A test of the extreme value type I assumption in the bus engine replacement model

Bradley J. Larsen [a,*], Florian Oswald [b], Gregor Reich [c], Dan Wunderli [d]

[a] Department of Economics, Massachusetts Institute of Technology, 50 Memorial Dr. E52-391, Cambridge, MA 02142, United States
[b] Department of Economics, University College London, Gower Street, London WC1E 6BT, United Kingdom
[c] Department of Business Administration, University of Zurich, Moussonstrasse 15, 8044 Zurich, Switzerland
[d] Department of Economics, University of Zurich, Wilfriedstrasse 6, 8032 Zurich, Switzerland

ARTICLE INFO

ABSTRACT

This note tests the assumption of dynamic discrete choice models that underlying utility shocks have an extreme value type I distribution. We find that extreme value type I shocks cannot be rejected in most specifications of the Rust (1987) bus engine replacement model.

© 2012 Elsevier B.V. All rights reserved.

1. Introduction

In dynamic discrete choice models, by far the most common distributional assumption for the unobserved, random component of utility is that it is drawn from the extreme value type I distribution (see the recent survey by Aguirregabiria and Mira, 2010). This assumption is not based on theory, but rather on computational convenience, and perhaps justifiably so: dynamic discrete choice models can be difficult to estimate, and the extreme value assumption reduces an important dimension of complexity by yielding closed-form solutions for choice probabilities as well as for value functions.

In spite of its prevalence, the extreme value assumption of discrete choice models has rarely been tested. It is our belief that it is important to test the sensitivity of econometric models to its assumptions, whether these assumptions are made for theoretical or computational reasons. This work contributes to the literature by testing the extreme value assumption in the now well-known engine replacement model of Rust (1987), a pioneering paper in dynamic discrete choice.

It is important to note that McFadden and Train (2000) show that any random utility model, such as in static demand estimation, can be approximated arbitrarily closely using the extreme value type I distribution for the idiosyncratic component of utility as long as the model includes a sufficiently flexible mixing distribution of random coefficients. With the Rust (1987) model, however, as with many other dynamic discrete choice settings, there are no random coefficients and the only source of uncertainty is the extreme value type I error, and hence the extreme value assumption may potentially be restrictive.

With the original data from Rust's paper, we use constrained optimization and numerical quadrature rules to test the extreme value distribution against several flexible distributions of the exponentiated generalized beta family, which nest the extreme value distribution as a special case. We find, surprisingly, that in most cases, the hypothesis of extreme value errors cannot be rejected, suggesting that, in Rust's data and model, the extreme value assumption, while useful from a computational standpoint, may also be a reasonable modeling assumption. When the assumption can be rejected at marginally significant levels, we demonstrate the implications which our adaptive estimation has for the model.

2. The bus engine replacement model

In the engine replacement model of Rust (1987), a single agent decides each period whether or not to replace a bus engine. The

* Corresponding author. Tel.: +1 617 286 2714; fax: +1 617 253 1330.
E-mail addresses: blarsen@mit.edu (B.J. Larsen), f.oswald@ucl.ac.uk (F. Oswald), gregor.reich@uzh.ch (G. Reich), dan.wunderli@econ.uzh.ch (D. Wunderli).

agent's per-period utility is

$$u(x_t, i, \theta) + \varepsilon_t(i) = \begin{cases} -RC - c(0, \theta_1) + \varepsilon_t(1) & \text{if } i = 1, \\ -c(x_t, \theta_1) + \varepsilon_t(0) & \text{if } i = 0, \end{cases}$$

where $i = 1$ represents the agent's decision to replace the engine and $i = 0$ the decision to not replace. x_t is the engine's mileage at time t, discretized into 90 mileage bins, and $\varepsilon_t(i)$ is a choice-specific shock to utility (observable to the agent but not to the econometrician). RC (replacement cost) and θ_1 (regular maintenance cost) are two of the parameters to be estimated. The function $c(x, \theta_1)$ represents the costs of maintenance on an engine with current mileage x. As in the main specification of Rust (1987), we take $c(x, \theta_1) = 0.001\theta_1 x$.

When the agent discounts future utility with discount factor β, the agent's value function can be written as

$$V_\theta(x_t, \varepsilon_t) = \max_{i \in \{0,1\}} u(x_t, i, \theta) + \varepsilon_t(i) + \beta E[V_\theta(x_{t+1}, \varepsilon_{t+1})|x_t, i]$$

where the subscript θ is used to denote the dependence of the value function on the full parameter vector, θ, which includes RC and θ_1, as well as θ_2, which represents any parameters of the unobserved error distribution, and θ_3, which represents the parameters of the mileage transition probability.

The expected continuation value can be written as

$$E[V_\theta(x_{t+1}, \varepsilon_{t+1})|x_t, i]$$
$$= \int_{x_{t+1}} \int_{\varepsilon_{t+1}} V_\theta(x_{t+1}, \varepsilon_{t+1}) P(dx_{t+1}, d\varepsilon_{t+1}|x_t, i, \theta) \quad (1)$$

where

$$P(x_{t+1}, \varepsilon_{t+1}|x_t, i, \theta) = q(\varepsilon_{t+1}|\theta_2) p(x_{t+1}|x_t, i, \theta_3)$$

with the mileage transition probability defined in θ_3, a three-element vector given by

$$p(x_{t+1}|x_t, i, \theta_3) = \begin{cases} \theta_{30} & \text{if } x_{t+1} = x_t \\ \theta_{31} & \text{if } x_{t+1} = x_t + 1 \\ \theta_{32} & \text{if } x_{t+1} = x_t + 2. \end{cases}$$

Now let $m_{it} \equiv u(x_t, i, \theta) + \beta E[V_\theta(x_{t+1}, \varepsilon_{t+1})|x_t, i]$. Also, let $\eta_t = \varepsilon_t(1) - \varepsilon_t(0)$. Then, under the assumption that ε_{t+1} has an extreme value type I distribution, η_t will be distributed logistically. Thus the probability that the agent chooses to not replace the engine in mileage state x_t is given by the familiar logit formula

$$\Pr(i_t = 0|x_t, \theta) = \frac{1}{1 + e^{-(m_{0t} - m_{1t})}}$$
$$= F(m_{0t} - m_{1t}, \theta_2).$$

$F(\cdot, \theta_2)$ denotes the logistic distribution, with scale and location parameters normalized, and hence θ_2 is empty. Below we will discuss more general distributions which do include parameters contained in θ_2.

Also note that

$$\int_{\varepsilon_t} V_\theta(x_t, \varepsilon_t) q(\varepsilon_t|\theta_2) d\varepsilon_t$$
$$= m_{0t} F(m_{0t} - m_{1t}, \theta_2) + m_{1t}(1 - F(m_{0t} - m_{1t}, \theta_2))$$
$$+ E[\eta_t|\eta_t > m_{0t} - m_{1t}](1 - F(m_{0t} - m_{1t}, \theta_2))$$
$$= \ln(e^{m_{0t}} + e^{m_{1t}}) \quad (2)$$

where the final line follows from the closed-form solution of the maximum of random variables with an extreme value type I distribution, familiar from the discrete choice literature.

The likelihood function of an observed sequence of mileage states and decisions, $(\{x_t\}_{t=1}^T, \{i_t\}_{t=1}^T)$, for a particular bus is given by

$$L\left(\{x_t\}_{t=1}^T, \{i_t\}_{t=1}^T|x_0, i_0, \theta\right) = \prod_{t=1}^T P(i_t|x_t, \theta) p(x_t|x_{t-1}, i_{t-1}, \theta_3) \quad (3)$$

which can be maximized using the constrained maximization procedure described below.

3. Estimation

We solve for the model's parameters following the "mathematical program with equilibrium constraints" (MPEC) approach proposed by Judd and Su (2011). That is, we maximize the likelihood in (3) subject to (1) holding at each state x. As in Rust (1987), we use a discount factor of $\beta = 0.9999$.

To test the assumption that $(\varepsilon_t(0), \varepsilon_t(1))$ have an extreme value type I distribution, we choose distributions of η_t which nest the logistic distribution. Specifically, we choose the family of distributions nested by the exponentiated generalized beta distribution of the second kind (EGB2). The EGB2 has the following CDF and density:

$$F_{EGB2}(\eta; p, q) = \mathcal{B}\left(\frac{1}{1 + e^{\eta + \psi(p) - \psi(q)}}; p, q\right)$$

$$f_{EGB2}(\eta; p, q) = \frac{e^{p(\eta + \psi(p) - \psi(q))}}{B(p, q)(1 + e^{\eta + \psi(p) - \psi(q)})^{p+q}}$$

where p and q are distributional parameters determining the shape (skewness and kurtosis) of the distribution, ψ is the digamma function (the first derivative of the log of the gamma function), and $B(p, q)$ is the beta function. Finally, $\mathcal{B}(\cdot; p, q)$ is the incomplete beta function.

Through the parameters p and q, the EGB2 distribution allows for a wide range of standardized skewness ($[-2, 2]$) and standardized kurtosis ($[3, 9]$), as shown in Hansen et al. (2007). The logistic distribution is a special case of the EGB2, with $p = q = 1$, and allows for zero skewness and only one level of kurtosis (a value of 4.2). Formulas for the skewness and kurtosis of the EBG2 are found in McDonald and Xu (1995).

Two special cases of the EGB2 which nest the logistic distribution while still being flexible are the exponentiated Burr type 3 distribution (EBurr3), given by $F_{EGB2}(\cdot; p = 1, q)$, and the exponentiated Burr type 12 distribution (EBurr12), given by $F_{EGB2}(\cdot; p, q = 1)$. These latter two distributions have convenient closed-form CDFs found in Hansen and McDonald (2002).

However, unlike with the logistic distribution, the truncated expectation, $E[\eta_t|\eta_t > m_{0t} - m_{1t}]$, found in (2), does not have a closed-form solution under these more flexible distributions. We evaluate this object using Gauss–Laguerre quadrature (see Kythe and Schäferkotter, 2005 for details). As demonstrated below, we find this method to be highly accurate.

4. Results

We use the original data from Rust (1987). This data set consists of many buses, each with monthly data on mileage readings as well as the agent's decision to replace the engine or not. The buses are divided into four groups based on their characteristics. See Rust (1987) for more details on the data.

We first replicated the Rust (1987) results using the closed-form which arises from the extreme value type I assumption. We then performed the same estimation but with Gauss–Laguerre quadrature, using 100 Laguerre nodes, rather than using the closed-form solution. The results of these replication exercises for bus groups 1–3 are displayed in Table 1 in the columns labeled "Closed-form" and "Gauss–Laguerre". The results from Rust (1987, Table IX), are also displayed for comparison.[1] As can be seen, the numerical quadrature results agree with the closed-form solution, and our results replicate Rust's quite closely.

To test the extreme value assumption, we estimate the model using each distribution described above and perform a likelihood ratio test. Table 2 shows the results, with a separate panel for the sample containing bus groups 1–3, bus group 4 alone, and bus groups 1–4. Note that the mileage transition parameters, θ_{30} and

---

[1] Rust (1987) standard errors come from the outer product of the gradient. All other standard errors are bootstrapped at the bus level.

Table 1
Replication of the original results from Rust (1987, Table IX).

| | Bus groups 1–3 ($N = 3864$) | | |
	(Rust, 1987)	Closed-form	Gauss–Laguerre
RC	11.7270	11.7194	11.7176
	(2.602)	(1.975)	(1.992)
θ_1	4.8259	4.8204	4.8191
	(1.792)	(1.340)	(1.357)
θ_{30}	0.3010	0.3010	0.3010
	(0.007)	(0.009)	(0.009)
θ_{31}	0.6884	0.6884	0.6884
	(0.008)	(0.009)	(0.009)
L	−2708.366	−2708.366	−2708.366

θ_{31}, are omitted from Table 2 because they do not depend on the distributional assumption of η_t.

In Table 2, note that the parameter estimates vary slightly depending on the distributional assumption used. For example, in each panel, the replacement cost (RC) is nearly the same across distributions. The regular maintenance cost, θ_1, on the other hand, is approximately twice as large under the logistic assumption as it is under EBurr12. However, in each case a Hausman test fails to reject when comparing θ_1 and RC under the logistic assumption to each of the flexible distributions, as shown in the final column of each panel.

Observe that the flexible distributions pick up some positive skewness. In other words, the distribution of the unobserved component of utility appears to be skewed to the right, or have a long right tail, having relatively few high values and having more mass concentrated at the left of the distribution. Recalling that this is the distribution of $\eta_t = \varepsilon_t(1) - \varepsilon_t(0)$, this implies that large shocks to utility favoring engine replacement are *less* likely to occur than would be predicted by the assumption that $(\varepsilon_t(0), \varepsilon_t(1))$ has an extreme value type I distribution. Table 2 displays mild differences in kurtosis as well.

Observe, however, that the p value from the likelihood ratio (LR) test would not allow one to reject the extreme value type I (or here, the logistic) assumption at conventional (5%) levels. This suggests that the extreme value assumption may be appropriate for this data set and model.

Only in the final panel of Table 2, bus groups 1–4, when comparing the logistic and EBurr3 distributions, does the LR test yield a marginally significant p value of 0.0725. If the hypothesis of extreme value errors could be rejected, it would be interesting

Table 2
Estimation with flexible distributions.

| | Bus groups 1–3 ($N = 3864$) | | | |
	Logistic	EBurr3	EBurr12	EGB2
RC	11.7176	12.1668	11.9838	12.5549
	(1.992)	(2.047)	(2.024)	(2.758)
θ_1	4.8191	2.9529	4.5258	2.6408
	(1.357)	(1.316)	(1.310)	(1.359)
Skewness	0	0.6077	0.0747	0.6982
		(0.272)	(0.104)	(0.279)
Kurtosis	4.2	4.3652	4.2439	4.5422
		(0.226)	(0.042)	(0.393)
Log-likelihood	−2708.366	−2707.865	−2708.321	−2707.851
LR test p value		0.3169	0.7623	0.5971
Hausman test p value		0.8390	0.9872	0.1506
	Bus group 4 ($N = 4292$)			
RC	10.0732	10.1709	10.1599	10.0567
	(1.378)	(1.510)	(1.366)	(2.465)
θ_1	2.2923	1.4652	2.2107	1.5443
	(0.574)	(1.067)	(0.606)	(1.637)
Skewness	0	0.3143	0.0278	0.2844
		(0.343)	(0.044)	(0.511)
Kurtosis	4.2	4.1631	4.2149	4.1273
		(0.184)	(0.024)	(0.631)
Log-likelihood	−3304.155	−3304.054	−3304.145	−3304.048
LR test p value		0.6530	0.8900	0.8988
Hausman test p value		0.7744	1.0000	0.5100
	Bus groups 1–4 ($N = 8156$)			
RC	9.7541	10.0752	9.9865	10.2604
	(0.776)	(0.909)	(0.788)	(1.103)
θ_1	2.6266	0.9671	2.3790	0.8456
	(0.488)	(0.530)	(0.494)	(0.607)
Skewness	0	0.6653	0.0761	0.7166
		(0.130)	(0.022)	(0.202)
Kurtosis	4.2	4.4341	4.2448	4.5361
		(0.158)	(0.014)	(0.355)
Log-likelihood	−6055.250	−6053.638	−6055.099	−6053.629
LR test p value		0.0725	0.5818	0.1975
Hausman test p value		0.9693	0.9995	0.6799

to identify the effects of this distributional assumption on the model. For this purpose, we examine more closely the estimated conditional choice probabilities for bus groups 1–4 under the logistic distribution and under EBurr3, shown in Fig. 1.
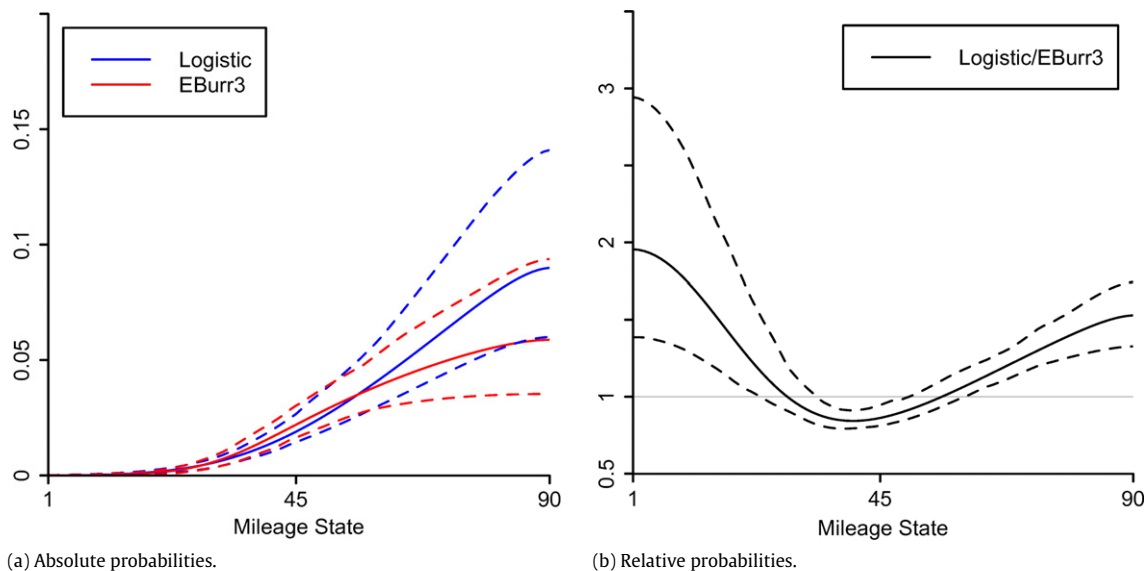

(a) Absolute probabilities.

(b) Relative probabilities.

Fig. 1. Conditional choice probabilities of engine replacement at each mileage state under the logistic versus the EBurr3 distribution (bus groups 1–4).

Observe in Fig. 1, panel (a), that at low mileage states, the two distributions predict similar absolute probabilities of engine replacement. At high mileage states (which are not reached as often in the data), the logistic distribution predicts a much higher likelihood of engine replacement than does EBurr3. This result may arise because engine replacement rarely occurs, and, being a tail occurrence, is predicted differently by the flexible EBurr3 distribution and by the logistic distribution. The plot of the relative probabilities in panel (b) demonstrates that at some mid-level mileage states, EBurr3 does predict a slightly higher replacement probability. Panel (b) also reveals that at many mileage states, particularly low and high mileage states, the 95% confidence interval of the relative probabilities (marked by the dashed lines) does not contain unity, meaning that the choice probabilities differ significantly from one another.

5. Conclusion

This work tests an assumption which is extremely common in dynamic (and static) discrete choice estimation: that the unobserved component of utility is distributed according to an extreme value type I distribution. This assumption is not made for theoretical reasons but rather because it yields closed-form solutions for choice probabilities and for expected continuation values. If the assumption is harmless, its usefulness makes it an excellent modeling choice. If, however, the model is quite sensitive to the distribution of unobserved shocks, researchers should be aware of this and consider these sensitivities.

Although the results of this work are only for one particular example, the procedure could be followed for other dynamic binary choice models. In particular, we demonstrate that without the closed-form solutions provided by the extreme value type I distribution, Gauss–Laguerre integration can be easily and accurately performed. Also, this work only considers a dynamic binary choice model, but similar approaches using numerical cubature rules could be applied to dynamic choice models with multiple discrete choices.

References

Aguirregabiria, V., Mira, P., 2010. Dynamic discrete choice structural models: a survey. Journal of Econometrics 156, 38–67.
Hansen, J.V., McDonald, J.B., 2002. A generalized model for predictive data mining. Information Systems Frontiers 4 (2), 179–186.
Hansen, C.B., McDonald, J.B., Theodossiou, P., 2007. Some flexible parametric models for partially adaptive estimators of econometric models. Economics 7, (The open access, open assessment e-journal).
Judd, K.L., Su, C.-L., 2011. Constrained optimization approaches to estimation of structural models. Working Paper, University of Chicago Booth School of Business.
Kythe, P., Schäferkotter, M., 2005. Handbook of Computational Methods for Integration, vol. 1. Chapman & Hall, CRC.
McDonald, J.B., Xu, Y.J., 1995. A generalization of the beta distribution with applications. Journal of Econometrics 66, 133–152.
McFadden, D., Train, K., 2000. Mixed MNL models for discrete response. Journal of Applied Econometrics 15 (5), 447–470.
Rust, J., 1987. Optimal replacement of GMC bus engines: an empirical model of Harold Zurcher. Econometrica 55 (5), 999–1033.