Estimation of Linear Regression Models from Bid-Ask Data by a Spread-Tolerant estimator*

Oliver Linton†
London School of Economics

January 1, 2001

Abstract

We investigate a class of estimators for linear regression models where the dependent variable is subject to bid-ask censoring. Our estimation method is based on a definition of error that is zero when the predictor lies between the actual bid price and ask price, and linear outside this range. Our estimator minimizes a sum of such squared errors; it is nonlinear, and indeed the criterion function itself is non-smooth. We establish its asymptotic properties using the approach of Pakes and Pollard (1989). We compare the estimator with mid-point OLS.

1 Introduction

Suppose that

$$y_i = \beta' x_i + u_i,$$

where $E(u_i|x_i) = 0$ with probability one. We observe x_i but never observe y_i; instead, we observe an upper and lower bound y_i^L, y_i^U with $y_i^L \leq y_i \leq y_i^U$, i.e., we observe a sample $\{x_i, y_i^L, y_i^U\}_{i=1}^n$ and wish to estimate β from this data. This sort of sampling scheme arises sometimes with financial data where only bid and ask price quotes are available, see for example Linton, Mammen, Nielsen, and Tanggaard (2000) and Campbell, Lo and MacKinlay (1997). It is easy to see that β cannot be identified without additional structure. We introduce an additional structure that ensures identifiability and yet is

*This research was supported by the National Science Foundation.

†Department of Economics, London School of Economics, Houghton Street, London WC2A 2AE, United Kingdom. Tel. 0207 955-7864; Fax. 0207 831–1840; E-mail address: lintono@lse.ac.uk

somewhat plausible. Specifically, we suppose that

$$y_i^L = y_i - \eta_{1i}$$

$$y_i^U = y_i + \eta_{2i},$$

where η_{1i}, η_{2i} are mutually independent realizations from the same distribution on $[0, \infty)$. The realizations of η_{1i}, η_{2i} can be quite different so that the spread $y_i^U - y_i^L = \eta_{2i} - \eta_{1i}$ can take a big range of values. We discuss estimation of β in this model.

One plausible estimation strategy here is to define

$$y_i^* = \frac{y_i^L + y_i^U}{2}$$

and to regress y_i^* on x_i. Because η_{1i}, η_{2i} come from the same distribution this provides consistent estimates of β, since

$$y_i^* = \beta' x_i + \varepsilon_i,$$

where $\varepsilon_i = u_i + (\eta_{2i} - \eta_{1i})/2$ is mean zero given x_i. This is true even if the distribution of η_{1i}, η_{2i} depends on x_i since they cancel each other out. Therefore, the OLS estimator is consistent and asymptotically normal.

A number of authors have proposed to calculate residuals in our model to be zero when the predicted value lies inside the observed spread, and to be the deviation from the closest of y_i^L, y_i^U otherwise, see for example Bliss (1997). That is, define the residual to be

$$\widehat{\epsilon}_i = \begin{cases} y_i^L - \widehat{y}_i & \text{if } y_i^L \geq \widehat{y}_i \\ y_i^U - \widehat{y}_i & \text{if } y_i^U \leq \widehat{y}_i \\ 0 & \text{else.} \end{cases} \tag{1}$$

This way of calculating residuals differs from the 'mid-point' based approach referred to above in which $\widehat{\epsilon}_i = y_i^* - \widehat{y}_i$. The definition (1) seems well justified because the actual value of y_i can lie anywhere in the interval $[y_i^L, y_i^U]$ and so predicted values that lie inside this range should be taken as plausible values. With this definition of error, we can take as measure of fit the sum of squared errors $\sum_i \widehat{\epsilon}_i^2$. Bliss (1997) uses this criterion to measure the performance of various methods of fitting the term structure from bid and ask quotes of coupon bond prices. We use this notion of error to generate an estimator of β. We establish the consistency and asymptotic normality of our estimator and make a comparison between it and the OLS estimator. We draw heavily on results of Pakes and Pollard (1989).

We use the notation $\|A\| = \sqrt{\text{tr}(A'A)}$ for any real matrix A, and let $\lambda_{\min}(A)), \lambda_{\max}(A)$ denote the smallest and largest eigenvalues of a real symmetric matrix A. We also let $1(B)$ be the indicator function of the event B.

2 The Estimator

Following on from (1), define the criterion function

$$Q_n(\beta) = \frac{1}{n}\sum_{i=1}^{n}\widehat{\epsilon}_i^2(\beta) = \frac{1}{n}\sum_{i=1}^{n}(y_i^L - \widehat{y}_i(\beta))^2 1(y_i^L \geq \widehat{y}_i(\beta)) + \frac{1}{n}\sum_{i=1}^{n}(y_i^U - \widehat{y}_i(\beta))^2 1(y_i^U \leq \widehat{y}_i(\beta)),$$

where $\widehat{y}_i(\beta) = \beta' x_i$. Define also the almost sure derivative of $Q_n(\beta)$,

$$G_n(\beta) = \frac{-1}{n}\sum_{i=1}^{n} x_i(y_i^L - \widehat{y}_i(\beta))1(y_i^L \geq \widehat{y}_i(\beta)) + \frac{-1}{n}\sum_{i=1}^{n} x_i(y_i^U - \widehat{y}_i(\beta))1(y_i^U \leq \widehat{y}_i(\beta)).$$

We define our estimator $\widehat{\beta}$ to be any sequence that satisfies

$$G_n(\widehat{\beta}) = \inf_{\beta\in B}\|G_n(\beta)\| + o_p(n^{-1/2}) \tag{2}$$

where B is some given compact set. We shall assume throughout that such a sequence exists even though G_n is not continuous everywhere. This is generally reasonable - just like the standard LAD estimator one finds multiple solutions to (2) and some simple rule like take the mean of the set of solutions ensures uniqueness. The discontinuities disappear rapidly as sample size increases. In high dimensions, it is necessary to use some iterative method like Nelder-Mead to find the solution to (2); in this case, good starting values maybe provided by the OLS estimator of y_i^* on x_i. In the next section we discuss the asymptotic properties of $\widehat{\beta}$.

3 Asymptotic Properties

We make the following assumptions.

A1. $(x_i, u_i, \eta_{1i}, \eta_{2i})$ are i.i.d., mutually independent, and have a distribution that is absolutely continuous with respect to Lebesgue measure. Denote by f_X, f_u, f_η the corresponding marginal densities, and F_X, F_u, F_η the c.d.f.'s.

A2. We suppose that u_i is symmetric about zero with support contained in $\mathbb{R}$, while η_{ji} has support contained in $[0,\infty)$. The supports of u_i and η_{ji} have an intersection that has positive Lebesgue measure.

A3. The density function f_u is continuously differentiable and $f_u'(u) \to 0$ as u approaches the boundary of its support.

A4. $\sigma_u^2 = E(u_i^2) < \infty$, $E(\eta_{1i}^2) < \infty$, and $0 < \lambda_{\min}(E(x_i x_i')) \leq \lambda_{\max}(E(x_i x_i')) < \infty$. Let $\mu_\eta = E(\eta_{ji})$ and $\sigma_\eta^2 = \mathrm{var}(\eta_{ji})$.

A5. The true parameter β_0 lies in the interior of the compact parameter set B.

REMARKS

1. In assumption A2 we are ruling out the possibility that $y_i^L \leq \beta_0' x_i \leq y_i^U$ with probability one. This might occur if for example the support of η_{1i} was $[1,2]$ while the support of u_i was $[-1,1]$ because then $u_i - \eta_{1i} < 0$ with probability one and so $y_i^L < \beta_0' x_i$ always. There will also therefore exist some other β close to β_0 for which this is true, and which is consequentially indistinguishable from the true one. In practice, this is not likely to be an onerous restriction since it seems plausible that the spread not be always much greater than the pricing error.

2. The assumptions can be weakened in various directions. Specifically, we can allow the error distributions to depend on x_i, provided $u_i|x_i$ is symmetric about zero, but at the cost of a more complicated limiting variance. It is possible to allow some, but not all, variables in x_i to be discrete. In some applications it may be too strong to require the data to be independent over time. This assumption can also be weakened.

3. The assumption that u_i is symmetric about zero is quite strong and is not required by the OLS estimator.

The following result is proven in the appendix.

THEOREM. *Suppose that assumptions A1-A5 hold. Then,*

$$\sqrt{n}(\widehat{\beta} - \beta_0) \Longrightarrow N(0, \Omega),$$

where

$$\Omega = [E(x_i x_i')]^{-1} \frac{\int_0^\infty \int_0^\infty v^2 f_u(v+\eta) f_\eta(\eta) d\eta dv}{2\left(\int_0^\infty \int_0^\infty f_u(v+\eta) f_\eta(\eta) d\eta dv\right)^2}.$$

REMARKS

1. We can write

$$\Omega = [E(x_i x_i')]^{-1} \frac{E[v_i^2 1(v_i \geq 0)]}{2\left(\Pr(v_i \geq 0)\right)^2},$$

where $v_i = u_i - \eta_{1i}$. By our assumption A2, there exists some set of positive values that both u_i and η_{1i} can take and so $\Pr[u_i \geq \eta_{1i}] > 0$, which guarantees that Ω is finite.

2. We can construct consistent standard errors from $\widehat{\Omega} = \widehat{A}^{-1} \widehat{B} \widehat{A}^{-1}$, where

$$\begin{aligned}
\widehat{B} &= \frac{2}{n} \sum_{i=1}^n x_i x_i' \widehat{\epsilon}_i^2(\widehat{\beta}) \\
\widehat{A} &= \frac{2}{n} \sum_{i=1}^n x_i x_i' [1(y_i^L \geq \widehat{y}_i(\widehat{\beta})) + 1(y_i^U \leq \widehat{y}_i(\widehat{\beta})],
\end{aligned}$$

where $\widehat{\epsilon}_i(\widehat{\beta})$ was defined in (1).

3. We can use the sum of squared residuals to measure the fit of the model and also to test hypotheses about β.

4. It is straightforward to extend our analysis to nonlinear regression functions, instrumental variables, and to LAD criterion functions.

4 Comparison with OLS

Here, we compare Ω with the variance of the OLS estimator of y_i^* on x_i, i.e.,

$$\Sigma = \left[E(x_i x_i')\right]^{-1} (\sigma_u^2 + \frac{1}{2}\sigma_\eta^2).$$

So the question is whether

$$\frac{E[v_i^2 1(v_i \geq 0)]}{\left(\Pr(v_i \geq 0)\right)^2} <>= 2\sigma_u^2 + \sigma_\eta^2.$$

Here, v_i is a random variable with mean $-\mu_\eta < 0$ and variance $\sigma_v^2 = \sigma_u^2 + \sigma_\eta^2$. We compare the two estimators in a special case where u_i is standard normal and η_{ji} are uniform on $[0, a]$ for some parameter a. When $a = 0$ the two estimators are actually the same and of course have the same variance. The relative efficiency of the two procedures as a function of a is shown in figure 1. It is non-monotonic in a: first, as a increases the OLS estimator is more efficient, but this increase peaks at approximately $a = 6$ [at which point OLS has slightly less than half the variance of $\widehat{\beta}$], and then decreases to the extent that when $a > 20$ the OLS estimator has larger variance. Thereafter, the inefficiency of OLS gets worser and worser. Of course, this is reflecting the fact that the composite error term in the y^* regression is becoming less and less normal, so the inefficiency of OLS should be no surprise.

5 Concluding Remarks

We close with some comments and suggestions for future work. This estimator seems to be eminently plausible, and so it is a bit of a surprise that it requires stronger conditions than mid-point OLS to ensure consistency. It would be of interest to find sampling schemes in which the mid-point OLS estimator is inconsistent, while our estimator is consistent. This might involve looking at LAD versions of our procedure [which is well-justified in any case]. Heteroskedasticity and asymmetry of the error terms are to be expected as well as dependence of the spread on the covariates is to be expected, and any reputable estimator should be able to deal with such things.

A Appendix: Proof of Theorem

The proof is based on verifying the conditions of Theorems 3.1-3.3 of Pakes and Pollard (1989).

Proof of Consistency. For each β define the i.i.d. random variables

$$v_i(\beta) = u_i - \eta_{1i} + x_i'(\beta_0 - \beta) \text{ and } w_i(\beta) = u_i + \eta_{2i} + x_i'(\beta_0 - \beta).$$

Then, by the law of iterated expectation

$$\begin{aligned} G(\beta) = E[G_n(\beta)] &= -E\left[x_i v_i(\beta) 1(v_i(\beta) \geq 0)\right] - E\left[x_i w_i(\beta) 1(w_i(\beta) \leq 0)\right] \\ &= -E\left[x_i E[v_i(\beta) 1(v_i(\beta) \geq 0) | x_i]\right] - E\left[x_i E[w_i(\beta) 1(w_i(\beta) \leq 0) | x_i]\right]. \end{aligned}$$

In the special case that $\beta = \beta_0$, $v_i = v_i(\beta_0) = u_i - \eta_{1i}$ and $w_i = w_i(\beta_0) = u_i + \eta_{2i}$, and

$$G(\beta_0) = -E(x_i) E\left[v_i 1(v_i \geq 0)\right] - E(x_i) E\left[w_i 1(w_i \leq 0)\right].$$

In this special case we also have that the densities of v_i, w_i are

$$\begin{aligned} f_v(v) &= \int_0^\infty f_u(v + \eta) f_\eta(\eta) d\eta \\ f_w(w) &= \int_0^\infty f_u(w - \eta) f_\eta(\eta) d\eta. \end{aligned}$$

By assumption A2, $f_v(v) > 0$ for some non trivial subset of $[0, \infty)$ and $f_w(w) > 0$ for some non trivial subset of $(-\infty, 0]$. Therefore, since f_u is symmetric about zero

$$\begin{aligned} & E\left[v_i 1(v_i \geq 0)\right] + E\left[w_i 1(w_i \leq 0)\right] \\ &= \int_0^\infty v f_v(v) dv + \int_{-\infty}^0 w f_w(w) dw \\ &= \int_0^\infty \int_0^\infty v f_u(v + \eta) f_\eta(\eta) d\eta dv + \int_{-\infty}^0 \int_0^\infty w f_u(w - \eta) f_\eta(\eta) d\eta dw \\ &= \int_0^\infty \int_0^\infty v f_u(v + \eta) f_\eta(\eta) d\eta dv + \int_0^\infty \int_0^\infty -w f_u(-w - \eta) f_\eta(\eta) dw d\eta \\ &= \int_0^\infty \int_0^\infty v f_u(v + \eta) f_\eta(\eta) d\eta dv - \int_0^\infty \int_0^\infty w f_u(w + \eta) f_\eta(\eta) d\eta dw \\ &= 0 \end{aligned}$$

by Fubini's theorem and a change of variables $w \mapsto -w$. Thus, we have shown that $G(\beta_0) = 0$. We now turn to the more general β case. By our assumption A1 the conditional densities of $v_i(\beta)$ and $w_i(\beta)$ given $x_i = x$ are

$$
\begin{aligned}
f_{v|x}(v) &= \int_0^\infty f_u(v + \eta - x'(\beta_0 - \beta)) f_\eta(\eta) d\eta \\
f_{w|x}(w) &= \int_0^\infty f_u(w - \eta - x'(\beta_0 - \beta)) f_\eta(\eta) d\eta,
\end{aligned}
$$

so that

$$
\begin{aligned}
G(\beta) &= -\int_{\mathbb{R}^d} \int_0^\infty \int_0^\infty xv f_u(v + \eta - x'(\beta_0 - \beta)) f_\eta(\eta) f_X(x) d\eta dv dx \\
&\quad - \int_{\mathbb{R}^d} \int_{-\infty}^0 \int_0^\infty xw f_u(w - \eta - x'(\beta_0 - \beta)) f_\eta(\eta) f_X(x) d\eta dw dx \\
&= \int_{\mathbb{R}^d} \int_0^\infty \int_0^\infty xv[f_u(v + \eta + x'(\beta_0 - \beta)) - f_u(v + \eta - x'(\beta_0 - \beta))] f_\eta(\eta) f_X(x) d\eta dv dx
\end{aligned}
$$

by the same change of variables and symmetry argument. We must show that for all $\varepsilon > 0$ there exists $\delta > 0$ such that $\inf_{\|\beta - \beta_0\| > \varepsilon} \|G(\beta)\| > \delta$. This is guaranteed by our assumption A2 because this implies that there exists a set of positive probability such that the term in square brackets is bounded away from zero. Specifically, we have using integration by parts that

$$
\begin{aligned}
&\int_0^\infty \int_0^\infty v[f_u(v + \eta + x'(\beta_0 - \beta)) - f_u(v + \eta - x'(\beta_0 - \beta))] f_\eta(\eta) d\eta dv \\
&= \int_0^\infty v[f_v(v + x'(\beta_0 - \beta)) - f_v(v - x'(\beta_0 - \beta))] dv \\
&= -\int_0^\infty [F_v(v + x'(\beta_0 - \beta)) - F_v(v - x'(\beta_0 - \beta))] dv
\end{aligned}
$$

where F_v is the c.d.f. of f_v, because

$$
[v[F_v(v + x'(\beta_0 - \beta)) - F_v(v - x'(\beta_0 - \beta))]]_0^\infty = 0
$$

for all x because $vf_v(v) \to 0$ as $v \to \infty$ by A4. By the mean value theorem

$$
\begin{aligned}
G(\beta) &= -\int_{\mathbb{R}^d} \int_0^\infty [F_v(v + x'(\beta_0 - \beta)) - F_v(v - x'(\beta_0 - \beta))] x f_X(x) dv dx \\
&= -\int_{\mathbb{R}^d} \int_0^\infty xx'(\beta_0 - \beta) f_X(x) f_v(v + \delta(x'(\beta_0 - \beta))) dv dx
\end{aligned}
$$

for some $\delta(x'(\beta_0-\beta))$ lying between $+x'(\beta_0-\beta)$ and $-x'(\beta_0-\beta)$. By bounding $f_v(v+\delta(x'(\beta_0-\beta)))$ away from zero on a non-trivial set using assumption A2 we can now write $||G(\beta)|| \geq C||\beta_0-\beta||$ for some positive constant C by the well known matrix inequality $\|Bx\| \geq C\|x\|$ for full rank B.

The first order condition is a sum of piecewise linear functions of β that are i.i.d., i.e.,

$$G_n(\beta) = \frac{1}{n}\sum_{i=1}^{n} x_i \rho(z_i,\beta),$$

where $\rho(z_i,\beta) = -\{v_i(\beta)1(v_i(\beta) \geq 0) + w_i(\beta)1(w_i(\beta) \leq 0)\}$ and $z_i = (x_i, y_i^L, y_i^U)$. By A1 and A4, $G_n(\beta)$ satisfies a weak law of large numbers. By standard results on indicator functions with linear indexes inside [see Pakes and Pollard (1989) and Sherman (1993)], this convergence can be made uniform over compacts, thus

$$\sup_{\beta\in B} \|G_n(\beta) - G(\beta)\| \xrightarrow{p} 0. \tag{3}$$

Combining this with the identification result given above gives consistency. ■

Proof of Asymptotic Normality. We must show that: (i) $\sqrt{n}G_n(\beta_0)$ is asymptotically normal; (ii) $\partial G(\beta)/\partial\beta$ is non-singular at $\beta = \beta_0$; (iii) a stochastic equicontinuity condition given below holds.

First of all,

$$\sqrt{n}G_n(\beta_0) = \frac{1}{\sqrt{n}}\sum_{i=1}^{n} x_i\rho(z_i,\beta_0).$$

where $\rho(z_i,\beta_0) = -\{v_i 1(v_i \geq 0) + w_i 1(w_i \leq 0)\}$ is asymptotically normal with mean zero and variance $E[x_i x_i']E[\rho(z_i,\beta_0)^2]$, since $\rho(z_i,\beta_0)$ is independent of x_i and mean zero. We have

$$\begin{aligned} E[\rho(z_i,\beta_0)^2] &= E[v_i^2 1(v_i \geq 0)] + E[w_i^2 1(w_i \leq 0)] + 2E[v_i w_i 1(v_i \geq 0)1(w_i \leq 0)] \\ &= 2\int_0^\infty \int_0^\infty f_u(v+\eta) f_\eta(\eta)d\eta dv \end{aligned}$$

because the two square terms are the same and the cross-product is zero, as we now show. First, note that v_i and $-w_i$ have the same marginal distribution

$$-(u_i - \eta_{1i}) = -u_i + \eta_{1i} \stackrel{d}{=} u_i + \eta_{2i}$$

by virtue of the symmetry of u_i and the common distribution of η_{1i}, η_{2i}. Second, since v_i is independent of w_i given u_i we have by the law of iterated expectation

$$E[v_i w_i 1(v_i \geq 0)1(w_i \leq 0)] = E\left[E[v_i 1(v_i \geq 0)|u_i]E[w_i 1(w_i \leq 0)|u_i]\right].$$

Then, note that

$$E[v_i 1(v_i \geq 0)|u_i = u] = uF_\eta(u) - \int_0^u \eta f_\eta(\eta)d\eta,$$

which is non-zero if and only if $u > 0$. Likewise,

$$E[w_i 1(w_i \leq 0)|u_i = u] = uF_\eta(-u) + \int_0^{-u} \eta f_\eta(\eta) d\eta$$

is non-zero if and only if $u < 0$. Therefore either one of these terms are zero so that $E[v_i w_i 1(v_i \geq 0) 1(w_i \leq 0)] = 0$ as required. This concludes the proof of (i).

Regarding (ii), by A2 and A3,

$$\begin{aligned} \frac{\partial G}{\partial \beta}(\beta_0) &= -\int_{\mathbb{R}^d} \int_0^\infty \int_0^\infty xx' v f_u'(v+\eta) f_\eta(\eta) f_X(x) d\eta dv dx \\ &\quad + \int_{\mathbb{R}^d} \int_{-\infty}^0 \int_0^\infty xx' w f_u'(w-\eta) f_\eta(\eta) f_X(x) d\eta dw dx \\ &= 2E(x_i x_i') \int_0^\infty \int_0^\infty f_u(v+\eta) f_\eta(\eta) dv, \end{aligned}$$

because by integration by parts

$$\begin{aligned} \int_0^\infty v f_u'(v+\eta) dv &= [v f_u(v+\eta)]_0^\infty - \int_0^\infty f_u(v+\eta) dv \\ &= -\int_0^\infty f_u(v+\eta) dv. \end{aligned}$$

It follows that $\partial G(\beta_0)/\partial\beta$ is non-singular because we assumed that $E(x_i x_i')$ was, and clearly $\int_0^\infty \int_0^\infty f_u(v+\eta) f_\eta(\eta) dv > 0$.

A sufficient stochastic equicontinuity condition for (iii) above is that: for all sequences $\delta_n \to 0$ we have

$$\sup_{\|\beta-\beta_0\| \leq \delta_n} \left\| \sqrt{n}[G_n(\beta) - G(\beta)] - \sqrt{n}[G_n(\beta_0) - G(\beta_0)] \right\| = o_p(1). \tag{4}$$

This condition is satisfied under our conditions because of the structure of G_n. See Pakes and Pollard (1989) and Sherman (1993) for further discussion.

Finally, we have by the arguments of Pakes and Pollard [which make use of A5] that

$$\sqrt{n}(\widehat{\beta} - \beta) = -\left[\frac{\partial G}{\partial \beta}(\beta_0) + o_p(1)\right]^{-1} \sqrt{n} G_n(\beta_0) + o_p(1),$$

so that $\sqrt{n}(\widehat{\beta} - \beta)$ is asymptotically normal with the stated variance.

■

References

[1] Bliss, R. (1997). Testing term structure estimation methods. *Advances in Futures and Options Research* 9, 197-231.

[2] Campbell, J.Y., A. Lo, and C. MacKinlay (1997). *The Econometrics of Financial Markets.* Princeton University Press; Princeton, New Jersey.

[3] Linton, O., E. Mammen, J.P. Nielsen, and C. Tanggaard (2000). Estimating Yield curves by Kernel Smoothing. Forthcoming in *The Journal of Econometrics*.

[4] Pakes, A., and D. Pollard (1989). Simulation and the asymptotics of optimization estimators. *Econometrica* **57**, 1027-1057.

[5] Sherman, R.P. (1993). The limiting distribution of the maximum rank correlation estimator. *Econometrica* **61**, 123-138.