Consistent Welfare Analysis when Some Heterogeneity is Unobserved

Amil Petrin
Graduate School of Business
University of Chicago
and NBER

May 24, 2003

Preliminary/Incomplete

Abstract

Most policy/welfare questions are non-linear functions (or functionals) of the observed and unobserved factors entering the model. In these cases all of the factors must be accounted for when the policy analysis takes place because mistakes do not "average out." In particular, if some heterogeneity is not observed and thus not conditioned on during the analysis, the policy estimate is biased and inconsistent. This paper describes this bias and provides methods to alleviate it. A major theme is that jointly determined variables contain information on common unobserved factors. A consequence is that errors from simultaneous equations systems can be used to construct estimates for these underlying factors. The empirical application uses data from the telecommunications industry: market-level observations on the adoption of cable television and on the average amount of cable television watched. The simultaneous determination of cable television adoption and the amount that it is watched is exploited to uncover otherwise unobserved taste heterogeneity. For the linear, log-linear, log-log, and logit demand models, the null of consistent welfare estimates without utilization data is rejected. The magnitude of the policy-estimate mistake is large. For some infra-marginal markets, welfare is misstated by several magnitudes without the correction.

*Thanks to Ernie Berndt, Steve Berry, Austan Goolsbee, Tom Holmes, Al Madansky, Kevin Murphy, Ariel Pakes, Wendy Petropoulos, Peter Reiss, Matthew White, the University of Chicago, the 2002 Society for Economic Dynamics, Yale, the NBER Productivity lunch, Wharton, and Harvard for helpful suggestions. Financial support from the Center for the Study of Industrial Organization and from the National Bureau of Economic Research is greatly appreciated.

Correspondence to: National Bureau of Economic Research, 1050 Massachusetts Ave., Cambridge, MA, 02139. email: amil.petrin@gsb.uchicago.edu.

1 Introduction

Demand and supply estimates are the fundamental input to applied welfare/policy analysis. These policy estimates are typically some function (or functional) of the dependent variable measured in levels. When the statistical model is not linear in the levels of the dependent variable, observed and unobserved factors enter the policy function(al) non-linearly. Perhaps the most common example in economics is the log transformation, which is a favorite for skewed dependent variables,[1] and is almost universally applied in policy analysis that uses regressions of (log) quantity on price and other controls, (log) wage on factors like education and ability, and (log) output on inputs for production functions.[2]

In the case of such non-linearity, the full model must be used to evaluate welfare/policy change with all relevant factors accounted for when the analysis takes place. Consistent counterfactual forecasts for policy analysis are constructed by first (re)transforming the estimated equation so that predictions from it are in the original units of the dependent variable (the relevant units for the analysis). Then, the policy calculation is performed for each agent using the "levels" function. Finally, integration over the distribution of agent types yields the aggregate change (different types will experience different welfare changes). Following this approach is of particular importance when the environmental change is not "infinitesimal."[3]

On its face, this calculation appears deceptively simple. The difficulties become apparent when one recognizes that the re-transformed model is not linear in the relevant unobserved factors. As Hausman and Newey (1995) note, this non-linearity means that "one can ignore the residual if it is all measurement error but not if it contains individual heterogeneity." Put another way,

[1] In these cases it can confer certain statistical benefits, like linearity and/or homoskedasticity, which makes unbiased and consistent estimation of parameter estimates and their standard errors straightforward.

[2] Sometimes the more general Box-Cox transformation is used, which nests the linear, log, and a large collection of other non-linear transformations.

[3] Most questions revolve around more substantial changes in the economic environment, like the surplus generated by the introduction of a new product or the increase in earnings arising from reducing class sizes by 30%.

all relevant factors must be accounted for (and all measurement error must be ignored) when the policy analysis takes place because the relevant unobserved factors do not "average out." Applications where this bias arises are common in economics, including estimation of consumer/producer surplus, the effect on earnings of schooling inputs or military service, and questions regarding GDP and productivity growth.

This paper explains the theory behind these biases. All unobserved heterogeneity must be accounted for to obtain consistent policy analysis estimates, and often the econometrician cannot distinguish in the estimated error the part attributable to unobserved heterogeneity.

This paper provides methods to alleviate this bias. The theme of the methods is that jointly determined variables contain information on the underlying common unobserved factors. Using insights from Madansky (1964), one implication is that the errors from structural simultaneous equations systems can be used to construct estimates of unobserved factors.[4] A number of different cases common to empirical work are analyzed.

The application exploits the fact that, for many goods, a consumer's valuation is reflected in two decisions: whether the good is adopted and how much it is used. These two decisions are made simultaneously and typically reflect some underlying factors that are similar and some that may differ.[5] Thus,

[4]There is a growing literature that uses control functions to obtain consistent estimates of model parameters associated with observed factors (see Heckman and Robb (1986), Pudney (1981), Pudney (1982), and Heckman and Scheinkman (1987), for example. The emphasis here is on estimating the unobserved factors.

[5]Research on demand estimation has recognized the simultaneity of a durable's adoption and utilization since the introduction of discrete/continuous models by Hausman (1979), Dubin and McFadden (1984), Hanemann (1984), and Mannering and Winston (1985). They describe related choices made by consumers that have both discrete (the adoption) and continuous (the use) aspects to them. One example is the choice of television-viewing medium (e.g. cable) and the amount of television watched conditional on this choice. Other examples include the purchase of a car and the mileage it is driven, and the adoption of household appliances (like air conditioners or heaters) and the use of electricity or gas to power them. See also Goldberg (1998), Hendel(1999), and Dube(2001). In every case the discrete and the continuous choice are determined simultaneously by some underlying factors that are similar and some that may differ.

this correction is available for many important products, including durables like cars, appliances, televisions, and houses (which make up more than 30% of expenditures in the U.S.)[6], and goods which have a subscription nature to them (e.g. cell phones).

Observing use is helpful if, after conditioning on the other observed variables like demographics and/or product characteristics, utilization rates still have explanatory power for demand. In particular, the correction is helpful when there are many infra-marginal consumers, as the empirical example illustrates.[7] While the information in the utilization decision can improve the efficiency of demand estimates, it can more importantly help cure the inconsistency of welfare estimates. Again, these corrections are a consistency issue for welfare estimates because unobserved taste heterogeneity does not enter the demand equation linearly, so taste errors do not average out in the welfare calculation.

The empirical application in this paper uses data from the telecommunications industry. Many policy questions arise in this industry regarding the effects of mergers (AOL/Time-Warner in 2000, ATT/Comcast in 2001, and Echostar-DirecTV in 2001), new product introductions (Satellite dish, DSL and cable modem internet access), regulation of cable franchise monopolies (price deregulation, promotion of entry), programming, and other related issues. The application here uses variability in the amount of television watched – a measure of otherwise unobserved taste heterogeneity – to improve demand estimates for the adoption of cable television. Two individuals that look the same "demographically" may differ quite substantially in their unobserved tastes for television, and their amount of television viewing in part reflects this otherwise unobserved difference.

This idea extends immediately to any pair (or more) of consumer choices

The purpose here is different from previous works. They are concerned with finding a functional form for indirect utility that has a closed-form solution for demand via Roy's identity, or with controlling for the endogeneity of the adoption decision when estimating the demand for use.

[6]See The Economic Report of the President, 2002.

[7]There are many examples from the demand literature of goods for which demographics explains only a small part of the variability in the adoption decision.

that reflect common underlying tastes. For example, whether cable television is adopted may be affected by the same underlying taste that determines whether a large television is purchased. If both decisions are observed, methods outlined in this paper are available for alleviating bias.

To show the relevance of the methods for an applied case, the application uses market-level data, the most commonly available type of data, to evaluate the change in welfare from a very large price increase in the expanded basic cable price. In particular, market-level averages of television viewing are shown to have, conditional on other market-level observables, significant explanatory power in the market-level adoption equation for cable television. For the linear, log-linear, log-log, and logit demand models, the null of consistent welfare estimates without utilization data is rejected. The median welfare change decreases between 11% and 76% with utilization in the demand equation. For some infra-marginal markets, the welfare change is understated by several magnitudes.

Sections 2 and 3 of the paper explore the econometrics behind the consistency and efficiency issues. Section 4 contains the application. Section 5 concludes.

2 The Bias Problem

The following simple model illustrates the generality of the bias problem. An observation is given by (Y_i, Z_i) and is assumed to be generated by

$$Y_i = f(Z_i, \omega_i, \varepsilon_i; \theta, \tau),$$

where Y_i is measured in levels, Z_i represents a vector of observed controls, ω_i is a (possible vector) of unobserved factors, ε_i is a measurement-type error, and θ and τ are (possibly vectors of) parameters.

For any (Z_i, ω_i) the answer to the policy question is based in large part on evaluating $f(\cdot)$ at these arguments. When this is done, ε_i, the error that does not represent heterogeneity, is evaluated at its population average of zero (this is the favored practice). Thus, for any (Z_i, ω_i),

$$f(Z_i, \omega_i, 0; \theta, \tau)$$

is an important part of the welfare/policy analysis.

Let $W(\cdot)$ denote the relevant operation on $f(\cdot)$ that yields the policy estimate.[8] Then, for any agent (Z_i, ω_i) the relevant computation is given by

$$W(f(Z_i, \omega_i, 0; \theta, \tau)).$$

For example, when a measurement of consumer surplus is desired, $f(\cdot)$ represents demand and the surplus operation is given by integrating this demand function over the price dimension. The result for the aggregate is based on the integral over the population, or

$$\int_{Z,\omega} W(f(Z, \omega, 0; \theta, \tau)) dP(Z, \omega) \tag{1}$$

where $dP(Z, \omega)$ gives the distribution in the population of observed and unobserved factors, and ε is held constant at its mean value of zero during the integration.

The bias in policy estimates arises because the distribution of ω is not typically known. Because it is hard to distinguish in the residual the relevant heterogeneity from the measurement-type error ε, the standard approach has been to set both parts of the residual to their population averages of zero when the policy analysis is undertaken. This policy estimate is given by

$$\int_{Z} W(f(Z, 0, 0; \theta, \tau)) dP(Z) \tag{2}$$

where ω is (like ε) evaluated at its mean of zero, instead of being integrated over as in (1).

The non-linear manner in which ω enters $f(\cdot)$ (and thus $W(\cdot)$) means (2) will not generally equal (1). The difference between (1) and (2) is the bias. This is the problem, and often there is nothing about the economics of the situation to suggest this bias is small. In fact, given the low explanatory power of some regressions, the econometrics of the situation suggests the bias could be very large.

The most popular transformation – the natural log – provides an illustrative example. When transformed the dependent variable is written as linear in

[8] If it is also not linear in $f(\cdot)$, this only exacerbates the bias problem.

regressors, parameters, and error, or:

$$log(Y_i) = Z_i\theta + \tau\omega_i + \varepsilon_i,$$

so linear techniques can be used in the estimation of the parameters θ. Since ω_i is unobserved, τ is not typically estimated. Instead, $\tau\omega_i$ is either treated as a random effect, or, if correlated with regressors, is addressed in some other econometrically appropriate manner.[9]

To show how poorly estimates that ignore heterogeneity can perform (and to make the point simply), assume $\theta = 0$ and ω is distributed normally with mean zero and variance σ^2. The approach that ignores the heterogeneity (equation (2)) applies the exponential function to the random variable ω evaluated at its mean value, or e^0, and then integrates across the population (since $\int_\omega dP(\omega) = 1$, $\int_\omega 1dP(\omega) = 1$). The correct approach (equation (1)), which appropriately accounts for heterogeneity, is obtained by integrating e^ω over the distribution of ω, or $\int_\omega e^\omega dP(\omega)$. The normality assumption means this integral equals $e^{\sigma_\omega^2/2}$. The only case for which (2) is correct is in the economically uninteresting case of no taste heterogeneity, when ω is a degenerate random variable (i.e. $\sigma_\omega^2 = 0$). Unsurprisingly, as σ_ω^2 increases, (with the mean held constant at zero), the bias becomes worse. This example suggests that empirical cases where this bias is worst are likely to arise from regression models where the estimated variance in the error is large (i.e. where little of the variance in the dependent variable is explained in the regression).

The following is a partial list – meant to be illustrative – of the wide range of cases when this bias problem arises:

(i) Consumer (or producer) surplus – When the policy question requires an estimate of consumer surplus, demand estimation that recognizes heterogeneity among agents is usually the important input. A simple example of the bias arising from ignoring heterogeneity is provided figures 1a and 1b. Two types of agents face an increase in price for a good they consume. The high demand type (A) and the low demand type (B) differ only in their intercept; their slope's are assumed to be identical in log(q)-p space (figure 1a).

[9] In the case where ω_i is correlated with Z_i, a model for ω_i may be available from the underlying economics (see, for example, Levinsohn and Petrin (2003)), allowing a proxy for the error to be constructed. Alternatively, a replicate observation may be observed.

The econometrician, ignoring the heterogeneity because it is unobserved or because it does not affect the consistency of the demand estimates from the log(q)-p equation - estimates the "representative" consumer's demand curve as line C. The aggregate surplus change induced by a price increase from p_0 to p_1 is computed using the levels equations in figure 1b (which are the transformed log-linear demand curves from figure 1a and given by *). Surplus equals the sum of areas under A^* and B^* and is equal to D+E+2F. The representative consumer model uses twice the area under C^* to estimate welfare, which equals 2E+2F. The error in the surplus estimate is D-E. This error worsens as the non-linearity becomes more severe, and it does not converge to zero as the sample size increases.

(ii) Inputs to schooling (or military service) and earnings – The impact of schooling inputs on future earnings is at the heart of education policy.[10] The policy analysis is often based on an educational production function which relates the log of wage to the inputs of schooling and other controls. From a policy planner's perspective, a necessary condition for policy approval is that increases in future (depreciated) earnings from additional schooling inputs exceed the costs of providing the additional inputs. The role of the estimated wage function is to forecast the counterfactual wage, and since wage functions are almost always estimated in log-levels, these forecasts require the full model. In particular, when the changes in schooling inputs are large – those often considered by policy makers – all observed and unobserved factors (e.g. ability) must be accounted for during the re-transformation to wage levels because mistakes in the forecasted wage levels do not average out in the aggregation. In the same way, these biases arise when measuring the effects of military service on earnings.

(iii) Productivity growth – Understanding the effects of policies on firm input choices and productivity growth is fundamental to comprehending their effects on GDP growth and its prospects. These questions require estimates of firm production functions, which are then used to forecast the effects of the policy changes. The potential for bias in these estimated effects is large for two

[10]These include investigations of the effects of class size, teacher quality, available resources, and the like.

reasons. First, production functions are almost always estimated in log-levels. Second, the recent focus on firm-level data has brought to light substantial heterogeneity in output at the firm-level even after conditioning on skilled and unskilled labor, capital, materials, and energy inputs.[11]

where $Z_{i\cdot}$'s denote the observed exogenous factors and the θ's their associated parameters, and the errors decompose as

$$\begin{aligned} \epsilon_{i1} &= \tau_1 \omega_i + \varepsilon_{i1} && (5) \\ \epsilon_{i2} &= \tau_2 \omega_i + \varepsilon_{i2}, && (6) \end{aligned}$$

where ω_i is the common unobserved factor (with coefficients τ_1 and τ_2) reflecting "shifters" that must be accounted for in the analysis, and $\varepsilon_{i\cdot}$'s reflect mean zero independent errors that do not enter the policy analysis. Consistent estimation of these two equations usually requires that at least one observed factor from Z_{i1} is excluded from Z_{i2} (and vice versa). The discussion in this paper assumes these equations are identified in every case considered and proceeds directly to methods for estimation of unobserved factors.[13]

Rewriting Y_{i1} in the terms of the reduced form yields

$$Y_{i1} = \frac{1}{1 - \theta_{1,2} * \theta_{2,1}} * [Z_{i1}\theta_1 + \theta_{1,2} Z_{i2}\theta_2 + (\tau_1 + \theta_{1,2}\tau_2)\omega_i + \varepsilon_{i1} + \theta_{1,2}\varepsilon_{i2}]. \quad (7)$$

For policy analysis, the relevant prediction requires the unobserved component $\tau_1\omega_i + \theta_{1,2}\tau_2\omega_i$.[14] This raises the main problem that the researcher faces (assuming consistent estimation of parameters on observed factors): she observes the sum of $\epsilon_{i1} + \theta_{1,2}\epsilon_{i2}$ (a scalar value), but must condition on $\tau_1\omega_i + \theta_{1,2}\tau_2\omega_i$.[15]

One way of entirely eliminating the bias from the policy estimate is available for many empirical questions. This case is when Y_{i2} is observed with no measurement error. An example is when the adoption and the use of a good are being considered simultaneously (as in the application), and the use of the good is directly metered. This holds for many goods, like local/long-distance/cell-phone use, automobile use (via the odometer), appliance use (via electricity monitoring), television viewing (via Nielsen, say), and the like.

[13] There is a large literature on identification in simultaneous equations. See, for example, Chapters 4 and 7 in the Handbook of Econometrics, Volume One.

[14] Of course, all other observables and parameters in the model are necessary for this computation too.

[15] In finite samples there will be an extra component to the error generated by the difference in estimated parameters and the true values.

When Y_{i2} has no measurement error, the variance of ε_{i2} is zero. This means that the ratio $\frac{\tau_1}{\tau_2}$ is consistently estimated from the regression of ϵ_{i1} on ϵ_{i2}, or

$$plim \frac{\sum_{i=1}^{N} \epsilon_{i1}\epsilon_{i2}}{\sum_{i=1}^{N} \epsilon_{i2}^2} = \frac{\tau_1}{\tau_2}.$$

With no measurement error ϵ_{i2} exactly equals $\tau_2\omega_i$. By multiplying ϵ_{i2} with the estimate of $\frac{\tau_1}{\tau_2}$ one obtains a consistent estimator for $\tau_1\omega_i$. Since $\tau_2\omega_i$ is observed directly (and $\theta_{1,2}$ is identified), a consistent estimate for $\tau_1\omega_i+\theta_{1,2}\tau_2\omega_i$ is given by "plugging-in."

A second scenario with two equations and one unobserved factor is that which arises when two observations on the same unobserved factor are available from different times (a "replicate" observation). The error structure is now given by

$$\begin{aligned} \epsilon_{is} &= \tau_1\omega_i + \varepsilon_{is}, \qquad (8) \\ \epsilon_{it} &= \tau_1\omega_i + \varepsilon_{it} \qquad (9) \end{aligned}$$

where i indexes agents and 1 and 2 have been replaced with s and $t, s \neq t$, to emphasize the times series nature of the data. Since the same agent is observed at a different time, the *same* unobserved effect is present (it is $\tau_1\omega_i$ in each case). If the variance in the errors is time-invariant, the simple average $\frac{1}{2}(\epsilon_{is} + \epsilon_{it})$ provides a minimum variance unbiased estimate of the unobserved factor. Note the importance of the replicate observation; without it, $\tau_1\omega_i$ is not separately identified from $\varepsilon_{i\cdot}$.

Unlike the first case, where an exact estimate of $\tau_1\omega_i$ is available, now only an unbiased estimate for $\tau_1\omega_i$ is available. Denote this estimate as $\hat{\omega}_i$, and let $dP(Z,\hat{\omega})$ denote the joint distribution of observed factors and unbiased estimates for the unobserved factor. The welfare calculation is now given by

$$\int_{Z,\hat{\omega}} W(f(Z,\hat{\omega},0;\theta,\tau))dP(Z,\hat{\omega}). \qquad (10)$$

Since the asymptotics usually arise from the number of agents i increasing and not from the number of equations per agent increasing, the unbiased estimate for $\tau_1\omega_i$ does not fully rid the welfare estimate from bias; $\hat{\omega}_i$ never

fully converges to $\tau_1\omega_i$. However, the calculation given by (10) is likely to provide an estimator with much less bias than that provided for by (2), where σ_ω^2 is assumed to be zero, *so* $\omega_i = 0$ *for all agents* i.

The last case in this subsection returns to the two cross-sectional equations for i. The question arises in the cross-section as to how one proceeds when the variance of ε_{i2} is not zero. Without the replicate observation that the time series provides, one can still proceed by regressing ϵ_{i1} on ϵ_{i2}, but in this case the least squares estimate of $\frac{\tau_1}{\tau_2}$ is not consistent. Specifically,

$$plim \frac{\sum_{i=1}^{N} \epsilon_{i1}\epsilon_{i2}}{\sum_{i=1}^{N} \epsilon_{i2}^2} == \frac{\tau_1\tau_2\sigma_\omega^2}{\tau_2^2\sigma_\omega^2 + \sigma_{\varepsilon_2}^2} = \frac{\tau_1}{\tau_2 + \sigma_{\varepsilon_2}^2/\tau_2\sigma_\omega^2}, \tag{11}$$

so the asymptotic bias in the estimate of the ratio $\frac{\tau_1}{\tau_2}$ is given by

$$\frac{-\tau_1}{(\tau_2^3\sigma_\omega^2/\sigma_{\varepsilon_2}^2) + \tau_2}.$$

The bias will be small when the variance in the unobserved factor is large relative to the variance in the measurement error. However, it is not possible to get rid of this bias without further information (like an additional equation).

This does not preclude one from using a simple average (for example) of ϵ_{i1} and $\frac{\hat{\tau_1}}{\tau_2}\epsilon_{i2}$ as an estimate for $\tau_1\omega_i$. This estimate has bias for $\tau_1\omega_i$ given by

$$1/2 * \frac{-\tau_1\omega_i}{(\tau_2^2\sigma_\omega^2/\sigma_{\varepsilon_2}^2) + 1}.$$

Even in cases where the variance of the measurement error is not small, the biased estimate for $\tau_1\omega_i$ may provide a significant improvement over assuming $\omega_i = 0$ for all i. Thus, unless it is believed that the second instrument is measured with such a great amount of error as to be completely useless in the analysis, the policy estimate should be constructed both ways to see if it differs substantially; if not, one might feel a little more confident that the effects of unobserved heterogeneity are not so pernicious.

3.2 Three Equations/One Unobserved Factor

When a third equation is added to the system of two equations/one unobserved factor, there will always be a minimum variance unbiased estimate available

for the unobserved factor. Three observations with uncorrelated errors are available for it. To see this, let the three equation system be given by

$$\epsilon_{i1} = \tau_1\omega_i + \varepsilon_{i1} \tag{12}$$

$$\epsilon_{i2} = \tau_2\omega_i + \varepsilon_{i2}, \tag{13}$$

$$\epsilon_{i3} = \tau_3\omega_i + \varepsilon_{i3}, \tag{14}$$

with ϵ_{i3} the new residual from the third structural equation. Here ϵ_{i3} (ϵ_{i2}) can be regressed on ϵ_{i2} (ϵ_{i3}). ϵ_{i1} can then be regressed on the predicted values from this regression to obtain a consistent estimate for $\frac{\tau_1}{\tau_3}$ ($\frac{\tau_1}{\tau_2}$). Finally the "unweighted" average

$$(\epsilon_{i1} + \frac{\tau_1}{\tau_2}\epsilon_{i2} + \frac{\tau_1}{\tau_3}\epsilon_{i3})/3 \tag{15}$$

provides an unbiased estimate for $\tau_1\omega_i$. A similar approach can be used to recover $\tau_2\omega_i$ and $\tau_3\omega_i$. Optimal weighting schemes based on the differences in variance of each component in the averages are available (as they are in the two equation cases).

With panel data, not only are unbiased estimates for $\tau_1\omega_i$ (for example) available, but the rate at which the variance in the error of the estimate falls is multiplicative in the number of time varying observations. For example, if there are two time series observations on each equation, there are six observations on $\tau_1\omega_i$. With four time series observations, there are twelve observations on $\tau_1\omega_i$. Thus, the availability of even a few replicate observations on any equation can substantially increase the precision of the unbiased estimate for $\tau_1\omega_i$.

3.3 The General Case

The logic of the IV approach extends to more general cases. See Madansky (1964).

4 Application: Cable Demand and TV Use

The application uses data on television viewing to improve the fit of the demand curve for cable television. The basis of demand estimation begins with

the (reduced form) regression of market-level shares on prices, average consumer demographics in the market, and product characteristics of the available goods (a market-level representative agent model, the kind most popular in the demand literature). The structural equation adds the market-level average of weekly hours watched of television to the reduced form equation. Estimates of demand and welfare are compared across the linear, log-linear, log-log, and logit demand models.

4.1 Demand and Utilization

For goods that have a utilization aspect to them, the amount demanded of the good is determined simultaneously with the decision of how much the good will be used. The simultaneity of the decisions does not imply either one "causes" the other. Instead, both decisions are caused by observed and unobserved factors.

The structural equation for demand is given by

$$\tilde{Q} = Z_Q \theta_Q + \theta_{Q,P} P + \theta_{Q,U} U + \epsilon_Q, \tag{16}$$

where $\tilde{Q} = g(Q)$ is the transformation that achieves linearity, Z_Q is the vector of relevant observed product and household characteristics (and interactions) after conditioning on usage U, θ_Q is the vector of associated parameters, $\theta_{Q,P}$[16] is the sensitivity of demand to the adoption price *conditional* on use U, $\theta_{Q,U}$ gives the change in quantity demand if utilization increases one unit holding adoption price and Z_Q constant, and ϵ_Q reflects both unobserved demographics, idiosyncratic household tastes, and errors of measurement.[17]

The structural equation for utilization is given by

$$U = Z_U \theta_U + \theta_{U,Q} \tilde{Q} + \theta_{U,P_U} P_U + \epsilon_U, \tag{17}$$

[16] Generally, subscripts of the form (A, B) denote a coefficient from the structural equation with the left-hand side variable given by A (that is, with A's coefficient normalized to one) and a right-hand side endogenous variable given by B.

[17] One can interpret these equations as the first order conditions for a representative agent model derived from a utility function given by $U^* = (Y_i - P_Q Q - P_U U) + V(Q, U, Z, \theta) + Q\epsilon_Q + U\epsilon_U$.

where θ_U is a vector of parameters that relates exogenous factors Z_U to use, $\theta_{U,Q}$ is the increase in use due to a one unit increase in demand, θ_{U,P_U} relates the amount of use undertaken to its unit price,[18] and ϵ_U represents unobserved factors affecting use.[19]

The two identification restrictions are imposed in the above equations. First, conditional on use, the price of use does not enter the adoption equation. Similarly, the second restriction is that, conditional on adoption, the adoption price does not enter the use equation. These two restrictions are sufficient to identify all of the parameters in the structural equation setting.

By solving the structural equations for $\tilde{Q}$, the reduced form demand equation can be fully characterized in terms of the structural parameters, exogenous variables, and structural errors:

$$\tilde{Q} = \frac{1}{1 - \theta_{Q,U} * \theta_{U,Q}} * [Z_Q\theta_Q + \theta_{Q,P}P + \theta_{Q,U}Z_U\theta_U + \theta_{Q,U}\theta_{U,P_U}P_U + \theta_{Q,U}\epsilon_U + \epsilon_Q]\,. \tag{18}$$

The coefficient on price – of particular importance because price is the variable over which the integration will take place in the surplus estimation – is the function of structural parameters given by

$$\gamma_P = \frac{1}{1 - \theta_{Q,U} * \theta_{U,Q}} * \theta_{Q,P}, \tag{19}$$

where price sensitivity conditional on usage $\theta_{Q,P}$ is grossed up by $\frac{1}{1-\theta_{Q,U}*\theta_{U,Q}}$ to account for the fact that adoption enters the use equation. The reduced form parameters are the statistics of interest for many questions because they

[18] In the electricity demand and appliance choice models, the price of utilization is the price per hour of electricity times the amount of electricity per hour the appliance uses. In the vehicle demand literature, the price of use is the miles per dollar the vehicle obtains. In the application later use is amount of television watched. The price of an hour of use is the opportunity cost of that hour for the individual.

[19] Economists have a long history of estimating equations similar to (17). The parameters of interest are usually (the equivalent of) $\theta_{U,P}$ or $\theta_{U,Q}$, and the econometric discussion revolves around the fact that the demand decision (for appliances say) is endogenous (see Dubin and McFadden (1984)). A common solution is to plug-in a predicted probability $\hat{Q}$ for Q, usually estimated from a reduced form equation where only exogenous variables are used to estimate $\hat{Q}$. Alternatively, one can instrument for Q with predicted probabilities $\hat{Q}$ that are constructed only from exogenous variables.

do not condition on usage, that is, because the reduced form demand equation is *the* equation from utility theory.[20]

If Z_Q are the only included regressors in the reduced form, the error from (18) is given by

$$\frac{1}{1-\theta_{Q,U} * \theta_{U,Q}} * \left[\theta_{Q,U} Z_{U-Q} \theta_{U-Q} + \theta_{Q,U} \epsilon_U + \epsilon_Q\right], \tag{20}$$

where Z_{U-Q} are demographics/characteristics in U but not in Q. This is the error that enters non-linearly into the welfare analysis. Assuming for now that Z_U is contained in Z_Q, so no relevant Z's go unobserved, the question turns on how much of each of the ϵ's belongs in the welfare calculation.

4.2 Welfare

The literature on demand and welfare estimation is largely dominated by four functional forms for $g(\cdot)$: linear, log-linear, log-log, and logit (or logit-like discrete choice models). Transformed back to levels, the quantity (here share) equation is given by

$$Q(P; Z, \omega, \theta, \tau).$$

The change in welfare for any agent that occurs when price increases from P_0 to P_1 is

$$\int_{P_0}^{P_1} Q(v; Z, \omega, \theta, \tau)\, dv \tag{21}$$

where, in the earlier notation, $Q(\cdot)$ is $f(\cdot)$ and $W(\cdot)$ is the integral operation over price. For simplicity, income effects are ignored.[21] Summing over the distribution of (Z, ω) pairs yields the measure of aggregate welfare:

$$\int_{Z,\omega} \int_{P_0}^{P_1} Q(v, Z, \omega, \theta, \tau)\, dv\ dP(Z, \omega). \tag{22}$$

Except for the case of the logit model, the approach in the demand literature is to assume all of the error $\theta_{Q,U} \epsilon_U + \epsilon_Q$ is measurement error. Thus, the

[20] The reduced form parameters are constructed from the structural estimates. The delta method can be used to compute standard errors.

[21] They can be added back to the framework at a cost to exposition.

measure used by practitioners for aggregate welfare is

$$\int_Z \int_{P_0}^{P_1} Q(v, Z, 0, \theta, \tau)\, dv\ dP(Z). \tag{23}$$

In the next section the use equation is employed to construct estimates for the unobserved heterogeneity, and welfare is estimated using

$$\int_{Z,\hat{\omega}} \int_{P_0}^{P_1} Q(v, Z, \hat{\omega}, \theta, \tau)\, dv\ dP(Z, \hat{\omega}). \tag{24}$$

One alternative to the log-log, log-linear, and linear approximations to demand is to take the other extreme viewpoint on the error, assuming it is all unobserved taste heterogeneity. This approach is usually advocated from within characteristics-based frameworks originating with Lancaster and McFadden. In these frameworks, decisions are modeled as directly determined from the latent variable "utility", a function of underlying tastes parameters and observed and unobserved product characteristics. Deviations between observed and predicted outcomes are explained entirely by differences in unobserved utility (or unobserved taste heterogeneity). Thus, in these frameworks, *all* of the error is compensated in the welfare calculation. From a practitioner's standpoint, one might view results from both approaches, when taken together, as providing a loose bound on the true change in welfare. Here the most popular form of these characteristics-based models, the logit demand model, is added to provide an alternative along these lines. Using this model is also useful because it explicitly recognizes that the dependent variable only varies between zero and one.

4.3 Data

The definition of goods in this market follows that in Goolsbee and Petrin (2001), with the overall set of goods approximated by the four main choices: antenna only, expanded basic cable, expanded basic cable plus some premium television (like the a la carte movie channel Home Box Office), and satellite dish. The focus of this paper will be on demand for expanded basic cable, which includes the local channels available via "antenna only" like ABC, NBC,

CBS, Fox, and PBS, plus additional channels like ESPN (a sports channel), TNT (primarily movies), and others, with an average cable franchise providing 62 channels on expanded basic. Premium channels are available a la carte if and only if expanded basic cable has been adopted. Finally, satellite dish is a multi-channel video option like cable that has a relatively small market share of 11% in 2001.

The basis of estimation will be data on market-level shares and demographics along with the cable characteristics of expanded basic that households face in their cable franchise market.[22] This market-level data typifies the kind most often available to practitioners. Average demographics and market shares for expanded basic come from Forrester Technographics, 2001, a household-level survey which is meant to be nationally representative.[23] To minimize sampling error problems in market share estimates from less populated areas, the analysis is restricted to 265 cable franchise markets for which at least 30 respondents exist. For these markets, 70% subscribe to either expanded basic or expanded basic plus premium cable, which compares closely to the 68% reported as the aggregate cable share in Federal Communications Commission 01-389, the 2001 annual report on the status of competition in multichannel video markets. Table 1 reports summary statistics for all of the aggregates used in the analysis.

The exercise here will focus on estimating demand for expanded basic for two reasons. First, with an average market share of 47%, it is the most popular of the four choices. Second, any one premium option rarely costs more than $10 per month, providing an upper bound on expanded basic welfare estimates that is not imposed during the estimation.

The utilization variable is derived from a Forrester survey question which asks *individuals* whether they (not the household) watch 0, 1-2, 3-5, 6-10, 11-15, or 16+ hours of television a week. Averaging these responses across households in every market (evaluating each bin except the largest at the midpoint) yields the variable television viewing hours (tvhours). This estimate is

[22]Each cable franchise is considered its own market, and almost all consumers in 2001 have one and only one cable company in their market.

[23]More details about it can be found in Goolsbee and Petrin (2001).

a lower bound on household television viewing because the top-coded bin is evaluated at 17 hours and the data is only reported for one individual in the household.[24] Except for the variables "all adults work" and "free time", which are excluded from the demand equation conditional on utilization, all of the other variables listed in Table 1 enter the cable adoption equation.

Table 1

Summary Statistics: Market Shares and Household Averages

Variable	Mean	Std. Dev.
Expanded Basic*	0.47	0.11
Expanded Basic + Premium*	0.23	0.10
Television Watched (Hrs. Weekly)	8.07	0.81
Free Time (Hrs. Weekly)	20.17	2.04
All adults work*	0.29	0.08
Married*	0.69	0.10
Household Size	2.61	1.23
Household Income	$56.0K	$10.5K
Household Assets	$220.8K	$92.5K
Observations	265	

Source: Forrester Technographics, 2001.

* designates a share variable.

Since television viewing is endogenous, an exclusion restriction is necessary to consistently identify its effect on the adoption decision. The excluded variables are constructed from survey answers to the questions "How many hours of <u>free time</u> (time which excludes work, chores, errands) do you have <u>per week</u>, including weekend hours?" and "Do all adults in the household work full time?" Thus, the identification assumption is that these variables, which affect the amount of time available to watch television, only affect cable television adoption decisions through their effect on television viewing. In the

[24] The exact underlining and question is "How may <u>hours per week</u> do <u>you</u> spend watching TV?".

data television viewing across markets increases with increases in free time, and decreases with increases in the fraction of households that have both adults working.

Table 2 summarizes the prices and characteristics of cable companies used to estimate demands. Consumer preferences can be affected by the channel capacity of a cable system, the price of expanded basic for the system, the price of expanded basic plus premium, the number of premium channels available, and the number of over the air channels available in the market area. The Factbook also provides the local franchise tax as a percent of revenue paid by the cable company to the local community.

Table 2
Summary Statistics: Television Markets

Variable	Mean	Std. Dev.
Monthly Expanded Basic Price	$27.17	$5.68
Monthly HBO Price	$11.44	$1.38
Channel Capacity	66.13	22.46
Premium Channels	5.56	1.45
Over-Air Channels	10.91	4.39
City Fixed Fee (percent)	4.29	1.14
Observations	265	

Source: Warren Publishing Television and Cable Factbook, 2001.

Since prices for antenna and satellite dish do not vary across markets, only the prices of expanded basic and expanded basic plus some premium enter the demand equation.[25] There is substantial price variation for the monthly cost of expanded basic around the mean price of $27.10, ranging from a minimum of $15 to a maximum of $45. In the data respondents only answer whether they have consumed some premium television. Expanded basic price plus the price of the most popular premium channel, Home Box Office (HBO), is used as the premium price proxy; consumers are paying at least this much to consume the

[25] This presents no problems for estimating cable demands, but does pose problems for satellite demand estimates (see Goolsbee and Petrin (2001)).

most popular premium television option.

One concern when estimating a demand curve using this variation directly is the potential for price endogeneity.[26] Three instruments for the two endogenous prices are used. The first is the local franchise tax paid by the cable company to the local community. This is a percent of gross revenue that varies by market and is reported in Warren Publishing (2001). It is positively correlated with prices. It is also not correlated with any of the observed characteristics, suggesting it may be uncorrelated with any unobserved characteristics that lead to price endogeneity problems (like customer service).

The last two instruments follow Hausman (1997) and Crawford (2001), averaging over the prices of the cable companies with the same multiple system operators (MSOs) but operating in different markets. These average prices reflect common cost side factors like programming costs shared by companies owned by the same MSO.[27] These instruments should exclude some of the idiosyncratic features of demand in each market that are correlated with prices in the market (like service).

4.4 Results

Following Hausman and Newey (1995), the focus begins with results from the log-log model, perhaps the most popular model used in demand estimation. Table 3 presents three different sets of demand estimates for the log-log model. The explanatory variables that enter the demand equation include channel capacity of the system, the number of premium channels available, the number of over-the-air channels, and market-level averages of household size, an indicator for marital status, and four indicator variables for five income groups and six indicator variables for seven wealth groups. To account for differences

[26] If there are some characteristics of the local cable franchise that are known by both the consumers and the suppliers and if cable prices respond to these factors, the price elasticity will typically be biased toward zero. For example, a cable system with the same observables but with relatively good service will tend to be more desirable and have higher prices, making it seem as though consumer demand does not respond to high prices. Goolsbee and Petrin (2001) show this problem is a pronounced one for cable markets.

[27] Larger MSOs with the ability to reach more advertising markets typically receive lower programming prices.

in sampling error, each market-level observation is weighted by the number of households observed in that market from the Forrester data.

The first column contains the OLS results without television hours entering the regression (i.e. the reduced form equation). Both price coefficients are very close to zero, which is consistent with the existence of unobserved characteristics like service that are positively correlated with price. The point estimates imply an elasticity of -0.39 for expanded basic cable, suggesting either an omitted variables problem or many cable monopolists pricing on the very inelastic part of the demand curve.

Column two presents the two stage least squares (2SLS) results for the no-utilization (reduced form) equation. The instruments for price are the market franchise tax and the average prices of same-MSO-but-different-market cable franchises (for both expanded basic and expanded basic plus premium). Both price coefficients increase by a factor of six. At the new point estimates, the aggregate elasticity for expanded basic cable is -1.91, comparable to previous expanded basic cable estimates found in the literature from Hazlett and Spitzer (1997), U.S. General Accounting Office (2000), Crawford (2000), and Goolsbee and Petrin (2001).Over-the-air channels and premium channels enter with the expected sign; increases in each are likely to increase antenna only and expanded basic plus premium shares respectively at the expense of expanded basic shares. There is one overidentification restriction that can be tested, and the validity of instruments/model is not rejected at the usual levels of significance (p-value=0.22). Finally, while the difference in elasticity estimates between the OLS and 2SLS results is 1.52 – large in economic terms – a Hausman test for a significant difference cannot reject the null of no difference (p-value=0.17).

Column three contains the 2SLS results conditional on utilization, with amount of free time and whether both adults work as instruments for the endogenous regressor tvhours. The unconditional price elasticity of demand implied by these conditional estimates is -2.17, similar to that obtained without utilization. The coefficient on market television viewing is significant at 5%, rejecting the demand specification without utilization. Conditional on

Table 3

Expanded Basic Demand Estimates

Dependent Variable: Expanded Basic Market Share

	OLS Coefficient (Std. Error)	2SLS Coefficient (Std. Error)	2SLS Coefficient (Std. Error)
expanded basic price	-0.396	-1.911	-1.222
	(0.299)	(1.146)	(1.198)
premium price	0.678	2.899	1.717
	(0.399)	(1.575)	(1.677)
tvhours			0.137
			(0.062)
channel capacity	-0.0002	-0.0001	0.0001
	(0.0006)	(0.0007)	(0.0008)
premium channels	-0.043	-0.038	-0.042
	(0.010)	(0.011)	(0.011)
over the air channels	-0.007	-0.005	-0.010
	(0.003)	(0.004)	(0.004)
household size	-0.273	-0.265	-0.067
	(0.088)	(0.094)	(0.130)
married	1.444	1.459	1.198
	(0.243)	(0.259)	(0.287)
constant	-2.257	-5.487	-4.729
	(0.635)	(2.257)	(2.303)
Root MSError	0.225	0.240	0.242
Observations	265	265	265

Note: All regressions include extensive controls for income and wealth, including four indicator variables for five income groups and six indicator variables for seven wealth groups. All of the income indicators and four of the six wealth indicators enter with t-statistics of at least one; six are significant at 5%. Income is negatively correlated with cable adoption conditional on wealth and wealth is positively correlated with cable adoption conditional on income.

expanded basic and expanded basic plus premium prices, a one hour increase in market television viewing (approximately one standard deviation) leads to an increase in the market share of expanded basic equal to 13%, In the levels, this translates into an increase in share of 6% on the base share of 47%.

Table 4

TVhours and Cable Elasticity Estimates, and Two Overidentification Tests

Method	TVhours Estimate (Std. Error)	Implied Cable Price Elasticity	P-Value for Over-ID Test 1	P-Value for Over-ID Test 2
Log-Log	0.13 (0.06)	-2.17	0.22	0.78
Log	0.13 (0.06)	-2.38	0.25	0.78
Logit	0.22 (0.11)	-2.26	0.33	0.79
Linear	0.53 (0.26)	-2.10	0.36	0.80

Table 4 compares point estimates and standard errors, implied elasticities, and p-values for two overidentification tests, one for each equation. For every functional form television hours is significant at the 5% level. All of the estimates imply an elasticity close to 2. In column three Test 1 asks whether the two overidentification restrictions for the cable demand equation can reject the specification. The lowest p-value is 0.22, so no rejection of the null of correct specification/valid instruments obtains at any meaningful level of significance. The second overidentification test in column four asks if the tvhours equation is well-specified. Again, no rejection obtains at any meaningful level of significance. The welfare implications of these estimates are considered next.

The average welfare change is equal to the average amount of household surplus in each market for expanded basic subscribers. Alternatively, for adopters of expanded basic in a market, it is the average difference in the reservation price minus the existing price of expanded basic, where a household's reservation price is defined as the price at which it substitutes to one of the other television products: expanded basic plus premium, satellite dish, or antenna

only.

Cable monopolies' bundling of expanded basic with premium implies a specification test that has power to uncover overstatements of welfare, like those that often arise when one extrapolates the surplus measure to regions outside the observed price variation. For the surplus question the price change of expanded basic increases to ∞ with all other product prices held constant. Since expanded basic is bundled with premium in the expanded basic plus premium choice, surplus for expanded basic cannot, in theory, exceed the price for the a la carte premium channel; no rational consumer would purchase expanded basic when they could get expanded basic *plus* premium for a lower price. This is not imposed during the estimation of welfare.

The initial analysis focuses on the region of observed price variation, beginning with the monthly welfare change for the price increase to the maximum expanded basic cable price observed in the data. This is a best case scenario for the no use case using the most popular functional form. Percentage changes in surplus associated with using utilization relative to not using it are considered first. The median percentage change in the log-log case is a decrease in welfare of 8.5% relative to no use data; ignoring use tends on average to overstate welfare. The inframarginal markets are starkly affected by the use data; 5th and 95th percentiles of this distribution of percentage changes are equivalent to -38.3% and 50.2% respectively. Thus, without the use data, low use markets have their welfare substantially overstated, while high use markets have their welfare substantially understated.

Monthly welfare estimates for the log-log specification with the tvhours regressor are reported in Figure 1. The distribution of welfare for each of four television viewing groups is shown: lowest use, low-mid use, mid-high use, and high use. Only two of 265 markets have an average surplus estimate that exceeds the price of HBO in the market, suggesting the log-log specification using only observed price variation is a reasonable one. The economic significance of including television viewing in the cable demand equation is apparent; as viewing increases, the distribution of surplus shifts to the right, from a mean of \$3.63 per month for the low use group to a mean of \$5.46 per month for the

high use group. These differences are statistically significant. Thus, the answer to any question related to the *distribution* of welfare or to infra-marginal markets improves with utilization data.

Figure 2 plots the distribution of absolute percentage changes in surplus associated with using utilization relative to not using it. The median absolute percentage change in the log-log case is 20.2% relative to no use data, with the 90th percentiles equal to 50.2% and the maximum equal to 120%. In only 25% of markets are these absolute percentage changes less than 10%, suggesting very different welfare implications for most markets once utilization is conditioned on.

Table 5 summarizes results for all of the functional forms under consideration. The table has estimates for median surplus, the percent of markets failing the welfare specification test, and the median percentage welfare increases relative to not incorporating use data. Welfare estimates for the same specifications but for price increases only up to \$44 a month, the largest price observed in the data for expanded basic, are reported in the top half of Table 5. These numbers provide a lower bound on welfare that does not suffer from the extrapolating outside the region of observed price variation. Using this variation, the surplus numbers across specifications are similar, ranging from the minimum linear case of \$0.95 to the maximum log-log case of \$4.28. Only two markets of the 265 fail the specification test, and only for the log-linear and log-log case. Even for this modest price increase, median percentage decreases range from 8% to 75%.

The bottom part of the table contains results for price increases to the reservation price. Median monthly surplus estimates vary across the estimators from \$2.89 for the linear case to \$9.05 for the log-log case. Welfare estimates using log-log demands exceed the price of HBO in almost 30% of markets, suggesting that functional form outside the region of observed price variation is driving these numbers. Median welfare estimates decrease between 11% and 76% over the no use case. When the distribution of absolute percentage changes is considered, the median change ranges from 23% to 76%. Overall, the use data plays an important role in the welfare estimates, and the results

Figure 1: Average Monthly Surplus: Expanded Basic Adopters

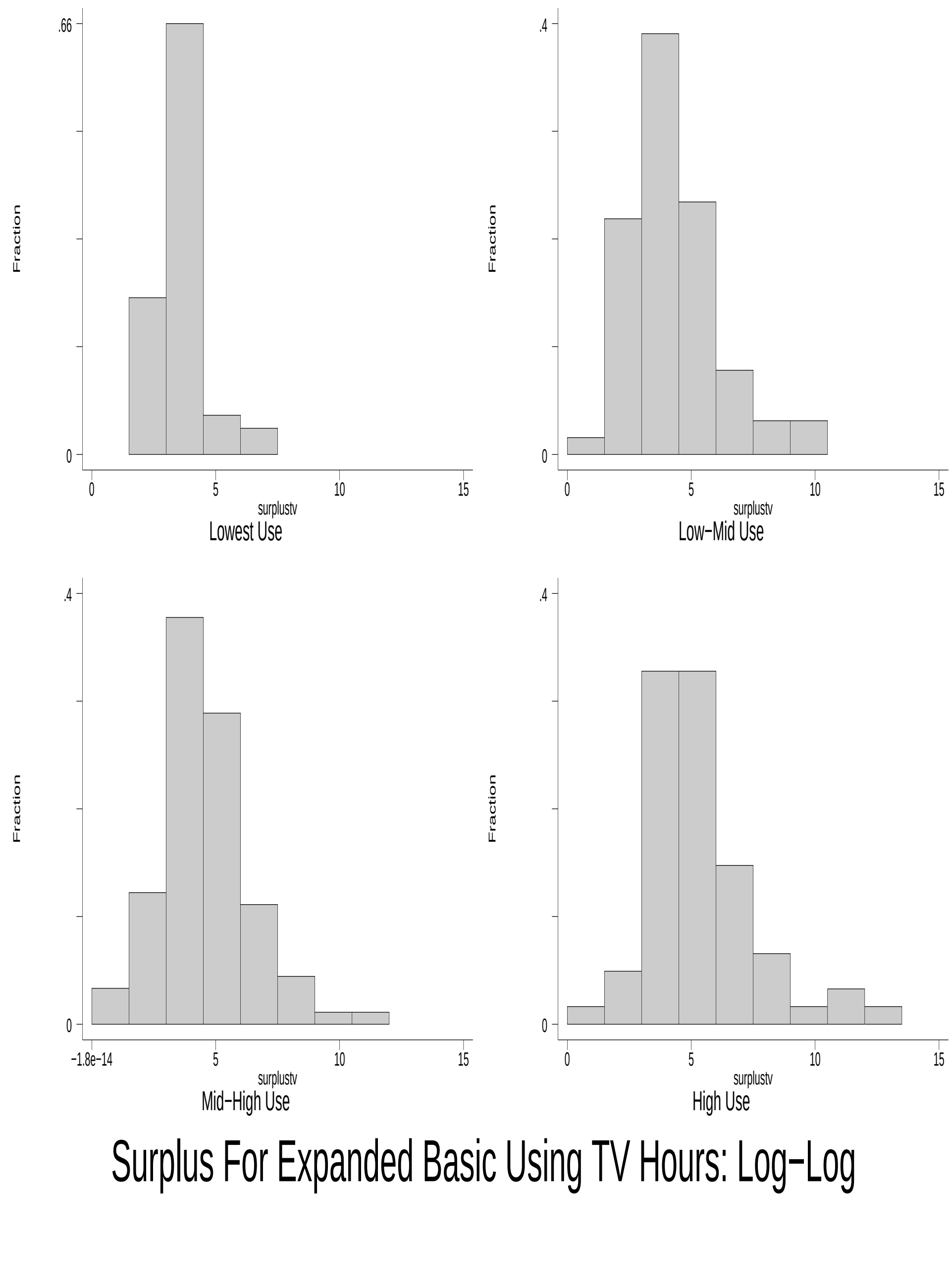
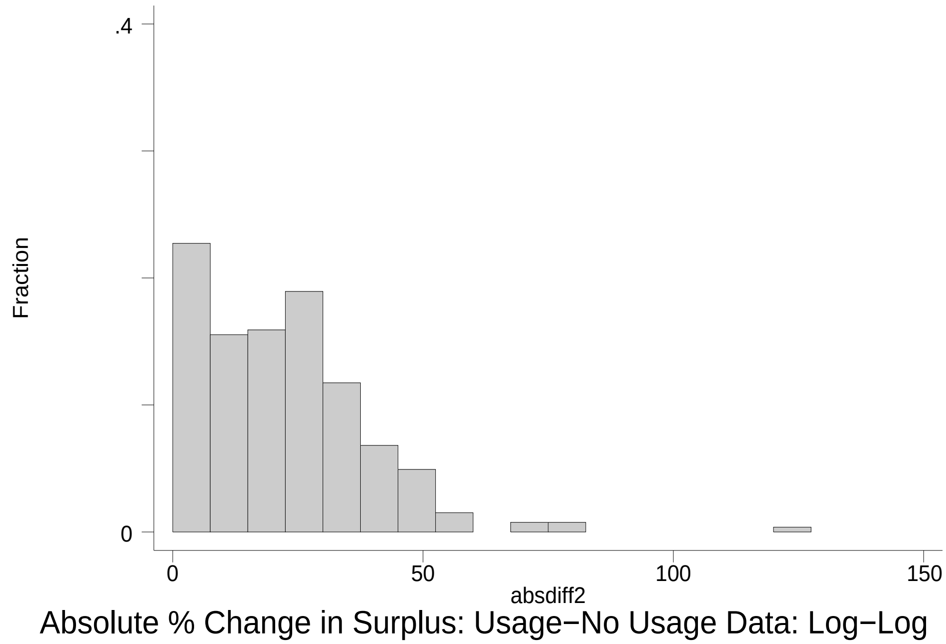


Figure 2: Surplus Changes With Use Data



suggest that households are realizing perhaps \$3-\$5 per month in surplus with their expanded basic choice, with the higher use households receiving more surplus.

5 Conclusions

Many policy/welfare questions are non-linear functions of the observables and errors entering the model. In these cases, all the factors must be accounted for when the policy analysis takes place, because errors do not "average out." Thus, from the applied person's perspective, the problem is that some important heterogeneity – perhaps a lot – is not observed, and thus cannot be conditioned on during the analysis, leading to biased and inconsistent policy estimates.

This paper describes this bias and provides methods to alleviate it. The theme of the methods is that jointly determined variables contain information

Table 5

Median Surplus and Percentage Changes in Surplus (Use-NoUse)/Use for Price Increases to...

...$44, the Maximum Expanded Basic Price:

Method	Median Surplus	% Surplus $> P_{HBO}$	Median Change %	Abs(Median) Change %
Log-Log	$4.28	0.7%	-08.5%	20.2%
Log-Linear	$3.88	0.7%	-11.6%	19.9%
Logit	$0.95	0.0%	-75.7%	75.7%
Linear	$2.11	0.0%	-20.2%	33.1%

...Estimated Reservation Prices:

Method	Median Surplus	% Surplus $> P_{HBO}$	Median % Change	Abs(Median) % Change
Log-Log	$9.05	27.9%	-29.9%	31.3%
Log-Linear	$5.11	1.5%	-21.2%	25.0%
Logit	$1.02	0.0%	-76.9%	76.9%
Linear	$2.89	0.0%	-11.9%	23.6%

on underlying and unobserved factors. The implication - which has many applications - is that the errors from structural simultaneous equations systems can be used to construct estimates of unobserved factors. A number of different cases common to empirical work are analyzed.

The application in the paper employs product utilization rates to account for otherwise unobserved taste heterogeneity when estimating demand and/or welfare. Bias is alleviated if, after conditioning on the observed, taste-related variables (usually demographics and product characteristics), utilization rates have explanatory power for what remains of durable demand variation. When they do, including utilization as an additional regressor improves the fit of the demand equation.

Utilization corrections are a consistency issue because unobserved taste heterogeneity does not enter the demand equation linearly, so these taste errors do not average out in the welfare calculation. As Hausman and Newey (1995) note, ignoring unobserved taste heterogeneity is "consistent with current practice in applied econometrics, and is difficult to improve without more information about the residual." The role of utilization data is precisely to provide this information; if observed, it can be used to condition out some of the unobserved taste heterogeneity that would otherwise be attributed to measurement error (and not conditioned on in the welfare calculation).

The empirical application in this paper uses data from the telecommunications industry. Variability in the amount of television watched – a measure of otherwise unobserved taste heterogeneity – is used to improve demand estimates for the adoption of cable television. The reason for doing so is straightforward; two individuals that look the same "demographically" may differ quite substantially in their unobserved tastes for television, and their amount of television viewing in part reflects this otherwise unobserved difference. To show the relevance of the methods for an applied case, the application uses market-level data, the most commonly available type of data, to evaluate the change in welfare from a very large price increase in the expanded basic cable price. In particular, market-level averages of television viewing are shown to have, conditional on other market-level observables, significant explanatory

power in the market-level adoption equation for cable television. For the linear, log-linear, log-log, and logit demand models, the null of consistent welfare estimates without utilization data is rejected. The median welfare change decreases between 11% and 76% with utilization in the demand equation. For some infra-marginal markets, the welfare change is understated by several magnitudes.

This idea extends immediately to any pair (or more) of consumer choices that reflect common underlying tastes. For example, whether cable television is adopted may be affected by the same underlying taste that determines whether a large television is purchased. If both decisions are observed, methods outlined in this paper are available for improving estimators.

References

Crawford, G. (2000): "The Impact of the 1992 Cable Act on Household Demand and Welfare," *RAND Journal of Economics*, 31(3), 422–449.

Crawford, G. (2001): "The Discriminatory Incentives to Bundle: The Case of Cable Television," Working Paper.

Dubin, J. A., and D. L. McFadden (1984): "An Econometric Analysis of Residential Electric Appliance Holdings and Consumption," *Econometrica*, 52(2), 345–62.

Goldberg, P. (1998): "The Effects of the Corporate Average Fuel Efficiency Standards in the U.S.," *Journal of Industrial Economics*, 46(1), 1–33.

Goolsbee, A., and A. Petrin (2001): "The Consumer Gains from Direct Broadcast Satellites and the Compeitition with Cable Television," *NBER Working Paper 8317*.

Hanemann, W. M. (1984): "Discrete/Continuous Models of Consumer Demand," *Econometrica*, 52(3), 541–62.

Hausman, J. (1979): "Individual Discount Rates and the Purchase and Utilization of Energy-Using Durables," *The Bell Journal of Economics*, 10(1), 33–54.

——— (1997): "Valuation of New Goods Under Perfect and Imperfect Competition," in *The Economics of New Goods*, ed. by R. Gordon, and T. Bresnahan. Chicago: University of Chicago Press, chap. 5, pp. 209–237.

Hausman, J. A., and W. K. Newey (1995): "Nonparametric Estimation of Exact Consumers Surplus and Deadweight Loss," *Econometrica*, 63(6), 1445–76.

Hazlett, T., and M. Spitzer (1997): *Public Policy Toward Cable Television*. The MIT Press and the AEI Press (Cambridge, Massachusetts and Washington, D.C.).

HECKMAN, J., AND R. ROBB (1986): *Alternative Methods for Solving the Problem of Selection Bias in Evaluating the Impact of Treatments on Outcomes, in Drawing Inferences from Self-Selected Samples*Springer-Verlag, chap. 1, pp. 63–113.

HECKMAN, J., AND J. SCHEINKMAN (1987): "The Importance of Bundling in a Gorman-Lancaster Model of Earnings," *The Review of Economic Studies*, 54(2), 243–255.

LEVINSOHN, J., AND A. PETRIN (2003): "Estimating Production Functions Using Inputs to Control for Unobservables," *forthcoming, Review of Economic Studies.*

MADANSKY, A. (1964): "Instrumental Variables in Factor Analysis," *Psychometrika*, 29(2), 105–113.

MANNERING, F., AND C. WINSTON (1985): "A Dynamic Empirical Analysis of Household Vehicle Ownership and Utilization," *RAND Journal of Economics*, 16(2), 215–236.

PUDNEY, S. (1981): "Instrumental Variable Estimation of a Characteristics Model of Demand," *The Review of Economic Studies*, 48(3), 417–433.

——— (1982): "Estimating Latent Variable Systems When Specification is Uncertain: Generalized Component Analysis and the Eliminant Method," *Journal of the American Statistical Association*, 77(380), 883–889.

U.S. GENERAL ACCOUNTING OFFICE (2000): "The Effect of Compeitition from Satellite Providers on Cable Rates," *Report to Congressional Requesters*, (GAO/RCED-00-164).

WARREN PUBLISHING (2001): *Warren Publishing Television and Cable Factbook*. Warren Publishing Inc (Washington, D.C.).