Technical Appendix

to

The Impact of the Americans with Disabilities Act on the Entry and Exit of Retail Firms

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

September 17, 2004

This appendix contains technical information on implementation of the estimation of the $CM_t/CM_t/\infty$ model.

1 Derivation of the likelihood

A sketch of the derivation of the likelihood of the $CM_t/CM_t/\infty$ model is given in the appendix of the paper. A more complete derivation is presented here. Begin with the penultimate equation in the appendix for the density of the number of firms conditional on lagged value and the mixing terms (u_t, v_t):

$$f\left(n_t|n_{t-1}, u_t, v_t\right) = \sum_{m=0}^{M_t} \exp\left[-\kappa_t\left(1-e^{-\mu_t}\right)\right]\tilde{B}_{mt}, \tag{A.1}$$

where $M_t \equiv \min\{n_{t-1}, n_t\}$, $\kappa_t = \lambda_t/\mu_t$, and

$$\tilde{B}_{mt} \equiv \begin{pmatrix} n_{t-1} \\ m \end{pmatrix} \frac{\kappa_t^{n_t-m}}{(n_t-m)!} e^{-\mu_t m}\left(1-e^{-\mu_t}\right)^{n_t+n_{t-1}-2m} \tag{A.2}$$

Note that each element of the sum in (A.1) is

$$\left\{\frac{\left[\kappa_t\left(1-e^{-\mu_t}\right)\right]^{n_t-m}}{(n_t-m)!}\exp\left[-\kappa_t\left(1-e^{-\mu_t}\right)\right]\right\}\left[\begin{pmatrix} n_{t-1} \\ m \end{pmatrix} e^{-\mu_t m}\left(1-e^{-\mu_t}\right)^{n_{t-1}-m}\right] \tag{A.3}$$

which has elements of the density of a Poisson random variable (although neither $f\left(n_t|n_{t-1}, u_t, v_t\right)$ nor $f\left(n_t|n_{t-1}, v_t\right)$ is in fact Poisson). In particular, the expression in the curly brackets is the Poisson probability of observing $n_t - m$ events with Poisson rate $\left(\kappa_t\left(1-e^{-\mu_t}\right)\right)$.

Finding $f(n_t|n_{t-1})$ requires integrating out the unobserved heterogeneity terms (u_t, v_t):

$$f\left(n_t|n_{t-1}\right) = E_{u,v}\left[f\left(n_t|n_{t-1}, u_t, v_t\right)\right] = E_v\left\{E_{u|v}\left[f\left(n_t|n_{t-1}, u_t, v_t\right)\right]\right\}$$

Begin with the inner expectation and integrate out u holding v fixed. Due to the assumption that $u|v$ has a gamma distribution the inner expectation may be found in closed form. The conditional expectation of each term (A.3) can be written

$$
\begin{aligned}
E_{u|v}\{\cdot\} &= \begin{pmatrix} n_{t-1} \\ m \end{pmatrix} e^{-\mu_t m}\left(1-e^{-\mu_t}\right)^{n_{t-1}-m} \\
&\cdot E_{u|v}\left\{\frac{\left(u_t\frac{\lambda_{0t}}{\mu_t}\left(1-e^{-\mu_t}\right)\right)^{n_t-m}}{(n_t-m)!}\exp\left[-u_t\frac{\lambda_{0t}}{\mu_t}\left(1-e^{-\mu_t}\right)\right]\right\}
\end{aligned}
\tag{A.4}
$$

The $E_{u|v}\{\cdot\}$ part is the same expression one gets when deriving the marginal distribution of a Poisson random variable with gamma mixing (here, u_t is the mixing term multiplying the Poisson rate $\lambda_{0t}\left(1-e^{-\mu_t}\right)/\mu_t$. The marginal distribution of a Poisson random variable with gamma mixing is negative binomial, and so

$$
\begin{aligned}
&E_{u|v}\left\{\frac{\left(u_t\frac{\lambda_{0t}}{\mu_t}\left(1-e^{-\mu_t}\right)\right)^{n_t-m}}{(n_t-m)!}\exp\left[-u_t\frac{\lambda_{0t}}{\mu_t}\left(1-e^{-\mu_t}\right)\right]\right\} = \\
&\qquad \frac{\Gamma\left(n_t-m+\gamma\right)}{\Gamma\left(\gamma\right)\Gamma\left(n_t-m+1\right)}\left(\frac{D_t}{D_t+1}\right)^{n_t-m}\left(\frac{\mu_t}{\lambda_{0t}\sigma_u^2 v^{\rho}\left(1-e^{-\mu_t}\right)+\mu_t}\right)^{\gamma}
\end{aligned}
$$

where $D_t \equiv \sigma_u^2 v_t^{\rho-1}\kappa_{0t}\left(1-e^{-\mu_t}\right)$, with $\kappa_0 = \lambda_0/\mu_0$. Add in the other terms from (A.4) and summing over m as in (A.1) leads to

$$
\begin{aligned}
f(n_t|n_{t-1},v_t) &= \left(\frac{\mu_t}{\lambda_{0t}\sigma_u^2 v_t^{\rho}\left(1-e^{-\mu_t}\right)+\mu_t}\right)^{\gamma} \\
&\cdot\sum_{m=0}^{M_t}\begin{pmatrix} n_{t-1} \\ m \end{pmatrix} e^{-\mu_t m}\left(1-e^{-\mu_t}\right)^{n_{t-1}-m}\left[\frac{\Gamma\left(n_t-m+\gamma\right)}{\Gamma\left(\gamma\right)\Gamma\left(n_t-m+1\right)}\left(\frac{D_t}{D_t+1}\right)^{n_t-m}\right]
\end{aligned}
$$

which can be written

$$
\begin{aligned}
f(n_t|n_{t-1},v_t) &= \left[\kappa_{0t}\sigma_u^2 v_t^{\tau-1}\left(1-e^{-\mu_t}\right)+1\right]^{-\gamma}\sum_{m=0}^{M_t} C_{mt}, \\
C_{mt} &\equiv B_{mt}\left[\frac{\Gamma\left(n_t-m+\gamma\right)}{\Gamma\left(\gamma\right)}\left(\frac{\sigma_u^2 v_t^{\tau-1}}{\kappa_{0t}\sigma_u^2 v_t^{\tau-1}\left(1-e^{-\mu_t}\right)+1}\right)^{n_t-m}\right], \\
B_{mt} &\equiv \tilde{B}_{mt}\left(\frac{v_t}{u_t}\right)^{n_t-m},
\end{aligned}
\tag{A.5}
$$

as in the main text. $\tilde{B}_{mt}$ in (A.5) is defined in (A.2).

Integrating v_t out of the likelihood to eliminate the final mixing term must be done numerically, because

$$f(n_t|n_{t-1}) = E_v f(n_t|n_{t-1}, v_t) = \int_0^\infty f(n_t|n_{t-1}, v_t)\mathcal{G}(\delta, \sigma_v^2; v)dv_t \tag{A.6}$$

does not have a closed form solution. Here $\mathcal{G}$ is the gamma pdf

$$\mathcal{G}(a, b; x) = \frac{x^{a-1}e^{-x/b}}{b^a\Gamma(a)}$$

2 Evaluation of the likelihood

The log likelihood is given in the main text, and is reproduced here:

$$l_{\boldsymbol{\theta}}\left(\boldsymbol{\theta}|n_{k0}, ((n_{kt}, \mathbf{X}_{kt}, \mathbf{Z}_{kt})_{t=1}^T)_{k=1}^K\right) = \sum_{k=1}^K\sum_{t=1}^T \log f\left(n_{kt}|n_{kt-1}\right)$$

where $f\left(n_{kt}|n_{kt-1}\right)$ is from (A.6), dropping the k subscripts, and

$$\lambda_t = \exp\left(\mathbf{X}_t'\boldsymbol{\alpha}\right)u_t = \lambda_{0t}u_t \tag{A.7}$$

$$\mu_t = \exp\left(\mathbf{Z}_t'\boldsymbol{\beta}\right)v_t = \mu_{0t}v_t$$

Restrictions

$$\begin{aligned} \delta &= \sigma_v^{-2} \\ \gamma &= \frac{\Gamma\left(\delta\right)}{\sigma_u^2\sigma_v^{2\tau}\Gamma\left(\tau+\delta\right)} \end{aligned}$$

are imposed for identification.

Evaluation of the likelihood requires computing the integral in (A.6). Although one could do simulated maximum likelihood, since the integral is unidimensional it is a good candidate for Gaussian quadrature. Direct numerical integration of well-behaved functions avoids simulation error at the cost of (usually much smaller) approximation error, and avoids the finite-sample bias that afflicts simulated maximum likelihood (see sec. 4.2 of (Hajivassiliou and Ruud, 1994), for example). Gaussian quadrature with R evaluation points is often much more precise than a simulation method (such as Monte Carlo integration) with R draws, because with quadrature one chooses the evaluation points to maximize accuracy as opposed to randomly drawing evaluation points in simulation methods.[1] To set up the quadrature, define

$$g(v_t) \equiv \frac{\Gamma\left(n_t - m + \gamma\right)}{\Gamma\left(\gamma\right)} B_{mt} \left[\kappa_{0t} w_t \left(1 - e^{-\mu_t}\right) + 1\right]^{-\gamma} \left(\frac{w_t}{\kappa_{0t} w_t \left(1 - e^{-\mu_t}\right) + 1}\right)^{n_t - m}$$

so that

$$f(n_t|n_{t-1}) = \sum_{m=0}^{M_t} E_{v_t} g(v_t)$$

Suppressed in the notation is that g is also a function of $\boldsymbol{\theta}$. Writing out the integral and the density, we have

$$f(n_t|n_{t-1}) = \sum_{m=0}^{M_t} \int_0^{\infty} g(v_t)\mathcal{G}(\delta, \sigma_v^2; v_t) = \sum_{m=0}^{M_t} \int_0^{\infty} g(v_t) \frac{v_t^{\delta-1} e^{-v_t/\sigma_v^2}}{\left(\sigma_v^2\right)^{\delta} \Gamma(\delta)} dv_t$$

Do change of variable $\xi = v_t/\sigma_v^2$. Then

$$f(n_t|n_{t-1}) = \sum_{m=0}^{M_t} \int_0^{\infty} g(\sigma_v^2 \xi) \frac{\xi^{\delta-1}}{\Gamma(\delta)} e^{-\xi} d\xi$$

Let $h(\xi) = g(\sigma_v^2 \xi)\xi^{\delta-1}/\Gamma(\delta)$. Then the integral is in the form

$$\sum_{m=0}^{M_t} \int_0^{\infty} h(\xi) e^{-\xi} d\xi \tag{A.8}$$

which is suited to Gauss-Laguerre quadrature, whence

$$f(n_t|n_{t-1}) \cong \sum_{m=0}^{M_t} \sum_{r=1}^{R} h(\xi_r) w_r$$

where the quadrature weights w_r and abscissae ξ_r are calculated using the FORTRAN Numerical Recipe `gaulag` (Press et al., 1992) with $R = 20$. One thing to note about calculation of the

[1] In particular, R point Gaussian quadrature is exact if the integrand is an polynomial of degree R. See Press, Teukolsky, Vetterling and Flannery (1992), sec.4.5. Monte Carlo methods compare increasingly favorably to deterministic methods as the dimension of the problem increases. Geweke and Keane (2001) state that quadrature is infeasible for problems higher than dimension three or four, although this threshold will increase as computing speeds increase. All integrals needed in the present work are unidimensional.

likelihood is that because of the sums over m in the expressions above (e.g. in (A.8)), calculation time increases with the mean of the dependent variable, not just with the number of observations. In practice, I found that when the average dependent variable (number of firms) was around 50, and 10 years $\times$ 3058 counties were observed, the estimations with full heterogeneity could take up to a week to run on a PC with an AMD (Intel clone) 1666 Mhz processor.

Doing as many calculations in logs as possible led to the most accurate results. In log form, $h(\xi)$ was calculated as

$$\log h(\xi) = \log g(\sigma_v^2 \xi) + (\delta - 1)\log \xi - \log \Gamma(\delta)$$

where

$$\begin{aligned} \log g(v) \quad = \quad & \log\left(\Gamma\left(n_t - m + \gamma\right)\right) - \log\left(\Gamma\left(\gamma\right)\right) + \log\left(\Gamma\left(n_{t-1} + 1\right)\right) \\ & - \log\left(\Gamma\left(m + 1\right)\right) - \log\left(\Gamma\left(n_{t-1} - m + 1\right)\right) - \log\left(\Gamma\left(n_t - m + 1\right)\right) \\ & + \left(n_t - m\right)\left(\log \kappa_{0t} + \log w_t\right) + \left(n_t + n_{t-1} - 2m\right)\log\left(1 - e^{-\mu_t}\right) \\ & - \mu_t m - \left(\gamma + n_t - m\right)\left[\log\left(\kappa_{0t} w_t\left(1 - e^{-\mu_t}\right) + 1\right)\right] \end{aligned}$$

Note that in the above expression, several of the Γ function arguments are integers, so they may be computed as factorials. Also, $\log \kappa_{0t}$ and $\log w_t$ can be simplified as

$$\log \kappa_{0t} = \log \lambda_{0t} - \log \mu_{0t} = \mathbf{X}_t' \boldsymbol{\alpha} - \mathbf{Z}_t' \boldsymbol{\beta}$$

$$\log w_t = \log \sigma_u^2 + (\rho - 1)\log v$$

3 Evaluation of the Derivatives

The derivatives are complicated enough that a good case could be made for numerical integration. I used analytical derivatives instead, which took longer to program but resulted in quicker estimation.

3.1 Derivatives with Heterogeneity Terms Excluded

When $\sigma_u^2 = \sigma_v^2 = 0$, the log likelihood is:

$$\ln L_{kt} = \log \sum_{m=0}^{M_{kt}} \exp\left(-\kappa_{kt}\left(1 - e^{-\mu_{kt}}\right)\right) B_{mkt}$$

where B_{mkt} is B_{mt} from above with the k subscript made explicit, so that

$$\begin{aligned}\log B_{mkt} &= \log\left(\Gamma\left(n_{kt-1}+1\right)\right) - \log\left(\Gamma\left(m+1\right)\right) - \log\left(\Gamma\left(n_{kt-1}-m+1\right)\right)\\ &\quad - \log\left(\Gamma\left(n_{kt}-m+1\right)\right) + \left(n_{kt}-m\right)\log\kappa_{0kt}\\ &\quad -\mu_{0kt}m + \left(n_{kt}+n_{kt-1}-2m\right)\log\left(1-e^{-\mu_{kt}}\right)\end{aligned}$$

Letting $\boldsymbol{\theta}$ be the parameter vector as above, the gradient of the log likelihood function is

$$\nabla_{\boldsymbol{\theta}} \ln L = \nabla_{\boldsymbol{\theta}} \sum_t \sum_k \ln L_{kt} = \sum_t \sum_k \frac{\sum_{m=0}^{M_{kt}} \nabla_{\boldsymbol{\theta}}\left[\exp\left(-\kappa_{kt}\left(1-e^{-\mu_{kt}}\right)\right) B_{mkt}\right]}{\exp\left(-\kappa_{kt}\left(1-e^{-\mu_{kt}}\right)\right)\sum_{m=0}^{M_{kt}} B_{mkt}}$$

Note that the denominator, common to all derivatives, is $f(n_{kt}|n_{kt-1})$ which is already calculated by the time the MLE algorithm requires the derivatives. The interior of the denominator is best found from the formula $\nabla_{\boldsymbol{\theta}} f_{kt} = f_{kt} \cdot \nabla_{\boldsymbol{\theta}} \log f_{kt}$. Here that is:

$$\begin{aligned}&\nabla_{\boldsymbol{\theta}}\left[\exp\left(-\kappa_{kt}\left(1-e^{-\mu_{kt}}\right)\right) B_{mkt}\right]\\ &\qquad = \left[\exp\left(-\kappa_{kt}\left(1-e^{-\mu_{kt}}\right)\right) B_{mkt}\right] \cdot \nabla_{\boldsymbol{\theta}}\left[-\kappa_{kt}\left(1-e^{-\mu_{kt}}\right) + \log B_{mkt}\right]\end{aligned}$$

3.2 Derivatives with Heterogeneity Terms Included

When $\sigma_u^2 = \sigma_v^2$ are non-zero, the gradient of the log likelihood function is

$$\nabla_{\boldsymbol{\theta}} \sum_t \sum_k \log\left(\sum_{m=0}^{M_t} \int_0^\infty h_{ktm}(\xi) e^{-\xi} d\xi\right) = \sum_t \sum_k \frac{\sum_{m=0}^{M_t} \int_0^\infty \left[\nabla_{\boldsymbol{\theta}} h_{ktm}(\xi)\right] e^{-\xi} d\xi}{\sum_{m=0}^{M_t} \int_0^\infty h_{ktm}(\xi) e^{-\xi} d\xi} \tag{A.9}$$

As above, the denominator is $f(n_{kt}|n_{kt-1})$ and need not be calculated again. For the numerator, define $d(\xi) \equiv \nabla_{\boldsymbol{\theta}} h(\xi)$. These derivatives of h were calculated using the analytical derivative feature of TSP, which prints results (conveniently) in FORTRAN-like syntax. The results are not reproduced here due to their length, but are available in the computer code. Once $d(\xi)$ is calculated, the numerator of (A.9) is approximated as $\sum_{m=0}^{M_t} \sum_{r=1}^{R} d(\xi_r) w_r$ with the same quadrature ascissae and weights as before.

4 The MLE Routine

Because the Hessian of the likelihood is complicated and expensive to calculate, maximization techniques and variance estimators that require only the gradient are an appealing choice here. I use the BFGS variant of the DFP algorithm in the application. The routine is a modified version of the Numerical Recipe `dfpmin` (Press et al., 1992). The convergence criterion was that each element of the gradient had to be less than 0.00001 and that the scaled step for each parameter element $\min\{|\Delta\theta_i|, |\Delta\theta_i/\theta_i|\}$, $i = 1, \ldots, \dim(\boldsymbol{\theta})$, was less than 0.001.

5 The Variance of the Estimates

Given the possibility of serial correlation among (u_{kt}, v_{kt}) within a county the clustered version of the Huber/White sandwich estimator of the variance is used (Wooldridge, 2001, sec.13.8.2). In particular, the asymptotic variance of $\boldsymbol{\theta}$ is calculated as

$$\hat{V} \left(\sum_{k=1}^{K} m_k' m_k \right) \hat{V}$$

where $\hat{V}$ is the usual BHHH estimate of the variance and m_k is the contribution of the kth county to the score:

$$m_k = \nabla_{\boldsymbol{\theta}} \sum_{t=1}^{T} \ln L_{kt}$$

References

Geweke, John and Keane, Michael (2001), Computationally Internsive Methods for Integration in Econometrics, *in* J. Heckman and E. Leamer, eds, 'Handbook of Econometrics, Volume V', Elsevier Science, Amsterdam, chapter 56, pp. 3463–3568.

Hajivassiliou, Vassilis A. and Ruud, Paul A. (1994), Classical Estimation Methods for LDV Models Using Simulation, *in* R. F. Engle and D. L. McFadden, eds, 'Handbook of Econometrics, Volume IV', Elsevier Science, Amsterdam, chapter 40, pp. 2384–2443.

Press, William H., Teukolsky, Saul A., Vetterling, William T. and Flannery, Brian P. (1992), *Numerical Recipes in FORTRAN: The Art of Scientific Computing*, 2nd edn, Cambridge Univ. Press, Cambridge.

Wooldridge, Jeffrey M. (2001), *Econometric Analysis of Cross Section and Panel Data*, MIT Press, Cambridge, MA.