



UNIVERSITY OF  
ILLINOIS LIBRARY  
AT URBANA-CHAMPAIGN  
BOOKSTACKS

Digitized by the Internet Archive  
in 2011 with funding from  
University of Illinois Urbana-Champaign



330  
B385  
No. 1222 COPY 2  
cup. 2



# BEBR

FACULTY WORKING  
PAPER NO. 1222

Asymptotic Theory and Econometric Practice

*Roger Koenker*

THE LIBRARY OF THE  
APR 01 1986  
UNIVERSITY OF ILLINOIS  
-CHAMPAIGN

College of Commerce and Business Administration  
Bureau of Economic and Business Research  
University of Illinois, Urbana-Champaign



# BEBR

FACULTY WORKING PAPER NO 1222

College of Commerce and Business Administration

University of Illinois at Urbana-Champaign

February 1986

Asymptotic Theory and Econometric Practice

Roger Koenker, Professor  
Department of Economics

Research supported by NSF Grant SES-8408567. A preliminary version of this work was presented at the 5th World Congress of the Econometric Society in Cambridge, MA., August 1985. This is a preliminary draft and should not be cited without the permission of the author.





## Asymptotic Theory and Econometric Practice

Roger Koenker

The classical paradigm of asymptotic theory rests on the following "willing suspension of disbelief." We are asked to imagine a colleague with an extremely diligent research assistant in the throes of specifying an econometric model. Daily, the RA arrives with buckets full of independent new observations, but our colleague is so uninspired by curiosity and convinced of the validity of his original model, that each day he simply reestimates this original model without alteration using larger and larger samples.

We estimate a poisson model of the specification of wage equations in the econometric literature based on a sample of 733 equations from 156 papers. The results strongly suggest that the classical paradigm is seriously flawed: the number of parameters estimated in wage equations, say  $p$  tends to infinity as the sample size tends to infinity, and, roughly,  $p^4/n$  tends to a constant. Should we abandon our cherished beliefs in consistency and asymptotic normality to the dustbin of irrelevance?

On the contrary, the forthright admission that in realistic econometric settings  $p \rightarrow \infty$  with  $n$ , offers an opportunity for an even more challenging (and informative) asymptotic theory. Huber(1973) was apparently the first to observe that under rather mild regularity conditions on the sequence of designs consistency and asymptotic normality of the least squares estimator in linear models was possible if  $p/n \rightarrow 0$ . Portnoy(1984,1985) extending results of Huber and others has shown that similar results may be established for a broad class of M-estimators for linear models when  $p \log(p)/n \rightarrow 0$ . We survey these results and report on some similar results for L-estimators for the linear model of the type proposed in Koenker and Bassett(1978).



## 1. Introduction

The classical paradigm of asymptotic theory in econometrics rests on the following "willing suspension of disbelief." We must imagine a colleague in the throes of specifying an econometric model. Daily, an extremely diligent research assistant arrives with hundreds of (independent) new observations, but our imaginary colleague is so uninspired by curiosity and convinced of the validity of his original model, that each day he simply reestimates his primal model--without alteration--employing his ever-larger samples.

Is this a plausible meta-model of econometric model building? Casual observation suggests that it is not. The parametric dimension of econometric models seems to expand inexorably as larger samples tempt the researcher to ask new questions and refine old ones. Indeed, this natural temptation is formally justified by the extensive literature on pre-testing and model selection. As larger samples improve the precision of our estimates, our willingness to accept bias in exchange for further improvements in precision inevitably declines.

In the next section we propose a simple, yet we hope plausible, meta-model of the econometric model specification process. And we present some empirical evidence on the specification of models of wage determination. We conclude from this exercise that the parametric dimension of wage models grows roughly like the fourth root of the sample size. The hypothesis of classical asymptotic theory that parametric dimension is fixed, i.e., independent of sample size, is decisively rejected.

Should we abandon our cherished beliefs in the consistency and asymptotic normality of econometric methods? Are the approximations suggested by fixed- $p$  asymptotic theory "irrelevant" to the "real world" of econometric practice? In Section 3 we argue, on the contrary, that the forthright admission that  $p \rightarrow \infty$  with  $n$ , offers an opportunity for a challenging and much more informative new form of asymptotic theory. We begin by reviewing results of Huber (1973) on the large sample theory of the least squares estimator in linear models with  $p \rightarrow \infty$ . The recent results of Yohai and Marrona (1979) and Portnoy (1984) on large- $p$  asymp-

otics for other  $m$ -estimators are then surveyed. And we conclude with some remarks about extending these results to the  $l$ -estimators for linear models introduced in Koenker and Bassett (1978).

## 2. Econometric Practice: A Meta-Model of Wage Determination Models

Models of wage determination offer an unusually rich and revealing source of data on the practice of model specification in econometrics. The "wage equation" pervades the applied econometrics literature; models of discrimination in employment, the effects of unions, returns to education and of wage determination. The development of several large scale panel surveys of labor market experience has facilitated the rapid growth of this empirical literature.

A meta-model is, of course, a model of models. As suggested in the previous section, we are primarily interested in modeling the dependence of the parametric dimension of models, say  $p$ , on the sample size of the available data, say  $n$ . Since the proposed dependent variable,  $p$ , is inherently a positive integer it is natural to begin with Poisson models in which the intensity (or rate) is taken to be some parametric function of the sample size and perhaps other characteristics of the research.

The data which we will analyze consists of 733 wage equations reported in 156 papers in mainstream economics journals and essay collections over the period 1970 to 1980. For each equation we observe the number of parameters estimated, the sample size, date of publication, and subject classified into four categories. We also record the number of equations reported in each paper which is used to weight the observations. Inevitably, there are ambiguities in interpretation of the data. What constitutes an equation? Usually, this is quite straightforward, however, occasionally one sees samples split by age, race, sex, etc., and estimated with and without homogeneity constraints on the coefficients. Our policy in these cases was to interpret the disaggregated form of the equation as a single equation with say,  $mp$ , parameters, not as  $m$  distinct equations with  $p$  parameters. Frequently, there are non-wage equations in the surveyed papers; these are remorselessly ignored. Equations must have wage, or some

function of wage as the dependent variable. Throughout, we have weighted observations on equations by the reciprocal of the number of equations appearing in the published paper. This tends to alleviate the problem of over-representation in the sample by a few (candid) "fishing" enthusiasts who report a large number of equations in a single paper.

It would be barbaric in the extreme to adopt a notation in which  $p$  was regressed on  $n$ , so we will revert to the more civilized convention of denoting our observed dependent variable by  $y$ , the sample size variable will be denoted  $z$ , and the vector of explanatory variables will be denoted  $x$ . Our meta sample size, 733, may thus be denoted simply as  $n$ , and the dimension of  $x$  by  $p$ . This notational recursion makes the world safe for meta-meta-econometrics.

For the Poisson model we may write, for a typical observation

$$P(Y=y) = e^{-\lambda} \lambda^y / y!$$

with the rate parameter  $\lambda$  is expressed, e.g., as,

$$\lambda = \exp(x\beta) = \exp \beta_1 + \beta_2 \log z$$

In this form, the expectation and variance of the random variable  $Y$  are both equal to the value  $\lambda$ . This is not entirely implausible since we might expect that the dispersion of model size would increase with its expectation. The Poisson hypothesis is obviously much stronger than this vague presumption of monotonicity and may be subjected to explicit test. This problem is addressed below.

The simplest, and therefore perhaps the most compelling, of our estimated meta-models yields<sup>1</sup>

---

<sup>1</sup> All estimation of Poisson models reported in this paper was carried out in the GLIM (Generalized Linear Interactive Modeling/System Release 3 Baker and Nelder (1978) see also McCullough and Nelder (1983). Reported standard errors beneath the coefficients in all poisson models are based on the GLIM quasi-likelihood model in which  $V(Y) = \sigma^2 E(Y)$  with  $\sigma^2$  a free parameter. It should be emphasized that in cases on overdispersion ( $\sigma^2 > 1$ ) strict adherence to the Poisson assumption that  $\sigma^2 = 1$  can seriously bias standard errors toward zero.

$$\log \lambda = \underset{(0.149)}{1.336} + \underset{(0.017)}{0.235} \log z \quad (2.1)$$

Thus, roughly speaking, a 1% increase in the sample size of a wage determination model induces a 1/4% increase in the number of parameters of the model. This parsimony elasticity, or for the sake of brevity, "parsity," is, perhaps, *the* critical parameter of meta-econometrics. It will be denoted as  $\pi$  below. To put it slightly differently,  $p/n^{1/4}$  is roughly constant ( $= \log 1.336 \approx 4$ ) over the range of observed wage equation models. It should be emphasized that the hypothesis of classical asymptotic theory that the dimension of parametric models is independent of sample size:  $\beta_2 = 0$  in (2.1) is decisively rejected by the data.

Our simple bivariate model is unsatisfactory in several respects:

- 1.) It predicts poorly for small  $n$ , implying extravagantly prodigal models for  $n < 100$ , and negative degrees of freedom for  $n < 10$ .
- 2.) The model, in GLIM terminology, is seriously overdispersed, i.e., the Poisson hypothesis that  $V(Y) = E(Y)$  is not supported by the data. The usual GLIM diagnostic is the estimated scale parameter

$$\bar{\sigma} = (n-p)^{-1} \sum (y_i - \hat{\lambda}_i)^2 / \hat{\lambda}_i$$

is 4.73 in this case and significantly different from the hypothesized value of one.

- 3.) There are a few highly influential observations with  $z_i$ 's (sample sizes) above 500,000.

Thus the narrow confidence interval on the coefficient of  $\log z$  in (2.1) constructed conditional on this specification of the meta-model is far too optimistic. We have experimented with several alternate functional forms of the model for the conditional expectation of model size. The obvious tactic of introducing a log quadratic term is (unfortunately) extremely sensitive to the observations alluded to in point (3.) above. With those observations, we obtain,

$$\log \lambda = \underset{(.512)}{-.438} + \underset{(.118)}{.663} \log z - \underset{(.0067)}{.0245} (\log z)^2 \quad (2.2)$$

while without them we have,

$$\log \lambda = \underset{(.512)}{1.737} + \underset{(.128)}{.0581} \log z + \underset{(.0078)}{.01543} (\log z)^2 \quad (2.3)$$

In the former the model predicts that model size declines after roughly  $n = 100,000$ , whereas

the latter implies smoothly increasing parsity. In both cases the parsity at mean<sup>2</sup> sample size ( $n \approx 1000$ ) is roughly comparable to our simple model,  $\pi = .32$  for (2.2) and  $\pi = .27$  for (2.3). It is admittedly disturbing to find that the rise and fall of parsity is so sensitive to a few large-sample observations from our meta-sample. However, such sensitivity, especially in quadratic models, is often inevitable. Further, one may wish to question whether the observations with  $n > 250,000$  are really drawn from the same population as the other observations of our meta-sample. For these cases, computational considerations enter the model specification process in a nontrivial way and may eventually come to dominate the "scientific" considerations which we emphasized in Section 1.<sup>3</sup> Thus we believe that there should be some *a priori* preference for (2.3) over (2.2).

We have also experimented with models in  $\log(\log n)$ . The estimated model

$$\log \lambda = \underset{(.315)}{-.777} + \underset{(.148)}{1.947} \log \log z \quad (2.4)$$

yields a slightly better fit than our simple meta-model (2.1) and at mean sample size it implies a parsity of  $\pi = .28$ . The log-log form has the attractive feature that the parsity parameter is proportional to the reciprocal of  $\log(\text{sample size})$ , and therefore tends to zero as  $n \rightarrow \infty$ . Figure 2.1 illustrates the differences among the four models reported above with respect to parsity as a function of sample size. One sees clearly in the figure that the differences between the functional forms are primarily in the extremes of the observed sample sizes.

We have emphasized above that all of the Poisson models suffer from over-dispersion, that is the estimated variance of dependent variable is roughly 3-4 times the mean that is predicted by the Poisson model. One interpretation of this overdispersion in Poisson models is that there is some inherent variability in the rate parameter  $\lambda$  around its hypothesized ( $\log$ )

---

<sup>2</sup> Since sample sizes are logged this mean ( $\bar{n} = 1041$ ) is geometric.

<sup>3</sup> This comment may seem to undercut our contention that  $p \rightarrow \infty$  with  $n$ , which if taken absolutely literally is evidently asymptotically computationally infeasible. Of course, what is relevant is what happens in the range of practical experience for which conventional asymptotic theory is expected to provide a guide; in the case of wage equations this seems to be roughly sample sizes in the range 50-500,000. Here the evidence seems overwhelming that  $p$  increasing gradually with  $n$ .

linear form. The classical approach to treating this (common) syndrome is to hypothesize a gamma distribution for the intercept of the rate equation, and on integrating out this random parameter one obtains a negative binomial model for the dependent variable. See e.g. Johnson and Kotz(1969) and the references cited there. This approach may be traced to Anscombe (1949) who applied it in entomology, a recent application in econometrics is Hausman and Griliches (1983), and an extremely insightful view of this problem and parametric heterogeneity in general is provided by Cox (1984). This interpretation is also set forth in Cheshire (1984).

Tests for parametric heterogeneity in Poisson models may be developed along the lines suggested by Lancaster (1984) based on Cheshire (1984), White (1982), Cox (1984) and others. The basic information identity

$$D = E \frac{\partial^2 \log f}{\partial \theta \partial \theta'} + E \left[ \frac{\partial \log f}{\partial \theta} \frac{\partial \log f'}{\partial \theta} \right] = 0$$

and its extensions may be used to construct tests which are readily computed as  $nR^2$  from a regression of a column of ones on a matrix of  $n$  by  $p(p+1)/2$  elements of  $D$  augmented by the matrix of gradient "observations"  $g = \partial \log f / \partial \theta$  evaluated at the mle. "Explanatory power" in this regression suggests systematic departures in the fitted model from the hypothesis that  $D$  and  $g$  have zero expectation. We have conducted a number of these tests restricting attention to the components of  $[Dg]$  corresponding to the intercept parameter in the log  $\lambda$  equation. Here the test is particularly simple since

$$\hat{d}_i = (y_i - \hat{\lambda}_i)^2 - \lambda_i$$

and

$$\hat{g}_i = y_i - \hat{\lambda}_i$$

where  $\hat{\mu}_i = e^{z_i \hat{\beta}}$ . The test statistic is 133.1 for meta-model (2.1) for example, which is clearly an implausible value for a  $\chi^2$  on 2 degrees of freedom variable.



Unfortunately, the negative binomial model while quite attractive from a number of perspectives is quite unwieldy computationally. Some initial forays have been made using the remarkable quasi-mle software of Spady (1984). This approach is capital intensive, but avoids the difficulties of coding analytical derivatives, and has the singular virtue of producing numerically reliable standard errors.<sup>4</sup> In the simple log linear model, we obtain

$$\log \alpha_i = 1.039 + .272 \log z$$

(22)            (.033)

with  $\hat{\beta} = 1.51$  (.13). The parsity parameter in this model is independent of  $z$  and at .272 roughly the same as in the simple Poisson model. Negative binomial models using other specifications of the conditional mean function also produce results closely resembling their poisson counterparts.

### 3. Asymptotic Theory: A Practical Paradigm

We are faced with a great dialectical discrepancy. Theory offers us a static view of the econometric model, a model "cast in concrete," unperturbed by the influx of new data. The practice of econometrics, however, offers quite a different, more plastic, view: models gradually expanding and elaborating themselves in response to the availability of new data. How are these views to be reconciled?

The answer, of course, is to expand the paradigm of classical asymptotic theory. Huber (1973) was apparently the first to observe that, under rather mild regularity conditions on the sequence of designs, consistency and asymptotic normality of the least-squares estimator in linear models was possible if  $p/n \rightarrow 0$ . These results are quite elementary, on the same level as the fixed  $p$  asymptotics which are done in introductory graduate courses, and therefore should be better known. To my knowledge, only the recent text of Amemiya (1985) treats any of these questions and even there the implications are only implicit.

---

<sup>4</sup> Standard errors are computed by numerical approximations to the general quasi-mle formula  $V = J^{-1} I J^{-1}$  where  $I$  denotes  $E \partial \log f / \partial \theta \partial \log f / \partial \theta'$  and  $J$  denotes  $E \partial^2 \log f / \partial \theta \partial \theta'$ .

To illustrate the general approach consider the simplest application the classical linear model with iid disturbances, and the asymptotic behavior of the least-squares estimator. For fixed  $p$ , and error distributions with finite variance, we know that  $\hat{\beta} \rightarrow \beta_0$ , strongly iff  $(XX)^{-1} \rightarrow 0$ . See Lai, Robbins and Wei (1979), for a proof this is a surprisingly delicate and difficult result. For  $p \rightarrow \infty$  with  $n$ , consider the "hat" matrix<sup>5</sup>  $H = X(X'X)^{-1}X'$  We know the following:  $h_{ii} \in [0,1]$ ,  $tr(H) = p$ ,  $HH = H$  Thus, since  $\hat{y} = Hy$ , we have

$$Var(\hat{y}_i) = \sum_{k=1}^n h_{ik}^2 \sigma^2 = h_{ii} \sigma^2 \quad (3.1)$$

so by Chebyshev's inequality

$$P[|\hat{y}_i - Ey_i| \geq h_i \frac{\sigma^2}{\epsilon^2}] \quad (3.2)$$

*Proposition 3.1.* (Huber)  $\hat{y}_i$  is weakly consistent, i.e.,  $\hat{y}_i \rightarrow^p x_i \beta$  iff  $h_i \rightarrow 0$ .

*Proof.* Sufficiency above. Necessity:

$$\hat{y}_i - Ey_i = h_i u_i + \sum_{k \neq i} h_{ik} u_k \quad (3.3)$$

For independent random variables,  $X, Y$ ,

$$P[|X+Y| > \epsilon] \geq P[X \geq \epsilon]P[Y > 0] + P[X \leq -\epsilon]P[Y < 0] \geq \min\{P[X \geq \epsilon], P[X \leq -\epsilon]\}$$

so

$$P[|\hat{y}_i - Ey_i| > \epsilon] \geq \min\{P[u_i \geq \epsilon/h_i], P[u_i \leq -\epsilon/h_i]\}$$

and if  $h_i \rightarrow 0$ , then the rhs is bounded away from zero ■

Note that  $\bar{h} = \max_i h_{ii}$ ,  $|h_{ii}| \geq \frac{1}{n} \sum h_{ii} = \frac{1}{n} Tr(H) = p/n$ , so  $h \rightarrow 0 \rightarrow p/n \rightarrow 0$  so  $p/n > 0$  is necessary, but not sufficient for weak consistency.

---

<sup>5</sup> This terminology is due to Tukey and may be attributed to the fact that  $H$  "puts the hat on  $y$ ", i.e.,  $\hat{y} = Hy$ .

Now consider an arbitrary linear function of  $\hat{\beta}$ , say  $\alpha' \hat{\beta}$ ,  $\|\alpha\| = 1$ . Assume  $F$  isn't Gaussian, and reparameterize so that

$$X'X = I_p$$

Hence,

$$\hat{\beta} = X'y$$

and

$$a = \alpha' \hat{\beta} = \alpha' X'y \equiv s'y$$

where

$$s's = \alpha' X'X\alpha = 1$$

so

$$\text{Var}(a) = \sigma^2$$

*Proposition 3.2.* (Huber)  $a$  is asymptotically Gaussian, iff  $\bar{s} = \max_i |s_i| \rightarrow 0$ .

*Proof.* If  $\bar{s} \rightarrow 0$  then either  $a$  doesn't have a limiting distribution or it is a convolution of two parts: one of which is  $F$ , thus not Gaussian, by hypothesis. If  $\bar{s} = \max |s_i| \rightarrow 0$ , then the Lindeberg condition is,

$$\begin{aligned} \frac{1}{\sigma^2} \sum E s_i^2 u_i^2 I(|s_i u_i| \leq \epsilon) &\leq \frac{1}{\sigma^2} \sum s_i^2 E u_i^2 I(|u_i| \geq \epsilon \sigma / \bar{s}) \\ &= \frac{1}{\sigma^2} E u^2 I(|u| \geq \epsilon \sigma / \bar{s}) \quad (\text{since } s's = 1) \end{aligned}$$

$$\rightarrow 0 \quad \text{since } \bar{s} \rightarrow 0$$

*Proposition 3.3* (Bickel) Estimable functions  $\alpha' \hat{\beta}$ , are asymptotically Gaussian with natural parameters iff the fitted values are consistent.

*Proof.* (Huber) Since  $X'X = I$

$$s_i^2 = (\sum x_{ik} \alpha_k)^2 \leq (\sum x_{ik}^2)(\sum \alpha_k^2) = h_i$$

so

$$\bar{h} \rightarrow 0 \Rightarrow \bar{s} \rightarrow 0 \Rightarrow a \rightarrow^D G$$

This establishes the "if", the "only if" follows from the decomposition (3.3) and the hypothesis that  $u_i$  is not, itself, Gaussian ■

These results for the least squares estimator are extremely encouraging. What happens in nonlinear cases? The simplest nonlinear case is robust regression for linear models. Here all the nonlinearity seems to be very well circumscribed, however, already, serious difficulties arise. Huber (1973), on the basis of informal expansions and Monte Carlo, conjectured that  $p^2/n \rightarrow 0$  was necessary to achieve a uniform normal approximation for a typical m-estimator in the absence of any symmetry conditions on the error distribution. Subsequently, Yohai and Marrona (1979) showed that  $p^{3/2}\bar{h} \rightarrow 0$  implied a uniform normal approximation, but this means, since  $\bar{h} = O(p/n)$ , that  $p^{5/2}/n$  would be sufficient. Huber (1981) conjectured that  $p\bar{h} \rightarrow 0$  was sufficient and that  $\sqrt{p}\bar{h} \rightarrow 0$  was necessary if the error distribution was permitted to be asymmetric. For symmetric errors one might hope that  $\bar{h} \rightarrow 0$  was sufficient as in the least-squares case. Huber (1981) contains an elementary proof for the case  $p^2\bar{h} \rightarrow 0$ .

Portnoy (1984, 1985) has recently improved these results and verified an important conjecture of Huber. In particular, he shows that under reasonably mild conditions on  $X$ ,<sup>6</sup>  $p(\log n)/n \rightarrow 0$ , suffices for norm consistency of m-estimators based on (smoothly) monotone  $\psi$  functions. Asymptotic normality is more problematic, and under slightly stronger regularity conditions, Portnoy shows that if  $(p \log p)^{3/2}/n \rightarrow 0$  then a uniform normal approximation is possible. Note that this essentially, except for the factor  $(\log p)^{3/2}$ , verifies Huber's conjecture. Unfortunately, Portnoy's arguments which are based on density expansions are extremely

<sup>6</sup> Conditions which roughly require that  $|x_i|/||x_i||$  be smoothly distributed on the unit sphere in  $R^p$ . As would be the case if they were iid and had a nice multivariate density.

complex and delicate. The situation is somewhat better for monotone  $\psi$ , but even there the sledding is rough.

Perhaps here we should pause to reconsider the implications for the wage equation literature considered in the previous section. Recall that our empirical meta-model of wage equations implied that  $p^4/n$  was roughly constant over the observed range of sample sizes. Thus, the Huber-Portnoy results would appear to be extremely encouraging. However, we should be careful to remember that they rely on certain regularity conditions on the sequence of designs in addition to the rate conditions on the growth of  $p$ . These conditions as Portnoy shows are satisfied by design sequences drawn at random from a distribution "not too concentrated in any fixed directions." This, in a simpler form, already arose in the case of least squares where  $\bar{h} \rightarrow 0$  implied  $p/n \rightarrow 0$  as a necessary condition, but clearly the  $\bar{h}$  condition, is much more stringent. For example in the  $p$  sample design it requires that the number of observations in each cell tends to infinity as  $n \rightarrow \infty$ .

## References

- Amemiya, T. (1985). *Advanced Econometrics*. Harvard.
- Anscombe, F (1949). The statistical analysis of insect counts based on the negative binomial distribution. *Biometrics* 5, 165-173.
- Baker, R.J. and Nelder, J.A. (1978) *The GLIM System: Generalized Linear Interactive Modeling* Numerical Algorithms Group
- Bassett,G.W. Koenker,R.W. (1978). The asymptotic distribution of the least absolute error estimator. *Journal of the American Statistical Association* 73, 618-622 .
- Cheshire, A. (1984). Testing for Neglected Heterogeneity. *Econometrica* 52, 865-872.
- Cox, D.R. (1984). Some remarks on overdispersion. *Biometrika* 70, 269-274.
- Hausman, J., Hall B. and Griliches,Z. (1984). Econometric models for count data with an application to the patents R&D relationship. *Econometrica* 52, 909-938.
- Huber P.J. (1973 ). Robust regression, asymptotics, conjectures and monte-carlo. *Annals of Statistics* 1, 799-821 .
- Huber, P.J. (1981). *Robust Statistics*. New York : Wiley.
- Johnson, N.L. and Kotz, S. (1969). *Discrete Distributions*. Wiley.
- Koenker,R.W. and Bassett,G.W. (1978). Regression quantiles. *Econometrica* 46 , 33-50 .
- Lai, T.L. Robbins, H. and Wei, C.Z. (1978). Strong consistency of least squares estimates in multiple regression. *Proceedings of the National Academy* 75, 3034-3036 .
- Lancaster, A.B. (1984). The covariance matrix of the information matrix test. *Econometrica* 52, 1051-1054.
- McCullagh, P. and Nelder, J.A. (1983). *Generalized Linear Models*. London: Chapman and Hall .
- Portnoy, S. (1984). Asymptotic behavior of M-estimators of p regression parameters when  $p^2/n$  is large I: Consistency. *Annals of Statistics* 12, 1298-1309.
- Portnoy, S. (1985). Asymptotic behavior of M-estimators of p regression parameters when  $p^2/n$  is large I: Normal Approximation. *Annals of Statistics* 13, 1298-1309.
- White, H. (1982). Maximum likelihood estimation of misspecified models. *Econometrica* 50, 1-25.
- Yohai, V. and Marrona, R.A. (1979). Asymptotic behavior of M-estimators for linear models. *Annals of Statistics* 7, 258-268.











HECKMAN  
BINDERY INC.



JUN 95

Bound-To-Please® N MANCHESTER  
INDIANA 46962

UNIVERSITY OF ILLINOIS-URBANA



3 0112 038252612