Journal of Econometrics 149 (2009) 136–148

Contents lists available at ScienceDirect

Journal of Econometrics

journal homepage: www.elsevier.com/locate/jeconom

Bayesian analysis of random coefficient logit models using aggregate data

Renna Jiang [a], Puneet Manchanda [b], Peter E. Rossi [a,*]

[a] *University of Chicago, United States*
[b] *University of Michigan, United States*

ARTICLE INFO

Article history:
Received 20 August 2007
Received in revised form
17 November 2008
Accepted 2 December 2008
Available online 3 January 2009

JEL classification:
C11
M3

Keywords:
Random coefficient logit
Aggregate share models
Bayesian analysis

ABSTRACT

We present a Bayesian approach for analyzing aggregate level sales data in a market with differentiated products. We consider the aggregate share model proposed by Berry et al. [Berry, Steven, Levinsohn, James, Pakes, Ariel, 1995. Automobile prices in market equilibrium. Econometrica. 63 (4), 841–890], which introduces a common demand shock into an aggregated random coefficient logit model. A full likelihood approach is possible with a specification of the distribution of the common demand shock. We introduce a reparameterization of the covariance matrix to improve the performance of the random walk Metropolis for covariance parameters. We illustrate the usefulness of our approach with both actual and simulated data. Sampling experiments show that our approach performs well relative to the GMM estimator even in the presence of a mis-specified shock distribution. We view our approach as useful for those who are willing to trade off one additional distributional assumption for increased efficiency in estimation.

1. Introduction

Empirical researchers often build demand models using aggregate level sales data since individual level data are not always available. Berry et al. (1995) (hereafter BLP) introduced a particularly appealing formulation in which a common demand shock is introduced into a random coefficient logit model to provide a coherent aggregate demand specification. This aggregate share model uses a logit specification at the individual level coupled with a normal distribution of parameters over individuals. A large and growing body of research employs the generalized method of moments (GMM) technique due to Berry (1994) to estimate such models with aggregate data (see, for example, Chintagunta et al. (2003), Davis (2006), Goldfarb et al. (2005), Nevo (2000, 2001), Sudhir (2001) and Villas-Boas (2004)).

GMM estimators do not require distributional assumptions regarding the common demand shock. Our approach is to make one further distributional assumption concerning the common demand shock and derive the likelihood. Our model uses a normal distribution for the common demand shock. The resulting likelihood for aggregate share data is not in any closed form and may be quite irregular. Instead of relying on estimation procedures that require maximization, we consider Bayesian Markov Chain Monte Carlo (MCMC) methods that do not require a regular (or even a smooth) criterion function.

We apply our new Bayesian approach to both simulated and actual datasets. Our approach is relatively insensitive to simulation error in the estimates of integral terms in the density and Jacobian. This stands in marked contrast to the GMM approach. We conduct sampling experiments in which our Bayes estimator is shown to have lower mean squared error (MSE) than the GMM estimator. The GMM method is based on a model with a tightly specified logit demand at the individual level and a normal distribution of heterogeneity but without distributional assumptions regarding the common demand shock. One might argue that the improved performance of the Bayes estimator is due to the fact that an additional distribution assumption is used in formulating the likelihood function. Simulations with different shock distributions and violations of the *i.i.d.* assumption show that the Bayes estimator still performs well relative to the GMM estimator. This suggests that the reason for the improved sampling performance of the Bayes estimator is that it makes more efficient use of the data.

An additional benefit of the Bayesian approach is the ability to conduct inference for model parameters and functions of model parameters. A natural by-product of our MCMC simulation-based method is a way of constructing posterior distributions for any function of the model parameters. Indeed, it is possible to argue that price elasticities are a much more natural summary of the model parameters than the point estimates of utility weights and

* Corresponding address: University of Chicago, Graduate School of Business, 5807 S. Woodlawn Ave., 60637 Chicago, IL, United States. Tel.: +1 773 702 7513; fax: +1 773 834 2081.
E-mail address: peter.rossi@chicagogsb.edu (P.E. Rossi).

0304-4076/$ – see front matter
doi:10.1016/j.jeconom.2008.12.010

the covariance matrix of the random coefficient distribution. In contrast, the computation of asymptotic standard errors of the nonlinear functions of parameter estimates is not a natural by-product of the GMM inference procedure. For example, under the GMM framework, some researchers have used bootstrap methods to obtain standard errors of price elasticity, price–cost margin or other various quantity of interest (Nevo, 2001; Goldfarb et al., 2005). In the GMM framework, standard errors for these functions of model parameters require supplemental computations outside of the estimation algorithm. The Bayesian MCMC approach delivers the necessary computations in one unified computational framework.

We explore the adequacy of the asymptotic standard errors obtained by the GMM procedure. We find that these asymptotic standard errors understate the true sampling variance and result in confidence intervals with coverage much lower than the nominal level.

There is a literature on Bayesian approaches to estimation in aggregate share models. The existing Bayesian approaches use a data augmentation idea in which the parameters of "pseudo" consumers are added to the model. These pseudo consumer parameters are not used to integrate over the random coefficient distribution but are added to the set of parameters used in inference. Obviously, these parameters are not of interest, in and of themselves, but are merely a device to facilitate estimation. In standard data augmentation applications, the posterior, with these augmented parameters integrated out, is the same as the posterior from the likelihood function without the parameters. In the augmentation approaches for aggregate share models, this is not true. That is, the model specifies an infinite number of consumers and, thus, the existing data augmentation approaches can only approximate the correct likelihood-based inference. Moreover, this approach is limited to, at most, several hundred "pseudo" consumers in augmentation.

Chen and Yang (2007) propose a model without the common demand shock. Musalem et al. (in press) consider a model with an aggregate demand shock and an improved algorithm that uses the augmentation idea. Musalem et al. state that their method is a valid approximation method; that is, they show that, as the number of pseudo household "augmented" parameters goes to infinity, they can approximate the posterior of the model with a continuum of consumers. However, in practice, a finite and relatively small number of augmented parameters must be used. A simulation study would be required to ascertain the approximation properties of their approach.

Romeo (2007) is best viewed as a hybrid approach. He exponentiates the GMM criterion in the spirit of Chernozhukov and Hong (2005). Gallant and Hong (2007) argue that Romeo's pseudo-likelihood can be regarded as a likelihood. However, the main point is that the efficiency properties of this approach are likely to be very similar to the GMM estimator as the Romeo-style "likelihood" uses the same GMM criterion as studied here. The principal advantage of Romeo's approach is the use of a prior and the ability to simulate distributions of arbitrary functions of the model parameters (such as elasticities).

The remainder of the paper is organized as follows. Section 2 presents the main model. Section 3 outlines the MCMC algorithm. Section 4 discusses the computation of elasticities. We briefly review the GMM estimation procedure in Section 5. Section 6 conducts sampling experiments and also evaluates the adequacy of the asymptotic GMM standard errors. In Section 7, we provide an empirical example in which there is a material difference between the results obtained via the GMM and our Bayes procedure. We consider the extension to instrumental variables in Section 8 and conclude in Section 9.

2. Model

We assume that the latent indirect utility that a consumer i derives from consuming product j at time t takes the following standard form:

$$U_{ijt} = f\left(X_{jt}\left|\theta^i\right.\right) + \eta_{jt} + \varepsilon_{ijt} = X_{jt}\theta^i + \eta_{jt} + \varepsilon_{ijt} \tag{1}$$

where X_{jt} is a 1 by K vector that includes all the observed product attributes (e.g., brand intercepts and price P_{jt}). η_{jt} is the aggregate demand shock common across consumers/households (some interpret this as a time-varying unobserved product attribute). ε_{ijt} is an idiosyncratic shock that is distributed *i.i.d.* as a type I Extreme Value (0, 1). There are J products and an outside good, i.e., at any time t, a household has the option of not buying any of the J products. As is standard in the literature, we characterize the distribution of household preferences via a normal distribution, $\theta^i \sim N\left(\overline{\theta}, \Sigma\right)$.

The predicted share is then obtained by integrating s_{ijt} over the distribution of θ^i,

$$\begin{aligned} s_{jt} &= \int s_{ijt}\phi\left(\theta^i|\bar{\theta}, \Sigma\right) \mathrm{d}\theta^i \\ &= \int \frac{\exp\left(X_{jt}\theta^i + \eta_{jt}\right)}{1 + \sum_{k=1}^{J} \exp\left(X_{kt}\theta^i + \eta_{kt}\right)} \phi\left(\theta^i|\bar{\theta}, \Sigma\right) \mathrm{d}\theta^i \end{aligned} \tag{2}$$

where ϕ denotes the multivariate normal density. We can also write expected or predicted shares in terms of "mean utility" by using the identity $\theta^i = \bar{\theta} + v_i \quad v_i \sim N\left(\mathbf{0}, \Sigma\right)$.

$$s_{jt} = \int \frac{\exp\left(\mu_{jt} + X_{jt}v_i\right)}{1 + \sum_{k=1}^{J} \exp\left(\mu_{kt} + X_{kt}v_i\right)} \phi\left(v_i|\mathbf{0}, \Sigma\right) \mathrm{d}v$$

where $\mu_{jt} = X_{jt}\bar{\theta} + \eta_{jt}$. (2) shows that, at any time t, given the distribution of θ^i and observed covariates $X_t = \left(X'_{1t}, \ldots, X'_{Jt}\right)'$, share $s_t = (s_{1t}, \ldots, s_{Jt})'$ is only a function of the aggregate demand shock $\eta_t = (\eta_{1t}, \ldots, \eta_{Jt})'$. That is, aggregate shares inherit randomness solely from the aggregate demand shocks. We can therefore write the density of shares as a function of the density of the aggregate demand shocks. We denote the relationship between s_{jt} and η_t by $h\left(\cdot\right)$ as follows:

$$s_{jt} = h\left(\eta_t|X_t, \overline{\theta}, \Sigma\right). \tag{3}$$

The model so far is identical to that in BLP. We add one additional assumption necessary to specify the likelihood. We assume that the common demand shocks are independently distributed across all products with identical variances, i.e. $\eta_{jt} \sim N\left(0, \tau^2\right)$. The joint density of shares at time t can be obtained using the Change-of-Variable Theorem as follows:

$$\begin{aligned} &\pi\left(s_{1t}, \ldots, s_{Jt}\left|X_t, \overline{\theta}, \Sigma, \tau^2\right.\right) \\ &\quad = \phi\left(h^{-1}\left(s_{1t}, \ldots, s_{Jt}\left|X_t, \overline{\theta}, \Sigma\right.\right)\left|\tau^2\right.\right) J_{(\eta_t \to s_t)} \\ &\quad = \phi\left(h^{-1}\left(s_{1t}, \ldots, s_{Jt}\left|X_t, \overline{\theta}, \Sigma\right.\right)\left|\tau^2\right.\right)\left(J_{(s_t \to \eta_t)}\right)^{-1}. \end{aligned} \tag{4}$$

The likelihood is given by

$$L\left(\overline{\theta}, \Sigma, \tau^2\right) = \prod_{t=1}^{T} \pi\left(s_t\left|X_t, \overline{\theta}, \Sigma, \tau^2\right.\right). \tag{5}$$

To evaluate the likelihood, we need to invert the h function in (3) and evaluate the Jacobian (J) in (4).