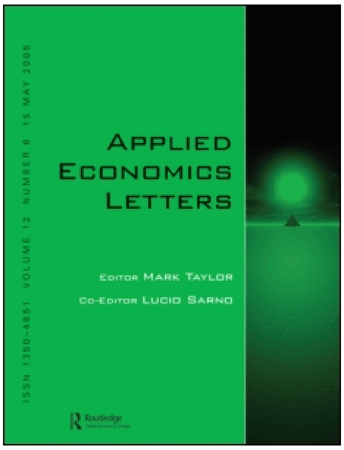

Applied Economics Letters
Publication details, including instructions for authors and subscription information:
http://www.informaworld.com/smpp/title~content=t713684190

A simple IM test for exponential distributions

Pablo Acosta [a]; Gabriel Montes Rojas [a]
[a] Department of Economics, University of Illinois at Urbana-Champaign, Champaign, IL, USA

First Published:February2009

To cite this Article Acosta, Pablo and Rojas, Gabriel Montes(2009)'A simple IM test for exponential distributions',Applied Economics Letters,16:2,109 — 112

To link to this Article: DOI: 10.1080/13504850601018221

URL: http://dx.doi.org/10.1080/13504850601018221

Applied Economics Letters, 2009, **16**, 109–112

A simple IM test for exponential distributions

Pablo Acosta and Gabriel Montes Rojas*

Department of Economics, University of Illinois at Urbana-Champaign, Champaign, IL 61820, USA

We construct a simple information matrix misspecification test for exponential distributions that can be applied in duration models. We evaluate the test performance using Monte Carlo simulation experiments. We found good empirical size properties and good power against Weibull and Gamma distributions.

I. Introduction

This article derives a misspecification test for exponential distribution. These distributions are quite used in duration models and survival analysis (Lancaster, 1990), including several applications in macroeconomics, finance and labour economics (optimal insurance policy, duration of unemployment spell, retirement behaviour, etc.). Quite often the data-generating process for estimating these types of models is assumed to behave as exponential. However, applied researchers fail to test if these assumptions are reasonable, especially when dealing with small samples where asymptotic results do not hold. This calls for developing tests for distributional assumptions in order to avoid misspecification.

We study the properties of an information matrix (IM) specification test, which was first developed in Chesher and Spady (1991). Using Monte Carlo simulations, the size and power of the test are computed. The simulation procedures show good performance in rejecting the null hypothesis of exponential distribution under distributional misspecification (we evaluate Weibull and Gamma alternatives).

The rest of the work this organized as follows. Section II provides a general set up for IM tests. Section III applies it to the exponential distribution. Section IV presents some Monte Carlo results. This article concludes with brief comments in Section V.

II. Information Matrix Tests

Let $X_1, X_2, \ldots$ be a sequence of independent random variables, each with distribution function G. Additionally let Ψ be a parametric family of distributions with parameter $\theta \in \Theta \subset R^q$, $q \geq 1$, and typical element $F(x;\ \theta)$. This family can also be characterized by the probability density function (pdf) $\partial F(x;\theta)/\partial x = f(x;\theta)$ which is twice continuously differentiable in θ and bounded. Define:

$$\theta^* = \arg\sup_{\theta\in\Theta} E_G \log f(X;\theta) \tag{1}$$

$$\hat{\theta}_n = \arg\max_{\theta\in\Theta} \frac{1}{n}\sum_{i=1}^{n} \log f(x_i,\theta) \tag{2}$$

where G refers to the true data-generation process.

*Corresponding author. E-mail: rmontes@uiuc.edu

Applied Economics Letters ISSN 1350–4851 print/ISSN 1466–4291 online © 2009 Taylor & Francis
http://www.informaworld.com
DOI: 10.1080/13504850601018221

Under certain regularity conditions exposed in Huber (1967) and White (1982), we will have strong consistency of the Quasi-maximum likelihood estimator (QMLE), i.e. $\hat{\theta}_n \to \theta^*$ almost surely. If the probability model is correctly specified (i.e. $g(x) = f(x;\, \theta_0)$, $\theta_0 \in \Theta$ using White, 1982, definition), then $E_G \log f(x;\, \theta)$ has a unique maximum, so that the QMLE is consistent for the 'true' parameter vector.

Now define:

$$S(\theta) = E_G\left(\frac{\partial \log f(x,\theta)}{\partial\theta}\right) \tag{3}$$

$$S_n(\theta) = \sum_{i=1}^{n} E_G\left(\frac{\partial \log f(x_i,\theta)}{\partial\theta}\right) \tag{4}$$

$$K(\theta) = E_G \frac{\partial^2 \log f(x,\theta)}{\partial\theta^2} \tag{5}$$

$$K_n(\theta) = \frac{1}{n}\sum_{i=1}^{n} E_G \frac{\partial^2 \log f(x_i,\theta)}{\partial\theta^2} \tag{6}$$

$$J(\theta) = E_G\left(\frac{\partial \log f(x,\theta)}{\partial\theta}\right)\left(\frac{\partial \log f(x,\theta)}{\partial\theta}\right)^T \tag{7}$$

$$J_n(\theta) = \frac{1}{n}\sum_{i=1}^{n}\left(\frac{\partial \log f(x_i,\theta)}{\partial\theta}\right)\left(\frac{\partial \log f(x_i,\theta)}{\partial\theta}\right)^T \tag{8}$$

Under additional mild conditions we also have consistency of (4), (6) and (8). Nevertheless, only in the case that $G \in \Psi$, we obtain the information matrix equality (IME), i.e. $E_G J(\theta) = -E_G H(\theta)$. Following White (1982), constructing a statistics based on $J + K$ is a natural way of testing distributional misspecification of the model.

Note that both J and K are $q \times q$ symmetric matrices. We need to check the $q(q+1)/2$ components, given by the vector:

$$D_n(\hat{\theta}_n) = \frac{1}{n}\sum_{i=1}^{n}\left\{\left(\frac{\partial \log f(x_i,\hat{\theta}_n)}{\partial\theta}\right)\left(\frac{\partial \log f(x_i,\hat{\theta}_n)}{\partial\theta}\right)^T + \frac{\partial^2 \log f(x_i,\hat{\theta}_n)}{\partial\theta^2}\right\} \tag{9}$$

White (1982) proposes the following statistic for testing misspecification:

$$\mathrm{IM} = nD_n(\hat{\theta}_n)'\Omega(\hat{\theta}_n)^{-1}D_n(\hat{\theta}_n) \tag{10}$$

where

$$\Omega(\hat{\theta}) = n^{-1}\sum_{i=1}^{n}\left[D_i(\hat{\theta}_n) - \nabla D_n(\hat{\theta}_n)K_n(\hat{\theta}_n)^{-1}S_n(\hat{\theta}_n)\right] \times \left[D_i(\hat{\theta}_n) - \nabla D_n(\hat{\theta}_n)K_n(\hat{\theta}_n)^{-1}S_n(\hat{\theta}_n)\right]^T$$

and

$$D_i(\hat{\theta}_n) = \left(\frac{\partial \log f(x_i,\hat{\theta}_n)}{\partial\theta}\right)\left(\frac{\partial \log f(x_i,\hat{\theta}_n)}{\partial\theta}\right)^T + \frac{\partial^2 \log f(x_i,\hat{\theta}_n)}{\partial\theta^2}.$$

Under the null hypothesis of correct misspecification, the statistic IM has a chi-squared limit law with $q(q+1)/2$ degrees of freedom.

III. Information Matrix Test for Exponential Distributions

We say that a random variable u has exponential distribution with parameter θ, i.e. $u \sim \varepsilon(\theta)$, if it has a pdf given by:

$$f(u;\theta) = \theta e^{-\theta u}, \quad \theta > 0, 0 < u < \infty \tag{11}$$

Consider:

$$E_G J(\theta) = E_G\left(\frac{\partial \ln f(u,\theta)}{\partial\theta}\right)^2 = \frac{1}{\theta^2} + E_G(u^2) - \frac{2E_G(u)}{\theta} = \frac{1}{\theta^2} + \mu_2' - \frac{2\mu_1'}{\theta} = \frac{1}{\theta^2}$$

$$E_G K(\theta) = E_G\left(-\frac{\partial^2 \ln f(u,\theta)}{\partial\theta^2}\right) = \frac{1}{\theta^2}$$

assuming $G = F(\theta) \in \Psi$, where Ψ is the family of exponential distributions. This shows that the IME holds.

After some algebra, the IM statistic can be expressed as:

$$\mathrm{IM} = \frac{n\left[1/n\sum_{i=1}^{n}\left(u_i^2 - 2u_i/\hat{\theta}_n\right)\right]^2}{1/n\sum_{i=1}^{n}\left[u_i^2 - 2u_i/\hat{\theta}_n - \left(1/\hat{\theta}_n - u_i\right)1/n\sum_{i=1}^{n}2u_i\right]^2} \xrightarrow{L} \chi_1^2 \tag{12}$$

or simplifying the above expression,

$$\mathrm{IM} = \frac{n\left(\hat{\mu}_2' - (2\hat{\mu}_1'/\hat{\theta}_n)\right)^2}{\hat{\mu}_4' - (4/\hat{\theta}_n)\hat{\mu}_3' + 4\hat{\mu}_3'\hat{\mu}_1' + (4/\hat{\theta}_n)\hat{\mu}_2' - (12/\hat{\theta}_n)\hat{\mu}_2'\hat{\mu}_1' + 4\hat{\mu}_2'\left(\hat{\mu}_1'\right)^2 - (8/\hat{\theta}_n)\left(\hat{\mu}_1'\right)^3 + 12/\hat{\theta}_n^2\left(\hat{\mu}_1'\right)^2} \xrightarrow{L} \chi_1^2 \qquad (13)$$

where $\hat{\mu}_k' = (1/n)\sum_{i=1}^n u_i^k$ are the sample raw moments for $k = 1, 2, 3, 4$ and $\hat{\theta} = n/\sum_{i=1}^n u_i$ is the QMLE. It can be easily seen that $\mathrm{IM} = 0$ when $\hat{\mu}_1' = 1/\theta$ and $\hat{\mu}_2' = 2/\theta^2$, i.e. they are equal to the raw moments of a random variable with exponential distribution.

IV. Monte Carlo Results

We evaluate Monte Carlo experiments based on 1000 random samples and we consider tests statistics of asymptotic level 0.10 under the null hypothesis. The sample sizes considered are $\{10, 20, 50, 100, 200, 500\}$.

First, we consider the empirical size properties of the IM test. We consider different parameter values $\theta \in \{1, 2, 3\}$. Table 1 reports the empirical size for these specifications: in all cases, the rejection rate is below the theoretical size, even for small sample sizes.

In order to analyse the power of the proposed test, we evaluate its performance when the data is generated from the most popular alternative in duration models, the Weibull distribution (Table 2). In this case, the pdf is given by

$$f(u) = \lambda\theta^\lambda u^{\lambda-1} e^{-(\theta u)^\lambda}, \quad \lambda > 0, \theta > 0, 0 < u < \infty \qquad (14)$$

Note that it becomes exponential with parameter θ if $\lambda = 1$.

Table 1. Empirical size

Sample size	$\theta = 1$	$\theta = 2$	$\theta = 3$
10	0.035	0.043	0.035
20	0.022	0.024	0.025
50	0.008	0.009	0.009
100	0.009	0.008	0.006
200	0.002	0.006	0.001
500	0.001	0.01	0.000

Notes: Monte Carlo experiments based on 1000 random samples. Theoretical size is 10%.

The exponential distribution can also be seen as a member of the gamma family of distribution functions. The pdf in this case is given by:

$$f(u) = \frac{\theta^\lambda}{\Gamma(\lambda)} u^{\lambda-1} e^{-(\theta u)}, \quad \lambda 0, \theta 0, 0 u \infty \qquad (15)$$

where $\Gamma(\cdot)$ is the gamma function. Again, it becomes exponential with parameter θ if $\lambda = 1$. We generate data from different gamma distributions and compute the power of the test (Table 3).

In both cases (Weibull and Gamma distributions), we obtain a good performance in terms of power, even for small sample sizes. High rejection rates are found for all parameter values and for sample sizes greater than 200. As expected, the tests are consistent, i.e. they have higher rejection rates for larger sample sizes, approaching 1 as more than 200 observations are used. Power also increases as the value of λ further departs from 1 (the baseline value for exponential distribution).

Table 2. Power, Weibull distribution

Sample size		$\lambda = 1.5$	$\lambda = 2$	$\lambda = 2.5$	$\lambda = 3$
$\theta = 1$	10	0.282	0.664	0.893	0.986
	20	0.412	0.905	1.000	1.000
	50	0.730	0.999	1.000	1.000
	100	0.956	1.000	1.000	1.000
	200	0.999	1.000	1.000	1.000
	500	1.000	1.000	1.000	1.000
$\theta = 2$	10	0.274	0.622	0.898	0.981
	20	0.413	0.900	0.993	1.000
	50	0.739	1.000	1.000	1.000
	100	0.947	1.000	1.000	1.000
	200	1.000	1.000	1.000	1.000
	500	1.000	1.000	1.000	1.000
$\theta = 3$	10	0.282	0.637	0.909	0.983
	20	0.415	0.902	0.992	1.000
	50	0.729	0.999	1.000	1.000
	100	0.949	1.000	1.000	1.000
	200	1.000	1.000	1.000	1.000
	500	1.000	1.000	1.000	1.000

Notes: Monte Carlo experiments based on 1000 random samples. Theoretical size is 10%.

Table 3. Power, Gamma distribution

Sample size		$\lambda = 1.5$	$\lambda = 2$	$\lambda = 2.5$	$\lambda = 3$
$\theta = 1$	10	0.140	0.278	0.393	0.529
	20	0.157	0.374	0.550	0.710
	50	0.205	0.597	0.842	0.956
	100	0.306	0.830	0.985	0.999
	200	0.548	0.971	1.000	1.000
	500	0.916	1.000	1.000	1.000
$\theta = 2$	10	0.113	0.255	0.411	0.533
	20	0.135	0.323	0.548	0.739
	50	0.203	0.588	0.843	0.954
	100	0.333	0.846	0.984	0.994
	200	0.545	0.977	0.999	1.000
	500	0.926	1.000	1.000	1.000
$\theta = 3$	10	0.127	0.232	0.384	0.507
	20	0.161	0.342	0.573	0.718
	50	0.197	0.580	0.870	0.951
	100	0.301	0.830	0.972	0.995
	200	0.554	0.982	1.000	1.000
	500	0.925	1.000	1.000	1.000

Notes: Monte Carlo experiments based on 1000 random samples. Theoretical size is 10%.

V. Conclusions

The IM test for exponential distribution is intended to be applied in empirical works using duration data. Experimental simulations show good size and power properties. Rejecting the null hypothesis alerts the researcher of the existence of a distributional misspecification. Other parametric and nonparametric models would be desirable in this case.

The IM test procedure can also be applied to other distributions (i.e. Weibull, Gamma), although the interpretation of the IM statistics is less straightforward. In the exponential case, only the first and second sample moments are needed for satisfying the IME, and it requires only four moments to compute the IM statistic.

Acknowledgement

We are grateful to Anil Bera for his helpful comments.

References

Chesher, A. D. and Spady, R. (1991) Asymptotic expansions of the information matrix test statistic, *Econometrica*, **59**, 787–815.

Huber, P. (1967) The behavior of maximum likelihood estimates under nonstandard conditions, in *Proceedings of the Fifth Berkeley Symposium on Mathematical Statistics and Probability*, University of California Press, Berkeley, CA, pp. 221–33.

Lancaster, T. (1990) *The Econometric Analysis of Transition Data*, Cambridge University Press, Cambridge.

White, H. (1982) Maximum likelihood estimation for misspecified models, *Econometrica*, **50**, 1–25.