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)

- $\beta_i$: price, distance, their interaction
- $\theta_i$: store indicator variables
- $\gamma$: 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)
- $\beta_i$: price, distance, their interaction
- $\theta_i$: store indicator variables
- $\gamma$: 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)
  - $\beta_i$: price, distance, their interaction
  - $\theta_i$: store indicator variables
  - $\gamma$: demographic individual characteristics

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

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

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

1. Motivation
   - Background on Count Data Models
   - Continuous-time Poisson Process

2. Model
   - Potential Continuous-time Utility
   - Linking Utility and Count Intensity
   - Count Probabilities in a new Mixed Poisson Model
   - Efficient Likelihood Evaluation Algorithm

3. Bayesian Analysis
   - Parametric vs Nonparametric Model
   - Dirichlet Process Prior

4. Application
   - Data and Variables
   - Results

5. Counterfactual Welfare Experiment

M. Burda, M. Harding, J. Hausman  
Mixed Poisson  
August 2010
## 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:
  \[ f(y = k) = \frac{\exp(-\lambda) \lambda^k}{k!}; \quad \lambda = \exp(X\beta) \]

- Mixed Poisson:
  \[ f(y = k) = \int_0^\infty \frac{\exp(-\lambda) \lambda^k}{k!} g(\lambda) d\lambda \]

- Negative Binomial: special case with \( \lambda \sim \text{gamma}(\delta, \delta) \)
  (Hausman, Hall, and Griliches 1984)
Background: Popular Count Data Models

- **Base-case Poisson:**
  \[
  f(y = k) = \frac{\exp(-\lambda) \lambda^k}{k!}; \quad \lambda = \exp(X\beta)
  \]

- **Mixed Poisson:**
  \[
  f(y = k) = \int_0^\infty \frac{\exp(-\lambda) \lambda^k}{k!} g(\lambda) d\lambda
  \]

- **Negative Binomial:** special case with \(\lambda \sim \text{gamma}(\delta, \delta)\)
  (Hausman, Hall, and Griliches 1984)
Background: Popular Count Data Models

- **Base-case Poisson:**
  \[
  f(y = k) = \frac{\exp(-\lambda) \lambda^k}{k!}; \quad \lambda = \exp(X\beta)
  \]

- **Mixed Poisson:**
  \[
  f(y = k) = \int_0^\infty \frac{\exp(-\lambda) \lambda^k}{k!} g(\lambda) \, d\lambda
  \]

- **Negative Binomial:** special case with \( \lambda \sim \text{gamma}(\delta, \delta) \)
  (Hausman, Hall, and Griliches 1984)
Background: Limits of a Continuous-time Poisson Process

1. **The probability of a unit addition to the count process $Y(t)$ within the interval $\Delta$ is given by**

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

2. **Allow for evolution of $\lambda$ over time to obtain the count process intensity $\tilde{\lambda}(t)$**:

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

3. **By the Poisson independence assumption, obtain the integrated intensity**

$$\lambda(t) = \int_{0}^{t} \tilde{\lambda}(s) ds$$

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 $\Delta$ is given by

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

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

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

- By the Poisson independence assumption, obtain the integrated intensity

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

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 $\Delta$ is given by

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

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

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

- By the Poisson independence assumption, obtain the integrated intensity

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

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

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

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

- $X_{itj}$ - key variables of interest (price, distance, and their interaction)
- $D_{itj}$ - store indicator variables
- $j \in \{1, \ldots, J\}$ - store alternatives
- $\tilde{\epsilon}_{itj}$ - 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 $j$:

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

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

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

$$\tilde{U}_{itc}(\tau) = \max_{j \in J} \{ \tilde{U}_{itj}(\tau) \}$$

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

$$\tilde{\lambda}_{itc}(\tau) = h(\tilde{U}_{itc}(\tau))$$

$$= \gamma' Z_{it}(\tau) + \omega_{1i} \tilde{\beta}'_i X_{itc}(\tau) + \omega_{2i} \tilde{\theta}'_i D_{itc}(\tau) + \omega_{3i} \tilde{\epsilon}_{itc}(\tau)$$

$$= \gamma' Z_{it}(\tau) + \beta'_i X_{itc}(\tau) + \theta'_i D_{itc}(\tau) + \epsilon_{itc}(\tau)$$

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

- Higher $\epsilon_{itj}(\tau)$ increases the probability of additional trip via increased count intensity $\tilde{\lambda}_{itj}(\tau)$

- Proportionality factors $\omega_{1i}$, $\omega_{2i}$, and $\omega_{3i}$ do not need to be separately identified
## Linking Utility and Count Intensity

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

\[
\tilde{U}_{ict}(\tau) = \max_{j \in J} \left\{ \tilde{U}_{itj}(\tau) \right\}
\]

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

\[
\tilde{\lambda}_{itc}(\tau) = h(\tilde{U}_{itc}(\tau)) \\
= \gamma' Z_{it}(\tau) + \omega_{1i} \tilde{\beta}'_i X_{itc}(\tau) + \omega_{2i} \tilde{\theta}'_i D_{itc}(\tau) + \omega_{3i} \tilde{\epsilon}_{itc}(\tau) \\
= \gamma' Z_{it}(\tau) + \beta'_i X_{itc}(\tau) + \theta'_i D_{itc}(\tau) + \epsilon_{itc}(\tau)
\]

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

- Higher $\epsilon_{itj}(\tau)$ increases the probability of additional trip via increased count intensity $\tilde{\lambda}_{itj}(\tau)$

- Proportionality factors $\omega_{1i}, \omega_{2i},$ and $\omega_{3i}$ do not need to be separately identified.
Linking Utility and Count Intensity

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

\[
\tilde{U}_{itc}(\tau) = \max_{j \in J} \left\{ \tilde{U}_{itj}(\tau) \right\}
\]

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

\[
\tilde{\lambda}_{itc}(\tau) = h(\tilde{U}_{itc}(\tau))
= \gamma' Z_{it}(\tau) + \omega_1 \tilde{\beta}_i X_{itc}(\tau) + \omega_2 \tilde{\theta}_i D_{itc}(\tau) + \omega_3 \tilde{\epsilon}_{itc}(\tau)
= \gamma' Z_{it}(\tau) + \beta'_i X_{itc}(\tau) + \theta'_i D_{itc}(\tau) + \epsilon_{itc}(\tau)
\]

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

Higher \( \epsilon_{itj}(\tau) \) increases the probability of additional trip via increased count intensity \( \tilde{\lambda}_{itj}(\tau) \).

Proportionality factors \( \omega_{1i}, \omega_{2i}, \) and \( \omega_{3i} \) do not need to be separately identified.
Linking Utility and Count Intensity

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

$$\tilde{U}_{itc}(\tau) = \max_{j \in J} \{ \tilde{U}_{itj}(\tau) \}$$

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

$$\tilde{\lambda}_{itc}(\tau) = h(\tilde{U}_{itc}(\tau))$$
$$= \gamma' Z_{it}(\tau) + \omega_1' \tilde{\beta}_i X_{itc}(\tau) + \omega_2' \tilde{\theta}_i D_{itc}(\tau) + \omega_3' \tilde{\epsilon}_{itc}(\tau)$$
$$= \gamma' Z_{it}(\tau) + \beta'_i X_{itc}(\tau) + \theta'_i D_{itc}(\tau) + \epsilon_{itc}(\tau)$$

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

- Higher $\epsilon_{itj}(\tau)$ increases the probability of additional trip via increased count intensity $\tilde{\lambda}_{itj}(\tau)$

- Proportionality factors $\omega_1i$, $\omega_2i$, and $\omega_3i$ do not need to be separately identified
Integrated Count Intensity for Discrete Data

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

\[
\tilde{U}_{itjk} = \tilde{\beta}'_i X_{itjk} + \tilde{\theta}'_i D_{itjk} + \tilde{\varepsilon}_{itjk}
\]

- Hence the integrated count intensity

\[
\lambda_{itc} = \int_{t-1}^{t} h(\tilde{U}_{itc}(\tau)) \, d\tau
\]

- Let

\[
\lambda_{itck} = \max \{ 0, \lambda^*_{itck} \}
\]

\[
\lambda^*_{itck} = \gamma' Z_{it} + \beta'_i X_{itck} + \theta_i D_{itck} + \varepsilon_{itck}
\]

and approximate the intensity integral by

\[
\lambda_{itc} = \frac{1}{y_{itc}} \sum_{k=1}^{y_{itc}} \lambda^*_{itck}
\]

\[
= \gamma' Z_{it} + \beta'_i X_{itc} + \theta_i D_{itc} + \varepsilon_{itc}
\]

\[
= V_{itc} + \varepsilon_{itc}
\]
Integrated Count Intensity for Discrete Data

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

$$\tilde{U}_{itjk} = \tilde{\beta}'_i X_{itjk} + \tilde{\theta}'_i D_{itjk} + \tilde{\epsilon}_{itjk}$$

- Hence the integrated count intensity

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

- Let

$$\lambda_{itck} = \max\{0, \lambda^*_{itck}\}$$

$$\lambda^*_{itck} = \gamma' Z_{it} + \beta'_i X_{itck} + \theta_i D_{itck} + \epsilon_{itck}$$

and approximate the intensity integral by

$$\lambda_{itc} = \frac{1}{y_{itc}} \sum_{k=1}^{y_{itc}} \lambda^*_{itck}$$

$$= \gamma' Z_{it} + \beta'_i X_{itc} + \theta_i D_{itc} + \epsilon_{itc}$$

$$= \overline{V}_{itc} + \overline{\epsilon}_{itc}$$
Integrated Count Intensity for Discrete Data

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

$$\tilde{U}_{itjk} = \tilde{\beta}_i^t X_{itjk} + \tilde{\theta}_i^t D_{itjk} + \tilde{\epsilon}_{itjk}$$

- Hence the integrated count intensity

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

- Let

$$\lambda_{itck} = \max \{0, \lambda_{itck}^*\}$$

$$\lambda_{itck}^* = \gamma^t Z_{it} + \beta_i^t X_{itck} + \theta_{ic} D_{itck} + \varepsilon_{itck}$$

and approximate the intensity integral by

$$\lambda_{itc} = \frac{1}{y_{itc}} \sum_{k=1}^{y_{itc}} \lambda_{itck}^*$$

$$= \gamma^t Z_{it} + \beta_i^t X_{itc} + \theta_{ic} D_{itc} + \varepsilon_{itc}$$

$$= \overline{V}_{itc} + \varepsilon_{itc}$$
Count Probabilities

- Denote by $\delta_{itj}$ the fraction of time period $t$ over which the alternative $j$ was maximizing the latent utility $\tilde{U}_{itj}(\tau)$ among other alternatives.

- The assumption of extreme value type 1 distribution on the residual $\tilde{\varepsilon}_{itjk}$ in

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

$$= \tilde{V}_{itc} + \tilde{\varepsilon}_{itjk}$$

yields

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

- Denote by $\delta_{itj}$ the fraction of time period $t$ over which the alternative $j$ was maximizing the latent utility $\tilde{U}_{itj}(\tau)$ among other alternatives.

- The assumption of extreme value type 1 distribution on the residual $\tilde{\varepsilon}_{itjk}$ in

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

yields

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

- The joint conditional trip count and store choice probability:

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

with

\[ \lambda_{itec} \propto \bar{\varepsilon}_{itec} = \frac{1}{y_{iteck}} \sum_{k=1}^{y_{iteck}} \varepsilon_{itck} \]

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

- The joint conditional trip count and store choice probability:

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

with

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

- Each \( \varepsilon_{itck} \) represents an \( J \)-order statistic (maximum) of \( \varepsilon_{itjk} \) 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:

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

with

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

- Each \( \varepsilon_{itck} \) represents an \( J \)-order statistic (maximum) of \( \varepsilon_{itjk} \) 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)
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
  \[
P(y_{itc} | \delta_{itc}) = \int_{\mathcal{V}} \int_{\mathcal{V}} f(y_{it} | \bar{\epsilon}_{itc}, \bar{V}_{itc}(\xi)) g(\bar{\epsilon}_{itc} | \bar{V}_{itc}(\xi)) d\bar{\epsilon}_{itc} g(\bar{V}_{itc}(\xi)) d\bar{V}_{itc}
  \]

- Evaluate analytically
  \[
  E_{\bar{\epsilon}} f(y_{itc} | \bar{V}_{itc}(\xi)) = \sum_{r=0}^{\infty} \frac{(-1)^r}{y_{it}! r!} \delta_{itc}^{r+y_{itc}} \eta'_{y_{it}+r} (\bar{\epsilon}_{itc}; \bar{V}_{itc})
  \]

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

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

$$E_{\varepsilon} f(y_{it} | \bar{V}_{itc}(\bar{\xi}))$$

- Evaluate analytically

$$E_{\varepsilon} f(y_{itc} | \bar{V}_{itc}) = \sum_{r=0}^{\infty} \frac{(-1)^{r}}{y_{it}! r!} \delta^{r+y_{itc}}_{itc} \eta'_{y_{it}+r}(\bar{\varepsilon}_{itc}; \bar{V}_{itc})$$

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: $\eta'_0$

- Sample $\bar{\xi} \equiv (\gamma, \beta, \theta)$ 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

  $$P(y_{itc} | \delta_{itc}) = \int_{\epsilon} \int f(y_{it} | \bar{\epsilon}_{itc}, \bar{V}_{itc}(\xi)) g(\bar{\epsilon}_{itc} | \bar{V}_{itc}(\xi)) d\bar{\epsilon}_{itc} g(\bar{V}_{itc}(\xi)) d\bar{V}_{itc}$$

- Evaluate analytically

  $$E_{\bar{\epsilon}} f(y_{itc} | \bar{V}_{itc}) = \sum_{r=0}^{\infty} \frac{(-1)^r}{y_{it}! r!} \delta_{itc}^{r+y_{itc}} \eta'_{y_{it}+r}(\bar{\epsilon}_{itc}; \bar{V}_{itc})$$

- Obtain $\eta'_{y_{itc}+r}$ recursively from the cumulant-gen. function of $\bar{\epsilon}_{itc}(s)$
- McFadden (1974) choice probabilities: $\eta'_0$
- 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

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

- Evaluate analytically

$$E_{\bar{\epsilon}} f(y_{itc} | \bar{V}_{itc}) = \sum_{r=0}^{\infty} \frac{(-1)^r}{y_{it}! r!} \delta_{itc}^{r + y_{itc}} \eta'_{y_{it} + r}(\bar{\epsilon}_{itc}; \bar{V}_{itc})$$

- Uncentered moments of $\bar{\epsilon}_{itc}$

- Obtain $\eta'_{y_{itc} + r}$ recursively from the cumulant-gen. function of $\bar{\epsilon}_{itc}(s)$

- McFadden (1974) choice probabilities: $\eta'_0$

- Sample $\bar{\xi} \equiv (\gamma, \beta, \theta)$ 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

\[
P(y_{itc} | \delta_{itc}) = \int_{\mathcal{V}} \int_{\mathcal{E}} f(y_{it} | \bar{\epsilon}_{itc}, \bar{V}_{itc}(\xi)) g(\bar{\epsilon}_{itc} | \bar{V}_{itc}(\xi)) d\bar{\epsilon}_{itc} g(\bar{V}_{itc}(\xi)) d\bar{V}_{itc}
\]

\[E_{\bar{E}} f(y_{it} | \bar{V}_{itc}(\xi))\]

- Evaluate analytically

\[
E_{\bar{E}} f(y_{itc} | \bar{V}_{itc}) = \sum_{r=0}^{\infty} \frac{(-1)^r}{y_{it}!r!} \delta_{itc}^{r+y_{itc}} \eta'_{y_{it}+r}(\bar{\epsilon}_{itc}; \bar{V}_{itc})
\]

uncentered moments of \( \bar{\epsilon}_{itc}(s) \)

- Obtain \( \eta'_{y_{itc}+r} \) recursively from the cumulant-gen. function of \( \bar{\epsilon}_{itc}(s) \)
- McFadden (1974) choice probabilities: \( \eta'_{0} \)
- 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

\[
P(y_{itc} | \delta_{itc}) = \int_{\mathcal{V}} \int_{\mathcal{E}} f(y_{it} | \bar{\epsilon}_{itc}, \bar{V}_{itc}(\xi)) g(\bar{\epsilon}_{itc} | \bar{V}_{itc}(\xi)) d\bar{\epsilon}_{itc} g(\bar{V}_{itc}(\xi)) d\bar{V}_{itc}
\]

\[
E_{\bar{\epsilon}} f(y_{it} | \bar{V}_{itc}(\xi))
\]

- Evaluate analytically

\[
E_{\bar{\epsilon}} f(y_{itc} | \bar{V}_{itc}) = \sum_{r=0}^{\infty} \frac{(-1)^r}{y_{it}! r!} \delta_{itc}^{r + y_{itc}} \eta'_{y_{it} + r}(\bar{\epsilon}_{itc}; \bar{V}_{itc})
\]

uncentered moments of \( \bar{\epsilon}_{itc} \)

- Obtain \( \eta'_{y_{itc} + r} \) recursively from the cumulant-gen. function of \( \bar{\epsilon}_{itc}(s) \)
- McFadden (1974) choice probabilities: \( \eta'_{0} \)
- Sample \( \xi \equiv (\gamma, \beta, \theta) \) using Bayesian data augmentation
Recursive Updating: Example for $y_{it} = 4$

<table>
<thead>
<tr>
<th>$r$</th>
<th>$q$</th>
<th>$p : 1$</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
<th>7</th>
<th>8</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>0</td>
<td>$\kappa_1(\xi)\tilde{\eta}_0$</td>
<td>$B_{4,0,0}\tilde{\eta}'_0$</td>
<td>$B_{4,0,0}\tilde{\eta}_0$</td>
<td>$B_{4,0,0}\tilde{\eta}'_0$</td>
<td>$\frac{1}{1} B_{4,1,0}\tilde{\eta}_0$</td>
<td>$\frac{1}{1} B_{4,2,0}\tilde{\eta}_0$</td>
<td>$\frac{1}{1} B_{4,3,0}\tilde{\eta}_0$</td>
<td>$\frac{1}{1} B_{4,4,0}\tilde{\eta}_0$</td>
</tr>
<tr>
<td>0</td>
<td>1</td>
<td>$\kappa_1(\xi)\tilde{\eta}'_1$</td>
<td>$B_{4,0,1}\tilde{\eta}_1$</td>
<td>$B_{4,0,1}\tilde{\eta}'_1$</td>
<td>$\frac{1}{1} B_{4,1,1}\tilde{\eta}_1$</td>
<td>$\frac{1}{1} B_{4,2,1}\tilde{\eta}'_1$</td>
<td>$\frac{1}{1} B_{4,3,1}\tilde{\eta}_1$</td>
<td>$\frac{1}{1} B_{4,4,1}\tilde{\eta}_1$</td>
<td></td>
</tr>
<tr>
<td>0</td>
<td>2</td>
<td>$\tilde{\eta}'_2$</td>
<td>$\kappa_1(\xi)\tilde{\eta}'_2$</td>
<td>$B_{4,0,2}\tilde{\eta}_2$</td>
<td>$\frac{1}{1} B_{4,1,2}\tilde{\eta}_2$</td>
<td>$\frac{1}{1} B_{4,2,2}\tilde{\eta}_2$</td>
<td>$\frac{1}{1} B_{4,3,2}\tilde{\eta}_2$</td>
<td>$\frac{1}{1} B_{4,4,2}\tilde{\eta}_2$</td>
<td></td>
</tr>
<tr>
<td>0</td>
<td>3</td>
<td>$\tilde{\eta}_3$</td>
<td>$\kappa_1(\xi)\tilde{\eta}_3$</td>
<td>$\tilde{\eta}_4$</td>
<td>$\frac{1}{1} B_{4,1,3}\tilde{\eta}_3$</td>
<td>$\frac{1}{1} B_{4,2,3}\tilde{\eta}_3$</td>
<td>$\frac{1}{1} B_{4,3,3}\tilde{\eta}_3$</td>
<td>$\frac{1}{1} B_{4,4,3}\tilde{\eta}_3$</td>
<td></td>
</tr>
<tr>
<td>0</td>
<td>4</td>
<td>$\kappa_1(\xi)\tilde{\eta}_4$</td>
<td>$\tilde{\eta}_4$</td>
<td>$\frac{1}{1} B_{4,2,4}\tilde{\eta}_4$</td>
<td>$\frac{1}{1} B_{4,3,4}\tilde{\eta}_4$</td>
<td>$\frac{1}{1} B_{4,4,4}\tilde{\eta}_4$</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1</td>
<td>5</td>
<td>$\frac{1}{2} \kappa_1(\xi)\tilde{\eta}_5$</td>
<td>$\tilde{\eta}_5$</td>
<td>$\frac{1}{1} B_{4,3,5}\tilde{\eta}_5$</td>
<td>$\frac{1}{1} B_{4,4,5}\tilde{\eta}_5$</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2</td>
<td>6</td>
<td>$\frac{1}{3} \kappa_1(\xi)\tilde{\eta}_6$</td>
<td>$\tilde{\eta}_6$</td>
<td>$\frac{1}{1} B_{4,3,6}\tilde{\eta}_6$</td>
<td>$\frac{1}{1} B_{4,4,6}\tilde{\eta}_6$</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>3</td>
<td>7</td>
<td>$\frac{1}{4} \kappa_1(\xi)\tilde{\eta}_7$</td>
<td>$\tilde{\eta}_7$</td>
<td>$\frac{1}{1} B_{4,3,7}\tilde{\eta}_7$</td>
<td>$\frac{1}{1} B_{4,4,7}\tilde{\eta}_7$</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>4</td>
<td>8</td>
<td>$\frac{1}{4} \kappa_1(\xi)\tilde{\eta}_8$</td>
<td>$\tilde{\eta}_8$</td>
<td>$\frac{1}{1} B_{4,3,8}\tilde{\eta}_8$</td>
<td>$\frac{1}{1} B_{4,4,8}\tilde{\eta}_8$</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

- 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, \( f_{\text{max}}(\varepsilon_{itck}) \) is a Gumbel distribution with mean \( \log(\nu_{itck}) \) where

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

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

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

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

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

- Use it to derive:
  - Cumulant generating function $K_{\varepsilon_{itck}}(s)$ and cumulants $\kappa_w(\varepsilon_{itck})$ of $\varepsilon_{itck}$
  - Cumulant generating function $K_{\bar{\varepsilon}_{itc}}(s)$ and cumulants $\kappa_w(\bar{\varepsilon}_{itc})$ of $\bar{\varepsilon}_{itc} = y_{it}^{-1} \sum_{k=1}^{y_{it}} \varepsilon_{itck}$

- Use these to evaluate the scaled moments $\tilde{\eta}'_{\gamma + y_{itc}}(\bar{\varepsilon}_{itc}; V_{itc})$ in the expansion for $E_{\tilde{\varepsilon}} f(y_{itc} | V_{itc})$
Lemma (1)

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

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

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

\[ E_{\hat{\mathcal{F}} \mid V_{itc}}(y_{itc}) = \sum_{r=0}^{\infty} \delta_{itc}^{y_{itc}+r} \left[ Q_{itc}^{T}, \eta_{itc}^{r-2} + r^{-1} \kappa_{1}(\nu_{itc}(\zeta))(\eta_{itc}^{r-1}) \right] \]

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

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

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

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

The series representation of \( E \xi f(y_{itc} | \bar{V}_{itc}) \) in Lemma 2 is absolutely summable, with bounds on numerical convergence given by \( O(y_{itc}^{-r}) \) as \( r \) grows large.

- Useful fact: the Riemann zeta function is a well-behaved term bounded with \( \bar{\zeta}(j) < \frac{\pi^2}{6} \) for \( j > 0 \) and with \( \bar{\zeta}(j) \to 1 \) as \( j \to \infty \).
- 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)
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

- $\theta_i \sim \text{MVN}(b_\theta, \Sigma_\theta)$ parametric, with updates of $b_\theta, \Sigma_\theta$
  - Controls for levels in $\theta_i$ with flexible parsimonious parametrization
  - In our application on store dummies

"Fixed Effects" (shallow hierarchy)

- $\gamma$ 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 \( \log \text{price}, \log \text{distance} \), and their interaction

- \( \theta_i \sim \text{MVN}(b_{\theta}, \Sigma_{\theta}) \) parametric, with updates of \( b_{\theta}, \Sigma_{\theta} \)
  - Controls for levels in \( \theta_i \) with flexible parsimonious parametrization
  - In our application on store dummies

"Fixed Effects" (shallow hierarchy)

- \( \gamma \) 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 $b_\theta, \Sigma_\theta$
  - Controls for levels in $\theta_i$ with flexible parsimonious parametrization
  - In our application on store dummies

"Fixed Effects" (shallow hierarchy)

- $\gamma$ 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 \( \psi \):
    \[
    Q(\cdot; \psi, G_{0p}) \propto F(\cdot; \psi) G_{0p}
    \]

- **Nonparametric model:**
  - Priors: \( \psi|G \sim G, \ G \sim DP(\alpha, G_0) \)
  - The joint distribution of \( z \) and \( \psi \):
    \[
    Q(\cdot; \psi, G) \propto \int F(\cdot; \psi) dG(\psi)
    \]

- \( G_0 \) 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
  - \( \alpha \to 0 \iff \) kernel estimation (all weight on data)
  - \( \alpha \to \infty \iff G = G_0 \iff \) 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 \( \psi \):
    \[
    Q(\cdot; \psi, G_{0p}) \propto F(\cdot; \psi) G_{0p}
    \]

- **Nonparametric model:**
  - Priors: \( \psi | G \sim G, G \sim DP(\alpha, G_0) \)
  - The joint distribution of \( z \) and \( \psi \):
    \[
    Q(\cdot; \psi, G) \propto \int F(\cdot; \psi) dG(\psi)
    \]
  - \( G_0 \) 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)
    - \( \alpha \to \infty \implies G = G_0 \iff \) 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 \( \psi \):
    \[
    Q(\cdot; \psi, G_{0p}) \propto F(\cdot; \psi) G_{0p}
    \]

- **Nonparametric model:**
  - Priors: \( \psi|G \sim G, G \sim DP(\alpha, G_0) \)
  - The joint distribution of \( z \) and \( \psi \):
    \[
    Q(\cdot; \psi, G) \propto \int F(\cdot; \psi) dG(\psi)
    \]

- \( G_0 \) 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
  - \( \alpha \to 0 \iff \) kernel estimation (all weight on data)
  - \( \alpha \to \infty \iff G = G_0 \iff \) 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 $\Psi$, 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 $(\Psi, B)$ with a base measure $G_0$ and a concentration parameter $\alpha$, 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 $(\Psi, B)$ such that, for any finite measurable partition $\{\Psi_i\}_{i=1}^J$ of the sample space $\Psi$, 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 $\Psi$, 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 $(\Psi, B)$ with a base measure $G_0$ and a concentration parameter $\alpha$, 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 $(\Psi, B)$ such that, for any finite measurable partition $\{\Psi_i\}_{i=1}^J$ of the sample space $\Psi$, 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 $\mathbf{c} = (c_1, \ldots, c_n)$ and $\gamma = (\gamma_c : c \in \{c_1, \ldots, c_n\})$. Repeatedly sample as follows:

- For $i = 1, \ldots, n$, 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 $\gamma_{c^*}$ 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 $i = 1, \ldots, n$: 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 $\{c_1, \ldots, c_n\}$ using the following probabilities:

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

where $b$ is the appropriate normalizing constant.

- For all $c \in \{c_1, \ldots, c_n\}$: Draw a new value from $\gamma_c|z_i$ such that $c_i = c$, or perform some other update to $\gamma_c$ 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:**

\[
\begin{align*}
\theta_i & \sim N(\mu_\theta, \Sigma_\theta) \\
\gamma & \sim N(\mu_\gamma, \Sigma_\gamma) \\
\beta_i|\psi_i & \sim F(\psi_i) \\
\psi_i|G & \sim G \\
G & \sim DP(\alpha, G_0)
\end{align*}
\]

- **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 $K$, $K(\beta_i|\gamma, \theta, \delta, Z, X, D) \propto \prod_{t=1}^T E_\varepsilon f(y_{it}|\overline{V}_{itc}) k_\psi(\beta)$
- $\theta_i|\cdot$: analogously to $\beta_i$ but with $k(\theta)$
- $\gamma|\cdot$: from $K$, $K(\gamma|\beta, \theta, \delta, Z, X, D) \propto \prod_{i=1}^N \prod_{t=1}^T E_\varepsilon f(y_{it}|\overline{V}_{itc}) k(\gamma)$
- $\delta|\cdot$: 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:

\[ \theta_i \sim N(\mu_\theta, \Sigma_\theta) \]
\[ \gamma \sim N(\mu_\gamma, \Sigma_\gamma) \]
\[ \beta_i | \psi_i \sim F(\psi_i) \]
\[ \psi_i | G \sim G \]
\[ G \sim DP(\alpha, G_0) \]

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 \( K(\beta_i | \gamma, \theta, \delta, Z, X, D) \propto \prod_{t=1}^{T} E_{\xi} f(y_{it} | \overline{V}_{itc}) k_{\phi_i}(\beta) \)
- \( \theta_i | \cdot \): analogously to \( \beta_i \) but with \( k(\theta) \)
- \( \gamma | \cdot \) from \( K(\gamma | \beta, \theta, \delta, Z, X, D) \propto \prod_{i=1}^{N} \prod_{t=1}^{T} E_{\xi} f(y_{it} | \overline{V}_{itc}) k(\gamma) \)
- \( \delta | \cdot \): 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

- **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

- **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 $P(K(\beta_i|\cdot) \notin V_0) \rightarrow 0$ 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 $\beta_i$ 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:
1. Motivation

2. Model

3. Bayesian Analysis

4. Application

5. Counterfactual

### Variables

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

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

**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 $\gamma$: 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 $\ln Price_{itjk}$

---

M. Burda, M. Harding, J. Hausman
Variables

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

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

**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 $\theta_i \sim MVN(\theta, \Sigma)$: **Individual supermarket effects**

3. With $\gamma$: **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 $\ln Price_{itjk}$
1. Motivation

2. Model

3. Bayesian Analysis

4. Application

5. Counterfactual

Variables

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

<table>
<thead>
<tr>
<th>Product Category</th>
<th>Bread</th>
<th>Butter and Margarine</th>
<th>Canned Soup</th>
<th>Cereal</th>
<th>Chips</th>
</tr>
</thead>
<tbody>
<tr>
<td>Weight</td>
<td>0.0804</td>
<td>0.0405</td>
<td>0.0533</td>
<td>0.0960</td>
<td>0.0741</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Product Category</th>
<th>Coffee</th>
<th>Cookies</th>
<th>Eggs</th>
<th>Ice Cream</th>
<th>Milk</th>
</tr>
</thead>
<tbody>
<tr>
<td>Weight</td>
<td>0.0450</td>
<td>0.0528</td>
<td></td>
<td>0.0663</td>
<td>0.1437</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Product Category</th>
<th>Orange Juice</th>
<th>Salad Mix</th>
<th>Soda</th>
<th>Water</th>
<th>Yogurt</th>
</tr>
</thead>
<tbody>
<tr>
<td>Weight</td>
<td>0.0339</td>
<td>0.0387</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

**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 $\gamma$: **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
## Results

<table>
<thead>
<tr>
<th>Variable</th>
<th>Selective Flexible Poisson Mixture</th>
<th>Normal Poisson</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Mean</td>
<td>Median</td>
</tr>
<tr>
<td>Singleton</td>
<td>0.90</td>
<td>0.69</td>
</tr>
<tr>
<td>Children</td>
<td>1.04</td>
<td>0.85</td>
</tr>
<tr>
<td>Non-white</td>
<td>0.20</td>
<td>0.35</td>
</tr>
<tr>
<td>Hispanic</td>
<td>0.98</td>
<td>0.41</td>
</tr>
<tr>
<td>Unemployed</td>
<td>0.66</td>
<td>0.46</td>
</tr>
<tr>
<td>Education</td>
<td>0.81</td>
<td>0.68</td>
</tr>
<tr>
<td>Middle Age</td>
<td>0.86</td>
<td>1.12</td>
</tr>
<tr>
<td>High Age</td>
<td>1.97</td>
<td>1.91</td>
</tr>
<tr>
<td>Middle Income</td>
<td>2.15</td>
<td>2.41</td>
</tr>
<tr>
<td>High Income</td>
<td>2.53</td>
<td>2.61</td>
</tr>
<tr>
<td>$\log P \times$ Singleton</td>
<td>-1.63</td>
<td>-1.84</td>
</tr>
<tr>
<td>$\log P \times$ Children</td>
<td>-0.66</td>
<td>-0.45</td>
</tr>
<tr>
<td>$\log P \times$ Non-white</td>
<td>0.01</td>
<td>0.24</td>
</tr>
<tr>
<td>$\log P \times$ Hispanic</td>
<td>0.78</td>
<td>0.76</td>
</tr>
<tr>
<td>$\log P \times$ Unemployed</td>
<td>1.92</td>
<td>1.36</td>
</tr>
<tr>
<td>$\log P \times$ Education</td>
<td>-1.16</td>
<td>-0.75</td>
</tr>
<tr>
<td>$\log P \times$ M Age</td>
<td>4.19</td>
<td>2.60</td>
</tr>
<tr>
<td>$\log P \times$ H Age</td>
<td>2.03</td>
<td>1.33</td>
</tr>
<tr>
<td>$\log P \times$ M Income</td>
<td>0.02</td>
<td>0.44</td>
</tr>
<tr>
<td>$\log P \times$ H Income</td>
<td>-0.30</td>
<td>-0.29</td>
</tr>
</tbody>
</table>

Table: Coefficients $\gamma$ on demographic variables. $\log P$ denotes interaction term with price.
## Results

<table>
<thead>
<tr>
<th>Variable</th>
<th>Selective Flexible Poisson Mixture</th>
<th>Normal Poisson</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Mean</td>
<td>Median</td>
</tr>
<tr>
<td>Singleton</td>
<td>0.33</td>
<td>0.31</td>
</tr>
<tr>
<td>Children</td>
<td>0.81</td>
<td>0.81</td>
</tr>
<tr>
<td>Non-white</td>
<td>0.20</td>
<td>0.20</td>
</tr>
<tr>
<td>Hispanic</td>
<td>1.26</td>
<td>1.30</td>
</tr>
<tr>
<td>Unemployed</td>
<td>1.33</td>
<td>1.30</td>
</tr>
<tr>
<td>Education</td>
<td>0.41</td>
<td>0.39</td>
</tr>
<tr>
<td>Middle Age</td>
<td>2.31</td>
<td>2.30</td>
</tr>
<tr>
<td>High Age</td>
<td>2.67</td>
<td>2.66</td>
</tr>
<tr>
<td>Middle Income</td>
<td>2.16</td>
<td>2.16</td>
</tr>
<tr>
<td>High Income</td>
<td>2.42</td>
<td>2.44</td>
</tr>
</tbody>
</table>

Table: Marginal coefficients $\gamma$ on demographic variables.
Results

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Mean</th>
<th>Median</th>
<th>Std.Dev.</th>
<th>90% BCS</th>
</tr>
</thead>
<tbody>
<tr>
<td>$b_{\theta_1}$ (HEB)</td>
<td>7.672</td>
<td>7.708</td>
<td>0.301</td>
<td>(7.093, 8.112)</td>
</tr>
<tr>
<td>$b_{\theta_2}$ (Kroger)</td>
<td>5.651</td>
<td>5.838</td>
<td>1.016</td>
<td>(3.931, 7.127)</td>
</tr>
<tr>
<td>$b_{\theta_3}$ (Randalls)</td>
<td>8.225</td>
<td>8.365</td>
<td>0.937</td>
<td>(6.607, 9.369)</td>
</tr>
<tr>
<td>$b_{\theta_4}$ (Walmart)</td>
<td>4.830</td>
<td>4.915</td>
<td>0.877</td>
<td>(3.380, 6.177)</td>
</tr>
<tr>
<td>$b_{\theta_5}$ (Pantry Foods)</td>
<td>11.79</td>
<td>11.681</td>
<td>0.486</td>
<td>(11.168, 12.679)</td>
</tr>
<tr>
<td>$b_{\theta_6}$ (other)</td>
<td>4.689</td>
<td>4.897</td>
<td>0.808</td>
<td>(3.331, 5.739)</td>
</tr>
</tbody>
</table>

**Table:** Hyperparameters $b_{\theta}$ of store indicator variable coefficients.
Results

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Mean</th>
<th>Median</th>
<th>Std.Dev.</th>
<th>90% BCS</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\Sigma_{\theta_1 \theta_1}$ (HEB)</td>
<td>2.205</td>
<td>2.199</td>
<td>0.142</td>
<td>(1.983, 2.450)</td>
</tr>
<tr>
<td>$\Sigma_{\theta_1 \theta_2}$ (HEB &amp; Kroger)</td>
<td>-0.008</td>
<td>-0.009</td>
<td>0.084</td>
<td>(-0.146, 0.130)</td>
</tr>
<tr>
<td>$\Sigma_{\theta_1 \theta_3}$ (HEB &amp; Randalls)</td>
<td>0.594</td>
<td>0.594</td>
<td>0.101</td>
<td>(0.428, 0.763)</td>
</tr>
<tr>
<td>$\Sigma_{\theta_1 \theta_4}$ (HEB &amp; Walmart)</td>
<td>0.211</td>
<td>0.210</td>
<td>0.079</td>
<td>(0.078, 0.345)</td>
</tr>
<tr>
<td>$\Sigma_{\theta_1 \theta_5}$ (HEB &amp; Pantry Foods)</td>
<td>-1.105</td>
<td>-1.090</td>
<td>0.144</td>
<td>(-1.366, -0.889)</td>
</tr>
<tr>
<td>$\Sigma_{\theta_1 \theta_6}$ (HEB &amp; other)</td>
<td>-0.877</td>
<td>-0.872</td>
<td>0.109</td>
<td>(-1.067, -0.710)</td>
</tr>
<tr>
<td>$\Sigma_{\theta_2 \theta_2}$ (Kroger)</td>
<td>1.992</td>
<td>1.988</td>
<td>0.134</td>
<td>(1.779, 2.224)</td>
</tr>
<tr>
<td>$\Sigma_{\theta_2 \theta_3}$ (Kroger &amp; Randalls)</td>
<td>0.139</td>
<td>0.137</td>
<td>0.087</td>
<td>(-0.001, 0.283)</td>
</tr>
<tr>
<td>$\Sigma_{\theta_2 \theta_4}$ (Kroger &amp; Walmart)</td>
<td>0.060</td>
<td>0.059</td>
<td>0.073</td>
<td>(-0.060, 0.180)</td>
</tr>
<tr>
<td>$\Sigma_{\theta_2 \theta_5}$ (Kroger &amp; Pantry Foods)</td>
<td>-0.169</td>
<td>-0.168</td>
<td>0.087</td>
<td>(-0.312, -0.028)</td>
</tr>
<tr>
<td>$\Sigma_{\theta_2 \theta_6}$ (Kroger &amp; other)</td>
<td>0.086</td>
<td>0.084</td>
<td>0.081</td>
<td>(-0.047, 0.221)</td>
</tr>
<tr>
<td>$\Sigma_{\theta_3 \theta_3}$ (Randalls)</td>
<td>2.209</td>
<td>2.200</td>
<td>0.178</td>
<td>(1.933, 2.516)</td>
</tr>
<tr>
<td>$\Sigma_{\theta_3 \theta_4}$ (Randalls &amp; Walmart)</td>
<td>-0.002</td>
<td>-0.003</td>
<td>0.076</td>
<td>(-0.126, 0.125)</td>
</tr>
<tr>
<td>$\Sigma_{\theta_3 \theta_5}$ (Randalls &amp; Pantry Foods)</td>
<td>0.559</td>
<td>0.541</td>
<td>0.154</td>
<td>(0.341, 0.862)</td>
</tr>
<tr>
<td>$\Sigma_{\theta_3 \theta_6}$ (Randalls &amp; other)</td>
<td>0.392</td>
<td>0.391</td>
<td>0.096</td>
<td>(0.236, 0.555)</td>
</tr>
<tr>
<td>$\Sigma_{\theta_4 \theta_4}$ (Walmart)</td>
<td>1.747</td>
<td>1.743</td>
<td>0.113</td>
<td>(1.569, 1.941)</td>
</tr>
<tr>
<td>$\Sigma_{\theta_4 \theta_5}$ (Walmart &amp; Pantry Foods)</td>
<td>0.331</td>
<td>0.331</td>
<td>0.087</td>
<td>(0.186, 0.472)</td>
</tr>
<tr>
<td>$\Sigma_{\theta_4 \theta_6}$ (Walmart &amp; other)</td>
<td>0.038</td>
<td>0.037</td>
<td>0.076</td>
<td>(-0.084, 0.162)</td>
</tr>
<tr>
<td>$\Sigma_{\theta_5 \theta_5}$ (Pantry Foods)</td>
<td>2.311</td>
<td>2.303</td>
<td>0.154</td>
<td>(2.074, 2.585)</td>
</tr>
<tr>
<td>$\Sigma_{\theta_5 \theta_6}$ (Pantry Foods &amp; other)</td>
<td>-0.410</td>
<td>-0.409</td>
<td>0.096</td>
<td>(-0.572, -0.256)</td>
</tr>
<tr>
<td>$\Sigma_{\theta_6 \theta_6}$ (other)</td>
<td>2.180</td>
<td>2.173</td>
<td>0.138</td>
<td>(1.967, 2.421)</td>
</tr>
</tbody>
</table>

Table: Hyperparameters $\Sigma_\theta$ of store indicator variable coefficients.
The Hausman test strongly rejects mean equivalence with the Normal counterparts.
Figure: Joint posterior density of draws of $\beta_i$
(logs price vs log distance)
Figure: Joint posterior density of draws of $\beta_i$
(log price $\times$ log distance vs log distance)
Figure: Joint posterior density of draws of $\hat{\beta}_i$
$log \text{ price } \times \log \text{ distance}$ vs log price
1. Motivation

2. Model

3. Bayesian Analysis

4. Application

5. Counterfactual

Figure: Posterior density of draws of DP hyperparameter $\alpha$

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:

$$
\Delta_{it} = \sum_{c=1}^{J} \delta_{itc}^{new} E[\lambda_{itc}^{new} \mid V_{itc}^{new}] - \sum_{c=1}^{J} \delta_{itc}^{old} E[\lambda_{itc}^{old} \mid V_{itc}^{old}]
$$

- Solve for the fixed-point additional income that offsets $\Delta_{it}$ in

$$
-\Delta_{it} = \sum_{c=1}^{J} \delta_{itc}^{new*} E[\lambda_{itc}^{new*} \mid V_{itc}^{new*}] - \sum_{c=1}^{J} \delta_{itc}^{new} E[\lambda_{itc}^{new} \mid V_{itc}^{new}]
$$

- Assume additional purchases split among alternatives by their expected proportions $\delta_{itc}^{new*}$ where $new^*$ denotes the state with additional income
# Counterfactual Welfare Experiment

<table>
<thead>
<tr>
<th>Walmart price increase Variable</th>
<th>10%</th>
<th>20%</th>
<th>30%</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Mean</td>
<td>Normal Mean</td>
<td>Mean</td>
</tr>
<tr>
<td>Pooled sample</td>
<td>5.96</td>
<td>17.76</td>
<td>8.57</td>
</tr>
<tr>
<td>Singleton = 1</td>
<td>9.84</td>
<td>13.05</td>
<td>12.22</td>
</tr>
<tr>
<td>Singleton = 0</td>
<td>4.93</td>
<td>19.12</td>
<td>7.61</td>
</tr>
<tr>
<td>Children = 1</td>
<td>3.88</td>
<td>12.50</td>
<td>5.58</td>
</tr>
<tr>
<td>Children = 0</td>
<td>6.49</td>
<td>19.11</td>
<td>9.34</td>
</tr>
<tr>
<td>Non-white = 1</td>
<td>8.78</td>
<td>21.62</td>
<td>9.71</td>
</tr>
<tr>
<td>Non-white = 0</td>
<td>5.27</td>
<td>17.00</td>
<td>8.27</td>
</tr>
<tr>
<td>Hispanic = 1</td>
<td>3.70</td>
<td>12.76</td>
<td>7.35</td>
</tr>
<tr>
<td>Hispanic = 0</td>
<td>6.18</td>
<td>18.41</td>
<td>8.68</td>
</tr>
<tr>
<td>Unemployed = 1</td>
<td>8.22</td>
<td>14.80</td>
<td>7.76</td>
</tr>
<tr>
<td>Unemployed = 0</td>
<td>5.79</td>
<td>18.07</td>
<td>8.63</td>
</tr>
<tr>
<td>Education = 1</td>
<td>7.01</td>
<td>17.29</td>
<td>9.11</td>
</tr>
<tr>
<td>Education = 0</td>
<td>4.77</td>
<td>18.17</td>
<td>7.95</td>
</tr>
<tr>
<td>Med Age = 1</td>
<td>5.31</td>
<td>18.17</td>
<td>7.41</td>
</tr>
<tr>
<td>Med Age = 0</td>
<td>6.71</td>
<td>17.05</td>
<td>9.93</td>
</tr>
<tr>
<td>High Age = 1</td>
<td>9.37</td>
<td>15.40</td>
<td>13.0</td>
</tr>
<tr>
<td>High Age = 0</td>
<td>4.59</td>
<td>18.41</td>
<td>6.77</td>
</tr>
<tr>
<td>Med Income = 1</td>
<td>3.31</td>
<td>13.55</td>
<td>4.99</td>
</tr>
<tr>
<td>Med Income = 0</td>
<td>6.88</td>
<td>19.92</td>
<td>9.77</td>
</tr>
<tr>
<td>High Income = 1</td>
<td>5.40</td>
<td>19.26</td>
<td>7.71</td>
</tr>
<tr>
<td>High Income = 0</td>
<td>6.64</td>
<td>16.18</td>
<td>9.63</td>
</tr>
</tbody>
</table>

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)
  - $\theta \sim \text{MVN}(b, \Sigma)$ (store indicator variables)
  - $\gamma$ (demographic individual characteristics)

- Application: supermarket choices of a panel of Houston households in 2004-2005, scanner data