Understanding Choice Intensity: A Poisson Mixture Model with Logit-based Random Utility Selective Mixing

Martin Burda — *University of Toronto*
Matthew Harding — *Stanford University*
Jerry Hausman — *MIT*

August 2010

Overview

- Model: New flexible mixed model for *count* data multinomial discrete choice, endogenizing count intensities
  - Key parameters interest: $\beta \sim F(\beta)$, flexible distribution
  - Other coefficients: $\theta, \gamma \sim MVN(b, \Sigma)$

- Application: supermarket choices of a panel of Houston households in 2004-2005, scanner data (Burda, Harding and Hausman 2008)
  - β_i : price, distance, their interaction
  - θ_i : store indicator variables
  - γ : demographic individual characteristics

- Estimation: Bayesian MCMC with a trivariate Dirichlet Process prior
  - Non-conjugate latent class sampling

Overview

- Model: New flexible mixed model for *count* data multinomial discrete choice, endogenizing count intensities
  - Key parameters interest: $\beta \sim F(\beta)$, flexible distribution
  - Other coefficients: $\theta, \gamma \sim MVN(b, \Sigma)$
- Application: supermarket choices of a panel of Houston households in 2004-2005, scanner data (Burda, Harding and Hausman 2008)
  - β_i : price, distance, their interaction
  - θ_i : store indicator variables
  - γ : demographic individual characteristics
- Estimation: Bayesian MCMC with a trivariate Dirichlet Process prior
  - Non-conjugate latent class sampling

Overview

- Model: New flexible mixed model for *count* data multinomial discrete choice, endogenizing count intensities
  - Key parameters interest: $\beta \sim F(\beta)$, flexible distribution
  - Other coefficients: $\theta, \gamma \sim MVN(b, \Sigma)$
- Application: supermarket choices of a panel of Houston households in 2004-2005, scanner data (Burda, Harding and Hausman 2008)
  - β_i : price, distance, their interaction
  - θ_i : store indicator variables
  - γ : demographic individual characteristics
- Estimation: Bayesian MCMC with a trivariate Dirichlet Process prior
  - Non-conjugate latent class sampling

Outline

Outline

1. Motivation
   1. Background on Count Data Models
   2. Continuous-time Poisson Process
2. Model
   1. Potential Continuous-time Utility
   2. Linking Utility and Count Intensity
   3. Count Probabilities in a new Mixed Poisson Model
   4. Efficient Likelihood Evaluation Algorithm
3. Bayesian Analysis
   1. Parametric vs Nonparametric Model
   2. Dirichlet Process Prior
4. Application
   1. Data and Variables
   2. Results
5. Counterfactual Welfare Experiment

Outline

Outline

Outline

1. Motivation
   1. Background on Count Data Models
   2. Continuous-time Poisson Process
2. Model
   1. Potential Continuous-time Utility
   2. Linking Utility and Count Intensity
   3. Count Probabilities in a new Mixed Poisson Model
   4. Efficient Likelihood Evaluation Algorithm
3. Bayesian Analysis
   1. Parametric vs Nonparametric Model
   2. Dirichlet Process Prior
4. Application
   1. Data and Variables
   2. Results
5. Counterfactual Welfare Experiment

Background: Popular Count Data Models

- Base-case Poisson:

$$f(y=k)=\frac{\exp(-\lambda)\,\lambda^k}{k!};\qquad \lambda=\exp(X\beta)$$

- Mixed Poisson:

$$f(y=k)=\int_0^\infty \frac{\exp(-\lambda)\,\lambda^k}{k!}g(\lambda)d\lambda$$

- Negative Binomial: special case with $\lambda \sim gamma(\delta,\delta)$ (Hausman, Hall, and Griliches 1984)

Background: Popular Count Data Models

- Base-case Poisson:

$$f(y = k) = \frac{\exp(-\lambda)\lambda^k}{k!}; \qquad \lambda = \exp(X\beta)$$

- Mixed Poisson:

$$f(y = k) = \int_0^\infty \frac{\exp(-\lambda)\lambda^k}{k!} g(\lambda) d\lambda$$

- Negative Binomial: special case with $\lambda \sim gamma(\delta, \delta)$ (Hausman, Hall, and Griliches 1984)

Background: Popular Count Data Models

- Base-case Poisson:

$$f(y=k)=\frac{\exp(-\lambda)\lambda^{k}}{k!};\qquad \lambda=\exp(X\beta)$$

- Mixed Poisson:

$$f(y=k)=\int_{0}^{\infty}\frac{\exp(-\lambda)\lambda^{k}}{k!}g(\lambda)d\lambda$$

- Negative Binomial: special case with $\lambda\sim gamma(\delta,\delta)$
(Hausman, Hall, and Griliches 1984)

Background: Limits of a Continuous-time Poisson Process

- The probability of a unit addition to the count process $Y(t)$ within the interval Δ is given by

$$P\{Y(t+\Delta) - Y(t) = 1\} = \lambda\Delta + o(\Delta)$$

- Allow for evolution of λ over time to obtain the count process intensity $\widetilde{\lambda}(t)$:

$$P\{Y(t+\Delta) - Y(t) = 1\} = \widetilde{\lambda}(t)\Delta + o(\Delta)$$

- By the Poisson independence assumption, obtain the integrated intensity

$$\lambda(t) = \int_0^t \widetilde{\lambda}(s)ds$$

yielding p.m.f. equivalent to the base-case Poisson.

Background: Limits of a Continuous-time Poisson Process

- The probability of a unit addition to the count process $Y(t)$ within the interval Δ is given by

$$P\{Y(t+\Delta) - Y(t) = 1\} = \lambda\Delta + o(\Delta)$$

- Allow for evolution of λ over time to obtain the count process intensity $\widetilde{\lambda}(t)$:

$$P\{Y(t+\Delta) - Y(t) = 1\} = \widetilde{\lambda}(t)\Delta + o(\Delta)$$

- By the Poisson independence assumption, obtain the integrated intensity

$$\lambda(t) = \int_0^t \widetilde{\lambda}(s)ds$$

yielding p.m.f. equivalent to the base-case Poisson.

Background: Limits of a Continuous-time Poisson Process

- The probability of a unit addition to the count process $Y(t)$ within the interval Δ is given by

$$P\{Y(t+\Delta)-Y(t)=1\}=\lambda\Delta+o(\Delta)$$

- Allow for evolution of λ over time to obtain the count process intensity $\widetilde{\lambda}(t)$:

$$P\{Y(t+\Delta)-Y(t)=1\}=\widetilde{\lambda}(t)\Delta+o(\Delta)$$

- By the Poisson independence assumption, obtain the integrated intensity

$$\lambda(t)=\int_0^t \widetilde{\lambda}(s)ds$$

yielding p.m.f. equivalent to the base-case Poisson.

Background: Sub-divisibility of the Poisson pmf

- The p.m.f. of a Poisson count variable Y whose counts y_s are observed on time intervals $(a_s, b_s]$ for $s = 1, \ldots, T$ with $a_s < b_s \leq a_{s+1} < b_{s+1}$ is given by

$$P(\{Y_s = y_s\}_{s=1}^{T}) = \prod_{s=1}^{T} \frac{\exp\left(-\lambda(b_s - a_s)\right)\left[\lambda(b_s - a_s)\right]^{y_s}}{y_s!}$$

Outline

1. Motivation
   1. Background on Count Data Models
   2. Continuous-time Poisson Process
2. Model
   1. Potential Continuous-time Utility
   2. Linking Utility and Count Intensity
   3. Count Probabilities in a new Mixed Poisson Model
   4. Efficient Likelihood Evaluation Algorithm
3. Bayesian Analysis
   1. Parametric vs Nonparametric Model
   2. Dirichlet Process Prior
4. Application
   1. Data and Variables
   2. Results
5. Counterfactual Welfare Experiment

Potential Continuous-time Utility

- Application: household choice of supermarket chain and count of monthly trips
- Continuous-time joint decision process on store selection and trip count intensity
- Latent continuous-time potential utility of an individual i at time instant $\tau \in (t-1, t]$ derived from the alternative j :

$$\widetilde{U}_{itj}(\tau) = \widetilde{\beta}_i' X_{itj}(\tau) + \widetilde{\theta}_i' D_{itj}(\tau) + \widetilde{\varepsilon}_{itj}(\tau)$$

- X_{itj} - key variables of interest (price, distance, and their interaction)
- D_{itj} - store indicator variables
- $j \in \{1, \ldots, J\}$ - store alternatives
- $\widetilde{\varepsilon}_{itj}$ - disturbance with extreme value type 1 marginal density

Potential Continuous-time Utility

- Application: household choice of supermarket chain and count of monthly trips
- Continuous-time joint decision process on store selection and trip count intensity
- Latent continuous-time potential utility of an individual i at time instant $\tau \in (t-1, t]$ derived from the alternative j :

$$\widetilde{U}_{itj}(\tau) = \widetilde{\beta}_i' X_{itj}(\tau) + \widetilde{\theta}_i' D_{itj}(\tau) + \widetilde{\varepsilon}_{itj}(\tau)$$

- X_{itj} - key variables of interest (price, distance, and their interaction)
- D_{itj} - store indicator variables
- $j \in \{1, \ldots, J\}$ - store alternatives
- $\widetilde{\varepsilon}_{itj}$ - disturbance with extreme value type 1 marginal density

Linking Utility and Count Intensity

- Denote the potential utility of the preferred choice (subscript c) by

$$\widetilde{U}_{itc}(\tau) = \max_{j \in \mathcal{J}} \left\{ \widetilde{U}_{itj}(\tau) \right\}$$

- The trip count intensity $\widetilde{\lambda}_{itc}(\tau)$ is linked by

$$\begin{aligned} \widetilde{\lambda}_{itc}(\tau) &= h(\widetilde{U}_{itc}(\tau)) \\ &= \gamma' Z_{it}(\tau) + \omega_{1i}\widetilde{\beta}_i' X_{itc}(\tau) + \omega_{2i}\widetilde{\theta}_i' D_{itc}(\tau) + \omega_{3i}\widetilde{\varepsilon}_{itc}(\tau) \\ &= \gamma' Z_{it}(\tau) + \beta_i' X_{itc}(\tau) + \theta_i' D_{itc}(\tau) + \varepsilon_{itc}(\tau) \end{aligned}$$

  for $\widetilde{\lambda}_{itc}(\tau) \geq 0$.
- Higher $\varepsilon_{itj}(\tau)$ increases the probability of additional trip via increased count intensity $\widetilde{\lambda}_{itj}(\tau)$
- Proportionality factors ω_{1i}, ω_{2i}, and ω_{3i} do not need to be separately identified

Linking Utility and Count Intensity

- Denote the potential utility of the preferred choice (subscript c) by

$$\widetilde{U}_{itc}(\tau) = \max_{j \in \mathcal{J}} \left\{ \widetilde{U}_{itj}(\tau) \right\}$$

- The trip count intensity $\widetilde{\lambda}_{itc}(\tau)$ is linked by

$$\begin{aligned} \widetilde{\lambda}_{itc}(\tau) &= h(\widetilde{U}_{itc}(\tau)) \\ &= \gamma' Z_{it}(\tau) + \omega_{1i}\widetilde{\beta}_i' X_{itc}(\tau) + \omega_{2i}\widetilde{\theta}_i' D_{itc}(\tau) + \omega_{3i}\widetilde{\varepsilon}_{itc}(\tau) \\ &= \gamma' Z_{it}(\tau) + \beta_i' X_{itc}(\tau) + \theta_i' D_{itc}(\tau) + \varepsilon_{itc}(\tau) \end{aligned}$$

for $\widetilde{\lambda}_{itc}(\tau) \geq 0$.

- Higher $\varepsilon_{itj}(\tau)$ increases the probability of additional trip via increased count intensity $\widetilde{\lambda}_{itj}(\tau)$
- Proportionality factors ω_{1i}, ω_{2i}, and ω_{3i} do not need to be separately identified

Linking Utility and Count Intensity

- Denote the potential utility of the preferred choice (subscript c) by

$$\widetilde{U}_{itc}(\tau) = \max_{j \in \mathcal{J}} \left\{ \widetilde{U}_{itj}(\tau) \right\}$$

- The trip count intensity $\widetilde{\lambda}_{itc}(\tau)$ is linked by

$$\begin{aligned} \widetilde{\lambda}_{itc}(\tau) &= h(\widetilde{U}_{itc}(\tau)) \\ &= \gamma' Z_{it}(\tau) + \omega_{1i}\widetilde{\beta}_i' X_{itc}(\tau) + \omega_{2i}\widetilde{\theta}_i' D_{itc}(\tau) + \omega_{3i}\widetilde{\varepsilon}_{itc}(\tau) \\ &= \gamma' Z_{it}(\tau) + \beta_i' X_{itc}(\tau) + \theta_i' D_{itc}(\tau) + \varepsilon_{itc}(\tau) \end{aligned}$$

  for $\widetilde{\lambda}_{itc}(\tau) \geq 0$.

- Higher $\varepsilon_{itj}(\tau)$ increases the probability of additional trip via increased count intensity $\widetilde{\lambda}_{itj}(\tau)$
- Proportionality factors ω_{1i}, ω_{2i}, and ω_{3i} do not need to be separately identified

Linking Utility and Count Intensity

- Denote the potential utility of the preferred choice (subscript c) by

$$\widetilde{U}_{itc}(\tau) = \max_{j \in \mathcal{J}} \left\{ \widetilde{U}_{itj}(\tau) \right\}$$

- The trip count intensity $\widetilde{\lambda}_{itc}(\tau)$ is linked by

$$\begin{aligned} \widetilde{\lambda}_{itc}(\tau) &= h(\widetilde{U}_{itc}(\tau)) \\ &= \gamma' Z_{it}(\tau) + \omega_{1i} \widetilde{\beta}_i' X_{itc}(\tau) + \omega_{2i} \widetilde{\theta}_i' D_{itc}(\tau) + \omega_{3i} \widetilde{\varepsilon}_{itc}(\tau) \\ &= \gamma' Z_{it}(\tau) + \beta_i' X_{itc}(\tau) + \theta_i' D_{itc}(\tau) + \varepsilon_{itc}(\tau) \end{aligned}$$

  for $\widetilde{\lambda}_{itc}(\tau) \geq 0$.
- Higher $\varepsilon_{itj}(\tau)$ increases the probability of additional trip via increased count intensity $\widetilde{\lambda}_{itj}(\tau)$
- Proportionality factors ω_{1i}, ω_{2i}, and ω_{3i} do not need to be separately identified

Integrated Count Intensity for Discrete Data

- For discrete y_{it} the realizations of $\widetilde{U}_{itj}(\tau)$ for $\tau \in (t-1, t]$ are given by

$$\widetilde{U}_{itjk} = \widetilde{\beta}_i' X_{itjk} + \widetilde{\theta}_i' D_{itjk} + \widetilde{\varepsilon}_{itjk}$$

- Hence the integrated count intensity

$$\lambda_{itc} = \int_{t-1}^{t} h(\widetilde{U}_{itc}(\tau)) d\tau$$

- Let

$$\begin{aligned} \lambda_{itck} &= \max\{0, \lambda_{itck}^*\} \\ \lambda_{itck}^* &= \gamma' Z_{it} + \beta_i' X_{itck} + \theta_{ic} D_{itck} + \varepsilon_{itck} \end{aligned}$$

and approximate the intensity integral by

$$\begin{aligned} \lambda_{itc} &= \frac{1}{y_{itc}} \sum_{k=1}^{y_{itc}} \lambda_{itck}^* \\ &= \gamma' Z_{it} + \beta_i' \overline{X}_{itc} + \theta_i \overline{D}_{itc} + \overline{\varepsilon}_{itc} \\ &= \overline{V}_{itc} + \overline{\varepsilon}_{itc} \end{aligned}$$

Integrated Count Intensity for Discrete Data

- For discrete y_{it} the realizations of $\widetilde{U}_{itj}(\tau)$ for $\tau \in (t-1, t]$ are given by

$$\widetilde{U}_{itjk} = \widetilde{\beta}_i' X_{itjk} + \widetilde{\theta}_i' D_{itjk} + \widetilde{\varepsilon}_{itjk}$$

- Hence the integrated count intensity

$$\lambda_{itc} = \int_{t-1}^{t} h(\widetilde{U}_{itc}(\tau)) d\tau$$

- Let

$$\begin{aligned} \lambda_{itck} &= \max\{0, \lambda_{itck}^*\} \\ \lambda_{itck}^* &= \gamma' Z_{it} + \beta_i' X_{itck} + \theta_{ic} D_{itck} + \varepsilon_{itck} \end{aligned}$$

and approximate the intensity integral by

$$\begin{aligned} \lambda_{itc} &= \frac{1}{y_{itc}} \sum_{k=1}^{y_{itc}} \lambda_{itck}^* \\ &= \gamma' Z_{it} + \beta_i' \overline{X}_{itc} + \theta_i \overline{D}_{itc} + \bar{\varepsilon}_{itc} \\ &= \overline{V}_{itc} + \bar{\varepsilon}_{itc} \end{aligned}$$

Integrated Count Intensity for Discrete Data

- For discrete y_{it} the realizations of $\widetilde{U}_{itj}(\tau)$ for $\tau \in (t-1, t]$ are given by

$$\widetilde{U}_{itjk} = \widetilde{\beta}_i' X_{itjk} + \widetilde{\theta}_i' D_{itjk} + \widetilde{\varepsilon}_{itjk}$$

- Hence the integrated count intensity

$$\lambda_{itc} = \int_{t-1}^{t} h(\widetilde{U}_{itc}(\tau)) d\tau$$

- Let

$$\begin{aligned} \lambda_{itck} &= \max\{0, \lambda_{itck}^*\} \\ \lambda_{itck}^* &= \gamma' Z_{it} + \beta_i' X_{itck} + \theta_{ic} D_{itck} + \varepsilon_{itck} \end{aligned}$$

and approximate the intensity integral by

$$\begin{aligned} \lambda_{itc} &= \frac{1}{y_{itc}} \sum_{k=1}^{y_{itc}} \lambda_{itck}^* \\ &= \gamma' Z_{it} + \beta_i' \overline{X}_{itc} + \theta_i \overline{D}_{itc} + \overline{\varepsilon}_{itc} \\ &= \overline{V}_{itc} + \overline{\varepsilon}_{itc} \end{aligned}$$

Count Probabilities

- Denote by δ_{itj} the fraction of time period t over which the alternative j was maximizing the latent utility $\widetilde{U}_{itj}(\tau)$ among other alternatives
- The assumption of extreme value type 1 distribution on the residual $\widetilde{\varepsilon}_{itjk}$ in

$$\begin{aligned}\widetilde{U}_{itjk} &= \widetilde{\beta}_i' X_{itjk} + \widetilde{\theta}_i' D_{itjk} + \widetilde{\varepsilon}_{itjk} \\ &= \widetilde{V}_{itc} + \widetilde{\varepsilon}_{itjk}\end{aligned}$$

yields

$$\delta_{itc} = \frac{\exp\left(\widetilde{V}_{itc}\right)}{\sum_{j=1}^{J} \exp\left(\widetilde{V}_{itj}\right)}$$

Count Probabilities

- Denote by δ_{itj} the fraction of time period t over which the alternative j was maximizing the latent utility $\widetilde{U}_{itj}(\tau)$ among other alternatives
- The assumption of extreme value type 1 distribution on the residual $\widetilde{\varepsilon}_{itjk}$ in

$$\begin{aligned}\widetilde{U}_{itjk} &= \widetilde{\beta}_i' X_{itjk} + \widetilde{\theta}_i' D_{itjk} + \widetilde{\varepsilon}_{itjk} \\ &= \widetilde{V}_{itc} + \widetilde{\varepsilon}_{itjk}\end{aligned}$$

yields

$$\delta_{itc} = \frac{\exp\left(\widetilde{V}_{itc}\right)}{\sum_{j=1}^{J} \exp\left(\widetilde{V}_{itj}\right)}$$

Count Probabilities

- The joint conditional trip count and store choice probability:

$$P(Y_{itc} = y_{itc}|\delta_{itc}) = \int \frac{\exp(-\delta_{itc}\lambda_{itc})(\delta_{itc}\lambda_{itc})^{y_{itc}}}{y_{itc}!} g(\lambda_{itc}) d(\lambda_{itc})$$

with

$$\lambda_{itc} \propto \bar{\varepsilon}_{itc} = \frac{1}{y_{itck}} \sum_{k=1}^{y_{itck}} \varepsilon_{itck}$$

- Each ε_{itck} represents an $J-$order statistic (maximum) of ε_{itjk} with mean V_{itjk} from utility maximization
- The density of $\bar{\varepsilon}_{itc}$ is the convolution of y_{itck} densities of $J-$order statistics (analytically intractable except for few special cases)

Count Probabilities

- The joint conditional trip count and store choice probability:

$$P(Y_{itc} = y_{itc} | \delta_{itc}) = \int \frac{\exp\left(-\delta_{itc}\lambda_{itc}\right)\left(\delta_{itc}\lambda_{itc}\right)^{y_{itc}}}{y_{itc}!} g(\lambda_{itc}) d(\lambda_{itc})$$

with

$$\lambda_{itc} \propto \bar{\varepsilon}_{itc} = \frac{1}{y_{itck}} \sum_{k=1}^{y_{itck}} \varepsilon_{itck}$$

- Each ε_{itck} represents an $J-$order statistic (maximum) of ε_{itjk} with mean V_{itjk} from utility maximization
- The density of $\bar{\varepsilon}_{itc}$ is the convolution of y_{itck} densities of $J-$order statistics (analytically intractable except for few special cases)

Count Probabilities

- The joint conditional trip count and store choice probability:

$$P(Y_{itc} = y_{itc} | \delta_{itc}) = \int \frac{\exp\left(-\delta_{itc}\lambda_{itc}\right)\left(\delta_{itc}\lambda_{itc}\right)^{y_{itc}}}{y_{itc}!} g(\lambda_{itc}) d(\lambda_{itc})$$

with

$$\lambda_{itc} \propto \overline{\varepsilon}_{itc} = \frac{1}{y_{itck}} \sum_{k=1}^{y_{itck}} \varepsilon_{itck}$$

- Each ε_{itck} represents an $J-$order statistic (maximum) of ε_{itjk} with mean V_{itjk} from utility maximization
- The density of $\overline{\varepsilon}_{itc}$ is the convolution of y_{itck} densities of $J-$order statistics (analytically intractable except for few special cases)

Likelihood Evaluation

- The joint count probability of the observed sample $y = \{y_{itc}\}$ is

$$P(Y = y) = \prod_{i=1}^{N} \prod_{t=1}^{T} \prod_{c=1}^{C_{it}} P(y_{itc}|\delta_{itc})$$

- Partition

$$P(y_{itc}|\delta_{itc}) = \int_{\mathcal{V}} \underbrace{\int_{\varepsilon} f(y_{it}|\overline{\varepsilon}_{itc}, \overline{V}_{itc}(\xi)) g(\overline{\varepsilon}_{itc}|\overline{V}_{itc}(\xi)) d\overline{\varepsilon}_{itc}}_{E_{\overline{\varepsilon}} f(y_{it}|\overline{V}_{itc}(\xi))} g(\overline{V}_{itc}(\xi)) d\overline{V}_{itc}$$

- Evaluate analytically

$$E_{\overline{\varepsilon}} f(y_{itc}|\overline{V}_{itc}) = \sum_{r=0}^{\infty} \frac{(-1)^r}{y_{it}! r!} \delta_{itc}^{r+y_{itc}} \underbrace{\eta'_{y_{it}+r}(\overline{\varepsilon}_{itc}; \overline{V}_{itc})}_{\text{uncentered moments of } \overline{\varepsilon}_{itc}}$$

- Obtain $\eta'_{y_{itc}+r}$ recursively from the cumulant-gen. function of $\overline{\varepsilon}_{itc}(s)$
- McFadden (1974) choice probabilities: η'_0
- Sample $\xi = (\gamma, \beta, \theta)$ using Bayesian data augmentation

Likelihood Evaluation

- The joint count probability of the observed sample $y = \{y_{itc}\}$ is

$$P(Y = y) = \prod_{i=1}^{N} \prod_{t=1}^{T} \prod_{c=1}^{C_{it}} P(y_{itc} | \delta_{itc})$$

- Partition

$$P(y_{itc}|\delta_{itc}) = \int_{\mathcal{V}} \underbrace{\int_{\varepsilon} f(y_{it}|\bar{\varepsilon}_{itc}, \overline{V}_{itc}(\xi)) g(\bar{\varepsilon}_{itc}|\overline{V}_{itc}(\xi)) d\bar{\varepsilon}_{itc}}_{E_{\bar{\varepsilon}} f(y_{it}|\overline{V}_{itc}(\xi))} g(\overline{V}_{itc}(\xi)) d\overline{V}_{itc}$$

- Evaluate analytically

$$E_{\bar{\varepsilon}} f(y_{itc}|\overline{V}_{itc}) = \sum_{r=0}^{\infty} \frac{(-1)^r}{y_{it}! r!} \delta_{itc}^{r+y_{itc}} \underbrace{\eta'_{y_{it}+r}(\bar{\varepsilon}_{itc}; \overline{V}_{itc})}_{\text{uncentered moments of } \bar{\varepsilon}_{itc}}$$

- Obtain $\eta'_{y_{itc}+r}$ recursively from the cumulant-gen. function of $\bar{\varepsilon}_{itc}(s)$
- McFadden (1974) choice probabilities: η'_0
- Sample $\xi = (\gamma, \beta, \theta)$ using Bayesian data augmentation

Likelihood Evaluation

- The joint count probability of the observed sample $y = \{y_{itc}\}$ is

$$P(Y = y) = \prod_{i=1}^{N} \prod_{t=1}^{T} \prod_{c=1}^{C_{it}} P(y_{itc} | \delta_{itc})$$

- Partition

$$P(y_{itc}|\delta_{itc}) = \int_{\mathcal{V}} \underbrace{\int_{\varepsilon} f(y_{it}|\overline{\varepsilon}_{itc}, \overline{V}_{itc}(\xi)) g(\overline{\varepsilon}_{itc}|\overline{V}_{itc}(\xi)) d\overline{\varepsilon}_{itc}}_{E_{\overline{\varepsilon}} f(y_{it}|\overline{V}_{itc}(\xi))} g(\overline{V}_{itc}(\xi)) d\overline{V}_{itc}$$

- Evaluate analytically

$$E_{\overline{\varepsilon}} f(y_{itc}|\overline{V}_{itc}) = \sum_{r=0}^{\infty} \frac{(-1)^r}{y_{it}!r!} \delta_{itc}^{r+y_{itc}} \underbrace{\eta'_{y_{it}+r}(\overline{\varepsilon}_{itc}; \overline{V}_{itc})}_{\text{uncentered moments of } \overline{\varepsilon}_{itc}}$$

- Obtain $\eta'_{y_{itc}+r}$ recursively from the cumulant-gen. function of $\overline{\varepsilon}_{itc}(s)$
- McFadden (1974) choice probabilities: η'_0
- Sample $\xi = (\gamma, \beta, \theta)$ using Bayesian data augmentation

Likelihood Evaluation

- The joint count probability of the observed sample $y = \{y_{itc}\}$ is

$$P(Y = y) = \prod_{i=1}^{N} \prod_{t=1}^{T} \prod_{c=1}^{C_{it}} P(y_{itc}|\delta_{itc})$$

- Partition

$$P(y_{itc}|\delta_{itc}) = \int_{\mathcal{V}} \underbrace{\int_{\varepsilon} f(y_{it}|\bar{\varepsilon}_{itc}, \overline{V}_{itc}(\xi)) g(\bar{\varepsilon}_{itc}|\overline{V}_{itc}(\xi)) d\bar{\varepsilon}_{itc}}_{E_{\bar{\varepsilon}} f(y_{it}|\overline{V}_{itc}(\xi))} g(\overline{V}_{itc}(\xi)) d\overline{V}_{itc}$$

- Evaluate analytically

$$E_{\bar{\varepsilon}} f(y_{itc}|\overline{V}_{itc}) = \sum_{r=0}^{\infty} \frac{(-1)^r}{y_{it}!r!} \delta_{itc}^{r+y_{itc}} \underbrace{\eta'_{y_{it}+r}(\bar{\varepsilon}_{itc}; \overline{V}_{itc})}_{\text{uncentered moments of } \bar{\varepsilon}_{itc}}$$

- Obtain $\eta'_{y_{itc}+r}$ recursively from the cumulant-gen. function of $\bar{\varepsilon}_{itc}(s)$
- McFadden (1974) choice probabilities: η'_0
- Sample $\xi \equiv (\gamma, \beta, \theta)$ using Bayesian data augmentation

Likelihood Evaluation

- The joint count probability of the observed sample $y = \{y_{itc}\}$ is

$$P(Y = y) = \prod_{i=1}^{N} \prod_{t=1}^{T} \prod_{c=1}^{C_{it}} P(y_{itc}|\delta_{itc})$$

- Partition

$$P(y_{itc}|\delta_{itc}) = \int_{\mathcal{V}} \underbrace{\int_{\varepsilon} f(y_{it}|\overline{\varepsilon}_{itc}, \overline{V}_{itc}(\xi)) g(\overline{\varepsilon}_{itc}|\overline{V}_{itc}(\xi)) d\overline{\varepsilon}_{itc}}_{E_{\overline{\varepsilon}} f(y_{it}|\overline{V}_{itc}(\xi))} g(\overline{V}_{itc}(\xi)) d\overline{V}_{itc}$$

- Evaluate analytically

$$E_{\overline{\varepsilon}} f(y_{itc}|\overline{V}_{itc}) = \sum_{r=0}^{\infty} \frac{(-1)^r}{y_{it}! r!} \delta_{itc}^{r+y_{itc}} \underbrace{\eta'_{y_{it}+r}(\overline{\varepsilon}_{itc}; \overline{V}_{itc})}_{\text{uncentered moments of } \overline{\varepsilon}_{itc}}$$

- Obtain $\eta'_{y_{itc}+r}$ recursively from the cumulant-gen. function of $\overline{\varepsilon}_{itc}(s)$
- McFadden (1974) choice probabilities: η'_0
- Sample $\xi = (\gamma, \beta, \theta)$ using Bayesian data augmentation

Likelihood Evaluation

- The joint count probability of the observed sample $y = \{y_{itc}\}$ is

$$P(Y = y) = \prod_{i=1}^{N} \prod_{t=1}^{T} \prod_{c=1}^{C_{it}} P(y_{itc}|\delta_{itc})$$

- Partition

$$P(y_{itc}|\delta_{itc}) = \int_{\mathcal{V}} \underbrace{\int_{\varepsilon} f(y_{it}|\overline{\varepsilon}_{itc}, \overline{V}_{itc}(\xi))g(\overline{\varepsilon}_{itc}|\overline{V}_{itc}(\xi))d\overline{\varepsilon}_{itc}}_{E_{\overline{\varepsilon}}f(y_{it}|\overline{V}_{itc}(\xi))} g(\overline{V}_{itc}(\xi))d\overline{V}_{itc}$$

- Evaluate analytically

$$E_{\overline{\varepsilon}}f(y_{itc}|\overline{V}_{itc}) = \sum_{r=0}^{\infty} \frac{(-1)^r}{y_{it}!r!} \delta_{itc}^{r+y_{itc}} \quad \underbrace{\eta'_{y_{it}+r}(\overline{\varepsilon}_{itc}; \overline{V}_{itc})}_{\text{uncentered moments of } \overline{\varepsilon}_{itc}}$$

- Obtain $\eta'_{y_{itc}+r}$ recursively from the cumulant-gen. function of $\overline{\varepsilon}_{itc}(s)$
- McFadden (1974) choice probabilities: η'_0
- Sample $\xi \equiv (\gamma, \beta, \theta)$ using Bayesian data augmentation

Recursive Updating: Example for $y_{it} = 4$

r	q	$p:1$	2	3	4	5	6	7	8
0	0	$\kappa_1(\xi)\widetilde{\eta}'_0$	$B_{4,0,0}\widetilde{\eta}'_0$	$B_{4,0,0}\widetilde{\eta}'_0$	$B_{4,0,0}\widetilde{\eta}'_0$	$\frac{1}{r_1}B_{4,1,0}\widetilde{\eta}'_0$	$\frac{1}{r_1}\frac{1}{r_2}B_{4,2,0}\widetilde{\eta}'_0$	$\frac{1}{r_1}\frac{1}{r_2}\frac{1}{r_3}B_{4,3,0}\widetilde{\eta}'_0$	$\frac{1}{r_1}\frac{1}{r_2}\frac{1}{r_3}\frac{1}{r_4}B_{4,4,0}\widetilde{\eta}'_0$
0	1	$=\widetilde{\eta}'_1$	$\kappa_1(\xi)\widetilde{\eta}'_1$	$B_{4,0,1}\widetilde{\eta}'_1$	$B_{4,0,1}\widetilde{\eta}'_1$	$\frac{1}{r_1}B_{4,1,1}\widetilde{\eta}'_1$	$\frac{1}{r_1}\frac{1}{r_2}B_{4,2,1}\widetilde{\eta}'_1$	$\frac{1}{r_1}\frac{1}{r_2}\frac{1}{r_3}B_{4,3,1}\widetilde{\eta}'_1$	$\frac{1}{r_1}\frac{1}{r_2}\frac{1}{r_3}\frac{1}{r_4}B_{4,4,1}\widetilde{\eta}'_1$
0	2		$=\widetilde{\eta}'_2$	$\kappa_1(\xi)\widetilde{\eta}'_2$	$B_{4,0,2}\widetilde{\eta}'_2$	$\frac{1}{r_1}B_{4,1,2}\widetilde{\eta}'_2$	$\frac{1}{r_1}\frac{1}{r_2}B_{4,2,2}\widetilde{\eta}'_2$	$\frac{1}{r_1}\frac{1}{r_2}\frac{1}{r_3}B_{4,3,2}\widetilde{\eta}'_2$	$\frac{1}{r_1}\frac{1}{r_2}\frac{1}{r_3}\frac{1}{r_4}B_{4,4,2}\widetilde{\eta}'_2$
0	3			$=\widetilde{\eta}'_3$	$\kappa_1(\xi)\widetilde{\eta}'_3$	$\frac{1}{r_1}B_{4,1,3}\widetilde{\eta}'_3$	$\frac{1}{r_1}\frac{1}{r_2}B_{4,2,3}\widetilde{\eta}'_3$	$\frac{1}{r_1}\frac{1}{r_2}\frac{1}{r_3}B_{4,3,3}\widetilde{\eta}'_3$	$\frac{1}{r_1}\frac{1}{r_2}\frac{1}{r_3}\frac{1}{r_4}B_{4,4,3}\widetilde{\eta}'_3$
0	4				$=\widetilde{\eta}'_4$	$\frac{1}{r_1}\kappa_1(\xi)\widetilde{\eta}'_4$	$\frac{1}{r_1}\frac{1}{r_2}B_{4,2,4}\widetilde{\eta}'_4$	$\frac{1}{r_1}\frac{1}{r_2}\frac{1}{r_3}B_{4,3,4}\widetilde{\eta}'_4$	$\frac{1}{r_1}\frac{1}{r_2}\frac{1}{r_3}\frac{1}{r_4}B_{4,4,4}\widetilde{\eta}'_4$
1	5					$=\widetilde{\eta}'_5$	$\frac{1}{r_2}\kappa_1(\xi)\widetilde{\eta}'_5$	$\frac{1}{r_2}\frac{1}{r_3}B_{4,3,5}\widetilde{\eta}'_5$	$\frac{1}{r_2}\frac{1}{r_3}\frac{1}{r_4}B_{4,4,5}\widetilde{\eta}'_5$
2	6						$=\widetilde{\eta}'_6$	$\frac{1}{r_3}\kappa_1(\xi)\widetilde{\eta}'_6$	$\frac{1}{r_3}\frac{1}{r_4}B_{4,4,5}\widetilde{\eta}'_6$
3	7							$=\widetilde{\eta}'_7$	$\frac{1}{r_4}\kappa_1(\xi)\widetilde{\eta}'_7$
4	8								$=\widetilde{\eta}'_8$

- The weight terms in green are pre-computed and stored in a memory array before the MCMC run.
- The one (first) cumulant term in violet is updated with each MCMC draw.
- The scaled moment terms in red are computed by recursively *summing up the columns*.
- Result: rapid likelihood evaluation for Markov chain!

Lemma (1)

Under our model assumptions, $f_{\max}(\varepsilon_{itck})$ is a Gumbel distribution with mean $\log(\nu_{itck})$ where

$$\nu_{itck}(\xi) = \sum_{j=1}^{J} \exp\left[-\left(V_{itck}(\xi) - V_{itjk}(\xi)\right)\right]$$

where $V_{itck} = \gamma' Z_{it} + \beta_i' X_{itck} + \theta_i D_{itck}$ and $\xi \equiv (\gamma, \beta, \theta)$

- Use it to derive:
  - Cumulant generating function $K_{\varepsilon_{itck}}(s)$ and cumulants $\kappa_w(\varepsilon_{itck})$ of ε_{itck}
  - Cumulant generating function $K_{\overline{\varepsilon}_{itc}}(s)$ and cumulants $\kappa_w(\overline{\varepsilon}_{itc})$ of $\overline{\varepsilon}_{itc} = y_{it}^{-1} \sum_{k=1}^{y_{it}} \varepsilon_{itck}$
- Use these to evaluate the scaled moments $\widetilde{\eta}'_{r+y_{itc}}(\overline{\varepsilon}_{itc}; \overline{V}_{itc})$ in the expansion for $E_{\overline{\varepsilon}} f(y_{itc}|\overline{V}_{itc})$

Lemma (1)

Under our model assumptions, $f_{\max}(\varepsilon_{itck})$ is a Gumbel distribution with mean $\log(\nu_{itck})$ where

$$\nu_{itck}(\xi) = \sum_{j=1}^{J} \exp\left[-\left(V_{itck}(\xi) - V_{itjk}(\xi)\right)\right]$$

where $V_{itck} = \gamma' Z_{it} + \beta_i' X_{itck} + \theta_i D_{itck}$ and $\xi \equiv (\gamma, \beta, \theta)$

- Use it to derive:
  - Cumulant generating function $K_{\varepsilon_{itck}}(s)$ and cumulants $\kappa_w(\varepsilon_{itck})$ of ε_{itck}
  - Cumulant generating function $K_{\overline{\varepsilon}_{itc}}(s)$ and cumulants $\kappa_w(\overline{\varepsilon}_{itc})$ of $\overline{\varepsilon}_{itc} = y_{it}^{-1} \sum_{k=1}^{y_{it}} \varepsilon_{itck}$
- Use these to evaluate the scaled moments $\widetilde{\eta}'_{r+y_{itc}}(\overline{\varepsilon}_{itc}; \overline{V}_{itc})$ in the expansion for $E_{\overline{\varepsilon}} f(y_{itc} | \overline{V}_{itc})$

Lemma (1)

Under our model assumptions, $f_{\max}(\varepsilon_{itck})$ is a Gumbel distribution with mean $\log(\nu_{itck})$ where

$$\nu_{itck}(\xi) = \sum_{j=1}^{J} \exp\left[-\left(V_{itck}(\xi) - V_{itjk}(\xi)\right)\right]$$

where $V_{itck} = \gamma' Z_{it} + \beta_i' X_{itck} + \theta_i D_{itck}$ and $\xi \equiv (\gamma, \beta, \theta)$

- Use it to derive:
  - Cumulant generating function $K_{\varepsilon_{itck}}(s)$ and cumulants $\kappa_w(\varepsilon_{itck})$ of ε_{itck}
  - Cumulant generating function $K_{\overline{\varepsilon}_{itc}}(s)$ and cumulants $\kappa_w(\overline{\varepsilon}_{itc})$ of $\overline{\varepsilon}_{itc} = y_{it}^{-1} \sum_{k=1}^{y_{it}} \varepsilon_{itck}$
- Use these to evaluate the scaled moments $\widetilde{\eta}'_{r+y_{itc}}(\overline{\varepsilon}_{itc}; \overline{V}_{itc})$ in the expansion for $E_{\overline{\varepsilon}} f(y_{itc}|\overline{V}_{itc})$

Theorem (1)

$$E_{\overline{\varepsilon}} f(y_{itc} | \overline{V}_{itc}) = \sum_{r=0}^{\infty} \delta_{itc}^{y_{itc}+r} \left[\mathbf{Q}^T_{y_{itc},r} \widetilde{\eta}'_{y_{it},r-2} + r^{-1} \kappa_1 \left(\nu_{itc}(\xi) \right) \widetilde{\eta}'_{y_{itc}+r-1} \right]$$

$$Q_{y_{itc},r,q} = \frac{1}{r!} B_{y_{itc},r,q} \quad \text{for } p \leq y_{itc}$$

$$= \frac{1}{r!^{(q-y_{itc})}} B_{y_{itc},r,q} \quad \text{for } y_{it} < p \leq r + y_{itc} - 2$$

$$B_{y_{itc},r,q} = (-1)^r \frac{(y_{itc} + r - 1)!}{q!} \left(\frac{1}{y_{itc}} \right)^{y_{itc}+r-q-1} \zeta(y_{itc} + r - q)$$

for $p = 1, \ldots, r + y_{itc}$ *and* $q = 0, ..., r + y_{itc} - 2$, *where* $\zeta(j)$ *is the Riemann zeta function.*

Lemma (2)

The series representation of $E_{\overline{\varepsilon}} f\left(y_{itc} \middle| \overline{V}_{itc}\right)$ in Lemma 2 is absolutely summable, with bounds on numerical convergence given by $O\left(y_{itc}^{-r}\right)$ as r grows large.

- Useful fact: the Riemann zeta function is a well-behaved term bounded with $\widetilde{\zeta}(j) < \frac{\pi^2}{6}$ for $j > 0$ and with $\widetilde{\zeta}(j) \rightarrow 1$ as $j \rightarrow \infty$.
- A number of explosive terms cancel out due to scaling by $(y_{itc}!r!)^{-1}$, convergence for r growing large

Outline

1. Motivation
   1. Background on Count Data Models
   2. Continuous-time Poisson Process
2. Model
   1. Potential Continuous-time Utility
   2. Linking Utility and Count Intensity
   3. Count Probabilities in a new Mixed Poisson Model
   4. Efficient Likelihood Evaluation Algorithm
3. Bayesian Analysis
   1. Parametric vs Nonparametric Model
   2. Dirichlet Process Prior
4. Application
   1. Data and Variables
   2. Results
5. Counterfactual Welfare Experiment

Bayesian Analysis: Background

- All forms of uncertainty are expressed in terms of probability
- Random coefficient LDV models
  - Rossi, Allenby and McCulloch (2005); Imai and van Dyk (2005); Athey and Imbens (2007); Imai, Jain, and Ching (2009, ECTA)
- Dirichlet process prior
  - Beginnings: Freedman (1963); Ferguson (1973); Blackwell and MacQueen (1973).
  - Recent applications: Hirano (2002); Chib and Hamilton (2002); Jensen and Maheu (2007)

Bayesian Analysis: Background

- All forms of uncertainty are expressed in terms of probability
- Random coefficient LDV models
  - Rossi, Allenby and McCulloch (2005); Imai and van Dyk (2005); Athey and Imbens (2007); Imai, Jain, and Ching (2009, ECTA)
- Dirichlet process prior
  - Beginnings: Freedman (1963); Ferguson (1973); Blackwell and MacQueen (1973).
  - Recent applications: Hirano (2002); Chib and Hamilton (2002); Jensen and Maheu (2007)

Bayesian Analysis: Background

- All forms of uncertainty are expressed in terms of probability
- Random coefficient LDV models
  - Rossi, Allenby and McCulloch (2005); Imai and van Dyk (2005); Athey and Imbens (2007); Imai, Jain, and Ching (2009, ECTA)
- Dirichlet process prior
  - Beginnings: Freedman (1963); Ferguson (1973); Blackwell and MacQueen (1973).
  - Recent applications: Hirano (2002); Chib and Hamilton (2002); Jensen and Maheu (2007)

Our Approach

"Random Effects" (deeper hierarchy)

- $\beta_i \sim F(\beta)$ nonparametric (non-conjugate Dirichlet Process prior)
  - Locally adaptive density estimation of $F(\beta)$
  - Focus on local details and uncovering clustering structures
  - In our application on variables *log price*, *log distance*, and their *interaction*
- $\theta_i \sim MVN(b_\theta, \Sigma_\theta)$ parametric, with updates of b_θ, Σ_θ
  - Controls for levels in θ_i with flexible parsimonious parametrization
  - In our application on store dummies

" Fixed Effects" (shallow hierarchy)

- γ without hyperparameters
  - Not identified in a multinomial choice
  - Identified in the cross-section in likelihood for counts
  - In our application on demographic variables

Our Approach

"Random Effects" (deeper hierarchy)

- $\beta_i \sim F(\beta)$ nonparametric (non-conjugate Dirichlet Process prior)
  - Locally adaptive density estimation of $F(\beta)$
  - Focus on local details and uncovering clustering structures
  - In our application on variables *log price, log distance,* and their *interaction*
- $\theta_i \sim MVN(b_\theta, \Sigma_\theta)$ parametric, with updates of b_θ, Σ_θ
  - Controls for levels in θ_i with flexible parsimonious parametrization
  - In our application on store dummies

" Fixed Effects" (shallow hierarchy)

- γ without hyperparameters
  - Not identified in a multinomial choice
  - Identified in the cross-section in likelihood for counts
  - In our application on demographic variables

Our Approach

"Random Effects" (deeper hierarchy)

- $\beta_i \sim F(\beta)$ nonparametric (non-conjugate Dirichlet Process prior)
  - Locally adaptive density estimation of $F(\beta)$
  - Focus on local details and uncovering clustering structures
  - In our application on variables *log price*, *log distance*, and their *interaction*
- $\theta_i \sim MVN(b_\theta, \Sigma_\theta)$ parametric, with updates of b_θ, Σ_θ
  - Controls for levels in θ_i with flexible parsimonious parametrization
  - In our application on store dummies

" Fixed Effects" (shallow hierarchy)

- γ without hyperparameters
  - Not identified in a multinomial choice
  - Identified in the cross-section in likelihood for counts
  - In our application on demographic variables

Bayesian Parametric vs. Nonparametric Model

- Data: $z = \{z_i\}_{i=1}^{n}$; Parameters: $\psi \in \Psi \subset \mathbb{R}^d$
- Parametric model:
  - Prior: $\psi \sim G_{0p}$
  - The joint distribution of z and ψ:

$$Q(\cdot;\psi, G_{0p}) \propto F(\cdot;\psi) G_{0p}$$

- Nonparametric model:
  - Priors: $\psi | G \sim G$, $G \sim DP(\alpha, G_0)$
  - The joint distribution of z and ψ:

$$Q(\cdot;\psi, G) \propto \int F(\cdot;\psi) dG(\psi)$$

  - G_0 baseline prior distribution - first choice in a parametric model
  - G random measure, deviates stochastically from G_0
  - $\alpha \in \mathbb{R}_+$ concentration of G around G_0, sampled within the system
    - $\alpha \to 0 \Longrightarrow$ kernel estimation (all weight on data)
    - $\alpha \to \infty \Longrightarrow G = G_0 \Leftrightarrow$ parametric model (all weight on the prior)

Bayesian Parametric vs. Nonparametric Model

- Data: $z = \{z_i\}_{i=1}^{n}$; Parameters: $\psi \in \Psi \subset \mathbb{R}^d$
- Parametric model:
  - Prior: $\psi \sim G_{0p}$
  - The joint distribution of z and ψ:

$$Q(\cdot;\psi, G_{0p}) \propto F(\cdot;\psi)G_{0p}$$

- Nonparametric model:
  - Priors: $\psi|G \sim G,\ G \sim DP(\alpha, G_0)$
  - The joint distribution of z and ψ:

$$Q(\cdot;\psi, G) \propto \int F(\cdot;\psi)dG(\psi)$$

  - G_0 baseline prior distribution - first choice in a parametric model
  - G random measure, deviates stochastically from G_0
  - $\alpha \in \mathbb{R}_+$ concentration of G around G_0, sampled within the system
    - $\alpha \to 0 \implies$ kernel estimation (all weight on data)
    - $\alpha \to \infty \implies G = G_0 \Leftrightarrow$ parametric model (all weight on the prior)

Bayesian Parametric vs. Nonparametric Model

- Data: $z = \{z_i\}_{i=1}^n$; Parameters: $\psi \in \Psi \subset \mathbb{R}^d$
- Parametric model:
  - Prior: $\psi \sim G_{0p}$
  - The joint distribution of z and ψ:

$$Q(\cdot; \psi, G_{0p}) \propto F(\cdot; \psi) G_{0p}$$

- Nonparametric model:
  - Priors: $\psi | G \sim G,\ G \sim DP(\alpha, G_0)$
  - The joint distribution of z and ψ:

$$Q(\cdot; \psi, G) \propto \int F(\cdot; \psi) dG(\psi)$$

  - G_0 baseline prior distribution - first choice in a parametric model
  - G random measure, deviates stochastically from G_0
  - $\alpha \in \mathbb{R}_+$ concentration of G around G_0, sampled within the system
    - $\alpha \to 0 \implies$ kernel estimation (all weight on data)
    - $\alpha \to \infty \implies G = G_0 \Leftrightarrow$ parametric model (all weight on the prior)

Dirichlet Process prior

- $DP(\alpha, G_0)$ as a distribution over distributions:
  - $\mathcal{M}(\Psi)$: collection of all probability measures on Ψ, endowed with the topology of weak convergence.
  - $\mathcal{M}(\mathcal{M}(\Psi))$: collection of all probability measures on $\mathcal{M}(\Psi)$
  - $G_0 \in \mathcal{M}(\Psi)$, $\alpha \in \mathbb{R}_+$

Definition

A Dirichlet Process on (Ψ, B) with a base measure G_0 and a concentration parameter α, denoted by $DP(G_0, \alpha) \in \mathcal{M}(\mathcal{M}(\Psi))$, is a distribution of a random probability measure $G \in \mathcal{M}(\Psi)$ over (Ψ, B) such that, for any finite measurable partition $\{\Psi_i\}_{i=1}^{J}$ of the sample space Ψ, the random vector $(G(\Psi_1), ..., G(\Psi_J))$ is distributed as $(G(\Psi_1), ..., G(\Psi_J)) \sim Dir(\alpha G_0(\Psi_1), ..., \alpha G_0(\Psi_J))$ where $Dir(\cdot)$ denotes the Dirichlet distribution.

Dirichlet Process prior

- $DP(\alpha, G_0)$ as a distribution over distributions:
  - $\mathcal{M}(\Psi)$: collection of all probability measures on Ψ, endowed with the topology of weak convergence.
  - $\mathcal{M}(\mathcal{M}(\Psi))$: collection of all probability measures on $\mathcal{M}(\Psi)$
  - $G_0 \in \mathcal{M}(\Psi)$, $\alpha \in \mathbb{R}_+$

Definition

A Dirichlet Process on (Ψ, B) with a base measure G_0 and a concentration parameter α, denoted by $DP(G_0, \alpha) \in \mathcal{M}(\mathcal{M}(\Psi))$, is a distribution of a random probability measure $G \in \mathcal{M}(\Psi)$ over (Ψ, B) such that, for any finite measurable partition $\{\Psi_i\}_{i=1}^{J}$ of the sample space Ψ, the random vector $(G(\Psi_1), ..., G(\Psi_J))$ is distributed as $(G(\Psi_1), ..., G(\Psi_J)) \sim Dir(\alpha G_0(\Psi_1), ..., \alpha G_0(\Psi_J))$ where $Dir(\cdot)$ denotes the Dirichlet distribution.

Sampling Algorithm

Neal (2000), Algorithm 7: Let the state of the Markov chain consist of $\mathbf{c} = (c_1, ..., c_n)$ and $\gamma = (\gamma_c : c \in \{c_1, ..., c_n\})$. Repeatedly sample as follows:

- For $i = 1, ..., n$, update c_i as follows: If c_i is not a singleton (i.e. $c_i = c_j$ for some $j \neq i$), let c_i^* be a newly created component, with γ_{c^*} drawn from G_0. Set the new c_i to this c_i^* with probability

$$a(c_i^*, c_i) = \min \left[1, \frac{\alpha}{n-1} \frac{L(\gamma_{c_i^*}|z_i)}{L(\gamma_{c_i}|z_i)}\right].$$

- For $i = 1, ..., n$: If c_i is a singleton (i.e. $c_i \neq c_j$ for all $j \neq i$), do nothing. Otherwise, choose a new value for c_i from $\{c_1, ..., c_n\}$ using the following probabilities:

$$P(c_i = c | c_{-i}, y_i, \gamma, c_i \in \{c_1, ..., c_n\}) = b \frac{n_{-i,c}}{n-1} L(\gamma_c | z_i)$$

  where b is the appropriate normalizing constant.
- For all $c \in \{c_1, ..., c_n\}$: Draw a new value from $\gamma_c | z_i$ such that $c_i = c$, or perform some other update to γ_c that leaves this distribution invariant.

Simulated Density Estimation

Densities of Marron and Wand, 1992

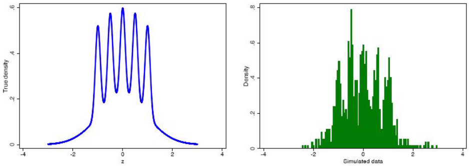


FIGURE 1. Left: trial true functional form of "the claw" posterior density of Marron and Wand (1992). Right: Histogram of a sample draw, $N = 1{,}000$.

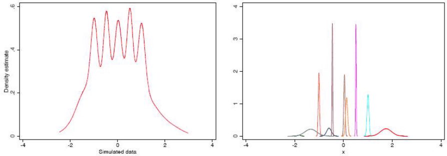


FIGURE 2. Left: DPM density estimate based on the sample in Figure 1, with 10,000 MC steps. Right: A typical snapshot of latent class positions scaled by the class membership intensity.

Simulated Density Estimation: latent classes

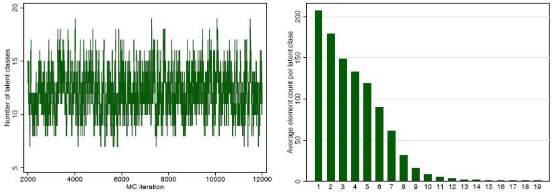


FIGURE 3. $\alpha = 1$. Left: Evolution of the number of latent classes over the MC chain. Right: Average number of latent class members, sorted by size.

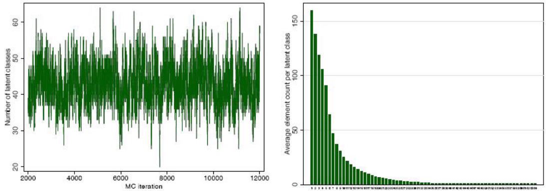


FIGURE 4. $\alpha = 10$. Left: Evolution of the number of latent classes over the MC chain. Right: Average number of latent class members, sorted by size.

Our Model: Priors and Posterior Draws

- Prior structure:

$$\begin{aligned} \theta_i &\sim N(\underline{\mu}_\theta, \underline{\Sigma}_\theta) \\ \gamma &\sim N(\underline{\mu}_\gamma, \underline{\Sigma}_\gamma) \\ \beta_i|\psi_i &\sim F(\psi_i) \\ \psi_i|G &\sim G \\ G &\sim DP(\alpha, G_0) \end{aligned}$$

- Gibbs blocks:
  - $\psi_i|\cdot$ DP hyperparameters (Neal 2000)
  - $\alpha|\cdot$ DP concentration parameter (Escobar and West, 1995)
  - $\beta_i|\cdot$ for each i from $K(\beta_i|\gamma, \theta, \delta, Z, X, D) \propto \prod_{t=1}^{T} E_{\overline{\varepsilon}} f(y_{it}|\overline{V}_{itc}) k_{\phi_i}(\beta)$
  - $\theta_i|\cdot$ analogously to β_i but with $k(\theta)$
  - $\gamma|\cdot$ from $K(\gamma|\beta, \theta, \delta, Z, X, D) \propto \prod_{i=1}^{N} \prod_{t=1}^{T} E_{\overline{\varepsilon}} f(y_{it}|\overline{V}_{itc}) k(\gamma)$
  - $\delta|\cdot$ as in Burda, Harding, and Hausman (2008)
  - Remaining hyperparameters (results A and B in Train, 2003, ch 12)

Our Model: Priors and Posterior Draws

- Prior structure:

$$\begin{aligned}\theta_i &\sim N(\underline{\mu}_\theta, \underline{\Sigma}_\theta) \\ \gamma &\sim N(\underline{\mu}_\gamma, \underline{\Sigma}_\gamma) \\ \beta_i|\psi_i &\sim F(\psi_i) \\ \psi_i|G &\sim G \\ G &\sim DP(\alpha, G_0)\end{aligned}$$

- Gibbs blocks:
  - $\psi_i|\cdot$ DP hyperparameters (Neal 2000)
  - $\alpha|\cdot$ DP concentration parameter (Escobar and West, 1995)
  - $\beta_i|\cdot$ for each i from $K(\beta_i|\gamma, \theta, \delta, Z, X, D) \propto \prod_{t=1}^{T} E_{\overline{\varepsilon}} f(y_{it}|\overline{V}_{itc}) k_{\phi_i}(\beta)$
  - $\theta_i|\cdot$ analogously to β_i but with $k(\theta)$
  - $\gamma|\cdot$ from $K(\gamma|\beta, \theta, \delta, Z, X, D) \propto \prod_{i=1}^{N} \prod_{t=1}^{T} E_{\overline{\varepsilon}} f(y_{it}|\overline{V}_{itc}) k(\gamma)$
  - $\delta|\cdot$ as in Burda, Harding, and Hausman (2008)
  - Remaining hyperparameters (results A and B in Train, 2003, ch 12)

Model Properties

- Identification
  - Property of the likelihood function - same from classical or Bayesian perspectives (Kadane 1974; Poirier 1998; Aldrich 2002)
  - Identification in discrete choice models: Bajari, Fox, Kim and Ryan (2009), Chiappori and Komunjer (2009), Lewbel (2000), Berry and Haile (2010), Briesch, Chintagunta, and Matzkin (2010), Fox and Gandhi (2010), among others
  - Proof of identifiability of infinite mixtures of Poisson distributions: Teicher (1960), Sapatinas (1995)
- Consistency
  - Under *iid* observations and identifiability, the posterior is consistent everywhere except possibly on a null set with respect to the prior (Doob 1949)
  - In the non-parametric context such null set may include cases of interest (Freedman 1963; Diaconis and Freedman 1986a,b, 1990)
  - Posterior consistency for the Dirichlet process prior holds under very general conditions (Ghosal 2008)

Model Properties

- Identification
  - Property of the likelihood function - same from classical or Bayesian perspectives (Kadane 1974; Poirier 1998; Aldrich 2002)
  - Identification in discrete choice models: Bajari, Fox, Kim and Ryan (2009), Chiappori and Komunjer (2009), Lewbel (2000), Berry and Haile (2010), Briesch, Chintagunta, and Matzkin (2010), Fox and Gandhi (2010), among others
  - Proof of identifiability of infinite mixtures of Poisson distributions: Teicher (1960), Sapatinas (1995)
- Consistency
  - Under *iid* observations and identifiability, the posterior is consistent everywhere except possibly on a null set with respect to the prior (Doob 1949)
  - In the non-parametric context such null set may include cases of interest (Freedman 1963; Diaconis and Freedman 1986a,b, 1990)
  - Posterior consistency for the Dirichlet process prior holds under very general conditions (Ghosal 2008)

Posterior Consistency

Theorem (2)

Under our model assumptions, for the posterior $K(\beta_i|\cdot)$ and an arbitrary neighborhood V_0 or the true posterior $K_0(\beta_i|\cdot)$ it holds that $P(K(\beta_i|\cdot) \notin V_0) \longrightarrow 0$ as the sample size tends to infinity.

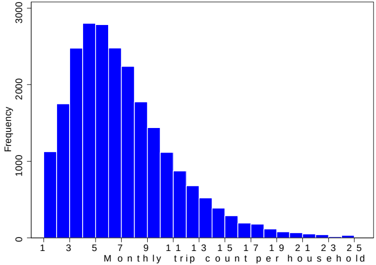
- The proof is based on Ghosal (2009) and Schwartz (1965):
  - A: The prior probability mass assigned to a complement of the sieve space implied by the model is exponentially small and the model sieve approaches the true population value of the parameter as the sample size grows without bound;
  - B: The model sieve satisfies an entropy condition binding the rate of growth of the sieve space in terms of its log $N(\epsilon/2)$-covering number;
  - C: The model likelihood for β_i is bounded in an appropriate sense;
  - D: The Kullback-Leibler positivity property of the prior is satisfied.

Outline

1. Motivation
    1. Background on Count Data Models
    2. Continuous-time Poisson Process
2. Model
    1. Potential Continuous-time Utility
    2. Linking Utility and Count Intensity
    3. Count Probabilities in a new Mixed Poisson Model
    4. Efficient Likelihood Evaluation Algorithm
3. Bayesian Analysis
    1. Parametric vs Nonparametric Model
    2. Dirichlet Process Prior
4. Application
    1. Data and Variables
    2. Results
5. Counterfactual Welfare Experiment

Data

- $N = 650$ households in the Houston area
- AC Nielsen store scanner data - we use 500K entries
- $T = 24$ months during the years 2004 and 2005
- Store chains: H.E. Butt, Kroger, Randall's, Walmart, PantryFoods, "other"
- Trip count:



Variables

1. With $\beta_i \sim F(\beta)$:
   - Price: based on a basket of goods in a given store-month

Product Category:	Bread	Butter and Margarine	Canned Soup	Cereal	Chips
Weight:	0.0804	0.0405	0.0533	0.0960	0.0741
Product Category:	Coffee	Cookies	Eggs	Ice Cream	Milk
Weight:	0.0450	0.0528	0.0323	0.0663	0.1437
Product Category:	Orange Juice	Salad Mix	Soda	Water	Yogurt
Weight:	0.0339	0.0387	0.1724	0.0326	0.0379

Table: Construction of the price index.

   - Distance: estimated driving to supermarket (GPS software to measure the arc distance from the centroid of the census tract in which a household lives to the centroid of the zip code in which a store is located).
   - Interaction: $\ln Price_{itjk} \times \ln Distance_{itjk}$
2. With $\theta_i \sim MVN(b_\theta, \Sigma_\theta)$: Individual supermarket effects
3. With γ : Demographic individual characteristics
   - Singleton (1 member household), Children, Non-white, Hispanic, Unemployed, Education (College +), Medium Age (>40 but <65 hshld head), High Age (>65), Medium Income (25K to 50K), High Income (>50K), and interactions of these with $\ln Price_{itjk}$

Variables

1. With $\beta_i \sim F(\beta)$:
   - Price: based on a basket of goods in a given store-month

Product Category:	Bread	Butter and Margarine	Canned Soup	Cereal	Chips
Weight:	0.0804	0.0405	0.0533	0.0960	0.0741
Product Category:	Coffee	Cookies	Eggs	Ice Cream	Milk
Weight:	0.0450	0.0528	0.0323	0.0663	0.1437
Product Category:	Orange Juice	Salad Mix	Soda	Water	Yogurt
Weight:	0.0339	0.0387	0.1724	0.0326	0.0379

Table: Construction of the price index.

   - Distance: estimated driving to supermarket
     (GPS software to measure the arc distance from the centroid of the census tract in which a household lives to the centroid of the zip code in which a store is located).
   - Interaction: $\ln Price_{itjk} \times \ln Distance_{itjk}$
2. With $\theta_i \sim MVN(b_\theta, \Sigma_\theta)$: Individual supermarket effects
3. With γ : Demographic individual characteristics
   - Singleton (1 member household), Children, Non-white, Hispanic, Unemployed, Education (College +), Medium Age (>40 but <65 hshld head), High Age (>65), Medium Income (25K to 50K), High Income (>50K), and interactions of these with $\ln Price_{itjk}$

Variables

1. With $\beta_i \sim F(\beta)$:
   - Price: based on a basket of goods in a given store-month

Product Category:	Bread	Butter and Margarine	Canned Soup	Cereal	Chips
Weight:	0.0804	0.0405	0.0533	0.0960	0.0741
Product Category:	Coffee	Cookies	Eggs	Ice Cream	Milk
Weight:	0.0450	0.0528	0.0323	0.0663	0.1437
Product Category:	Orange Juice	Salad Mix	Soda	Water	Yogurt
Weight:	0.0339	0.0387	0.1724	0.0326	0.0379

Table: Construction of the price index.

   - Distance: estimated driving to supermarket
     (GPS software to measure the arc distance from the centroid of the census tract in which a household lives to the centroid of the zip code in which a store is located).
   - Interaction: $\ln Price_{itjk} \times \ln Distance_{itjk}$
2. With $\theta_i \sim MVN(b_\theta, \Sigma_\theta)$: Individual supermarket effects
3. With γ : Demographic individual characteristics
   - Singleton (1 member household), Children, Non-white, Hispanic, Unemployed, Education (College +), Medium Age (>40 but <65 hshld head), High Age (>65), Medium Income (25K to 50K), High Income (>50K), and interactions of these with $\ln Price_{itjk}$

Price Index


Save $
with SAFEWAY Brands.
NATIONAL BRANDS
$157.15
Compare and Save!
SAFEWAY BRANDS
$110.55
transit passes

Results

Variable	Selective Flexible Poisson Mixture Mean	Median	S.D.	90% BCS	Normal Poisson Mean	Median	S.D.	90% BCS
Singleton	0.90	0.69	0.20	(0.64, 1.30)	1.89	1.92	0.25	(1.41, 2.27)
Children	1.04	0.85	0.10	(0.88, 1.25)	0.24	0.23	0.36	(-0.35, 0.77)
Non-white	0.20	0.35	0.13	(-0.03, 0.41)	-0.58	-0.64	0.38	(-1.17, 0.09)
Hispanic	0.98	0.41	0.28	(0.43, 1.37)	1.33	1.32	0.31	(0.82, 1.82)
Unemployed	0.66	0.46	0.20	(0.32, 0.98)	-0.61	-0.63	0.43	(-1.32, 0.15)
Education	0.81	0.68	0.15	(0.59, 1.11)	0.79	0.77	0.23	(0.40, 1.18)
Middle Age	0.86	1.12	0.12	(0.68, 1.09)	1.56	1.62	0.30	(0.91, 1.98)
High Age	1.97	1.91	0.18	(1.67, 2.28)	2.67	2.63	0.46	(1.97, 3.42)
Middle Income	2.15	2.41	0.12	(1.95, 2.36)	1.08	1.06	0.25	(0.64, 1.46)
High Income	2.53	2.61	0.20	(2.20, 2.89)	1.33	1.36	0.19	(0.96, 1.62)
$logP\times$ Singleton	-1.63	-1.84	0.42	(-2.36,-0.95)	-3.01	-3.08	0.69	(-3.95,-1.91)
$logP\times$ Children	-0.66	-0.45	0.44	(-1.35,-0.07)	1.14	1.09	0.70	(-0.24, 2.12)
$logP\times$ Non-white	0.01	0.24	0.37	(-0.42, 0.86)	4.93	5.51	1.24	(2.55, 6.43)
$logP\times$ Hispanic	0.78	0.76	0.28	(0.34, 1.31)	0.97	1.06	0.51	(0.05, 1.69)
$logP\times$ Unemployed	1.92	1.36	0.44	(1.40, 2.67)	3.74	3.96	0.63	(2.39, 4.48)
$logP\times$ Education	-1.16	-0.75	0.39	(-1.72,-0.60)	-0.69	-0.86	0.61	(-1.58, 0.38)
$logP\times$ M Age	4.19	2.60	0.69	(3.10, 5.15)	-0.67	-0.97	0.92	(-1.77, 1.38)
$logP\times$ H Age	2.03	1.33	0.18	(1.68, 2.27)	-3.39	-2.96	1.16	(-5.22,-1.97)
$logP\times$ M Income	0.02	0.44	0.51	(-0.88, 0.84)	1.66	1.66	0.45	(0.82, 2.48)
$logP\times$ H Income	-0.30	-0.29	0.42	(-1.16, 0.34)	1.29	1.36	0.65	(0.09, 2.35)

Table: Coefficients γ on demographic variables. $logP$ denotes interaction term with price.

Results

	Selective Flexible Poisson Mixture				Normal Poisson			
Variable	Mean	Median	S.D.	90% BCS	Mean	Median	S.D.	90% BCS
Singleton	0.33	0.31	0.13	(0.12,0.60)	0.85	0.85	0.18	(0.54,1.17)
Children	0.81	0.81	0.15	(0.55,1.05)	0.64	0.59	0.24	(0.27,1.06)
Non-white	0.20	0.20	0.12	(-0.02,0.43)	1.12	1.14	0.19	(0.76,1.39)
Hispanic	1.26	1.30	0.24	(0.74,1.58)	1.67	1.66	0.25	(1.25,2.08)
Unemployed	1.33	1.30	0.24	(0.97,1.79)	0.68	0.70	0.30	(0.14,1.14)
Education	0.41	0.39	0.17	(0.11,0.72)	0.55	0.54	0.17	(0.28,0.86)
Middle Age	2.31	2.30	0.20	(1.95,2.64)	1.32	1.33	0.16	(1.03,1.59)
High Age	2.67	2.66	0.17	(2.41,2.93)	1.50	1.50	0.19	(1.17,1.79)
Middle Income	2.16	2.16	0.17	(1.86,2.46)	1.65	1.66	0.20	(1.31,1.98)
High Income	2.42	2.44	0.15	(2.12,2.64)	1.78	1.85	0.23	(1.36,2.10)

Table: Marginal coefficients γ on demographic variables.

Results

Parameter	Mean	Median	Std.Dev.	90% BCS
$b_{\theta 1}$ (HEB)	7.672	7.708	0.301	(7.093, 8.112)
$b_{\theta 2}$ (Kroger)	5.651	5.838	1.016	(3.931, 7.127)
$b_{\theta 3}$ (Randalls)	8.225	8.365	0.937	(6.607, 9.369)
$b_{\theta 4}$ (Walmart)	4.830	4.915	0.877	(3.380, 6.177)
$b_{\theta 5}$ (Pantry Foods)	11.79	11.681	0.486	(11.168,12.679)
$b_{\theta 6}$ (other)	4.689	4.897	0.808	(3.331, 5.739)

Table: Hyperparameters b_θ of store indicator variable coefficients.

Results

Parameter	Mean	Median	Std.Dev.	90% BCS
$\Sigma_{\theta 1 \theta 1}$ (HEB)	2.205	2.199	0.142	(1.983, 2.450)
$\Sigma_{\theta 1 \theta 2}$ (HEB & Kroger)	-0.008	-0.009	0.084	(-0.146, 0.130)
$\Sigma_{\theta 1 \theta 3}$ (HEB & Randalls)	0.594	0.594	0.101	(0.428, 0.763)
$\Sigma_{\theta 1 \theta 4}$ (HEB & Walmart)	0.211	0.210	0.079	(0.078, 0.345)
$\Sigma_{\theta 1 \theta 5}$ (HEB & Pantry Foods)	-1.105	-1.090	0.144	(-1.366,-0.889)
$\Sigma_{\theta 1 \theta 6}$ (HEB & other)	-0.877	-0.872	0.109	(-1.067,-0.710)
$\Sigma_{\theta 2 \theta 2}$ (Kroger)	1.992	1.988	0.134	(1.779, 2.224)
$\Sigma_{\theta 2 \theta 3}$ (Kroger & Randalls)	0.139	0.137	0.087	(-0.001, 0.283)
$\Sigma_{\theta 2 \theta 4}$ (Kroger & Walmart)	0.060	0.059	0.073	(-0.060, 0.180)
$\Sigma_{\theta 2 \theta 5}$ (Kroger & Pantry Foods)	-0.169	-0.168	0.087	(-0.312,-0.028)
$\Sigma_{\theta 2 \theta 6}$ (Kroger & other)	0.086	0.084	0.081	(-0.047, 0.221)
$\Sigma_{\theta 3 \theta 3}$ (Randalls)	2.209	2.200	0.178	(1.933, 2.516)
$\Sigma_{\theta 3 \theta 4}$ (Randalls & Walmart)	-0.002	-0.003	0.076	(-0.126, 0.125)
$\Sigma_{\theta 3 \theta 5}$ (Randalls & Pantry Foods)	0.559	0.541	0.154	(0.341, 0.862)
$\Sigma_{\theta 3 \theta 6}$ (Randalls & other)	0.392	0.391	0.096	(0.236, 0.555)
$\Sigma_{\theta 4 \theta 4}$ (Walmart)	1.747	1.743	0.113	(1.569, 1.941)
$\Sigma_{\theta 4 \theta 5}$ (Walmart & Pantry Foods)	0.331	0.331	0.087	(0.186, 0.472)
$\Sigma_{\theta 4 \theta 6}$ (Walmart & other)	0.038	0.037	0.076	(-0.084, 0.162)
$\Sigma_{\theta 5 \theta 5}$ (Pantry Foods)	2.311	2.303	0.154	(2.074, 2.585)
$\Sigma_{\theta 5 \theta 6}$ (Pantry Foods & other)	-0.410	-0.409	0.096	(-0.572,-0.256)
$\Sigma_{\theta 6 \theta 6}$ (other)	2.180	2.173	0.138	(1.967, 2.421)

Table: Hyperparameters Σ_θ of store indicator variable coefficients.

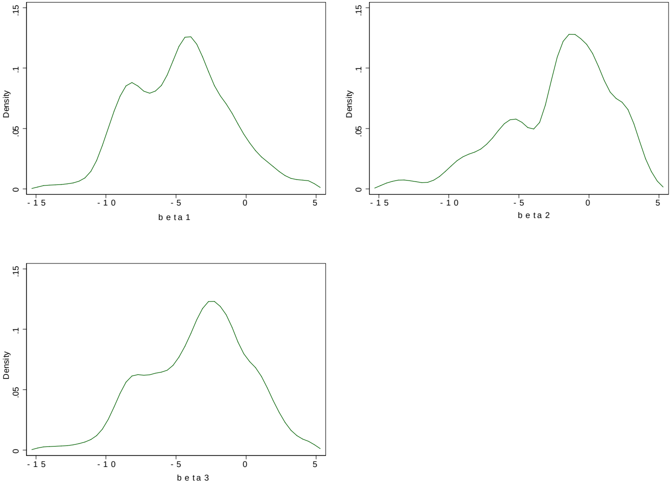
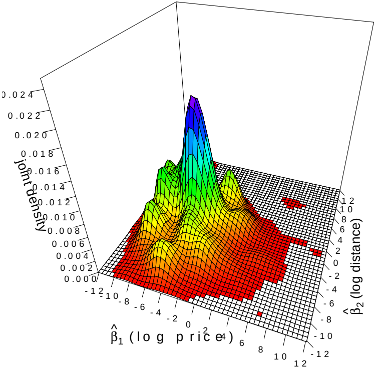


Figure: Posterior density of draws of β_i (logs price, distance, their interaction)

The Hausman test strongly rejects mean equivalence with the Normal counterparts



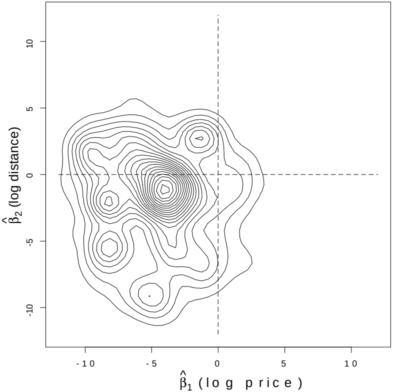


Figure: Joint posterior density of draws of β_i
(logs price vs log distance)

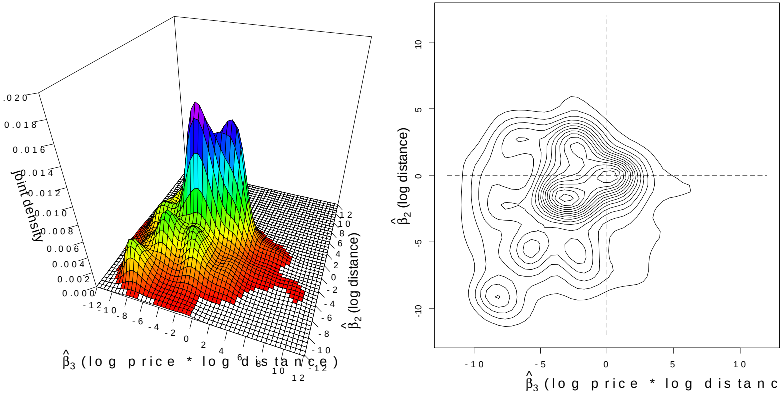


Figure: Joint posterior density of draws of β_i
(log price × log distance vs log distance)

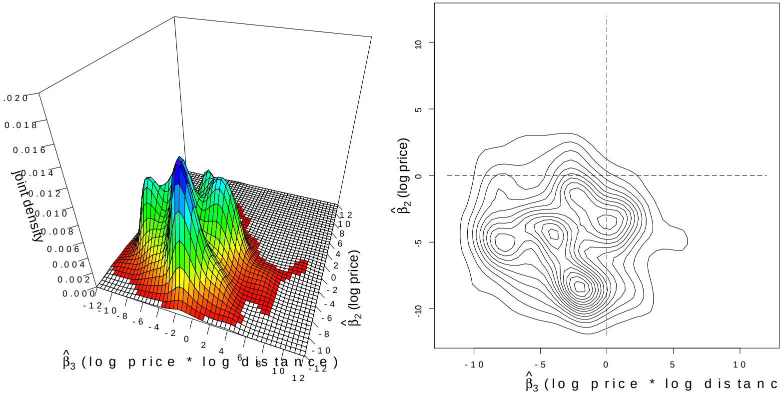


Figure: Joint posterior density of draws of β_i
(log price $\times$ log distance vs log price)

(animation)

(animation)

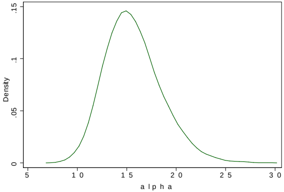


Figure: Posterior density of draws of DP hyperparameter α

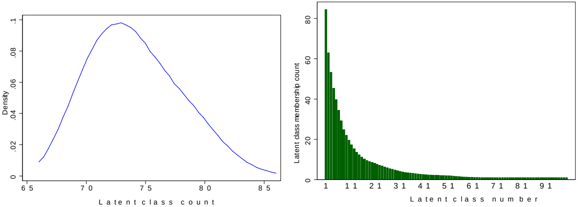


Figure: The number of latent classes density (left)
and ordered average latent class membership count (right)

Outline

1. Motivation
   1. Background on Count Data Models
   2. Continuous-time Poisson Process
2. Model
   1. Potential Continuous-time Utility
   2. Linking Utility and Count Intensity
   3. Count Probabilities in a new Mixed Poisson Model
   4. Efficient Likelihood Evaluation Algorithm
3. Bayesian Analysis
   1. Parametric vs Nonparametric Model
   2. Dirichlet Process Prior
4. Application
   1. Data and Variables
   2. Results
5. Counterfactual Welfare Experiment

Counterfactual Welfare Experiment

- Increase Walmart prices by 10%, 20%, 30%
- How much additional funding each i, t needs to achieve the same utility as before the price increase?
- The difference in count intensities after the price increase:

$$\Delta_{it} = \sum_{c=1}^{J} \delta_{itc}^{new} E[\lambda_{itc}^{new} | \overline{V}_{itc}^{new}] - \sum_{c=1}^{J} \delta_{itc}^{old} E[\lambda_{itc}^{old} | \overline{V}_{itc}^{old}]$$

- Solve for the fixed-point additional income that offsets Δ_{it} in

$$-\Delta_{it} = \sum_{c=1}^{J} \delta_{itc}^{new*} E[\lambda_{itc}^{new*} | \overline{V}_{itc}^{new*}] - \sum_{c=1}^{J} \delta_{itc}^{new} E[\lambda_{itc}^{new} | \overline{V}_{itc}^{new}]$$

- Assume additional purchases split among alternatives by their expected proportions δ_{itc}^{new*} where $new*$ denotes the state with additional income

Counterfactual Welfare Experiment

Walmart price increase	10%		20%		30%	
Variable	Mean	Normal Mean	Mean	Normal Mean	Mean	Normal Mean
Pooled sample	5.96	17.76	8.57	22.12	10.6	26.36
Singleton = 1	9.84	13.05	12.22	17.12	12.9	21.03
Singleton = 0	4.93	19.12	7.61	23.56	9.98	27.89
Children = 1	3.88	12.50	5.58	16.71	7.68	20.73
Children = 0	6.49	19.11	9.34	23.48	11.31	27.75
Non-white = 1	8.78	21.62	9.71	26.28	8.78	30.81
Non-white = 0	5.27	17.00	8.27	21.31	11.10	25.48
Hispanic = 1	3.70	12.76	7.35	16.33	12.49	20.16
Hispanic = 0	6.18	18.41	8.68	22.84	10.44	27.11
Unemployed = 1	8.22	14.80	7.76	19.21	3.86	23.25
Unemployed = 0	5.79	18.07	8.63	22.43	11.11	26.69
Education = 1	7.01	17.29	9.11	21.39	11.04	25.67
Education = 0	4.77	18.17	7.95	22.76	10.11	26.95
Med Age = 1	5.31	18.17	7.41	22.57	8.96	26.77
Med Age = 0	6.71	17.05	9.93	21.36	12.67	25.67
High Age = 1	9.37	15.40	13.0	19.98	16.35	24.72
High Age = 0	4.59	18.41	6.77	22.72	8.45	26.83
Med Income = 1	3.31	13.55	4.99	16.79	8.81	19.72
Med Income = 0	6.88	19.92	9.77	24.80	11.20	29.64
High Income = 1	5.40	19.26	7.71	23.39	8.19	27.63
High Income = 0	6.64	16.18	9.63	20.78	13.61	25.02

Monthly compensating variation in dollar amounts. The sample monthly average grocery food expenditure is $170 of which $84 is spent in Walmart. The Hausman test strongly rejects mean equivalence with the Normal counterparts.

Summary

- New flexible mixed model for count data multinomial discrete choice, endogenizing count intensities
- Derivation of count probabilities via cumulant representations of scaled moments
- Efficient iterative updating scheme
- Three types of parameters:
  - Key parameters interest: $\beta \sim F(\beta)$ (price, distance, their interaction)
  - $\theta \sim MVN(b, \Sigma)$ (store indicator variables)
  - γ (demographic individual characteristics)
- Application: supermarket choices of a panel of Houston households in 2004-2005, scanner data