$, where $\widehat{\sigma}_k^2$ is the sample standard deviation of characteristic k. Fan and Gijbels (1996) describe asymptotically appropriate methods for choosing the bandwidth. However, in our application, since the hedonics depend on several covariates, the asymptotic approximations are not very likely to be reliable. Therefore, we choose the bandwidth based on visual inspection.

Given the large number of covariates, we should not interpret these estimates as "nonparametric". However, compared to other flexible functional forms, such a high order polynomials, local linear methods appeared to give much more plausible estimates of the implicit prices.

3.2 Second Step: Applying the First Order Conditions.

After estimating the implicit prices, we next estimate the preferences for continuous characteristics. If household i chooses product j^*, equation (10) must hold. Using estimates of implicit prices obtained from the first step and the observed choice of $x_{j^*,k}$, an estimation $\widehat{\beta}_{i,k}$ of $\beta_{i,k}$ can be recovered as follows:

$$\widehat{\beta}_{i,k} = x_{j^*,k} \frac{\partial \widehat{\mathbf{p}}_m(x_{j^*}, \xi_{j^*})}{\partial x_{j,k}} \tag{20}$$

13

In equation (20), we recover household i's preference for characteristic k using our estimate of the (local) implicit price recovered from the first step, $\frac{\partial \widehat{\mathbf{p}}_m(x_{j*}, \xi_{j*})}{\partial x_{j,k}}$. Since we observe each household's choice, $x_{j*,k}$, we can recover an estimate of the unobserved taste parameter, $\widehat{\beta}_{i,k}$.

Since the preference parameter for every individual can be recovered using our first order condition, the population distribution of tastes in the market can also be estimated. If we observed a sufficiently large number of households, we would be able to nonparametrically recover the joint distribution of the $\beta_{i,k}$ for the continuous product characteristics. Therefore, can identify a nonparametric distribution of these random coefficients (see Bajari and Benkard 2002 for a further discussion). In the next step, we describe how to estimate the joint distribution of tastes and demographics.

3.3 Third Step: Modeling the Joint Distribution of Tastes and Demographics.

After household level preference parameters are recovered, we then estimate (7)-(8). We could easily do this using very flexible, local linear methods. However, for presentation purposes, it is more convenient to model the joint distribution of tastes and demographic characteristics using a linear model. For continuous characteristics we let:

$$\beta_{i,k} = \theta_{0,k} + \sum_s \theta_{k,s} d_{i,s} + \eta_{i,k} \tag{21}$$

We then simply estimate (21) using regression. The regression that we run is:

$$\widehat{\beta}_{i,k} = \theta_{0,k} + \sum_s \theta_{k,s} d_{i,s} + \eta_{i,k} \tag{22}$$

In equation (22), we have simply substituted our estimate of $\widehat{\beta}_{i,k}$ from the second stage into equation (21).

Given estimates $\widehat{\theta}_{k,s}$, the residuals can be interpreted as household specific taste shocks. Note that, once

again, we have not imposed any parametric restrictions on the $\eta_{i,k}$. Previous work, such as BLP (1995) and Petrin (2002) imposes parametric assumptions on taste shocks because of the computational demands of their estimator. Since our estimator is computationally light, such requirements are not necessary.

Equations (13) and (14) imply that preference parameters for dichotomous characteristics are not identified. We can only infer that the preferences for a particular household are above or below the threshold value equal to the implicit price of the discrete characteristic. Given this lack of identification, we will use a parametric model for these taste coefficients. We shall assume, as in equation (21), that preferences are a linear function of demographics. However, unlike the continuous case, we will impose a parametric distribution on $\eta_{i,k}$. For this application, we shall assume that they are normally distributed, although other distributions could also easily be estimated. If $\eta_{i,k}$ is normally distributed with mean zero and standard deviation σ, then by equation (13) and (14), the probability that household $i = 1, ..., I$ chooses CDROM is

$$1 - N(\theta_{0,k} + \sum_s \theta_{k,s} d_{i,s} - \frac{\Delta p}{\Delta CDROM}; \sigma) \tag{23}$$

where N is the normal cdf. The likelihood function for the population distribution of CDROM

$$L(\theta, \sigma) = \prod_{i=1}^{I} N(h(d_i; \theta_k) - \frac{\Delta p}{\Delta CDROM}; \sigma)^{1-CDROM_{j^*(i)}} (1 - N(h(d_i; \theta_k) - \frac{\Delta p}{\Delta CDROM}; \sigma))^{CDROM_{j^*(i)}}. \tag{24}$$

$$\text{where } h(d_i; \theta_k) = \theta_{0,k} + \sum_s \theta_{k,s} d_{i,s} \tag{25}$$

where in equation (24), $single_{j^*(i)}$ is an indicator variable that is equal to one if household i purchases single detached housing and zero otherwise. This is a version of the probit model where instead of normalizing $\sigma = 1$, we have normalized the coefficient on price equal to -1. We estimate the model above using

maximum likelihood.

In principal, we could model correlation between the taste coefficients for all discrete product characteristics using a multivariate normal distribution and estimate a more flexible model of how tastes for characteristics are correlated. However, for expositional clarity, we estimate the tastes for each product characteristic independently. An alternative approach, which does not require assuming that tastes lie in a parametric family, is to use the bounds approach described in Bajari and Benkard (2002).

4 RUMs.

One of our goals is to compare hedonic models to more conventional RUMs. Therefore, in this section, we describe a fairly general semi-parametric RUM. This model nests as special cases many commonly used RUMs such as the logit, nested logit, GEV, multinomial probit, as well as random coefficients versions of these models, such as BLP models.

In the model, each consumer chooses between J mutually exclusive alternatives. We index consumers by $i \in 1..I$ and products by $j \in 1...J$. Following the previous literature, we assume that individuals' utility functions can be written as a function of individual characteristics (describing individual tastes), product characteristics, and an additively separable random error term:

$$u_{ij} = u(x_j, y_i - p_j, \beta_i) + \epsilon_{ij} \qquad \text{for } j \in 1..J \tag{26}$$

In equation (26), $x_j \equiv (x_{j,1}, x_{j,2}, ..., x_{j,K})$ is a K-dimensional vector of characteristics associated with product j. We assume that $x \in \mathcal{X}$, where $\mathcal{X} \subseteq R^K$ is a compact set. In addition, p_j is the price of product j, y_i is the income of consumer i, β_i is a vector of individual taste parameters with support $\mathcal{B} \subseteq R^B$, and ϵ_{ij} is an individual and product specific random error term. The term $y_i - p_j$ represents consumption of all

16

other goods, which we treat as a composite commodity denoted as c.

We assume that the utility obtained from not purchasing any variety of the good is also a function of a random error term,

$$u_{i0} = u(\tilde{0}, y_i, \beta_i) + \epsilon_{i0}. \tag{27}$$

It is not necessary to include the outside good in the model for most of what follows. We include it because its presence underscores some of the undesirable properties of the model, and because much of the previous literature models the outside good similarly.

In the model, consumers are rational utility maximizers. Consumer i chooses product j if and only if j maximizes utility,

$$i \text{ chooses } j \iff u_{ij} \geq u_{ik} \text{ for all } k \neq j. \tag{28}$$

Let $s_{i,j} \equiv P_i(j|\beta_i, y_i)$ denote the probability that consumer i chooses product j conditional on β_i and y_i, and let $\epsilon_{i,-j}$ denote the vector of error terms for individual i excluding product j. By equation (28) it follows that:

$$P_i(j|\beta_i, y_i) = \int_{-\infty}^{\infty} \int_{-\infty}^{u_{i,j}-u_{i,0}} \cdots \int_{-\infty}^{u_{i,j}-u_{i,j-1}} \int_{-\infty}^{u_{i,j}-u_{i,j+1}} \cdots \int_{-\infty}^{u_{i,j}-u_{i,J}} f(\epsilon|\beta_i, y_i) d\epsilon_{i,-j} d\epsilon_{i,j}. \tag{29}$$

That is, the probability that consumer i chooses product j is the probability that the realization of ϵ makes choice j utility maximizing for consumer i. If the researcher has access to micro data containing individual level choices, equation (29) can be used to construct a likelihood function.

Let $s_j \equiv P(j)$ denote the market share of product j, i.e. the probability that j is chosen averaging over

the $i = 1, ..., I$ consumers. Then,

$$P(j) = \int P_i(j|\beta_i, y_i)dF(\beta_i, y_i). \tag{30}$$

In equation (30) we integrate out over the population distribution of β and y in order to compute the probability that product j is chosen by a randomly selected consumer. In product markets, $P(j)$ is typically interpreted as the demand for product j. If only aggregate market shares are observed, then equation (30) can be used to construct a likelihood function.

In its general form, the model in equations (26)-(30) is quite flexible. It nests as a special case many standard RUMs such as the logit, GEV, probit, and BLP. However, in applied work, the model is typically not estimated in full generality due to the complexity of evaluating the likelihood function. Instead, econometricians typically make restrictive functional form assumptions to simplify the numerical analysis of the model. For example, in the random coefficients logit model it is assumed that ϵ_{ij} is iid, independent of β_i and y_i, and is distributed extreme value. In that case, the integral (29) has a closed form solution. In the random coefficients probit model, it is assumed that ϵ_{ij} is independent of β_i and y_i and normally distributed. In that case, simulation methods such as Gibbs sampling can be used to compute (29).

5   Economic Implications of Standard RUMs.

In this section we discuss some criticisms of commonly used RUMs that have been discussed in the literature. In its general form, the model in equations (26)-(30) is quite flexible. It nests as a special case many standard RUMs such as the logit, GEV, probit, and BLP. It also nests the hedonic model of Section 2 as a special case. Estimating this model is complete generality is not computationally feasible. Functional form assumptions are used to make estimation computationally feasible. Below, we discuss some assumptions

18

that are commonly made in empirical applications.

5.1 Assumptions

The first assumption we make is a common assumption regarding the independence of the consumer taste coefficients and the error terms.

Assumption I The vector of errors, ε, is independent of consumer tastes and income. That is, $f(\beta, y, \varepsilon) = f_{\beta,y}(\beta, y)f_\epsilon(\epsilon)$.

This assumption is made for convenience only, while it may be unpalatable in many applications, it is not important to this paper. All of the results that follow can be shown to hold without it.

Next, we restrict the set of utility functions that we will consider.

Assumption U
(i)For all $(x, \beta) \in X \times B$, $u(x, c, \beta)$ is continuous in all its arguments, and $u(x, \cdot, \beta)$ is strictly increasing.
(ii)For every $(\beta, y) \in B \times R^+$ and every $0 < p < y$, $|u(., y - p, \beta)| < \infty$.

Assumption $U(i)$ says that the deterministic part of the utility functions is continuous and that individuals have monotone preferences with respect to the composite commodity. Assumption $U(ii)$ says that the utility function as defined over characteristics is bounded for every individual so long as the budget constraint is satisfied. Assumption U holds in all applications of RUMs in the previous literature that we are aware of.

The next assumption is the critical assumption driving our results.

Assumption R For all $M < \infty$, there exists a δ_M such that $Pr(\epsilon_{ij} < M|\epsilon_{i,-j}) < \delta_M < 1$ for all i, j pairs, all $\epsilon_{i,-j}$, and all $J \in \mathcal{Z}_+$.

Assumption R amounts to assuming that the conditional error distributions have unbounded upper support. All GEV models, the probit model, and all GEV- and probit-based random coefficients models satisfy R, so long as the errors have strictly positive variance, mean that is bounded below, and so long as the errors are not perfectly correlated. Assumption R guarantees that $\lim_{J \to \infty} \max_{j \in 1...J} \epsilon_{ij} = \infty$ a.s.[5]

---

[5] The proof is a straightforward application of the Borell-Cantelli Lemma. See appendix.

5.2 Property One: The Shape of the Demand Curve

The first property implied by these assumptions is that the demand curve is never bounded above.

> **1. Demand is positive for every price vector.** Suppose that assumptions I, R and U hold. Suppose further that, either: (i) $u(x, c, \beta)$ is linear in c, or (ii) for all $\bar{\beta} \in B$, $F(y|\bar{\beta})$ has full support on R^+. Then in the model described above, for every product, demand is strictly positive for every price.

Conditions (i) and (ii) are satisfied in much of the previous literature on RUMs. For example, condition (i) is typically satisfied in standard applications of logit and probit. Under condition (i), income does not affect individuals' choices and thus income can be omitted from the analysis. Condition (ii) is satisfied in BLP (1995) and every application of BLP style models that we are aware of that do not satisfy (i).

This first property can be convenient for applied work, since market shares always have positive probability regardless of the prices. While it is convenient, it may lead to a potentially undesirable implications for demand. Since demand is positive at any price, the model might be biased toward generating large amounts of consumer surplus from each product. Several authors have pointed out that consumer surplus calculations can be strongly influenced by the upper tail of the demand curve in the logit model.

5.3 Properties of the Model When the Number of Products Increases

We now show that the behavior of existing models can lead to potentially undesirable properties when the number of products in the market can be large.[6]

> **2. Share of the Outside Good** Let s_{i0} denote the probability that individual i chooses the outside good. If I, R, and U hold, then as $J \to \infty$ either $s_{i0} \to 0$, or $s_{i0} > 0$ and $u(x_j, y_i - p_j, \beta_i) \to -\infty$ for all but a finite set of goods.

If we are trying to describe choice behavior in a narrowly defined market, it seems unreasonable that the

---

[6] Note that in this section we have intentionally omitted the process by which the products are added when the limit is taken because the properties listed hold regardless of what process is generating the added products so long as the assumptions listed are satisfied.

share of the outside good tends to zero when there are many varieties of the good. Some characteristics are typically shared by all inside goods (e.g., cell-phones are typically used to make telephone calls, breakfast cereals are typically eaten for breakfast). If an individual has a strong negative taste for a common characteristic of the inside good (e.g., they have no cell-phone service in their area, they do not like cereal, or they do not eat breakfast), then no matter how many varieties are available we should not expect the individual to purchase the good. At very least it would be desirable for the structural demand model to be rich enough to allow for the possibility that the outside good retains positive share in the limit.

Note that this property is not limited to just the outside good. For any fixed set of parameters the model similarly implies that the share of every good goes to zero as products are added to the market, so long as the products added are of sufficiently high mean quality. This property also contradicts the intuition of many theoretical differentiated products models, which suggest that the location where products enter in characteristics space should matter in determining whether or not shares go to zero in the limit.

We now list one additional assumption regarding the error distribution:

Assumption H For each j, the limit as ϵ_j tends to the upper limit of its support of the hazard rate of $F_{\epsilon_j}(\cdot)$ is infinite, i.e., $\lim_{\epsilon \to b} \frac{f_{\epsilon_j}(\epsilon)}{1 - F_{\epsilon_j}(\epsilon)} = \infty$, where b is the upper end of the support of $F_{\epsilon_j}(\cdot)$ and b may equal infinity.

Assumption H is satisfied by all bounded distributions and the normal distribution (probit), but not extreme value distributions. It turns out that whether or not assumption H holds determines some important theoretical properties of the choice model.

3. Lack of Perfect Substitutes Suppose that ϵ_{ij} is iid and that I, R, and U hold, but H does not hold. Then each product almost surely does not have a perfect substitute even as $J \to \infty$. That is, even when the number of products is infinite, each individual would suffer utility losses that are almost surely bounded away from zero if her first choice product were removed from the choice set.

4. Lack of Perfect Competition Suppose that ϵ_{ij} is iid and that I, R, and U hold but H does not hold. Then in a symmetric Bertrand-Nash price setting equilibrium with single product firms, markups are

almost surely bounded away from zero when $J \to \infty$.

Properties three and four cover only the *iid* case for simplicity. They are closely related so we discuss them together. Property three implies that, even in the limiting case, the assumptions commonly maintained (e.g., in logit, GEV, and random coefficients logit models) are not sufficient to imply that individuals would be willing to switch to their second favorite product with zero compensation when the number of products becomes large. As a result, markups also remain bounded away from zero in the limit.[7]

Again, if we are considering a narrowly defined market, then we might expect that the product space should fill up eventually and products should become close substitutes in the limit. The functional form assumed in extreme value based models does not allow for this possibility.

Properties three and four do not hold necessarily, but depend on the shape of the distribution of ϵ_{ij}, and specifically the upper tails of the distribution. They hold for the GEV and the logit, including random coefficients logit models. This suggests that, even if independence is assumed, the probit model might have better economic properties than the logit model. In particular, probit may be preferable to logit in certain applications such as welfare studies, where there may be a tendency for logit to overvalue additional choices, and in studies of competition in differentiated products markets, where logit may tend to imply markups that are too high as a result of overestimating the differentiation between products. However, the practical importance of this result still needs to be investigated.

5. Contribution of Observed Characteristics Suppose that assumptions I, R and U hold. Then the contribution of the observed characteristics to utility almost surely goes to zero as the number of products becomes large. That is, $\lim_{J \to \infty} \frac{\epsilon_{ij^*}}{u_{ij^*}} = 1$ a.s., where $j^* = \arg \max_{j \in 0..J} u_{ij}$.

Property five shows that in RUMs the contribution to utility from observed variables changes depending

---

7   Anderson, de Palma, and Thisse also show this for the standard logit model.

on the number of products in the market, which also seems economically unintuitive. It seems more intuitive that the percentage of the utility explained by observables should remain more or less constant for any given market as the number of products in the market changes.

    6. Compensating Variation Suppose that assumptions I, R, and U hold. Then as the number of products becomes large the compensating variation for removing all of the inside goods almost surely tends to infinity for every individual.

Property six singles out a problem with using RUMs for welfare analysis. The model implies that with enough products to choose from every individual needs arbitrarily large amounts of income to be as well off with the outside good alone as with the inside good. The implication is that every individual is costlessly receiving arbitrarily large (relative to income or price) levels of utility from something about the good that we cannot observe.

5.4 Discussion

All of the properties above are driven by properties of the random error term, particularly through changes in its dimension driven by changes in the number of products. Caplin and Nalebuff (1991) show that one interpretation of the error terms in the standard RUM model is as a "taste for products", with the following construction:

$$\epsilon_{ij} = \lambda_i' \eta_j. \tag{31}$$

where λ_i is individual i's J-dimensional random vector of tastes for each product, and η_j is a vector of zeros with a one in the jth element. This construction makes it clear that the standard econometric models are special cases of pure characteristics models in which individuals have preferences (with a specific distribution) over product dummies.

We are not the first to raise many of these issues. For instance, Ackerberg and Rysman (2001), Berry and

23

Pakes (2000), Caplin and Nalebuff (1991), and Petrin (2002) note that the logit model tends to overstate the benefits from product variety. Also, Andersen, de Palma and Thisse (1992) establish that Bertrand competition does not converge to perfect competition in the logit model. However, to the best of our knowledge, our second, third and fifth undesirable property listed above are new to the literature. In addition, we believe that the result that Bertrand competition does converge to perfect competition for the probit model is new.

In the hedonic model of section 2, each consumer i has a unique taste coefficient for each characteristic. The hedonic model is similar to many commonly used RUMs except for the following two features:

1. There is no random idiosyncratic taste shock ε_{ij}.
2. No parametric or independence restrictions are imposed on the joint distribution of β_i except for the normality assumption on discrete product characteristics.[8]

It can easily be seen that the hedonic model does not impose many of the potentially undesirable assumptions of standard RUMs listed above. First, in the hedonic model, it is not always the case that demand is positive at any price. Consider the demand for product j, suppose that there exists a product j' such that:

$$x_{j',k} > x_{j,k} \text{ for } k = 1, ..., K. \tag{32}$$

If $p_{j'} < p_j$ then the demand for product j will be zero since product j' has a higher value of all characteristics but has a lower price.

Furthermore, for individuals with low preference for characteristics of the inside good relative to their preference for the composite commodity, only a very low price will induce them to purchase the good. If a consumer's willingness to pay for characteristics of the good is below the marginal cost of production, then it may be that no rational price is low enough to induce purchase. Thus, the share of the outside good does not necessarily tend to zero as more products enter the market.

---

[8] This assumption is required for point identification of the model. For an approach that uses set identification, see BB.

If the distance between the characteristics of product j and j' is small and preferences are Lipschitz continuous, then in the pure hedonic model these products will be close substitutes. As a result, as the number of products becomes infinite, all products will have a perfect substitutes and markups in Bertrand price competition will tend to zero.

Finally, the pure hedonic model does not imply that a continuum of products provides consumers with infinite utility relative to income or price. Thus, the compensating variation for removing all inside goods remains bounded even in that case.

However, the hedonic model also has limitations. In the hedonic model, not all products are required to be strong gross substitutes. The model may also impose strong assumptions on cross price elasticities. For instance, in the vertical model, a given product j only has positive cross price elasticities with it's two neighboring products. This is potentially unappealing in markets where there are only a handful of observed characteristics. (See Anderson, DePalma and Thisse (1992)).

In addition, the empirical results in BB suggest that if many products are included in the choice set, then the perfect information assumption in the pure hedonic model can lead to demand curves that are too elastic. To summarize their results, one reason for including the random error term in the utility function is that it could represent imperfect information (e.g. due to a cost of acquiring information about products). Leaving out this imperfect information may imply too high a degree of substitutability across products.

6 Application.

In this section, we compare the hedonic estimator to a standard RUM based on Berry (1994). The data comes from the PC Data Retail Hardware Monthly Report and includes quantity sold, average sales price, and a long list of machine characteristics for desktop computers. The data reportedly covers 75% of retail

computer sales in the United States. The raw data set contained 29 months of data, but we use only the

data for the last period, December 1999, covering 695 machines. We chose to use data from a single period

to keep the exposition simple. Please see Benkard and Bajari (forthcoming) for a more detailed discussion

of this data.

The raw data contained a large number of characteristics, including dummies for each individual processor type. We eliminated the processor dummies in favor of a CPU benchmark variable.[9] The final data set

contained 19 characteristics, including five operating system dummies (Win 3.1, NT 4.0, NT, Win 98, Win

95) plus CPU benchmark, MMX, RAM, hard drive capacity, SCSI, CDROM, DVD, modem, modem speed,

NIC, monitor dummy, monitor size (if monitor supplied), zip drive, desktop (versus tower), and refurbished.

Summary statistics for the product characteristics are given in table 1.

6.1 Logit Model with Omitted Product Characteristics

The first model we consider is a logit model with omitted attributes as in Berry (1994). In the Berry logit,

consumer utility takes the form:

$$
\begin{aligned}
u_{ij} \;=\; & \beta_1 \log(CPU_j) + \beta_2 \log(RAM_j) + \beta_3 \log(HD_j) + \beta_4 SCSI_j + \beta_5 MMX_j + \beta_6 CDROM_j \\
& \beta_7 DVD_j + \beta_8 NIC_j + \beta_9 MON_j + \beta_{10} ZIP_j + \beta_{11} DT_j + \beta_{12} REFURB_j + \beta_{13} MODEM_j \\
& \beta_{14} WINNT4_j + \beta_{15} WINNT_j + \beta_{16} \log WIN98_j + \beta_{17} WIN95 + \beta_{18} p_j + \xi_j + \varepsilon_{ij}
\end{aligned}
$$

This is analogous to the model of section 2, but it includes a random preference shock ε_{ij} and does not

---

[9] The CPU benchmark was obtained from www.cpuscorecard.com. A regression of CPU benchmark on chip speed interacted with chip dummies yielded an R^2 of 0.999.

permit random coefficients. We estimated the model using the techniques proposed in Berry (1994). For instance, an instrument for the price of product j, following Berry, is:

$$\sum_{k \neq j} CPU_k \tag{33}$$

That is, we sum up all of the CPU performance measures for all products excluding j. The intuition behind this instrument is that it is a measure of the isolation of j in product space. If j has a particularly fast CPU speed, then the term (33) will be particularly large for this product. We construct this instrument for all of the 17 product characteristics listed above. If product j is more isolated in product space, then there will be more market power associated with this product. Models of price competition suggest that this should shift price, all else held constant.[10]

The estimates are shown in Table 2. The estimation method followed Berry (1994). By transforming the dependent variable to logs, the omitted product characteristic enters into the model linearly. We then use straightforward 2SLS to estimate the model. Most of the coefficients that are significant have the expected sign. For instance, the price coefficient is negative, while the CPU benchmark and RAM coefficients are positive. The coefficients on SCSI and Zip Drive have a negative sign which is counter-intuitive. Many of the other coefficients are not significant at conventional levels, which is perhaps not surprising given the large number of characteristics included in the model. Accounting for price endogeneity, using the instruments (33) turned out to be quite important. In specifications without these instruments, the price coefficient was positive. This is consistent with the biases discussed in Berry (1994) and BLP.

---

[10] We did not include an outside good in either the logit or hedonic specification. This was to simplify the presentation of the results. All of the derived estimates should be interpreted as the preferences of the population of buyers who actually purchased computers in December 1999.

6.2 Hedonic Model of Preferences.

The hedonic model of preferences that we us is based on equation (6). In practice, it is not practical to use nonparametric techniques on a 18-dimensional system because of the curse of dimensionality. In order to reduce the dimensionality of our system, we model the hedonic as nonparametric in CPU Benchmark, RAM, HD and CD ROM.[11]

Following Benkard and Bajari (forthcoming), the other characteristics are assumed to enter into the hedonic linearly. Under this assumption, we are able to modify the estimator described in Section 3 as follows. First, we run a linear hedonic regression that includes all 18 product characteristics listed in Section 2. We then subtract the implicit prices of the 14 characteristics assumed to enter linearly into the hedonic from the price observed in the market. The nonparametric hedonic regression of section 3 is then performed on the 4 product characteristics CPU Benchmark, RAM, HD and CD ROM.[12]

In Table 3, we summarize our results from the hedonic regressions. The first column contains the coefficients from a linear regression and the second from our local linear regression. Most of the coefficients have the expected signs and plausible magnitudes with the exception of MMX. In the second column of the table, we summarize the distribution of the implicit prices from the local linear model. Recall that in the method described in Section 3, we find a vector of implicit prices for each product observed in the data. The second column summarizes the mean of these 703 implicit prices (the standard deviation is in

---

[11] CD ROM is not a particularly important contributor to the overall price. However, it affords us the opportunity to demonstrate how to estimate taste coefficients for discrete characteristics.

[12] We make the linearity assumption for analytical tractability. However, this assumption is probably not an unrealitic approximation since these characteristics include operating system dummies and components such as DVD Drives and Modems which peripheral devices which can easily be added or removed from the system.

In the results we report, the value of ξ_j includes the value of the 14 linear characteristics plus the residual from the local linear regression. This is done to simplify the presentation of the results and has little effect on our final results because these 13 charactersitics account for only a fairly small fraction of the observed prices.

parentheses). The average implicit price is similar to the OLS regression coefficient. However, we note that there is a fairly wide dispersion in these prices. This dispersion is useful in identifying the distribution of our random coefficients since this implies that in the data there is variation in the marginal willingness to pay.

In Table 4, we compute the mean willingness to pay for a 1 percent increase in the continuous product characteristics. Given the model of Section 2, this computation can be easily made. Consider the continuous product characteristic CPU. Suppose that household i purchases product j. Then, our estimate of i's random coefficient for CPU is recovered by evaluating:

$$\widehat{\beta}_i = \widehat{\alpha}_{CPU_j} \cdot CPU_j$$

where $\widehat{\alpha}_{CPU_j}$ is the implicit price of CPU speed for product j from the local linear regression. Given the functional form (6), i's willingness to pay for a 1 percent increase in CPU speed is a transfer (in terms of the composite commodity c) that keeps utility constant. Simple algebra implies this is equal to:

$$\widehat{\beta}_i \left(\log\left(1.01 CPU_j\right) - \log\left(CPU_j\right) \right)$$

We perform this calculation for every person in our sample and summarize the distribution of willingness to pay in the second column.[13] The marginal valuation of these product characteristics seems consistent with economic intuition. For instance, a 1 percent increase in RAM would lead to 0.74 extra MB. This would be valued at \$1.93 by an average consumer. Including a CD ROM drive would lead would valued at

---

[13] Note that people who purchase the same computer will have the random coefficients for continuous product characteristics.

$107.22 by an average consumer.

In Table 5, we summarize the joint distribution of the random coefficients. We find that the random coefficients are correlated in a manner that seems consistent with economic intuition. For instance, the first column indicates that the valuation of CPU is positively correlated with RAM, HD and ξ. This is a priori plausible since we would expect CPU intensive users will on average require more memory and more sophisticated peripherals (as reflected by ξ).

In applications of BLP, parametric assumptions about the distribution of random coefficients are made in order to make the estimator computationaly tractable. For instance, the random coefficients are commonly assumed to be independently distributed. The hedonic model does not impose these assumptions. The results of Table 5 suggests that these simplifying assumptions are violated in this market.

6.3   Comparison of Estimators.

Next, we attempt to assess the empirical relevance of the properties discussed in Section 5. We will compare the RUM and hedonic models in two ways. First, we compare the price elasticities predicted by each of the models and the shape of the demand curves. Second, we compute the welfare loss from removing a consumer's most favored product from the choice set.

In Table 6, we list the five products with the largest market share in our data set. These products have between 32 and 64MB of RAM, 4.3-13.0 GB of Hard Drive Space and are all equipped with a CDROM. The products range in price from $539.75-$858.18.

In Table 7, we compute the elasticity of the residual demand curves for each products.[14] The elasticities implied by the logit model ranged from -0.93 to -1.65. The hedonic model generated considerably more

---

[14]   In these computations, we dropped all products with sales of less than 5,000 units since these may not be readily available to all consumers.

elastic demand curves with elasticities ranging from -5.42 to -97.34. The elasticities from the hedonic model are always several times larger than the logit model. This finding is consistent with the discussion in section 5. The error term for the logit model introduces, by construction, more "taste for product". In the hedonic model, products substitute more readily because product specific taste shocks are not included in the model. A priori, some of the elasticities from the hedonic model might be criticized as too large. An elasticity of -97.34 is probably unrealistically large (although not all elasticities are this large). This finding is consistent with our earlier discussion that competition in hedonic models can be too localized. Regardless of which elasticity is most plausible a priori, it is clear that the elasticity of the residual demand curve depends crucially on the method that is used.

In principal, it might be possible to search for a hybrid between hedonic and RUM models. These two models differ in their treatment of the unobserved product characteristic. In RUM models, one typically assumes that there is an omitted product characteristic for each of the J products as illustrated in equation (31).[15] In Benkard and Bajari (forthcoming), we demonstrate that using factor analysis it is possible to identify the number of omitted product attributes if we have a panel of products. The intuition is simple. If two products both have a high, omitted product characteristics then, ceterus paribus, their prices will be positively correlated. Using standard arguments from factor analysis, the covariance matrix of residuals can therefore be used to estimate the number of omitted factors. We conjecture that such a model would generate elasticities between the hedonic and logit model. The estimates in Benkard and Bajari (forthcoming) suggest that the number of omitted product characteristics for our computer data set is small- perhaps two or three. If we accept this conclusion, then our hedonic elasticities will tend to be too large

---

[15] In our RUM model, since we use Berry's (1994) procedure, there are actually 2J omitted product characterstics. Those contained in the error term (31) and ξ_j.

and the logit elasticities will be too small. However, the number of omitted product characteristics in the hedonic model is far closer to the estimates in Benkard and Bajari than the logit model.

In figure 1, we plot the residual demand curves for each of these products. The logit model generates a lens shaped demand curve that asymptotes towards both the horizontal and vertical axis. The demand curves implied by the hedonic model intersects the horizontal axis. Unlike the logit model, demand is not always strictly positive. The hedonic demand curve also become very elastic near the vertical axis. After prices reach some threshold value, such as near $600 for the HP 6595, consumers simply refuse to purchase the product. Our elasticity estimates and figures are consistent with properties 1, 3 and 4 listed in section 5. The logit model generates more "taste for product" and makes product markets appear to be considerably less competitive than the hedonic model.

Property 6 suggests that the logit model may generate counter-intuitive implications for compensating variation. In Table 8, we summarize the welfare loss from removing each consumer's favorite product from the choice set. That is, we compute the value of the composite commodity, c, required to keep the consumer indifferent between her most preferred and second most preferred product. The average welfare loss from the logit model is $355 compared to $9.43 in the hedonic model. This number seems large given that the average price of our top five computers is between $470 and $858 dollars. If the welfare for the 650,000 consumers is $219 million in the logit model compared to $6 million in the hedonic model. The complete distribution of welfare losses is plotted in figure 2. This figure reveals that the distribution of welfare losses is skewed.

The intuition behind these results can be seen in Table 9. On average, 40 percent of the utility in the logit model comes from the error term. Recall that the error term is a product specific taste shock. Hence,

as suggested by property 5, when the number of products becomes sufficiently large, the contribution of the idiosyncratic taste shocks can completely dominate the observed product characteristics. Therefore, consumers can receive a large reduction in utility by substituting to their second most favored product because so much of their valuation for the good is captured in the taste shocks. A priori, it is not unreasonable that there are product specific taste shocks for computers. However, we do not find it very plausible that so much of total utility comes from these shocks.

7 Conclusions.

In this paper, we compared random utility and hedonic models of demand. We began by describing a simple method for estimating a structural, hedonic model of demand. The model is based on Bajari and Benkard (2004) but is simpler to compute while maintaining much of the flexibility of their framework. Next, we demonstrated that standard random utility models have some undesirable economic properties when viewed as structural models of demand. Some commonly used models introduce to much "taste for product". As a result, market power may be overstated and welfare experiments from these models may be misleading. The empirical analysis of Section 6 suggests that these magnitudes may be large in applications.

We cautiously conclude that these results support further research on alternative demand models. For example, the random utility framework of Ackerberg and Rysman (2001), and the pure hedonic model of Berry and Pakes (2000) and Bajari and Benkard (2003), do not necessarily have the undesirable properties derived in section 3. Aggregate demand models such as those used in Hausman (1997) also do not necessarily have these properties. We speculate that there may also be alternative RUM models that eliminate these properties. However, that is an area for further research.

8 Appendix

8.1 Proofs for Section 2

8.1.1 Proof of Property 1

Consider demand at any point $(x_j, p_j, x_{-j}, p_{-j})$. In the case of (i), we fix $\bar{\beta}_i$ arbitrarily and choose $\bar{y}_i$ such that $F(\bar{y}_i|\bar{\beta}) > 0$. In the case of (ii), we fix $\bar{\beta}_i \in B$ arbitrarily and choose $\bar{y}_i$ such that $F(\bar{y}_i|\bar{\beta}) > 0$ and $\bar{y}_i > p_j$. This can be done since under (ii) y_i has full support conditional on β. Set $\epsilon_{ik} = 0$ for all $k \neq j$. Conditional on these values, product j is preferred to all other products if and only if

$$\epsilon_{ij} > \max_{k \neq j}\{u(x_k, \bar{y}_i - p_k, \bar{\beta}_i)\} - u(x_j, \bar{y}_i - p_j, \bar{\beta}_i) \equiv \bar{u}_k - \bar{u}_j. \tag{34}$$

By R, the probability corresponding to (34) is strictly positive.

Let

$$A_j = \{(y, \beta, \epsilon) \in R_+ \times B \times R^{J+1}| \ u_{ij} \geq u_{ik} \ \forall k \in 0..J\} \tag{35}$$

A_j represents the set of consumer demographics, taste coefficients, and error terms that rationalize a consumer choosing choice j. In order to find total demand for product j, we simply integrate A_j over the distribution of unobservables to get market share, and then multiply by the market size, M.

$$q_j(x, p; \theta) = M \int_{R^{J+1}} \int_B \int_{R_+} A_j \quad dF(\beta, y)dF(\epsilon) \tag{36}$$

Thus, using the same point above,

$$q_j(x_j, p_j, x_{-j}, p_{-j}) > M * Prob[\epsilon_{ij} > \bar{u}_k - \bar{u}_j|\bar{\beta}, \bar{y}_i]f(\bar{\beta}, \bar{y}_i) > 0 \tag{37}$$

Since we chose the vector of prices arbitrarily, demand is positive for every good for every price vector.

8.1.2 Proof of Properties 2 and 5

Lemma 1 *Assumption R implies that* $\lim_{J \to \infty} \max_{j \in 1..J} \epsilon_i = \infty a.s.$

For any $M < \infty$, let A_n be the event $\{\epsilon_{i0} < M, ..., \epsilon_{in} < M\}$. Then,

$$Pr(A_n) = Pr(\epsilon_{i0} < M)Pr(\epsilon_{i1} < M | \epsilon_{i0} < M) * ... * Pr(\epsilon_{in} < M | \epsilon_{i,-n} < M) \tag{38}$$

By assumption R, there exists a δ_M such that each term in the above expression is less than δ_M. Therefore $Pr(A_n) < \delta_M^n$. Since this holds for all n, the sum $\sum_n Pr(A_n)$ must converge. By the Borell-Cantelli Lemma $Pr(\limsup A_n) = 0$.

Properties 2, and 5 hold as a direct consequence of this.

8.1.3 Property 6

By the previous result,

$$\lim_{J \to \infty} \max_{j \in 1..J} \epsilon_{ij} = \infty \tag{39}$$

For individual i, the compensating variation for removing the inside goods, CV, is the solution to,

$$u(\tilde{0}, y_i + CV, \beta_i) = \max_{j \in 1..J} \{u(x_j, y_i - p_j, \beta_i) + \epsilon_{ij}\} - \epsilon_{i0} \tag{40}$$

Because utility is bounded, for any given individual the right hand side tends to infinity with J. (Technically, we also need to assume that products are added in such a way that the number of products that are within consumer i's budget tends to infinity with J.) Since preferences for c are monotone, it must be that CV does too.

8.1.4 Properties 3 and 4

We show properties 3 and 4 for the iid case. This case provides the central intuition that the thickness of the tails of the distribution matters in determining the limiting properties of the demand system.

We show two proofs: 1) $\lim_{J \to \infty} E[\epsilon_1^J - \epsilon_2^J] = 0$, where ϵ_1^J is the highest of J draws on ϵ and ϵ_2^J is the second highest, if and only if H holds; 2) as the number of products becomes large the markup in a symmetric Bertrand-Nash price-setting equilibrium with single product firms tends to 0 if and only if H holds.

1. Rewrite the desired expression using iterated expectations and bring the limit into the integral to get $E_{\epsilon_2^J}[\lim_{J \to \infty} E(\epsilon_1^J - \epsilon_2^J \mid \epsilon_2^J)]$. Now, note that we have shown above that $\lim_{J \to \infty} \epsilon_2^J = \infty$ a.s. It is also easy to show that $\lim_{x \to \infty} E[y - x \mid x] = 0$ if and only if the hazard rate of the conditional distribution $y|x$ goes to infinity as x becomes large. But, the conditional distribution of $\epsilon_1^J \mid \epsilon_2^J$ is proportional to the distribution of ϵ. Thus $\lim_{J \to \infty} E[\epsilon_1^J - \epsilon_2^J | \epsilon_2^J] = 0$ if and only if the hazard rate of $F()$ goes to infinity in the upper tail. This proves property four.

2. Consider J identical single product firms facing a demand system generated by a discrete choice model where the utility function is $u_{ij} = p_j - \epsilon_{ij}$ and the errors are assumed to be iid. In a symmetric Bertrand-Nash price setting equilibrium, all firms' prices are the same and each firm has equal market share

$s_j = 1/J$. The markup is $\frac{s_j}{-\frac{\partial s_j}{\partial p_j}}$. We now consider $\frac{\partial s_j}{\partial p_j}$:

$$s_j = Pr(\epsilon_k \leq \epsilon_j + p_k - p_j \forall k \neq j) \tag{41}$$

$$= \int_{-\infty}^{\infty} Pr(\epsilon_k \leq \epsilon_j + p_k - p_j \forall k \neq j \mid \epsilon_j) dP(\epsilon_j) \tag{42}$$

$$= \int_{-\infty}^{\infty} \Pi_{k \neq j} F(\epsilon_j + p_k - p_j) f(\epsilon_j) d\epsilon_j \tag{43}$$

$$= \int_{-\infty}^{\infty} F^{J-1}(\epsilon_j) f(\epsilon_j) d\epsilon_j \tag{44}$$

$$= 1/J \tag{45}$$

$$\Rightarrow \tag{46}$$

$$\frac{\partial s_j}{\partial p_j} = -\int_{-\infty}^{\infty} \sum_{k \neq j} \left(f(\epsilon_j + p_k - p_j) \Pi_{l \neq k,j} F(\epsilon_j + p_k - p_j) \right) f(\epsilon_j) d\epsilon_j \tag{47}$$

$$= -\int_{-\infty}^{\infty} (J-1) F^{J-2}(\epsilon_j) f(\epsilon_j) f(\epsilon_j) d\epsilon_j \tag{48}$$

$$\Rightarrow \tag{49}$$

$$1/markup = J \int_{-\infty}^{\infty} (J-1) F^{J-2}(\epsilon_j) f(\epsilon_j) f(\epsilon_j) d\epsilon_j \tag{50}$$

For the markup to go to zero the last expression must go to infinity. Note that $(J-1)F^{J-2}(\epsilon)f(\epsilon)$ is the density of ϵ_2^J so that the whole expression can be written as $E_{\epsilon_2^J}[Jf(\epsilon)]$. By Markov's inequality, we have:

$$1/markup = E_{\epsilon_2^J}[Jf(\epsilon)] \tag{51}$$

$$\geq Jk_J Pr_{\epsilon_2^J}[Jf(\epsilon) \geq Jk_J] \tag{52}$$

$$= Jk_J Pr_{\epsilon_2^J}[f(\epsilon) \geq k_J] \tag{53}$$

$$= Jk_J \int_{-\infty}^{\infty} \{f(\epsilon) \geq k_J\} (J-1) F^{J-2}(\epsilon) f(\epsilon) d\epsilon \tag{54}$$

for any sequence k_J. This last expression makes it obvious that for any distribution whose density is bounded below (e.g. uniform) the markup does indeed converge to zero. We now show that this is also true

for densities satisfying H.

Fix any $M < \infty$. Then by H there exists an $\underline{\epsilon}_M < \infty$ such that $\frac{f(\epsilon)}{1-F(\epsilon)} \geq M$ for all $\epsilon \geq \underline{\epsilon}_M$. Thus, for any number k_J, we have that if $\epsilon \geq \underline{\epsilon}_M$ and $M(1 - F(\epsilon)) \geq k_J$ then it must be that $f(\epsilon) \geq k_J$. Now consider the integral above:

$$Jk_J \int_{-\infty}^{\infty} \{f(\epsilon) \geq k_J\}(J-1)F^{J-2}(\epsilon)f(\epsilon)d\epsilon \geq Jk_J \int_{\underline{\epsilon}_M}^{\infty} \{M(1-F(\epsilon)) \geq k_J\}(J-1)F^{J-2}(\epsilon)f(\epsilon)d\epsilon$$

(55)

However, we can now solve for the upper end of the region of integration as well because $F()$ is a monotonic function:

$$M(1 - F(\epsilon)) \geq k_J \tag{56}$$

$$\Leftrightarrow F(\epsilon) \leq 1 - \frac{k_J}{M} \tag{57}$$

$$\Leftrightarrow \epsilon \leq F^{-1}(1 - \frac{k_J}{M}) \tag{58}$$

Plugging this back into the integral gives:

$$Jk_J \int_{-\infty}^{\infty} \{f(\epsilon) \geq k_J\}(J-1)F^{J-2}(\epsilon)f(\epsilon)d\epsilon \geq Jk_J \int_{\underline{\epsilon}_M}^{F^{-1}(1-\frac{k_J}{M})}(J-1)F^{J-2}(\epsilon)f(\epsilon)d\epsilon \tag{59}$$

$$= Jk_J \left[(1 - \frac{k_J}{M})^{J-1} - F^{J-1}(\underline{\epsilon}_M) \right] \tag{60}$$

$$= Jk_J \left(1 - \frac{k_J}{M} \right)^{J-1} - Jk_J\delta^{J-1} \tag{61}$$

where $\delta = F(\underline{\epsilon}_M) < 1$. We now let $k_J = J^{-1/\gamma}$ where $\gamma > 1$. The second part of the expression goes to zero as J gets large (since the exponential portion goes to zero faster than J). The rate of convergence of k_J has been chosen such that the first part diverges. This proves property three.

38

9 References

[1] Ackerberg, D. and M. Rysman (2001), "Unobserved Product Differentiation in Discrete Choice Models: Estimating Price Elasticities and Welfare Effects," forthcoming, *RAND Journa of Economics*.

[2] Anderson, S., de Palma, A. and J. Thisse (1992), *Discrete Choice Theory of Product Differentiation*, Cambridge: MIT Press.

[3] Bajari, P., and C. L. Benkard (2004), "Demand Estimation with Heterogenous Consumers and Unobserved Product Characteristics: A Hedonic Approach," NBER Working Paper w10278.

[4] Bajari, P., and Kahn, M. (2004), "Estimating Housing Demand with an Application to Explaining Racial Segregation in Cities" forthcoming, *Journal of Business and Economic Statistics*.

[5] Benkard, C.L. and Bajari, P. (2004), "Hedonic Price Indexes with Unobserved Product Characteristics", forthcoming, *Journal of Business and Economic Statistics*.

[6] Berry, S. (1994), "Estimating Discrete-Choice Models of Product Differentiation," RAND Journal of Economics, 25:2, 242-262.

[7] Berry, S., J. Levinsohn, and A. Pakes (1995), "Automobile Prices in Market Equilibrium," Econometrica, 63:4, 841-89.

[8] Berry, S., and A. Pakes (2001), "Estimating the Pure Hedonic Discrete Choice Model," Working Paper, Yale University.

[9] Bresnahan, T. (1987), "Competition and Collusion in the American Automobile Market: The 1955 Price War," *Journal of Industrial Economics*, vol. XXXV, 457-482.

[10] Caplin, A. and B. Nalebuff (1991), "Aggregation and Imperfect Competition: On the Existence of Equilibrium," Econometrica, 59:1, 1-23.

[11] Domencich, T. A. and D. McFadden (1975), *Urban Travel Demand*, Amsterdam: North Holland Publishing Company.

[12] Epple, D. (1987), "Hedonic Prices and Implicit Markets: Estimating Demand and Supply Functions for Differentiated Products," *The Journal of Political Economy*, 95:1, 59-80.

[13] Fan, J. and I. Gijbels (1996), *Local Polynomial Modeling and Its Applications*, London: Chapman and Hall.

[14] Gorman,T.(1980),"A Possible Procedure for Analyzing Quality Differentials in the Egg Market," *Re-*

view of Economic Studies,47:5,843-856.

[15] Hausman, J. (1997), "Valuation of New Goods Under Perfect and Imperfect Competition," in Bresnahan, T. and R. Gordon (eds.), The Economics of New Goods, Studies in Income and Wealth, Vol. 58, Chicago: NBER.

[16] Lancaster, K. (1966), "A New Approach to Consumer Theory," *The Journal of Political Economy*, 74:2, 132-157.

[17] Lancaster, K. (1971), *Consumer Demand*, NewYork: Columbia University Press.

[18] Manski, C. (1977), "The Structure of Random Utility Models," Theory and Decision, Vol. 8, 229-254.

[19] Petrin, A. (2002), "Quantifying the Benefits of New Products: The Case of the Minivan," *Journal of Political Economy*, 100, 705-729.

[20] Rosen,S.(1974),"Hedonic Prices and Implicit Markets: Product Differentiation in Pure Competition," *The Journal of Political Economy*, 82, 34-55.

[21] Shaked, A. and J. Sutton (1987), "Product differentiation and industrial structure", *Journal of Industrial Economics*, 36, 131-146.