Bootstrapping the Conditional Moment Test for Parametric Duration Models

James E. Prieger*
Department of Economics
University of California
One Shields Avenue
Davis, CA 95616-8578
telephone: (530) 752-8727
fax: (530) 752-9382
jeprieger@ucdavis.edu

June 26, 2003

Working Paper #03-4

University of California, Davis, Department of Economics

Abstract

This letter evaluates the performance of auxiliary regression-based specification tests for parametric duration models estimated with censored data. The test using asymptotic critical values has poor size. Bootstrapping corrects the size problem but results in a biased power curve.

Keywords: conditional moment test, test size, right censoring, type I censoring, duration analysis, exponential distribution, Weibull distribution, specification test, power curve, bootstrap bias.

*This paper stems from the author's dissertation at the University of California, Berkeley. It is a pre-print version of a paper accepted for publication in *Applied Economics Letters*. Please cite the published version rather than this working paper.

1 Introduction

In empirical work in which observations are durations, economic theory rarely suggests which is the correct parametric duration model to use. Therefore when a parametric model is used, specification testing of the distributional assumption has become the norm. Specification testing of duration models is often done via conditional moment tests. These tests can be implemented easily via an auxiliary regression that obviates calculation of the variance matrix for the test statistic (Pagan and Vella, 1989). It is well known that test statistics from auxiliary regressions converge very slowly to their asymptotic distribution. Thus, they appear to be prime candidates for bootstrapping. In this note I derive suitable moment conditions for censored duration models. Although the bootstrap corrects the size problem of the auxiliary regression test, an unexpected results is that the power curve exhibits severe bias. The bias is exacerbated when the data are censored.

I proceed in the next section by introducing a conditional moment test for censored duration models. The Monte Carlo results are in section 3. For derivations and further details on the Monte Carlo exercise, see Prieger (2000).

2 Conditional Moment Tests

Consider an independent latent sample $\{y_i^*\}$ from the duration random variable $Y > 0$, $i = 1, \ldots, N$. Let the observed sample $\{y_i\}$ be censored, with fixed right censoring points $\{c_i\}$ and censoring indicators $\{d_i\}$ such that $y_i = \min\{y_i^*, c_i\}$ and $d_i = 1$ if $y_i = c_i$ and 0 if $y_i < c_i$. Let the hazard function of Y be $h(y|x, \theta_0) \equiv f(y|x, \theta_0) / [1 - F(y|x, \theta_0)]$, where f and F are the pdf and cdf of Y, respectively, θ_0 is a k-vector of parameters and x is an ℓ-vector of explanatory variables. Define $\varepsilon(y, x, \theta_0) \equiv \int_0^y h(t|x, \theta_0)\, dt$, the integrated hazard, to be the *generalized error*. Asterisks denote latent, uncensored quantities, so that $\varepsilon_i^* \equiv \varepsilon(y_i^*, x_i, \theta_0)$ and $\varepsilon_i \equiv \varepsilon(y_i, x_i, \theta_0)$. Then the log likelihood of the observed sample is

$$l(\theta) = \sum_{i=1}^{N} (1 - d_i) \log h(y_i | x_i, \theta) - \varepsilon(y_i, x_i, \theta). \tag{1}$$

The conditional moment approach to specification testing exploits the fact that if the model is correctly specified, the sample average moments of ε_i^* should be close to the population moment expectations.[1] Let m^*: $\mathbb{R}_{++} \to \mathbb{R}^q$ be a vector of conditional moments. Denote $m_i^{0*} \equiv m^*(\varepsilon_i^*)$ and construct m^* so that

$$E\left(m_i^{0*} | x_i\right) = 0, \quad i = 1, \ldots, N. \tag{2}$$

If the uncensored sample $\{y_i^*\}$ were observed, one could base a specification test on the sample analog of (2). Censoring complicates matters slightly, requiring the expectation of (2) conditional on the censoring. Let $w_i = (x_i, c_i, d_i)$ and $m_i^0 \equiv E[m^*(\varepsilon_i^*) | \varepsilon_i, w_i]$. By the law of iterated expectations and (2) we have:

$$E\left(m_i^0 | x_i, c_i\right) = E\left\{E\left[m^*(\varepsilon_i^*) | \varepsilon_i, w_i\right] | x_i, c_i\right\} = 0, \quad i = 1, \ldots, N, \tag{3}$$

where the outer expectation is taken over (ε_i, d_i). Thus specification tests for censored samples may be based on the sample analog of (3). These moment conditions hold for *all* duration distributions as long as the distribution is correctly specified.

Many tests using generalized residuals in the literature are based on raw moments of powers of ε^*. Because ε^* is distributed unit exponential for all duration distributions (Lancaster, 1985), if the $(p-1)$th row of m_i^{0*} is $(\varepsilon_i^{*p} - p!)$ then equality (2) holds. In a censored sample, the appropriate element of m_i^0 corresponding to the pth power of $\hat{\varepsilon}_i$ is

$$m_{ip}^0 = \hat{\varepsilon}_i^p - p! + d_i \sum_{j=0}^{p-1} \frac{p!}{j!} \hat{\varepsilon}_i^j. \tag{4}$$

[1] This section draws upon Pagan and Vella (1989).

One cannot test the first moment because it is exactly zero when evaluated with the ML estimate. The moments investigated in the Monte Carlo exercise are the second and third:[2]

$$m_{i2}^0 = \hat{\varepsilon}_i^2 - 2 + 2d_i(\hat{\varepsilon}_i + 1) \tag{5}$$

$$m_{i3}^0 = \hat{\varepsilon}_i^3 - 6 + 3d_i\left(\hat{\varepsilon}_i^2 + 2\hat{\varepsilon}_i + 2\right). \tag{6}$$

Let $\hat{\theta}$ be the MLE of θ_0, $\hat{\varepsilon}_i \equiv \varepsilon\left(y_i, x_i, \hat{\theta}\right)$, and $\hat{m}_i \equiv m\left(\hat{\varepsilon}_i, w_i\right)$. Define $\hat{g}_i$ to be the derivative of the ith log likelihood for observation i (i.e., $\hat{g}_i \equiv \nabla l_i|_{\theta=\hat{\theta}}$). For the auxiliary regression form of the test, regress $\hat{m}_i$ on $\hat{g}_i$ and a constant (this will be a seemingly unrelated regression [SUR] if $q > 1$) and test the constant(s) for significance with the usual t-test or Wald test.

3 Monte Carlo Results

In this section I examine the small sample performance of the auxiliary regression-based test applied to an exponential regression model. I consider the test with asymptotic critical values and with bootstrap critical values. Given the known size problems of auxiliary regression tests, they appear to be prime candidates for bootstrapping.

The Monte Carlo exercises have the following design. The data generating process, conditional on a heterogeneity term u, is exponential with hazard function $h = \exp\left(-x_i'\beta_0 - u\right)$. To ensure $E(h) = \exp\left(-x_i'\beta_0\right)$, u is $\mathcal{N}\left(\sigma^2/2, \sigma^2\right)$. The data are right-censored, with fixed censoring point c chosen to achieve a desired expected percentage of censoring in the data. The generalized residual under the null hypothesis that $u = 0$ is $\hat{\varepsilon}_i = y_i \exp(-x_i'\hat{\beta})$. The ℓ-vector of scores for the auxiliary regression are $\hat{g}_i = \left[\hat{\varepsilon}_i - (1 - d_i)\right] x_i$ for the ith observation. The moment conditions (5)–(6) are

[2] Condition m_{i2}^0 is numerically equal to the moment condition in Lancaster's (1985) LM test for unobserved heterogeneity when evaluated at the ML estimate of θ (Pagan and Vella, 1989; Prieger, 2000).

regressed via SUR on the scores and a constant and the constants are tested for joint significance with a Wald test. The critical values are either asymptotic based on the χ^2 distribution or from a parametric bootstrap.[3] The covariate vector x is composed of a constant and a standard normal random variable. The regressors and $\beta_0' = (1, 2)$ are fixed throughout all simulations.

Test statistics from auxiliary regressions have poor size in small samples. The first two columns of Table 1 present the actual size of the tests based on second moments, for various sample sizes and levels of censoring. The first column shows that the size of the raw moments test is far from the nominal 5% level when the asymptotic critical value is used, unless the sample sizes are large. The actual test size is about 11% when the sample size is 250 and about 7.5% when the sample size is 1,000. Censoring does not appear to make the distortion worse. When sample sizes increase to 10,000, the size drops to near the correct level, although the bias is still significant for the uncensored case. Thus, although the auxiliary regression method is convenient, it may lead to incorrect inference unless sample sizes are large. The use of the bootstrap (column two) clears up the size distortion quite well for all levels of censoring and sample sizes. None of the bootstrap sizes shows significant bias. When the second and third moments are used together (the last two columns of Table 1), the same general results are evident.

The power of the tests against the alternative of multiplicative lognormally-distributed heterogeneity is depicted in figure 1, for two levels of censoring (none and 25% of the sample). The second and third moment conditions are used and the sample size is 1000.[4] The amount of heterogeneity increases along the horizontal

[3] In the parametric bootstrap (Horowitz, 2001, pp.3165,3181), bootstrap samples are generated from an exponential distribution using the estimated parameters and the actual x's. Qualitatively similar results obtain when the paired (y, x) bootstrap is used.

[4] The bootstrap sample size is 99 and 100,000 iterations are performed. The power is evaluated at 15 points and curves are smoothed for plotting. The bias persists when more bootstrap iterations

axis, which is scaled in the graphs to be the percentage increase in the variance of the latent duration variable due to the heterogeneity.

The power curves reveal the following points. First, as one would expect, the power of a particular form of the test decreases as the amount of censoring in the sample increases. This is because as the censoring becomes more severe, there is less information in the sample. Second, and more novel, all these auxiliary regression tests are biased: for small amounts of heterogeneity there is a smaller chance of rejecting the null hypothesis when false than when true. The bias persists over a large range of alternatives in the bootstrapped test, particularly under censoring (the heavy dashed line). The bootstrap test is consistent[5] against all these alternative hypotheses because the auxiliary regression with the true critical value is consistent in this case (Horowitz, 1997, sec. 4.6), so the bias is purely a small sample phenomenon. These small sample results are unfortunate, however, given the convenience of auxiliary regression tests and the growing use of the bootstrap.

4 Conclusion

The raw moment specification test performed via auxiliary regression is one of the common specification tests for duration data. This note presents moments suitable for censored samples. As the bootstrap becomes more commonly used, it is natural to expect its application to these tests to clear up the size distortions when asymptotic critical values are used. The simulations in this paper show the size correction provided by the bootstrap comes at a cost: bias and low power. The applied economist will need to interpret a "failure to reject" the chosen specification with this caution in mind.

are used.

[5]A test is *consistent* against an alternative hypothesis if its power goes to one asymptotically.

References

Horowitz, Joel L. (1997), Bootstrap Methods in Econometrics, *in* D. Kreps and K. Wallis, eds, 'Advances in Economics and Econometrics: Theory and Applications: Seventh World Congress. Volume 3.', Econometric Society (Monographs, no.28), Cambridge: Cambridge U. Press.

Horowitz, Joel L. (2001), The Bootstrap, *in* J. J. Heckman and E. Leamer, eds, 'Handbook of Econometrics', Vol. 5, Elsevier, Amsterdam, chapter 52, pp. 3159–3228.

Lancaster, Tony (1985), 'Generalised Residuals and Heterogeneous Duration Models: With Applications to the Weibull Model', *Journal of Econometrics* **28**(1), 155–169.

Pagan, Adrian and Vella, Frank (1989), 'Diagnostic Tests for Models Based on Individual Data: A Survey', *Journal of Applied Econometrics* **4**(0), S29–S59.

Prieger, James E. (2000), 'Conditional Moment Tests for Parametric Duration Models', Working Paper 00-10, Department of Economics, University of California, Davis.

	Second Moments		Second and Third Moments	
Test	*Asymptotic Critical Values*	*Bootstrap Critical Values*	*Asymptotic Critical Values*	*Bootstrap Critical Values*
$N = 250$				
No censoring	0.116*	0.050	0.313*	0.050
25% censoring	0.107*	0.050	0.296*	0.052*
50% censoring	0.110*	0.050	0.311*	0.050
$N = 1,000$				
No censoring	0.075*	0.050	0.192*	0.050
25% censoring	0.073*	0.050	0.190*	0.051
50% censoring	0.074*	0.048	0.212*	0.052
$N = 10,000$				
No censoring	0.056*	0.048	0.087*	0.049
25% censoring	0.055	0.050	0.092*	0.050
50% censoring	0.052	0.047	0.104*	0.049

Table notes: Nominal size is 5%; * indicates significant (1% level) bias in the empirical size. N is sample size. Censoring is accomplished with a fixed right censoring point common to all observations. Bootstrap sample size = 999. Number of Monte Carlo trials is 100,000 for N=250, 25,000 for N=1,000, and 10,000 for N=10,000.

Table 1: Empirical Levels of the Auxiliary Regression Conditional Moment Tests

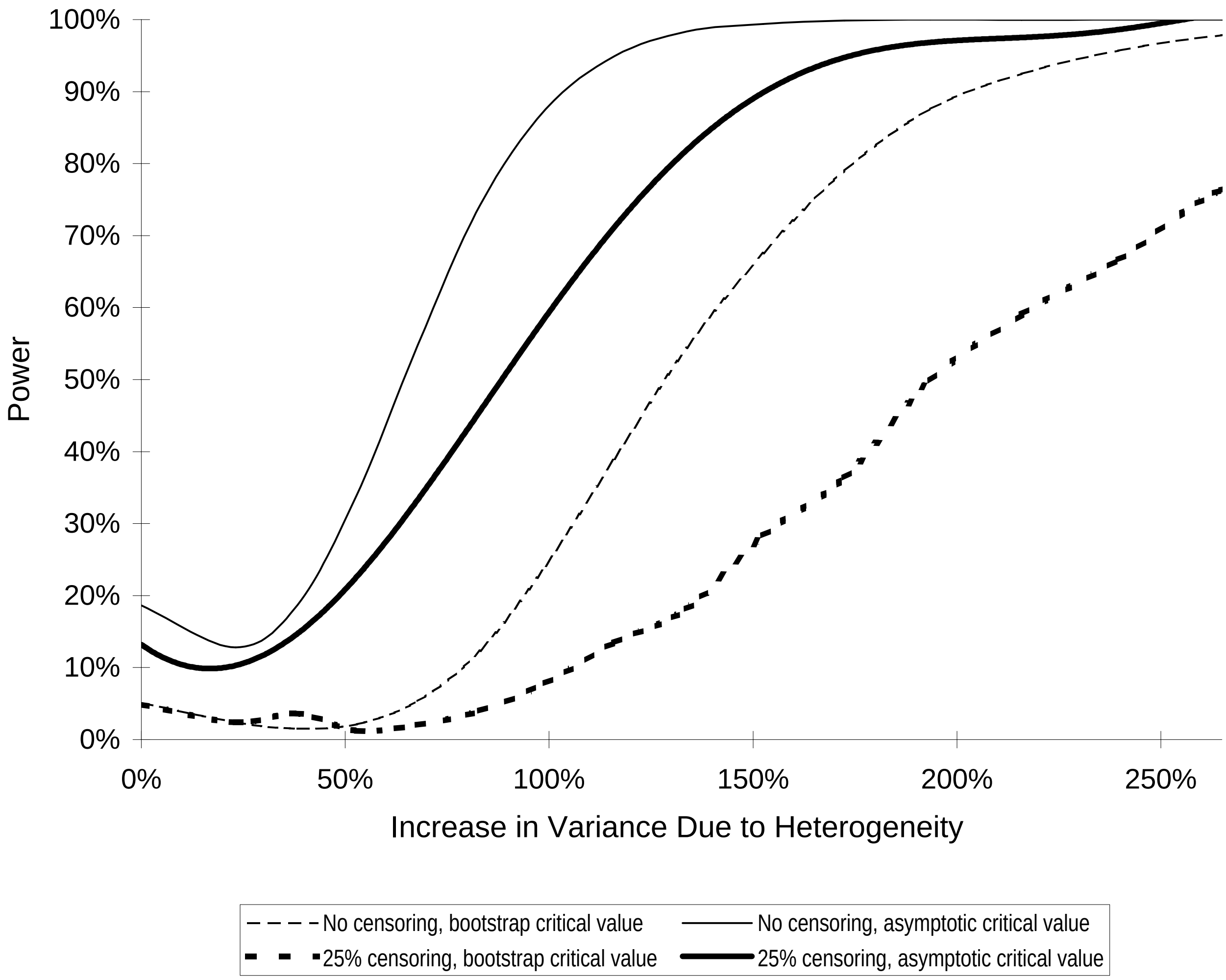

Figure 1: Power curves for the tests vs. lognormal multiplicative heterogeneity (2nd and 3rd moments, N=1,000)