ESTIMATING NEW PRODUCT DEMAND FROM BIASED SURVEY DATA

Roger Klein and Robert Sherman
Bellcore

Abstract

Market researchers often conduct surveys asking respondents to estimate their future demand for new products. However, projected demand may exhibit systematic bias. For example, the more respondents like a product, the more they may exaggerate their demand. We found evidence of such exaggeration in a recent survey of demand for a potential new video product. In this paper, we develop a computationally tractable procedure that corrects for a general form of systematic bias in demand projections. This general form is characterized by a monotonic transformation of projected demand, and covers exaggeration bias as a special case.

1. Introduction

Many popular econometric models have the form

$$\Lambda(Y) = X'\beta_0 + u \tag{1}$$

where Y is a response variable, X is a vector of predictor variables, β_0 is a vector of unknown parameters, u is an error term, and Λ is a monotonic transformation. The Box-Cox model (Box and Cox, 1964) is a famous example, where Λ is known up to a single real-valued parameter and u is normally distributed with mean zero and unknown variance. All unknown parameters are estimated using maximum likelihood. Horowitz (1992) presents a kernel-based method for estimating both Λ and the distribution function of u without making any parametric assumptions about their functional forms. His estimators are $\sqrt{n}$-consistent, converge to Gaussian processes, and can be used to estimate the quantiles of Y given X.

In this paper, we consider the following model:

$$\Lambda(Y) = (X'\beta_0 + u)\{X'\beta_0 + u > c\} \tag{2}$$

where the constant c is known to equal either 0 or $-\infty$. This model allows a monotonic function of Y to equal a censored regression when $c = 0$, and reduces to (1) when $c = -\infty$. We will argue that (2) with $c = 0$ is a useful model for new product demand based on survey data. In this context, Y denotes reported demand and $\Lambda(Y)$ denotes actual future demand. The inverse function, Λ^{-1}, may be interpreted as a reporting function mapping actual demand into reported demand, Y. We develop a procedure for estimating β_0, the variance of u, and Λ without making any parametric assumptions about the functional form of Λ. However, we assume that the distribution of u is known up to scale. We obtain $\sqrt{n}$-consistent estimates of the model parameters and show that the estimate of Λ converges to a Gaussian process at a $\sqrt{n}$ rate. These estimates can be used to produce reliable estimates of new product demand.

While our procedure places heavier demands on u than the kernel-based method of Horowitz, it places lighter demands on Λ, and the latter fact is crucial for the application we consider. The kernel estimators are based on a representation of Λ that holds only if each distribution function of Y (given $X'\beta_0$) is differentiable (Horowitz, p. 5). Further, each distribution function must have at least 3 derivatives for the estimator of Λ to control asymptotic bias and so achieve $\sqrt{n}$-consistency (Horowitz, pp. 10–11). This is needed even if the distribution of u is known. In our application, the response variable only takes on nonnegative integer values. Consequently, each conditional distribution function of Y is a step function and so not even continuous. Our procedure covers applications like this without sacrificing $\sqrt{n}$-consistency. We also note that the kernel-based estimator only applies when $c = -\infty$ (Horowitz, p. 21). For our application, we need a procedure that covers the case $c = 0$. The procedure we develop covers this case and generalizes immediately to cover the case $c = -\infty$. Finally, we note that kernel-based methods are much more computationally intensive than the procedure proposed here.

In order to motivate the use of model (2), we consider a recent survey of demand for a potential new video product. Respondents are asked to estimate the average number of times per month they would use the product if it were offered, and there is a charge for each use. They report nonnegative integer values. Figure 1 gives a histogram of the responses. Because of the proprietary nature of the data, reported quantities have been masked by dividing them by their median value.

From this figure, one would suspect that individuals reporting high levels of demand are exaggerating.[1] For example, 7% of those surveyed reported a level of demand more than three times the median level. Moving out in the tail of this distribution, there are people who report a level of demand exceeding median demand by factors of 6, 10, 13, and even 20. Given the nature of this particular new product, such levels of demand are highly suspect. Though it is not possible to discern from the masked data, median and lower levels of demand agree with what one might expect for the type of product being surveyed. We also note that nearly 20% of the respondents reported zero demand for the product.

Write Y for projected demand and Q for actual future demand (quantity). From the description of the data just given, it seems reasonable to assume that there exists a function Λ defined on $[0, \infty)$ and satisfying the following conditions:

(1) $\Lambda(Y) = Q$ where Λ is strictly increasing.

(2) $\Lambda(0) = 0$ and $\Lambda(s) = s$ for some known, positive number s.

Assumption (1) requires that large projected demands correspond to large actual demands, but imposes no further structure on Λ. By freeing Λ from parametric restrictions like those imposed by the Box-Cox formulation, we allow for different reporting regimes:

$$\Lambda(Y) \equiv \begin{cases} Y < \Lambda_1(Y)\,, & Y \in A_1 : \text{ Underreporting} \\ Y = \Lambda_2(Y)\,, & Y \in A_2 : \text{ Accurate Reporting} \\ Y > \Lambda_3(Y)\,, & Y \in A_3 : \text{ Overreporting} \end{cases}$$

The regime regions A_i must partition the real line, but need not be known or contiguous. In addition, Λ need not be continuous, even at transition points from one regime to another. This type of freedom in modeling Λ is desirable in

[1] Exaggeration is common in surveys on demand for new products. This phenomenon may be due to new product enthusiasm, an attempt to influence the decision to market the product, a desire to please the interviewer, or the tendency for people to be less sensitive to total costs in a survey than they would be if they were making actual purchases. There may be other plausible explanations. For whatever reason or combination of reasons, the tendency to exaggerate is an acknowledged problem in surveys of this type.

the context of our data: we are skeptical about the magnitude of high reported quantities but are reluctant to impose the form of the relationship between Y and $\Lambda(Y)$, beyond requiring that the relationship be monotonic. Note, however, that we do not assume that misreporting exists: $\Lambda(Y) = Y$ satisfies assumptions (1) and (2) given above.

Without assumption (2), Λ can only be identified up to location and scale. For full identification, we require that Λ be known *a priori* at two distinct nonnegative points. Zero is a natural point to choose in the context of new product demand: $\Lambda(0) = 0$ says that people who project zero demand will, in fact, not use the product in the future. The assumption $\Lambda(s) = s$ for s known and positive is a key identifying condition. It says that there is a known, nonzero "safety point" at which demand is accurately reported.

It is instructive to contrast the method of Horowitz with ours regarding assumptions needed to identify $\Lambda(Y)$. Both methods require location and scale assumptions to identify $\Lambda(Y)$. Horowitz (pp. 10–11) assumes that $\Lambda(Y)$ is known at one point and that one of the slope parameters in model (1) is known up to sign. For purposes of making inferences on Y, which is natural in the context of actual market data, for both methods, location and scale assumptions serve merely as convenient normalizations. Indeed, any location and scale normalizations would suffice. However, in the application we consider in this paper, we are not directly interested in reported demand, Y, but rather in actual demand, $\Lambda(Y)$. Assumption (2) is substantive in this context. Accordingly, we have developed graphical and formal safety point tests in Sections 3 and 4.

We now further sketch out the import of assumptions (1) and (2). As mentioned above, nearly 20% of the respondents in the survey projected zero demand for the product. It is quite common in surveys of this type to see a significant fraction of respondents report zero demand. Using this fact in conjunction with assumptions (1) and (2), we can correctly assign each reported quantity to one of three regions as follows:

$$\begin{array}{lll} \text{Region I}: & Y = 0 & \Longleftrightarrow Q = 0 \\ \text{Region II}: & 0 < Y \leq s & \Longleftrightarrow 0 < Q \leq s \\ \text{Region III}: & Y > s & \Longleftrightarrow Q > s \end{array}$$

In Section 2, we show how to estimate the parameters of a demand model for Q using only information on the region containing Y. The classification scheme above shows that this region contains the corresponding Q value. We do not use any other information about reported quantities, and consequently, provide an estimation method that is insensitive to reporting bias of the form characterized by assumptions (1) and (2).

The new product demand model we assume is the standard Tobit (Tobin (1958)) model:

$$Q = (X'\beta_0 + u)\{X'\beta_0 + u > 0\}\,. \tag{3}$$

The disturbance term u is assumed to be independent of X and normally distributed with mean zero and unknown positive variance σ_0^2.

The distributional assumption on u is important, because various estimated quantities of interest will be inconsistent if it does not hold. In the concluding section, we mention work on a semiparametric version of the procedure developed in the next section that makes no parametric assumptions on the error term.

Other assumptions implicit in (3) can also be relaxed. For example, we can permit quantities to depend nonlinearly on X. We can also replace Q in (3) with a known, strictly increasing function, $F(Q)$. However, in order to test whether F is correctly specified, more information is needed to distinguish F from Λ. For example, we could assume that F is the log transformation and that there is a known interval of accurate reporting.

Proceeding with assumptions (1), (2), and (3), in the next section we provide a method for estimating the parameters β_0 and σ_0. Using these parameter estimates, we then show how to recover an estimate of Λ.

Our focus will be on recovering an estimate of Λ. There are two main reasons for this. First, knowledge of this function should prove useful in survey design, by defining the nature and extent of misreporting. This, in turn, may suggest ways of designing future surveys to avert or at least minimize this problem. Second, by applying an estimate of Λ to reported quantities, we can recover estimates of actual demand. This can greatly facilitate making revenue forecasts for the product being surveyed.

In Section 3, we report the results of several simulation experiments illustrating the performance of the estimation method. We also provide graphical tests of the safety point assumption. Section 4 provides a formal statistical test of this assumption. In Section 5, we apply the method to the video application discussed at the beginning of this section. Section 6 collects results on the asymptotic properties of the parameter estimates, the estimate of Λ, and the safety point tests. In particular, we show that the parameter estimates and the pointwise estimates of Λ are $\sqrt{n}$-consistent and asymptotically normally distributed, that the estimate of Λ converges uniformly on compact intervals at rate $n^{1/2-\delta}$ for any $\delta > 0$, and that $\sqrt{n}(\hat{\Lambda} - \Lambda)$ converges in distribution to a mean-zero Gaussian process on compact intervals. Finally, Section 7 provides a summary and directions for future research.

2. The *Orbit* Procedure

In this section we present the *Orbit* procedure, so-called because it borrows features from an ordered choice model (Amemiya, 1985, Chapter 9) and the Tobit model defined in (3).[2] It is a 2-stage procedure in which we first estimate

[2]After completing this paper, we learned that Goldberger first used the name *Orbit* perhaps as far back as 1964 to refer to a method, due to Orcutt, for estimating a Tobit model subject to sample selection. See Kiefer (1989) for a reference.

the parameters of the Tobit model, and then use these estimates to recover an estimate of the function Λ at points of interest.

Let $(Y_1, X_1), \ldots, (Y_n, X_n)$ denote a sample of n independent observations from the model defined by assumptions (1) through (3) from the last section. Write Z_i for (Y_i, X_i) and $z \equiv (y, x)$ for an element of S_Z, the support of Z_i. Write θ for (β, σ), θ_0 for (β_0, σ_0), and Θ for a compact subset of $\mathbb{R}^k \otimes \mathbb{R}^+$. For each $t > 0$, z in S_Z, $\lambda > 0$, and θ in Θ, define

$$
\begin{aligned}
f_t(z, \lambda, \theta) &= \{y = 0\} \log \Phi\left(-\frac{x'\beta}{\sigma}\right) \\
&+ \{0 < y \leq t\} \log\left[\Phi\left(\frac{\lambda - x'\beta}{\sigma}\right) - \Phi\left(-\frac{x'\beta}{\sigma}\right)\right] \\
&+ \{y > t\} \log\left[1 - \Phi\left(\frac{\lambda - x'\beta}{\sigma}\right)\right]
\end{aligned}
$$

where Φ denotes the cumulative distribution function of a standard normal random variable. Write P_n for the empirical measure that places mass $\frac{1}{n}$ at each Z_i, and note that $P_n f_s(\cdot, s, \theta)$ defines a log-likelihood function for the data. Define

$$\hat{\theta}(s) = \underset{\Theta}{\operatorname{argmax}}\, P_n f_s(\cdot, s, \theta)\,.$$

We call $\hat{\theta}(s)$ an *Orbit* maximum likelihood estimator of θ_0. Standard arguments sketched out in Section 6 show that $\hat{\theta}(s)$ is $\sqrt{n}$-consistent for θ_0 and asymptotically normally distributed.

For each $t > 0$, let Λ_t denote a compact subset of $\mathbb{R}^+$ containing $\Lambda(t)$. We estimate $\Lambda(t)$ with

$$\hat{\Lambda}(t; s) = \underset{\Lambda_t}{\operatorname{argmax}}\, P_n f_t(\cdot, \lambda, \hat{\theta}(s))\,. \tag{4}$$

Straightforward arguments show that $\Lambda(t)$ maximizes $\mathbb{E} f_t(\cdot, \lambda, \theta_0)$, the function to which $P_n f_t(\cdot, \lambda, \hat{\theta}(s))$ converges uniformly. It then readily follows that $\hat{\Lambda}(t; s)$ consistently estimates $\Lambda(t)$. We call $\hat{\Lambda}(t; s)$ an *Orbit* estimator of $\Lambda(t)$ and the 2-stage procedure which produces both $\hat{\theta}(s)$ and $\hat{\Lambda}(t; s)$, the *Orbit* procedure.

Note that $\hat{\Lambda}(Y_i; s)$ is a natural estimate of Q_i, the actual demand of the ith individual in the sample. Repeating this calculation for each of the reported quantities yields estimates of actual quantities.

Remark 1. The *Orbit* procedure separates parameter estimation from estimating the function Λ. For this reason, the computational burden involved in using (4) to recover estimates of actual demand is slight. Even if $\hat{\Lambda}(Y; s)$ is computed for each positive value of Y in the sample, each of these optimizations is over a single variable, and so can be performed very quickly. For example, using the *MAXLIK* routine in GAUSS on a 486DX2/66 PC, we did 500 such optimizations well within an hour. Moreover, for some applications, even if the sample size is large, the number of distinct positive Y values in the sample may be quite small. In this situation, the procedure is extremely fast. For example, the sample size for the application considered in Section 5 is around 1000, but because of rounding, the number of distinct positive Y values is around 20. We produced the corresponding estimates of actual demand in less than a minute.

REMARK 2. There are at least two other ways to estimate the function Λ.[3] The first applies only when there are no explanatory variables in the model. It is based on the observation that

$$\Lambda(t) = \alpha_0 + \sigma_0 \, \Phi^{-1}(\mathbb{P}\{Y \le t\})$$

where α_0 is the intercept in the model. An estimate of $\Lambda(t)$ can be obtained by substituting *Orbit* estimates for α_0 and σ_0, and substituting the corresponding sample proportion for $\mathbb{P}[Y \le t]$. Since the sample proportion can be viewed as a (nonparametric) maximum likelihood estimator of $\mathbb{P}[Y \le t]$, it readily follows that this estimator is equivalent to the *Orbit* estimator. However, since no second-stage optimization is required, this alternative method is easier and faster to implement than *Orbit*.

The second method applies when explanatory variables are in the model and involves simultaneously estimating Λ at all the points of interest, say, $t_1, \ldots, t_k$, rather than one at a time. The underlying model is:

$$Z = j \quad \text{if} \quad t_{j-1} < Y \le t_j \qquad j = 0, 1, \ldots, k+1$$

where $t_{-1} = -\infty$, $t_0 = 0$, and $t_{k+1} = \infty$ and

$$\mathbb{P}[Z = j] = \Phi\left(\frac{\Lambda(t_j) - X'\beta_0}{\sigma_0}\right) - \Phi\left(\frac{\Lambda(t_{j-1}) - X'\beta_0}{\sigma_0}\right).$$

After substituting *Orbit* estimates for β_0 and σ_0, estimates of the $\Lambda(t_j)$ are obtained through the usual maximum likelihood procedure for ordered qualitative response models.

Like the first alternative, this method is equivalent to the *Orbit* procedure when only an intercept is fit. When explanatory variables are in the model, this second method is different from *Orbit*. While we have yet to determine the relative efficiencies of this second method and *Orbit*,[4] there is evidence that the *Orbit* procedure provides a significant computational edge, especially when the number of t_j values is large. In one simulation, *Orbit* estimated 15 $\Lambda(t_j)$ values in about 30 seconds using the GAUSS *MAXLIK* routine running on a 486DX2/66 machine. Using the *Orbit* estimates as starting values, the second method required over 45 minutes to converge, even though the final estimates were close to the *Orbit* estimates.

[3]We are grateful to two referees for suggesting these alternative methods.

[4]The most efficient procedure would involve simultaneously estimating β_0, σ_0, and Λ at points of interest. Such a method, however, would be computationally burdensome.

When the $\Lambda(t)$'s are estimated sequentially rather than simultaneously, the natural starting value for estimating $\Lambda(t_{(k)})$ is the estimate of $\Lambda(t_{(k-1)})$, where $t_{(k)}$ is the kth largest of the t_j's. When $\Lambda(t_{(k)})$ and $\Lambda(t_{(k-1)})$ are close, the kth maximization in the second stage of *Orbit* is very fast.

Remark 3. $\hat{\Lambda}(s; s)$ is constrained to equal s. To see this, for $t > 0$, $z \in S_Z$, $\mu > 0$, and $\gamma \in I\!R^k$, define

$$\begin{aligned} g_t(z, \mu, \gamma) &= \{y = 0\} \log \Phi(-x'\gamma) \\ &+ \{0 < y \leq t\} \log [\Phi(\mu - x'\gamma) - \Phi(-x'\gamma)] \\ &+ \{y > t\} \log [1 - \Phi(\mu - x'\gamma)] \ . \end{aligned}$$

Note that $g_t(\cdot, \lambda/\sigma, \beta/\sigma) = f_t(\cdot, \lambda, \theta)$. Since $P_n f_s(\cdot, s, \theta)$ is maximized at $\hat{\theta}(s)$, $P_n g_s(\cdot, \mu, \hat{\beta}(s)/\hat{\sigma}(s))$ is maximized at $\mu = s/\hat{\sigma}(s)$. Consequently, $P_n f_s(\cdot, \lambda, \hat{\theta}(s))$ is maximized at $\lambda = s$.

Remark 4. We show in Section 6 that if $\Lambda(t) = t$ in a neighborhood of a safety point s, then it is possible to replace s with any consistent estimator without affecting the asymptotic distributions of the *Orbit* estimators. For example, one may be willing to assume *a priori* that $\Lambda(t) = t$ in a neighborhood of a specified population quantile q of the marginal distribution of Y. The result says that $\hat{\theta}(q)$ and $\hat{\Lambda}(t; q)$ have the same respective asymptotic distributions as $\hat{\theta}(\hat{q})$ and $\hat{\Lambda}(t; \hat{q})$ where $\hat{q}$ is the sample quantile corresponding to q. This result is also useful when applying a test of the safety point assumption developed in Section 4.

Remark 5. After applying the *Orbit* procedure, one may find accurate reporting over an entire range of Y values. A likelihood-based procedure that exploits this extra information will yield a more efficient estimator of θ_0, which, in turn, will lead to a more efficient estimator of $\Lambda(t)$ in the second stage of the *Orbit* procedure. For example, after applying *Orbit* one may find that an interval of the form $[0, s']$ is safe. In this case, one may then estimate a truncated Tobit model to obtain a more efficient estimate of θ_0. However, to minimize assumptions on Λ, we chose not to assume *a priori* the existence of a safety interval.

Remark 6. For positive t values less than the minimum positive Y value in the sample, there are no observations for which $0 < Y \leq t$. Similarly, for t values greater than or equal to the maximum Y value in the sample, there are no observations for which $Y > t$. For such t values, the objective function $P_n f_t(\cdot, \lambda, \hat{\theta}(s))$ degenerates in the sense that it is maximized at $\pm\infty$. As a result, it is not possible to produce corresponding *Orbit* estimates of $\Lambda(t)$.

3. Simulation Results

In this section, we discuss the results of several simulations exploring various aspects of the estimator $\hat{\Lambda}$ defined in (4). The designs were chosen to facilitate comparison with results obtained for the application presented in Section 4.

For each of the simulations, the sample size is 1000 and

$$Q = (2 + 2X + 2u)\{2 + 2X + 2u > 0\} \tag{5}$$

where X and u are independent, each having a standard normal distribution. Thus, $\beta_0 = (2, 2)$, $\sigma_0 = 2$, and the distribution of Q is a mixture of a point mass at zero and a $N(2, 8)$ distribution truncated at zero. There are about 25% zeros in a typical sample from this mixture distribution. Also, for each simulation, we let the interval $[0, 4]$ be an "accuracy region" – a region of accurate reporting. The point 4 corresponds to about the 65th to 70th percentile of the positive Q values. This setup corresponds roughly to what we observed in the application in Section 4.

In the first simulation, we investigate the performance of the estimator when there is linear exaggeration beyond the point 4. Write M for the function Λ^{-1}. We take

$$Y = M(Q) = Q\{Q \le 4\} + (2Q - 4)\{Q > 4\}\,. \tag{6}$$

Results of this simulation appear in Figures 2 and 3. The plots in Figure 2 illustrate the performance of the estimator when the safety point, s, falls within the accuracy region at the 50th percentile, or median, of the positive Y values. The corresponding plots in Figure 3 illustrate performance when s falls outside the accuracy region, at the 90th percentile of the positive Y values.

Turn to Figure 2. In the upper left-hand corner, three quantities are plotted against the estimated quantities $\hat{Q} = \hat{\Lambda}(Y; s)$ where Y comes from equation (6) and the safety point s is approximately equal to 3: the points plot Y vs. $\hat{Q}$, giving an estimate of the function M; the piecewise-linear dashed curve is a plot of $M(\hat{Q})$ vs. $\hat{Q}$; the straight dashed line is the reference line $\hat{Q}$ vs. $\hat{Q}$. The vertical line indicates the position of the safety point, s. Notice the close correspondence between $M(\hat{Q})$ and Y values, with only a slight increase in variation associated with reported quantities in the extreme tails of the Y distribution. In the upper right-hand corner we see the point plot of $\hat{Q}$ vs. Q superimposed over the dashed reference plot of Q vs. Q, where Q comes from equation (5). We see that estimated quantities very accurately track actual quantities. In the lower left-hand corner we show a plot of centered first differences of Y vs. $\hat{Q}$. This plot gives an estimate of the derivative of the function M in equation (6). Notice that the estimated derivative at the safety point is very close to unity, as it should be. Finally, for each $t > 0$, define

$$\hat{\sigma}(t) \equiv \frac{\hat{\sigma}(s)}{\hat{\Lambda}(t; s)} \cdot t \tag{7}$$

and note that $\hat{\sigma}(t)$ should be close to σ_0 in a neighborhood of s if there is an accuracy region about s. This is confirmed by the simulation: the plot of $\hat{\sigma}(Y)$

vs. Y in the lower right-hand corner of Figure 2 shows a region of relative constancy about s, and this constant value is very close to $\sigma_0 = 2$.

Figure 3 shows what can happen when the safety point lies outside the accuracy region. The 4 plots have the same format as in Figure 2, and are based on the same simulated data. The only difference is that the point s is chosen to be the 90th percentile of the positive Y values - approximately equal to 8. The plot in the upper right-hand corner shows that $\hat{Q}$ overestimates Q everywhere. We see the corresponding problem in the upper left-hand plot. Notice, however, that the two diagnostic plots on the bottom of the page alert us that something is wrong. In the lower left-hand plot the estimate of the derivative at s is far from unity, suggesting that we select another s from a region where the derivative estimates appear roughly constant. Similarly, the lack of relative constancy about $\hat{\sigma}(s)$ in the lower right-hand plot suggests a similar course of action.

We performed a second simulation following exactly the same pattern as the first but with quadratic exaggeration beyond the point 4. Specifically, we took

$$Y = M(Q) = Q\{Q \leq 4\} + (Q^2 - 12)\{Q > 4\}\,. \tag{8}$$

The diagnostic plots told the same story as their counterparts in Figures 2 and 3: $M(\hat{Q})$ and $\hat{Q}$ did a very good job of estimating Y and Q, respectively, when s was chosen correctly. The plots clearly signaled a problem when s was misspecified. In another simulation of the same model, we considered the case of no exaggeration, namely,

$$Y = Q\,.$$

Once again, $\hat{Q}$ accurately tracked Q, with minor deviations in the extreme tails of the Y distribution. Because of the similarity of the results with linear exaggeration, the plots associated with quadratic and no exaggeration are not reproduced in this paper.

4. Safety Point Tests

Assumptions (1), (2), and (3) are sufficient to estimate β_0, σ_0, and Λ. However, since $\hat{\Lambda}(s;s)$ is constrained to equal s (Remark 3 in Section 2), it is impossible to test the assumption $\Lambda(s) = s$ without more information. This is a critical assumption, since the asymptotic bias incurred from estimating $\Lambda(t)$ with (4) can be shown to equal

$$\Lambda(t)\,[1 - s/\Lambda(s)]\ .$$

This bias is zero when $\Lambda(s) = s$, but can be substantial if $\Lambda(s)$ is not close to s. In order to test $\Lambda(s) = s$ we add the following condition to assumption (2):

$\Lambda(t) = t$ on an interval containing s.

Since the procedure used to estimate $\hat{\Lambda}(t;s)$ imposes no constraints on points near s, departures from the above condition can be easily diagnosed, both graphically, as in the last section, and with a formal statistical test. We present a formal test in this section.

Recall the definition of $\hat{\sigma}(t)$ given in (7). For each $t > 0$ define $\sigma(t) = \sigma_0 \cdot \frac{t}{\Lambda(t)}$. Let a and b be positive real numbers with $a < b$. We say that the interval $[a, b]$ is a safety interval if the following conditions hold:

(A) $\Lambda(t) = t$ on $[a, b]$.

(B) $\Lambda'(t) = 1$ on (a, b).

(C) $\sigma(t) = \sigma_0$ on $[a, b]$.

Of course, condition (A) implies (B) and is equivalent to (C). Still, because of the constraint on $\hat{\Lambda}(s;s)$, it is useful to separate these conditions.

We start by assuming that the point s and a set of neighboring points all lie in a safety interval. We then develop a χ^2 test of an implication of this assumption. For ease of exposition, we construct a statistic that tests an implication of condition (C). Simple functions of this statistic can be used to test corresponding implications of conditions (A) and (B).

Let $S \equiv (s_1, \ldots, s_k)'$ be a vector of positive real numbers hypothesized to be in a safety interval containing s. We exclude s from S. We wish to test the following hypothesis:

$$H_0 : \sigma(s) = \sigma(s_1) = \cdots = \sigma(s_k)\,.$$

Note that H_0 is implied by condition (C) with $[a, b]$ containing s and all the s_i's. However, H_0 is not equivalent to (C) since H_0 can hold without (C) holding. However, if H_0 does not hold, then (C) is violated.

Write $\hat{\sigma}(S) = (1/\hat{\sigma}(s_1), \ldots, 1/\hat{\sigma}(s_k))'$ and $\mathbf{1}$ for a column vector of k ones. Consider the statistic

$$C_n = \sqrt{n}[\hat{\sigma}(s) \cdot \hat{\sigma}(S) - \mathbf{1}]\,.$$

If s and the components of S all lie in a safety interval, then there exists a nonsingular matrix Ω such that

$$C_n \Longrightarrow N(\mathbf{0}, \Omega)$$

where the symbol $\Longrightarrow$ denotes convergence in distribution and $\mathbf{0}$ denotes the zero vector in $\mathbb{R}^k$. A proof of this result, exhibiting the explicit form of Ω, is given in Section 6. A consistent estimator of Ω, denoted $\hat{\Omega}$, is also presented there. Deduce that

$$C_n' \hat{\Omega}^{-1} C_n \Longrightarrow \chi^2_k\,.$$

The statistic $C_n' \hat{\Omega}^{-1} C_n$ can be used to test H_0.

Turn to Figure 5, and consider the upper right-hand plot. The data is a replication of a simulation of quadratic exaggeration defined in (8) in Section 3. For this data, we plot the square root of the test statistic $C_n\hat{\Omega}^{-1}C_n$ against *Orbit* estimates of actual quantities, denoted $\hat{Q}$. That is, the ordinate of the ith point in this plot is the square root of $C_n\hat{\Omega}^{-1}C_n$ where S is the single point $\hat{\Lambda}(Y_i; s)$. The dotted vertical line corresponds to the safety point s, approximately equal to 3. The dotted horizontal line is the line $y = 1.96$ and corresponds to the 95th percentile of the distribution of the square root of a χ_1^2 random variable.

As expected, the values of the test statistic for points within the accuracy region $[0, 4]$ are generally not significant at the 5% level, but are highly significant at this level for points much greater than 4.[5] Choosing $S = (1, 2, 4)'$, we get a test value of 1.57 for $C_n\hat{\Omega}^{-1}C_n$. This value is the 33rd percentile of the distribution of a χ_3^2 random variable.

Some care must be exercised in interpreting the results of a test based on $C_n\hat{\Omega}^{-1}C_n$. As noted above, if one rejects H_0 then one must reject condition (C). However, if one accepts H_0 one need not accept (C). This follows from the fact that having $\sigma(t)$ constant on an interval is necessary but not sufficient for that interval to be safe. To see this, consider $M(t) = t\{t \leq 4\} + 2t\{t > 4\}$. The function $\sigma(t)$ will equal σ_0 on $[0, 4]$ and $2\sigma_0$ on $(4, \infty)$. At this point, the practitioner must judge, based on knowledge of the application, whether an apparent region of constancy actually corresponds to a safety interval.

Finally, note that one can view the hypothesized safety points $s_1, \ldots, s_k$ as population quantiles of the unconditional distribution of Y. At the end of Section 6 we show that the s_i's can be replaced by the corresponding sample quantiles (or any consistent estimators) without affecting the asymptotic distribution of the χ^2 test. We use this fact in the next section when we test the safety point assumption in the context of survey data.

5. An Application

In this section, we present the results of applying the *Orbit* procedure developed in Section 2 to the survey data on demand for a potential new video product described in the introduction. Because of the proprietary nature of the data, we cannot, at present, identify either the new product or the exogenous variables entering the demand model.

In this survey, respondents are asked to estimate the average number of times per month they would use the product if it were offered, and there is a charge for each use. They report nonnegative integer values.

Let Y^* denote average projected monthly demand and let Q denote average actual monthly demand. We assume that there exists a strictly increasing function Λ such that

$$\Lambda(Y^*) = Q\,.$$

[5] This statement is made informally. We make no claims about the asymptotic distribution of the maximum of the individual test statistics.

As before, we assume that

$$Q = (X'\beta_0 + u)\{X'\beta_0 + u > 0\}$$

where X is a vector of explanatory variables, β_0 is a vector of unknown parameters, and the random variable u is normally distributed with mean zero and unknown variance σ_0^2, and is independent of X.

Since Y^* is an average, we assume that its positive part is continuously distributed. Therefore, we do not observe Y^*, but rather a rounded version of it, denoted Y. As a matter of convenience, we shall assume that respondents round their Y^*'s up to the nearest integer.[6] It follows that for each positive integer k,

$$Y \leq k \Longleftrightarrow Y^* \leq k \Longleftrightarrow Q \leq \Lambda(k)\,.$$

Therefore, with $\Lambda(Y) = \Lambda(Y^*)$ on the support of Y, we can proceed to estimate quantities of interest under assumptions (1), (2), and (3).

Table 1 gives *Orbit* estimates and t-ratios. The safety point for the masked quantities (reported quantities divided by their median) is chosen to be unity, corresponding to the median of the positive reported quantities. The results are based on a sample of the form $(Y_1, X_1), \ldots, (Y_n, X_n)$ where $n = 922$.

We experimented with adding other pertinent variables to the model but the improvement in fit was negligible. In addition, we tried fitting appropriately transformed variables, certain interaction terms, and higher order polynomial effects but with largely the same result. None of the alternative models we estimated led to a noticeable change in the estimate of Λ.

Given the inherent coarseness of the data ($\hat{\sigma}(1) = 2.39$), we feel that this final main effects model is a reasonable one. The fact that the signs of the estimated parameters as well as the correlation matrix for the estimated parameters are believable also supports this claim.

So, with the *Orbit* estimator $\hat{\theta}(1)$ in hand, we now apply (4) from the last section to estimate Q. The results appear in Figure 4. In the upper left-hand corner is a point plot of Y vs. $\hat{Q}$ superimposed over the dashed reference line $\hat{Q}$ vs. $\hat{Q}$, where Y stands for reported demand and $\hat{Q}$ stands for estimates of actual demand obtained from applying (4) to the Y values. The position of the safety point, $s = 1$, is indicated with a vertical line. This plot suggests that reported estimates of average usage are reliable from zero to around the point 1.3 or possibly 1.7 (about the 75th percentile of the Y values), but then begin to take off. The diagnostic plots also seem to confirm that unity is a proper choice of safety point. The estimate of $M'(1)$ in the upper right-hand plot is very close to one, and there appears to be stability about unity in the plot of $\hat{\sigma}(Y)$ vs. Y.

[6] We find that the choice of rounding scheme has little effect on estimation. Therefore, in framing our assumptions, we are guided by convenience: we assume respondents round up. By adopting this convention, we avoid having to make any changes in assumptions, objective functions, or interpretation of results from the last section. Very small, but annoying, changes have to be made for other rounding schemes.

The lower right-hand plot is a histogram of $\hat{Q}$ values where the range of the horizontal axis is the range of the corresponding Y values. The difference in ranges is striking. It is also interesting to compare the other three plots in this figure with the corresponding plots from the simulations in Section 3. In a rough, qualitative sense, it would appear that the type of exaggeration present in this data is somewhere between linear and quadratic beyond the point 1.7.

Finally, refer to Figure 5. In the lower left-hand corner, we plot the square root of the test statistic $A_n \hat{\Sigma}^{-1} A_n$ against *Orbit* estimates of actual quantities. The dotted vertical line indicates the point s, equal to unity. The dotted horizontal line indicates the 95th percentile of the distribution of the square root of a χ^2_1 random variable.

Write (s^{--}, s^{-}) for the points adjacent to unity from below and (s^{+}, s^{++}) for the points adjacent from above. That is, $s^{--} < s^{-} < 1 < s^{+} < s^{++}$. At the 5% level, the points s^{--}, s^{+}, and s^{++} individually and jointly pass the corresponding χ^2 tests. However, s^{-}, the point just below unity, fails the χ^2_1 test at the 5% level. Consequently, there is some doubt about unity as a safety point.

To further probe the matter, we repeat the entire *Orbit* procedure, this time taking the point s^{+} as a new potential safety point. The corresponding χ^2_1 test results appear in the lower right-hand plot of Figure 5. The vector $S = (1, s^{++})$ passes the χ^2_2 test at the 5% level with a p-value of .82. We conclude that there is not enough evidence to reject the hypothesis that

$$\sigma(1) = \sigma(s^{+}) = \sigma(s^{++}) \,.$$

Notice that there are no other apparent intervals of constancy in the lower left-hand plot in Figure 4. This and properties of the application suggest that it is reasonable to view the interval $[1, s^{++}]$ as a safety interval. That unity is a reasonable safety point is further supported by a p-value of .46 for its associated χ^2_1 test value.

6. Asymptotic Properties

In this section, we establish some asymptotic properties of the *Orbit* estimates and the safety point tests. In particular, we show that the estimates of θ_0 and $\Lambda(t)$ are $\sqrt{n}$-consistent and asymptotically normally distributed, that the estimate of Λ converges uniformly on compact intervals at rate $n^{1/2-\delta}$ for any $\delta > 0$, and that $\sqrt{n}(\hat{\Lambda} - \Lambda)$ converges in distribution to a mean-zero Gaussian process on compact intervals. We first present results for the case where the safety point s is known. Subsequently, we will slightly strengthen our assumptions to include a small safe interval about an unknown, but consistently estimable point s. For example, s might be chosen *a priori* to be the median of the marginal distribution of the positive Y values, as in the application in Section 5. For this case, we show that we may replace s with any consistent estimator without affecting the asymptotic properties of the *Orbit* estimates and safety point tests.

We begin with the case where the safety point s is known. Review the notation introduced at the beginning of Section 2. The following assumptions are sufficient for the consistency and asymptotic normality results:

A1. $Z_1, \ldots, Z_n$ is a sample of independent observations from the model described by assumptions (1), (2), and (3) in the introduction.
A2. The support of X, the vector of explanatory variables in (3), is bounded.
A3. θ_0 is an interior point of Θ, a compact subset of $\mathbb{R}^k \otimes \mathbb{R}^+$.
A4. $\Lambda(t)$ is an interior point of Λ_t, a compact subset of $\mathbb{R}^+$.

Assumption A1 describes the data and the model. Assumption A2 is made solely for convenience, and guarantees that probabilities that are arguments of the log function stay bounded away from zero. A3 and A4 are standard assumptions, ensuring consistency and a limiting normal distribution for the estimators.

If A1 through A3 hold, then standard arguments show that $\hat{\theta}(s)$ converges in probability to θ_0. For example, a simple piece of calculus shows that $\mathbb{E} f_s(\cdot, s, \theta)$ is uniquely maximized at θ_0. Andrews (1987) shows that $P_n f_s(\cdot, s, \theta)$ converges uniformly in probability to $\mathbb{E} f_s(\cdot, s, \theta)$. Consistency then follows from Amemiya (1985, pp.106–107).

Let λ denote a positive real number. Define the following derivative operators: $\nabla_\lambda \equiv \frac{\partial}{\partial \lambda}$; $\nabla_\theta \equiv \frac{\partial}{\partial \theta}$; $\nabla_{\lambda\theta} \equiv \nabla_\theta[\nabla_\lambda]$; $\nabla_{\theta\theta} \equiv \nabla_\theta[\nabla_\theta]$; $\nabla_{\lambda\lambda} \equiv \nabla_\lambda[\nabla_\lambda]$.

Theorem 1: *If A1 through A3 hold, then*

$$\sqrt{n}(\hat{\theta}(s) - \theta_0) = \sqrt{n} P_n g_s(\cdot, \theta_0) + o_p(1)$$

where

$$g_s(z, \theta) = -[H_s(\theta)]^{-1} \nabla_\theta f_s(z, s, \theta)$$

and

$$H_s(\theta) = \mathbb{E} \nabla_{\theta\theta} f_s(\cdot, s, \theta)\,.$$

Theorem 1 follows from standard Taylor expansion arguments. See, for example, Amemiya (1985, pp.111–114). Write $\mathbf{0}$ for the zero vector in $\mathbb{R}^k$. The symbol $\Longrightarrow$ denotes convergence in distribution.

Corollary: *If A1 through A3 hold, then*

$$\sqrt{n}[\hat{\theta}(s) - \theta_0] \Longrightarrow N(\mathbf{0}, -[H_s(\theta_0)]^{-1})\,.$$

Turn to $\hat{\Lambda}(t;s)$. If A1 through A4 hold, then straightforward arguments, similar to those referred to in relation to $\hat{\theta}(s)$, show that for each $t > 0$, $\hat{\Lambda}(t;s)$ converges in probability to $\Lambda(t)$.

For each $t > 0$, λ in Λ_t, and θ in Θ, define

$$m_t(\lambda,\theta) = \mathbb{E}\nabla_{\lambda\theta} f_t(\cdot,\lambda,\theta)$$

and

$$H_t(\lambda,\theta) = \mathbb{E}\nabla_{\lambda\lambda} f_t(\cdot,\lambda,\theta)\,.$$

THEOREM 2: *If A1 through A4 hold, then for each $t > 0$,*

$$\sqrt{n}[\hat{\Lambda}(t;s) - \Lambda(t)] = \sqrt{n}P_n h_t^s(\cdot,\Lambda(t),\theta_0) + o_p(1)$$

where

$$h_t^s(z,\lambda,\theta) = -[\nabla_\lambda f_t(z,\lambda,\theta) + [g_s(z,\theta)]' m_t(\lambda,\theta)]/H_t(\lambda,\theta)\,.$$

Theorem 2 follows from standard Taylor expansion arguments. The term $g_s(z,\theta_0)]' m_t(\Lambda(t),\theta_0)$ corrects the variance of $\hat{\Lambda}(t;s)$ for the fact that θ_0 is known with error.

We can use Theorem 2 to derive the distribution of the test statistic $C_n'\hat{\Omega}^{-1}C_n$ defined in Section 4. Note that for each $t > 0$,

$$[\hat{\sigma}(s)/\hat{\sigma}(t) - 1] = [\Lambda(t;s) - t]\,/t\,.$$

Recall from Section 4 that S denotes the set of k points under test. Let D denote the diagonal matrix with the components of S along the diagonal.

COROLLARY 2.1: *Suppose A1 through A4 hold. If the components of S lie within a safety interval, then*

$$C_n \Longrightarrow N(\mathbf{0},\Omega)$$

where

$$\Omega = D^{-1}\Sigma D^{-1}$$

and the ijth element of Σ is given by

$$\mathbb{E}h_{s_i}^s(\cdot,\Lambda(s_i),\theta_0)\,h_{s_j}^s(\cdot,\Lambda(s_j),\theta_0)\,.$$

We assume that the matrix Σ in Corollary 2.1 is invertible. The ijth element of Σ can be consistently estimated by replacing $\Lambda(t)$ with $\hat{\Lambda}(t)$, θ_0 with $\hat{\theta}(s)$, and expectations with the corresponding sample averages. Let $\hat{\Sigma}$ denote the corresponding matrix estimator and write $\hat{\Omega}$ for $D^{-1}\hat{\Sigma}D^{-1}$. Since $\hat{\Sigma}$ consistently estimates Σ, we have that

$$C_n' \hat{\Omega}^{-1} C_n \Longrightarrow \chi_k^2 \,.$$

This fact justifies the test developed in Section 4.

COROLLARY 2.2: *If A1 through A4 hold, then for each $t > 0$,*

$$\sqrt{n}[\hat{\Lambda}(t;s) - \Lambda(t)] \Longrightarrow N(0, V_t(\Lambda(t), \theta_0; s))$$

where

$$V_t(\lambda, \theta; s) = \mathbb{E}[h_t^s(\cdot, \lambda, \theta)]^2 \,.$$

A simple calculation shows that $V_t(\Lambda(t), \theta_0; s)$ is finite for each $t > 0$. Similar calculations show that $V_t(\Lambda(t), \theta_0; s)$ converges to zero as t converges to zero and converges to infinity as t goes to infinity. Also, $V_s(\Lambda(s), \theta_0; s) = 0$, since $\hat{\Lambda}(s;s) = \Lambda(s) = s$. These facts are illustrated in the upper left-hand plot in Figure 5. The data for the plot come from a simulation replicating quadratic exaggeration defined in (8) in Section 3. The vertical line indicates the safety point, approximately equal to 3. The square root of the sample analogue of $V_Y(\Lambda(Y), \theta_0, ; s)$ is plotted against $\hat{Q}$.

The shape of the function $\Lambda(t)$ defines the nature and extent of misreporting. Our next results concern the uniform convergence of $\hat{\Lambda}(t;s)$ to $\Lambda(t)$. Such uniform results are useful for inferring the shape of $\Lambda(t)$ from $\hat{\Lambda}(t;s)$.

A simple calculation shows that for each $t > 0$,

$$\hat{\Lambda}(t;s) - \Lambda(t) = -\frac{P_n \nabla_\lambda f_t(\cdot, \Lambda(t), \hat{\theta}(s))}{P_n \nabla_{\lambda\lambda} f_t(\cdot, \lambda_n, \hat{\theta}(s))}$$

for λ_n between $\Lambda(t;s)$ and $\Lambda(t)$. Fix positive numbers $r < \rho$. A straightforward calculation shows that

$$\inf_{n, r \le t \le \rho} \left| P_n \nabla_{\lambda\lambda} f_t(\cdot, \lambda_n, \hat{\theta}(s)) \right| > 0$$

and

$$\begin{aligned} P_n \nabla_\lambda f_t(\cdot, \Lambda(t), \hat{\theta}(s)) &= H_t(\Lambda(t), \theta_0)\, P_n h_t^s(\cdot, \Lambda(t), \theta_0) \\ &+ O_p(1/\sqrt{n}) \end{aligned} \tag{9}$$

uniformly over t in the set $[r, \rho]$. Another straightforward calculation shows that

$$\sup_{z \in S_Z, r \le t \le \rho} |H_t(\Lambda(t), \theta_0)\, h_t^s(z, \Lambda(t), \theta_0)| < \infty \,.$$

Thus, the average in (9) is of mean-zero, independent, identically distributed, bounded random variables.

THEOREM 3: *Fix positive numbers $r < \rho$. If A1 through A4 hold, then for each $\delta > 0$,*

$$n^{1/2-\delta} \sup_{r \le t \le \rho} \left| \hat{\Lambda}(t; s) - \Lambda(t) \right| = o_p(1) .$$

The proof of Theorem 3 follows from the preceding remarks together with Lemma 1 in Appendix A of Klein and Spady (1993). Their Lemma 1 is based on arguments in Lemma 2 of Bhattacharya (1967).

For each $t \ge 0$, write $\Gamma_n(t; s)$ for $\sqrt{n}(\hat{\Lambda}(t; s) - \Lambda(t))$. Fix positive numbers $r < \rho$. Apply Lemma 15 and Theorem 21 in Pollard (1984, Chapter VII) together with Theorem 2 and Theorem 3 to see that the process $\{\Gamma_n(t; s) : r \le t \le \rho\}$ converges in distribution to a Gaussian process $\{\Gamma(t; s) : r \le t \le \rho\}$ satisfying $\Gamma(s; s) = 0$ and having covariance kernel

$$C(t, \tau) = \mathbb{E} h_t^s(\cdot, \Lambda(t), \theta_0)\, h_\tau^s(\cdot, \Lambda(\tau), \theta_0) .$$

Finally, suppose we are willing to assume *a priori* that $\Lambda(t) = t$ in a neighborhood of, say, a specified population quantile of the marginal distribution of Y. Let s denote this population quantile and $\hat{s}$ the corresponding sample quantile. Note that $\hat{s}$ consistently estimates s. We now show that $\hat{\theta}(\hat{s})$ has the same asymptotic distribution as $\hat{\theta}(s)$. This, the fact that $\hat{\Lambda}(t; \hat{s})$ depends on $\hat{s}$ only through $\hat{\theta}(\hat{s})$, and standard uniform convergence results will imply that

$$\hat{\Lambda}(t; \hat{s}) = \Lambda(t) - \frac{P_n \nabla_\lambda f_t(\cdot, \Lambda(t), \theta_0) + (\hat{\theta}(\hat{s}) - \theta_0)\, m_t(\Lambda(t), \theta_0)}{H_t(\Lambda(t), \theta_0)} + o_p(\frac{1}{\sqrt{n}}) .$$

The distributional result for $\hat{\theta}(\hat{s})$ will then imply that $\hat{\Lambda}(t; \hat{s})$ has the same asymptotic distribution as $\hat{\Lambda}(t; s)$.

We first show that $\hat{\theta}(\hat{s})$ converges in probability to θ_0. Since $\hat{s}$ consistently estimates s, there must exist a sequence $\{\delta_n\}$ of positive numbers converging to zero as n tends to infinity for which $\mathbb{P}\{|\hat{s} - s| > \delta_n\} \to 0$. It follows that

$$|\hat{\theta}(\hat{s}) - \theta_0| \le \sup_{|t-s| \le \delta_n} |\hat{\theta}(t) - \theta_0| + o_p(1) .$$

A simple calculation shows that

$$\hat{\theta}(\hat{s}) - \theta_0 = -[P_n \nabla_{\theta\theta} f_{\hat{s}}(\cdot, \hat{s}, \theta^*)]^{-1} P_n \nabla_\theta f_{\hat{s}}(\cdot, \hat{s}, \theta_0)$$

where θ^* is between $\hat{\theta}(\hat{s})$ and θ_0. Note that for any point t satisfying $\Lambda(t) = t$,

$$\mathbb{E} \nabla_\theta f_t(\cdot, t, \theta_0) = 0 . \tag{10}$$

Since $\Lambda(t) = t$ near s, eventually $\Lambda(t) = t$ for all t within δ_n of s. The Euclidean property of the class of functions $\{\nabla_\theta f_t(\cdot, t, \theta_0) : |t-s| \le \delta_n\}$ follows easily from an application of Lemmas 2.4, 2.12, and 2.14 in Pakes and Pollard (1989). It then follows from their Lemma 2.8 that

$$\sup_{|t-s|\le\delta_n} |P_n \nabla_\theta f_t(\cdot, t, \theta_0)| = o_p(1)\,.$$

A similar argument shows that

$$\sup_{|t-s|\le\delta_n,\,\theta\in\Theta} |P_n \nabla_{\theta\theta} f_t(\cdot, t, \theta) - \mathbb{E}\nabla_{\theta\theta} f_t(\cdot, t, \theta)| = o_p(1)\,.$$

It follows from the continuity of $\mathbb{E}\nabla_{\theta\theta} f_t(\cdot, t, \theta)$ as a function of t that

$$\sup_{|t-s|\le\delta_n,\,\theta\in\Theta} |\mathbb{E}\nabla_{\theta\theta} f_t(\cdot, t, \theta) - \mathbb{E}\nabla_{\theta\theta} f_s(\cdot, s, \theta)| = o(1)\,.$$

The strict concavity of $\mathbb{E} f_s(\cdot, s, \theta)$ and the compactness of Θ ensure the boundedness of $\sup_\Theta[\mathbb{E}\nabla_{\theta\theta} f_s(\cdot, s, \theta)]^{-1}$. Deduce that $\hat\theta(\hat s)$ consistently estimates θ_0.

Apply the consistency result to see that $\sqrt{n}[\hat\theta(\hat s) - \hat\theta(s)]$ equals

$$-[\mathbb{E}\nabla_{\theta\theta} f_s(\cdot, s, \theta_0)]^{-1}\sqrt{n}[P_n\nabla_\theta f_{\hat s}(\cdot, \hat s, \theta_0) - P_n\nabla_\theta f_s(\cdot, s, \theta_0)] + o_p(1)\,.$$

Arguing as before, we have

$$\sqrt{n}|\hat\theta(\hat s) - \hat\theta(s)| \le \sup_{|t-s|\le\delta_n} \sqrt{n}|\hat\theta(t) - \hat\theta(s)| + o_p(1)\,.$$

Apply (10) once more along with Lemma 2.17 in Pakes and Pollard (1989) to see that $\sqrt{n}[\hat\theta(\hat s) - \hat\theta(s)]$ has order $o_p(1)$.

7. Conclusions

In this paper, we consider the problem of estimating demand for a new product in the presence of biased demand projections. The form of the bias is characterized by a strictly increasing function Λ of projected demand. We develop a two-stage procedure, called *Orbit*, for estimating (1) the parameters of a standard Tobit model for actual future demand, and (2) the function, Λ. We make no parametric assumptions about the functional form of Λ. Nor do we require that Λ be continuous. The *Orbit* estimates are $\sqrt{n}$-consistent and asymptotically normally distributed, the estimate of Λ converges uniformly on compact sets at rate $n^{1/2-\delta}$ for any $\delta > 0$, and $\sqrt{n}(\hat\Lambda - \Lambda)$ converges in distribution to a zero-mean Gaussian process on compact intervals. Moreover, the procedure is computationally tractable.

To apply the *Orbit* procedure, there must exist a positive safety point at which reported quantities equal actual quantities. This point must either be known *a priori*, or if unknown, must be consistently estimable and contained in a safe open interval. We provide graphical and formal tests of this assumption. When a given safety point is incorrect, the graphical tests can suggest a proper choice of safety point.

In our simulations, we examine reporting mechanisms for which reported quantities equal actual quantities on a given interval. Beyond a threshold point, we let reported quantities overstate actual quantities in a variety of ways. In each simulation, with a sample size of 1000 observations, *Orbit* accurately recovers the the relationship between reported and actual quantities.

We also apply the *Orbit* procedure to survey data on a potential new video product. Under our model assumptions, the survey respondents report demand projections that exaggerate actual future demand beyond 1.7 times the median level, corresponding to about 25% of the sample. This level of exaggeration falls somewhere between linear and quadratic exaggeration, as discussed in Section 4.

The standard Tobit model for actual future demand assumes normality of the model's error term. It is possible to test this assumption, and when it fails, to apply a semiparametric version of *Orbit* that does not require making any parametric assumptions about the distribution of the error term. We are currently investigating the theoretical properties of this procedure and its finite sample performance.

REFERENCES

AMEMIYA, T. (1985): *Advanced Econometrics.* Harvard University Press, Cambridge, Mass.

ANDREWS, D. W. K. (1987): "Consistency in Nonlinear Econometric Models: A Generic Uniform Law of Large Numbers," *Econometrica*, 55, 1465–1471.

BHATTACHARYA, P. K. (1967): "Estimation of a Probability Density Function and its Derivatives," *The Indian Journal of Statistics: Series A*, 373–383.

BOX, G. E. P. AND COX, D. R. (1964): "An Analysis of Transformations," *Journal of the Royal Statistical Society*, Series B, 34, 187–220.

HOROWITZ, J. L. (1992): "Semiparametric Estimation of a Regression Model with an Unknown Transformation of the Dependent Variable," Working Paper #92–28, Department of Economics, University of Iowa, Iowa City, IA.

KIEFER, N. M.(1989): "The ET Interview: Arthur S. Goldberger," *Econometric Theory*, 5, 133-160.

KLEIN, R. W. AND SPADY, R. H. (1993): "An Efficient Semiparametric Estimator for Binary Response Models," *Econometrica*, 61, 387–421.

PAKES, A. and POLLARD, D. (1989). Simulation and the Asymptotics of Optimization Estimators. *Econometrica*, 57, 1027–1057.

POLLARD D. (1984). *Convergence of Stochastic Processes.* Springer, New York.

TOBIN, J. (1958): "Estimation of Relationships for Limited Dependent Variables," *Econometrica*, 26, 24–36.