Binary Outcome Models with Extreme Covariates: Estimation and Prediction

Laura Liu
University of Pittsburgh

Yulong Wang
Syracuse University

July 2024

Motivation

- Binary models are common in micro and macro empirical research
  - e.g., employment, recession, default...
- Recently, there have been more extreme events occurred
  - e.g., Covid, higher inflation, extreme weather...
- Important to study the effects of extreme X on binary Y
- Existing methods
  - Parametric models (e.g., Logit, Probit): potential large misspecification bias in the tail, esp., as some economics data exhibit heavy tails...
  - Nonparametric models (e.g., kernel, sieve): limited information in the tail, resulting estimator may be largely inefficient...

This Paper

- Novel semiparametric approach
  - Based on Bayes' theorem and regularly varying (RV) functions
  - Pareto approx. in the tail while flexible beyond the tail
- Panel data
  - Effectively accommodate unobserved unit-specific tail thickness and RV fcns
  - Related to panel Logit on tail observations with $\log X$ as regressor
    - For small T, conditional MLE: cancel out unit-specific heterogeneity
    - For large T, bias correction: estimate unit-specific heterogeneity
  - Establish consistency and asymptotic normality for i.n.i.d. tail data
- Extensions
  - Additional non-extreme covariates Z
  - Dynamic panel data models

Related Literature

- Binary cross-sectional models
  - Parametric, e.g., Logit & Probit: Wooldridge (2010)
  - Nonparametric: Matzkin (1992)
  - Semiparametric: Klein & Spady (1993), Manski (1985), Horowitz (1992)
  - Bayes' representation: discriminant analysis (Amemiya, 1985), Klein & Spady (1993)
- Binary panel data models
  - Small T, (correlated) random effects and fixed effects: Wooldridge (2010)
  - Large T, fixed effects and bias corrections: Fernández-Val & Weidner (2018), Stammann, Heiss, & McFadden (2016)
  - Dynamic panel: Honoré and Kyriazidou (2000)
- Heavy tail and extreme value theory
  - Heavy tail: de Haan & Ferreira (2006), Gabaix (2009, 2016), Gomes & Guillou (2015), Fedotenkov (2020)
  - With covariates: Wang & Tsai (2009)

Related Literature (cont.)

- Unit-specific forecasts in panel data
  - Linear: Liu, Moon, & Schorfheide (2020), Liu (2023), Giacomini, Lee, & Sarpietro (2023)
  - Discrete outcome: Christensen, Moon, & Schorfheide (2020)
  - Also see: Fosten & Greenaway-McGrevy (2022), Qu, Timmermann, & Zhu (2023)
- Extreme events and economic consequences, esp. Covid
  - Exclude outliers: Schorfheide & Song (2021)
  - Jointly model outliers and ordinary observations: Carriero, Clark, Marcellino, & Mertens (2022), Lenza & Primiceri (2022)

Roadmap

1. Cross-sectional Data
2. Panel Data
3. Monte Carlo Simulations
4. Empirical Example: Housing Price and Bank Loan Charge-off

Cross-section: Simple Model and Objects of Interest

- $\{Y_i, X_i\}_{i=1}^N$: binary outcome $Y_i \in \{0, 1\}$ and continuous covariate $X_i \in \mathbb{R}$
- Potential objects of interest
  - Conditional probability: $\pi(x) = \mathbb{P}(Y = 1|X = x)$
  - Partial effects: $\partial\pi(x)/\partial x$
  - Extreme elasticity: $\delta(x) = \frac{\partial(\pi(x)(1-\pi(x)))}{\partial x}\frac{x}{\pi(x)(1-\pi(x))}$

Cross-section: Bayes' Theorem

- We observe extreme X, and essentially aim to analyze the comovement in the tail between X and Y.
- To facilitate this tail analysis, we use Bayes' theorem to reverse the conditioning set:

$$\mathbb{P}(Y=1|X=x)=\frac{f_{X|Y}(x|1)\,\mathbb{P}(Y=1)}{f_{X|Y}(x|1)\,\mathbb{P}(Y=1)+f_{X|Y}(x|0)\,\mathbb{P}(Y=0)}.$$

- Note: just characterization of data, allow for any causal directions.
- Similar Bayes' representation also used in discriminant analysis (Amemiya, 1985) and Klein & Spady's (1993) single-index estimator. Here we focus on tail properties.

Cross-section: Heavy Tail and RV function

- Heavy tails are common in economic and financial data (Gabaix, 2009, 2016).
- To assess heavy tails:
  - Visual inspection: e.g., log-log plot ⏩ Fig
  - Estimate Pareto exponent and conduct tests (Clauset, Shalizi, & Newman, 2009).
  - Data-driven evaluation: e.g., cross-validation and out-of-sample forecast
- Regularly varying (RV) function
  - Generic $g \in RV_{-\alpha}$: for some $\alpha > 0$, for all $x > 0$, $\lim_{\eta\to\infty} \frac{g(\eta x)}{g(\eta)} = x^{-\alpha}$.
  - Pareto exponent α: tail thickness. A smaller α indicates a heavier tail.
  - Equivalently, $g(x) = x^{-\alpha}\mathcal{L}(x)$, where $\mathcal{L}(\cdot)$ is a slowly varying function s.t. for all $x > 0$, $\lim_{\eta\to\infty} \frac{\mathcal{L}(\eta x)}{\mathcal{L}(\eta)} \to 1$.

Cross-section: Heavy Tail and RV function (cont.)

- Back to our binary model, assume $1 - F_{X|Y}(\cdot|y) \in RV_{-\alpha^{(y)}}$, for $y \in \{0, 1\}$.
  - Allow for different tail thickness for $y \in \{0, 1\}$.
  - Semiparametric setup
    - First-order Pareto approximation in the tail, for sufficiently large $\underline{x}^{(y)} > 0$,

$$f_{X|Y}\left(x \middle| y, x \geq \underline{x}^{(y)}\right) \sim \alpha^{(y)} \left(\frac{x}{\underline{x}^{(y)}}\right)^{-\alpha^{(y)}-1}$$

    - Flexible beyond the tail
  - RV condition is mild and satisfied by many common distributions, e.g., Pareto, Student-t, F, Cauchy (de Haan & Ferreira, 2006)
  - We can handle very heavy tails where $\alpha^{(y)} \in (0, 1)$ and no moments exist.
- For multidim. X, consider $v(X;\gamma)$: $v(\cdot;\cdot)$ is known, γ can be consistently estimated and converges fast enough, e.g., index model $v(X;\gamma) = X'\gamma$

Cross-section: Extreme Elasticity

In the tail, extreme elasticity is determined solely by the difference between $\alpha^{(y)}$.

Proposition (Cross-sectional data: extreme elasticity). *Suppose we have: (a)* $1 - F_{X|Y}(\cdot|y) \in RV_{-\alpha^{(y)}}$*, for* $y \in \{0, 1\}$*; (b)* $f_{X|Y}(x|y)$ *and* $f'_{X|Y}(x|y)$ *are non-increasing in* $x \geq \underline{x}$*, for some* $\underline{x} > 0$*; and (c)* $0 < \mathbb{P}(Y = 1) < 1$*. Then, as* $x \to \infty$*,*

$$\delta(x) = \frac{\partial(\pi(x)(1 - \pi(x)))}{\partial x} \frac{x}{\pi(x)(1 - \pi(x))} \to -\left|\alpha^{(1)} - \alpha^{(0)}\right|.$$

Cross-section: Comparison with Threshold-crossing

Proposition (Threshold-crossing model). $Y = \mathbf{1}\{X - \varepsilon \geq 0\}$.
Suppose we have: (a) $1 - F_X \in RV_{-\alpha_X}$, *and* $1 - F_\varepsilon \in RV_{-\alpha_\varepsilon}$; *(b)* $f_X(x)$ *is non-increasing in* $x \geq \underline{x}$, *for some* $\underline{x} > 0$; *and (c)* $\varepsilon \perp X$. *Then, for* $y \in \{0, 1\}$, $1 - F_{X|Y}(\cdot|y) \in RV_{-\alpha^{(y)}}$ *with*

$$\alpha^{(0)} = \alpha_X + \alpha_\varepsilon, \text{ and } \alpha^{(1)} = \alpha_X.$$

- Note: our method does not require a threshold-crossing structure.
- Threshold-crossing and uncond. RV tails imply our cond. RV tails.
- Difficult to estimate threshold-crossing model with unobserved RV errors ε.
- In contrast, our method uses cutoffs of observed X, and is easy to implement and scalable to more complicated models.

Cross-section: Additional Covariates Z

- Conditional probability

$$\pi(x,z) = \mathbb{P}(Y=1|X=x, Z=z)$$
$$= \frac{f_{X|Y,Z}(x|1,z)\,\mathbb{P}(Y=1|Z=z)}{f_{X|Y,Z}(x|1,z)\,\mathbb{P}(Y=1|Z=z) + f_{X|Y,Z}(x|0,z)\,\mathbb{P}(Y=0|Z=z)}$$

- Assume pseudo-linear Pareto exponent: $\alpha^{(y)}(z) = z'\theta^{(y)}$.
- First-order Pareto approximation in the tail. MLE for tail observations.

$$\hat{\theta}^{(y)} = \arg\max_{\theta\in\Theta^{(y)}} \sum_{i=1}^{N} \left(\log Z_i'\theta - Z_i'\theta \log \frac{X_i}{\underline{x}_N^{(y)}}\right) \mathbf{1}\left\{X_i^{(y)} \geq \underline{x}_N^{(y)}, Y_i = y\right\}$$

  - Related to exponential tail index in Wang & Tsai (2009).
  - Without Z, reduced to classic Hill (1975) estimator.

Cross-section: Additional Covariates Z (cont.)

- Choice of cutoff $\underline{x}_N^{(y)}$
  - In theory, upper and lower bounds. Larger $\underline{x}_N^{(y)}$ to eliminate asymptotic bias.
  - In practice, empirical quantiles (e.g., 90% or 95%) + log-log plot.
  - Also compared with estimated $\underline{x}_N^{(y)}$ (Guillou & Hall, 2001; Clauset, Shalizi & Newman, 2009). Robust wrt a range of $\underline{x}_N^{(y)}$.
- Asymptotic theory
  - $\underline{x}_N^{(y)} \to \infty$ while z remains constant.
  - Convergence rate is slower than $\sqrt{N}$ due to limited tail observations.

Roadmap

Panel Data: Baseline Model

- $i = 1, ..., N$ and $t = 1, ..., T$. For illustration, let $T = 2$.
- Assume that:
  - $\{Y_{it}, X_{it}\}$ are independent across i and stationary across t
  - Conditional independence across t: $\mathbb{P}_i(Y_{i1}, Y_{i2}|X_{i1}, X_{i2}) = \mathbb{P}_i(Y_{i1}|X_{i1})\mathbb{P}_i(Y_{i2}|X_{i2})$
- Unobserved unit-specific tail thickness and RV functions

$$\tilde{\alpha}_i^{(y)} = \alpha^{(y)} + \lambda_i, \text{ and } 1 - F_{i,X_{it}|Y_{it}}(x|y) = x^{-\tilde{\alpha}_i^{(y)}} \mathcal{L}_i^{(y)}(x),$$

for some slowly varying function $\mathcal{L}_i^{(y)}(x)$.

  - Allows heterogeneity in RV fcns w/o explicit modeling, similar to fixed effects.
  - Accommodates unit-specific scale parameters.
  - With Z, unit-specific tail thickness and RV fcns can flexibly depend on Z.
  - Need to handle i.n.i.d. case for theory.

Panel Data: Conditional Probability

- Let subscript i denote quantities given unit-specific heterogeneity, e.g., $\mathbb{P}_i(\cdot) = \mathbb{P}\left(\cdot\,;\ \lambda_i, \{\mathcal{L}_i^{(y)}\}\right)$.
- Conditional probability

$$\mathbb{P}_i\left(Y_{it} = 1 | X_{it} = x\right) \sim \frac{1}{1 + \frac{\mathbb{P}_i(Y_{it}=0)}{\mathbb{P}_i(Y_{it}=1)} \frac{\tilde{\alpha}_i^{(0)}}{\tilde{\alpha}_i^{(1)}} \frac{\mathcal{L}_i^{(0)}(x)}{\mathcal{L}_i^{(1)}(x)} \frac{x^{-\tilde{\alpha}_i^{(0)}}}{x^{-\tilde{\alpha}_i^{(1)}}}} \sim \frac{1}{1 + A_i \cdot x^{\alpha^{(1)} - \alpha^{(0)}}}.$$

- Without further assumptions, A_i cannot be consistently estimated in small-T panels due to incidental parameter problem.
- Instead, cancel them out by conditioning on $Y_{i1} + Y_{i2} = 1$, as $x_1, x_2 \to \infty$,

$$\mathbb{P}_i\left(Y_{i1} = 1 | Y_{i1} + Y_{i2} = 1, X_{i1} = x_1, X_{i2} = x_2\right) \sim \frac{1}{1 + \left(\frac{x_1}{x_2}\right)^{\alpha^{(1)} - \alpha^{(0)}}}.$$

Panel Data: Conditional MLE

- Let $\alpha^* = \alpha^{(1)} - \alpha^{(0)}$. Conditional MLE on tail observations.

$$\hat{\alpha}^* = \arg\max_{\alpha^*} \prod_{i=1}^{N} \left(\frac{1}{1 + \left(\frac{X_{i1}}{X_{i2}}\right)^{\alpha^*}} \right)^{Y_{i1}} \left(\frac{\left(\frac{X_{i1}}{X_{i2}}\right)^{\alpha^*}}{1 + \left(\frac{X_{i1}}{X_{i2}}\right)^{\alpha^*}} \right)^{1-Y_{i1}} \mathbf{1}\{\Xi_i\},$$

where $\Xi_i = \{Y_{i1} + Y_{i2} = 1, X_{i1} \geq \underline{x}_{1N}, X_{i2} \geq \underline{x}_{2N}\}$. For simplicity, one can set $\underline{x}_{1N} = \underline{x}_{2N} = \underline{x}_N$ in implementation.

Panel Data: Comparision with Panel Logit

- Let $\tilde{A}_i = -\log A_i$.

$$\mathbb{P}_i\left(Y_{it} = 1 | X_{it} = x\right) \sim \frac{1}{1 + A_i \cdot x^{\alpha^*}} = \frac{1}{1 + \exp\left(-\tilde{A}_i + \alpha^* \log x\right)}.$$

- $\tilde{A}_i$ can be viewed as fixed effects that flexibly depend on unobserved heterogeneity (and observed heterogeneity in the case with Z).
- Our objective function is asymptotically equivalent to panel Logit on tail observations, with $\log X$ as regressor.
- Justify and provide assumptions behind the intuitive practice of taking log of extreme covariates.
- Our semiparametric approach: focus on tail behavior, avoiding parametric assumptions on entire error distribution. More robust to misspecification, but reduced sample size.

Panel Data: with Z_i, Asymptotic Theory

Assumption (Panel data: model assumptions). *Suppose we have:*

1. $\{Y_{i1}, Y_{i2}, X_{i1}, X_{i2}, Z_i\}$ *are independent across* i.
2. *For each* i, $\{Y_{it}, X_{it}\}$ *are stationary across* $t = 1, 2$.
3. *For each* i, $Y_{i1} \perp Y_{i2} | X_{i1}, X_{i2}, Z_i$.
4. *For each* i, $\mathbb{P}_i\left(Y_{it} = 1 | X_{i1}, X_{i2}, Z_i\right) = \mathbb{P}_i\left(Y_{it} = 1 | X_{it}, Z_i\right)$ *for* $t = 1, 2$.
5. $\frac{1}{N}\sum_{i=1}^{N} \mathbb{P}_i\left[\mathbb{P}_i(Y_{it} = 1 | Z_i) \in (0,1)\right] > \underline{p}$ *for some* $\underline{p} > 0$*, for all* N *sufficiently large.*

Assumption (Panel data: tail approximation).

Assumption (Panel data: estimation).

Theorem (Panel data: common parameters).

Panel Data: with Z_i, Asymptotic Theory

Assumption (Panel data: model assumptions).

Assumption (Panel data: tail approximation). *For $i = 1, \cdots, N$, for $y \in \{0,1\}$, and almost surely for $z \in supp(\mathcal{Z}_i)$:*

1. *Let $\tilde{\alpha}_i^{(y)}(z) = z'\theta^{(y)} + \lambda_i$. The conditional cdf $F_{i,X_{it}|Y_{it},Z_i}(x|y,z)$ satisfies that*
$$1 - F_{i,X_{it}|Y_{it},Z_i}(x|y,z) = C_i^{(y)}(z)\, x^{-\tilde{\alpha}_i^{(y)}(z)} \left(1 + D_i^{(y)}(z)\, x^{-\beta_i^{(y)}(z)} + o\left(x^{-\beta_i^{(y)}(z)}\right)\right),$$
as $x \to \infty$, where functions $C_i^{(y)}(z) > 0$, $\left|D_i^{(y)}(z)\right| \leq \overline{D} < \infty$, $\tilde{\alpha}_i^{(y)}(z) > 0$, for some constants $\overline{D}$ and $\underline{\beta}$.
2. *The remainder term satisfies that $\sup_{z,i} \left\{x^{\beta_i^{(y)}(z)} o\left(x^{-\beta_i^{(y)}(z)}\right)\right\} \to 0$, as $x \to \infty$.*
3. *The conditional pdf $f_{i,X_{it}|Y_{it},Z_i}(x|y,z)$ is non-increasing in $x \geq \underline{x}$, for some $\underline{x} > 0$.*

Assumption (Panel data: estimation).

Theorem (Panel data: common parameters).

Panel Data: with Z_i, Asymptotic Theory

Assumption (Panel data: model assumptions).

Assumption (Panel data: tail approximation).

Assumption (Panel data: estimation). *Let $\theta^* \in \Theta$, where Θ is a compact convex subset of $\mathbb{R}^{d_\theta}$ and d_θ is the dimension of θ^*. For $y \in \{0,1\}$, suppose we have:*

1. $N\xi_N \to \infty$, *as* $N \to \infty$.
2. *Let* $M_i^{(y)}(Z_i) = |Z_i|\, C_i^{(1)}(Z_i)\, C_i^{(0)}(Z_i) \left| D_i^{(y)}(Z_i) \right| \left| 1 - \frac{\beta_i^{(y)}(Z_i)}{\tilde{\alpha}_i^{(y)}(Z_i)} \right|$. *Then,* $$\sqrt{\frac{N}{\xi_N}} \cdot \frac{1}{N} \sum_{i=1}^N \mathbb{E}_i \left[\left| \log \frac{X_{i1}}{X_i 2} \right| M_i^{(y)}(Z_i) \underline{x}_N^{-\tilde{\alpha}_i^{(1)}(Z_i) - \tilde{\alpha}_i^{(0)}(Z_i) - \beta_i^{(y)}(Z_i)} \right] \to 0, \text{ as } N \to \infty.$$
3. $$\Sigma_N = \frac{1}{N} \sum_{i=1}^N \mathbb{E}_i \left[\frac{\left(\frac{X_{i1}}{X_{i2}}\right)^{Z_i'\theta^*}}{\left(1+\left(\frac{X_{i1}}{X_{i2}}\right)^{Z_i'\theta^*}\right)^2} \left(\log \frac{X_{i1}}{X_{i2}}\right)^2 Z_i Z_i' \middle| \Xi_i \right] \to \Sigma(\theta^*), \text{ as } N \to \infty, \text{ where}$$ $\Sigma(\theta^*)$ *is a finite positive definite matrix.*
4. *For all* i, $\mathbb{E}_i \left[\left| \log \frac{X_{i1}}{X_{i2}} \right|^{2+\kappa} \|Z_i\|^{2+\kappa} \right] < M$ *for some* $\kappa, M > 0$.

Theorem (Panel data: common parameters).

Panel Data: with Z_i, Asymptotic Theory

Assumption (Panel data: model assumptions).

Assumption (Panel data: tail approximation).

Assumption (Panel data: estimation).

Theorem (Panel data: common parameters). *Let* $N_\Xi = \sum_{i=1}^{N} \mathbf{1}\{\Xi_i\}$. *Suppose the above assumptions hold. Then,*

$$\sqrt{N_\Xi}\Sigma_N^{1/2}\left(\hat{\theta}^* - \theta^*\right) \xrightarrow{d} \mathcal{N}\left(0, \mathcal{I}_{d_\theta}\right).$$

Panel Data: with Large T

- Recall $\tilde{A}_i = -\log A_i$ and the asymptotic equivalence

$$\mathbb{P}_i\left(Y_{it} = 1 | X_{it} = x, Z_i\right) \sim \frac{1}{1 + A_i \cdot x^{Z_i'\theta^*}} = \frac{1}{1 + \exp\left(-\tilde{A}_i + \log x \cdot Z_i'\theta^*\right)}.$$

- For large T, we can estimate $\tilde{A}_i$ via panel Logit estimators with bias corrections (Fernández-Val & Weidner, 2018; Stammann, Heiss, & McFadden, 2016).
- After estimating $\{\theta^*, \{\tilde{A}_i\}\}$, we can estimate the average partial effects (APEs) and provide unit-specific forecasts.

Panel Data: with Time-varying Z_{it}

- Now the unit-specific tail thickness becomes $\tilde{\alpha}_i^{(y)}(z_{it}) = z_{it}'\theta^{(y)} + \lambda_i$, and A_i in becomes $A_{it} = \frac{\mathbb{P}_i(Y_{it}=0|Z_{it})}{\mathbb{P}_i(Y_{it}=1|Z_{it})} \frac{\tilde{\alpha}_i^{(0)}(Z_{it})}{\tilde{\alpha}_i^{(1)}(Z_{it})} \mathcal{L}_i^*(Z_{it})$.
- Note that we cannot difference out A_{it} as before.
- Under regularity conditions, we can estimate θ^* via local (conditional) likelihood estimator (Tibshirani & Hastie, 1987; Honoré & Kyriazidou, 2000).

$$\hat{\theta}^* = \arg\max_{\theta^* \in \Theta} \sum_{i=1}^{N} k_h^{(d_z)}(Z_{i1}, Z_{i2}) \left\{ (1 - Y_{i1}) \log \frac{X_{i1}}{X_{i2}} \cdot Z_{i1}'\theta^* + \log \left(1 + \left(\frac{X_{i1}}{X_{i2}} \right)^{Z_{i1}'\theta^*} \right) \right\} \mathbf{1}\{\Xi_i\}$$

Dynamic Panel Data Models

- Suppose $X_{it}, Y_{it} \mid X_{i,1:t-1}, Y_{i,1:t-1}, Z_i = X_{it}, Y_{it} \mid Y_{i,t-1}, Z_i$.
- Partition data by $(Y_{it}, Y_{i,t-1}) = (y, y_-)$

$$\begin{aligned}&\mathbb{P}_i\left(Y_{it} = y | X_{it} = x, Y_{i,t-1} = y_-, Z_i = z\right)\\ &= \frac{f_{i,X_{it}|Y_{it},Y_{i,t-1},Z_i}\left(x|y, y_-, z\right)\mathbb{P}_i\left(Y_{it} = y | Y_{i,t-1} = y_-, Z_i = z\right)}{\sum_{y,y_-} f_{i,X_{it}|Y_{it},Y_{i,t-1},Z_i}\left(x|y, y_-, z\right)\mathbb{P}_i\left(Y_{it} = y | Y_{i,t-1} = y_-, Z_i = z\right)}.\end{aligned}$$

- Unit-specific tail thickness and RV functions

$$\begin{aligned}&\tilde{\alpha}_i^{(yy_-)}(z) = z'\theta^{(yy_-)} + \lambda_i,\\ &1 - F_{i,X_{it}|Y_{it},Y_{i,t-1},Z_i}\left(\cdot|y, y_-, z\right) \in RV_{\tilde{\alpha}_i^{(yy_-)}(z)}.\end{aligned}$$

Dynamic Panel Data Models (cont.)

- Normalize $\theta^{(00)} = \mathbf{0}$. Suppose that we have five periods of data $\mathbf{Y}_i = (Y_{i1,} Y_{i2,}, Y_{i3,}, Y_{i4}, Y_{i5})'$. Consider four events:

$$E_1 : \mathbf{Y}_i = (0,0,1,1,0)', \quad E_2 : \mathbf{Y}_i = (0,1,1,0,0)',$$
$$E_3 : \mathbf{Y}_i = (1,1,0,0,1)', \quad E_4 : \mathbf{Y}_i = (1,0,0,1,1)'.$$

- Given stationarity of the joint distribution $\{Y_{it}, Y_{i,t-1}, X_{it}\} \mid Z_i$, as $x_t \to \infty$ for $t = 1, \cdots, 5$,

$$\mathbb{P}_i \left(E_1 \middle| \cup_{e=1}^4 E_e, \{X_{it} = x_t\}_{t=1}^5, Z_i = z \right)$$
$$\sim \frac{x_3^{-z'\theta^{(01)}} x_4^{-z'\theta^{(11)}} x_5^{-z'\theta^{(10)}}}{\begin{pmatrix} x_3^{-z'\theta^{(01)}} x_4^{-z'\theta^{(11)}} x_5^{-z'\theta^{(10)}} + x_2^{-z'\theta^{(01)}} x_3^{-z'\theta^{(11)}} x_4^{-z'\theta^{(10)}} \\ + x_2^{-z'\theta^{(11)}} x_3^{-z'\theta^{(10)}} x_5^{-z'\theta^{(01)}} + x_2^{-z'\theta^{(10)}} x_4^{-z'\theta^{(01)}} x_5^{-z'\theta^{(11)}} \end{pmatrix}}.$$

- Similar conditional likelihood estimator for $\left(\theta^{(01)}, \theta^{(10)}, \theta^{(11)}\right)$.

Roadmap

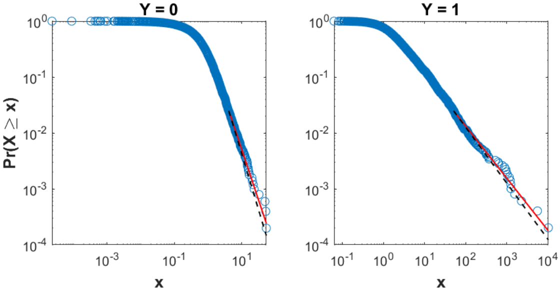
Alternative Estimators

- Cross-sectional data
  - Parametric: $Y_i = \mathbf{1}(\beta_0 + \beta_1 X_i - \varepsilon_i \geq 0)$
    - Logit, using all observations
    - Logit, using only tail observations
  - Nonparametric: $\mathbb{P}(Y_i|X_i) = \mathbb{E}[Y_i|X_i]$
    - Local linear reg.: $\hat{\beta}(x) = \arg\min_{\beta} \sum_{i=1}^{N} k_h(X_i - x)\left[Y_i - \beta_0 - \beta_1(X_i - x)\right]^2$
    - Local Logit: Let $p_i(\beta, x) = \frac{1}{1+\exp[-(\beta_0+\beta_1(X_i - x))]}$.
      $\hat{\beta}(x) = \arg\max_{\beta} \sum_{i=1}^{N} k_h(X_i - x)\left[Y_i \log p_i(\beta, x) + (1 - Y_i)\log(1 - p_i(\beta, x))\right]$
- Panel data
  - Parametric: $Y_{it} = \mathbf{1}(\beta_0 + \beta_1 X_{it} + C_i - \varepsilon_{it} \geq 0)$
  - Nonparametric: $\mathbb{P}_i(Y_{it}|X_{it}) = \mathbb{E}[Y_{it}|X_{it}, V_i]$, e.g., $V_i = \sum_t X_{it}$ (correlated random effects)
- Similar expressions for estimators with additional covariates Z_i.

Experiment 1: Cross-sectional Data

Model:	$Y_i = \mathbf{1}\left(X_i - \varepsilon_i \geq \text{med}_X - \text{med}_\varepsilon\right)$
Covariate:	$X_i \sim \lvert t_{\alpha_X} \rvert$, $\alpha_X = 0.5, 1, 1.5, 2$
Error term:	$\varepsilon_i \sim \lvert t_{\alpha_\varepsilon} \rvert$, $\alpha_\varepsilon = 0.5, 1, 1.5, 2$
Sample Size:	$N = 10000$
# Repetitions:	$N_{sim} = 1000$

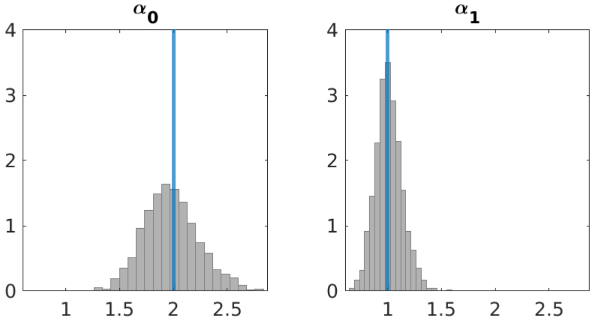
- Threshold-crossing model: extreme elasticity $\approx -\lvert\alpha^{(1)} - \alpha^{(0)}\rvert \approx -\alpha_\varepsilon$
- Log-log plot for one sample with $\alpha_X = 1$ and $\alpha_\varepsilon = 1$. Back



True asymptotic values: black dash lines. Estimated tail indices: red solid lines.

Parameter Estimates

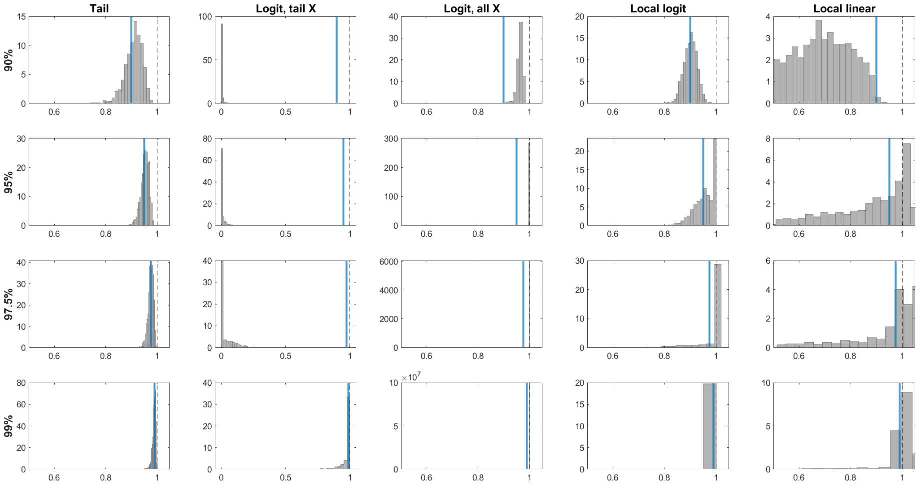
Experiment 1: Cross-sectional Data



Specification with $\alpha_X = 1$ and $\alpha_\varepsilon = 1$, so in the tail, $\alpha^{(0)} \to 2$ and $\alpha^{(1)} \to 1$, indicated by the blue vertical lines.

$\hat{P}(Y = 1|X = x)$

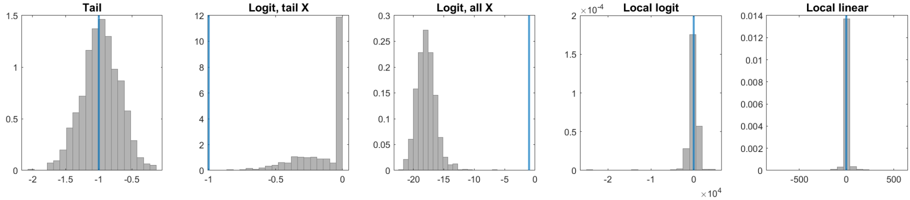
Experiment 1: Cross-sectional Data



Specification with $\alpha_X = 1$ and $\alpha_\varepsilon = 1$. Each row evaluates the probability at a specific percentile of the distribution of X (90%, 95%, 97.5%, and 99%). The true $P(Y = 1|X = x)$ is indicated by the blue vertical lines.

Extreme Elasticity

Experiment 1: Cross-sectional Data



Specification with $\alpha_X = 1$ and $\alpha_\varepsilon = 1$. Each row evaluates the probability at a specific percentile of the distribution of X (90%, 95%, 97.5%, and 99%). The true $P(Y = 1|X = x)$ is indicated by the blue vertical lines.

Roadmap

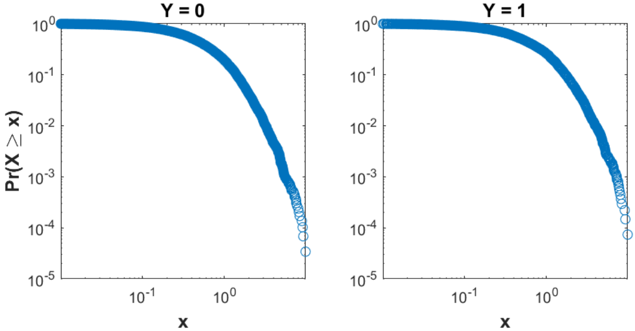
Housing Price and Bank Loan Charge-off

- Charge-off rates reflect bank losses. A bank could be a risker for a particular type of loan if its corresponding charge-off rates exceed a certain threshold.
- Consider a panel of small banks ($<$ \$1b in assets), similar to Liu, Moon & Schorfheide (2023).
- Assume small banks operate in local markets. When local housing price drops a lot, would small banks become riskier? (crucial in 2007-08 financial crisis)
- Variables:
  - Y_{it}: risk dummy based on loan charge off-rate of bank i in quarter t, $Y_{it} = 1$ if charge-off rate $> c$. Let $c = 0$ in the baseline model.
  - X_{it}: deflation rate of local housing price in the previous quarter.
  - Z_i: average change in local unemployment rate.

Data and Sample

- Data sources:
  - Bank balance sheet data: Call Reports from Chicago Fed
  - Currently: local market = county
  - Determine the local market for each bank: Summary of Deposits from FDIC
  - Housing price index: Federal Housing Finance Agency. Convert 3-digit zip code data to county level using HUD USPS ZIP Code Crosswalk.
  - Unemployment rate: Bureau of Labor Statistics
- Let us focus on the following sample
  - Residential real estate (RRE) charge-off rates
  - Sample period: 1999Q4 - 2009Q3: $N = 8538$, $T = 40$
  - Forecast period $(T + 1)$: 2009Q4

Log-log Plot



Baseline sample: RRE, forecasting period $=$ 2009Q4, $T = 40$.

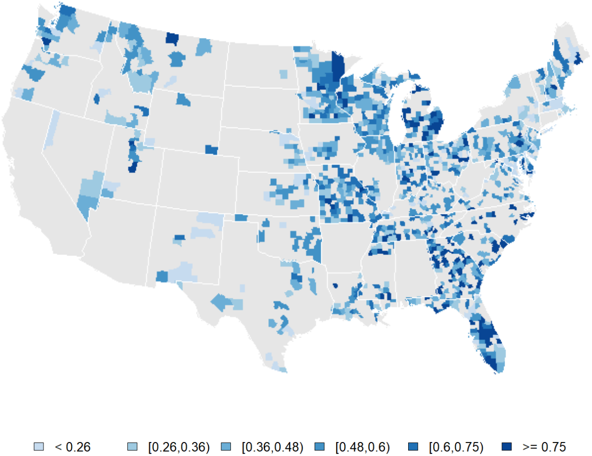
Forecast Evaluation: Log Predictive Score (LPS)

	RRE	CRE
Tail	*-1433.72*	*-1134.48*
Logit, tail X	-18.66 ***	-23.49 ***
Logit, all X	-160.25 ***	-54.26 ***
Local Logit	-429.20 ***	-516.43 ***

CRE: (non-farm) non-residential commercial real estate (CRE) charge-off rates.

Forecasting period = 2009Q4, $T = 40$. The forecasts are assessed by the LPS and a test integrating Amisano and Giacomini (2007) and Qu et al. (2020). For the tail estimator, the table reports the exact values of LPS $\cdot N_f^{\dagger}$. For other estimators, the table reports their differences from the tail estimator. The tests compare other estimators with the tail estimator, with significance levels indicated by *: 10%, **: 5%, and ***: 1%.

Predictive Probability of High Risk



Baseline sample: RRE, forecasting period = 2009Q4, $T = 40$. Average by county

Heterogeneity and Bank Characteristics

	Initial	Average
Log assets	0.20***	0.19***
	(0.03)	(0.03)
Loan frac.	0.28***	0.68***
	(0.10)	(0.11)
Capital-assets	-0.65	-2.05***
	(0.56)	(0.60)
Loan-assets	0.83***	1.77***
	(0.18)	(0.19)
ALLL-loan	2.61	15.01***
	(3.18)	(3.45)
Diversification	1.00**	2.79***
	(0.44)	(0.50)
Ret. on assets	-30.53***	-36.51***
	(9.62)	(11.66)
OCA	-19.34**	-60.19***
	(9.05)	(12.30)

Baseline sample: RRE, forecasting period = 2009Q4, $T = 40$. Regression of $\tilde{A}_i$ bank characteristics. In the "Initial" column, bank characteristics are given by their values at the initial period 1999Q4. In the "Average" column, bank characteristics are given by their time averages over the estimation sample. Significance levels are indicated by *: 10%, **: 5%, and ***: 1%.

Take-aways

- Novel semiparametric approach
  - Based on Bayes' theorem and regularly varying (RV) functions
  - Pareto approx. in the tail while flexible beyond the tail
- Panel data
  - Effectively accommodate unobserved unit-specific tail thickness and RV fcns
  - Related to panel Logit on tail observations with $\log X$ as regressor
    - For small T, conditional MLE: cancel out unit-specific heterogeneity
    - For large T, bias correction: estimate unit-specific heterogeneity
  - Establish consistency and asymptotic normality for i.n.i.d. tail data
- Extensions
  - Additional non-extreme covariates Z
  - Dynamic panel data models
- Many potential micro/macro applications, esp. given recent extreme events

Take-aways

- Novel semiparametric approach
  - Based on Bayes' theorem and regularly varying (RV) functions
  - Pareto approx. in the tail while flexible beyond the tail
- Panel data
  - Effectively accommodate unobserved unit-specific tail thickness and RV fcns
  - Related to panel Logit on tail observations with $\log X$ as regressor
    - For small T, conditional MLE: cancel out unit-specific heterogeneity
    - For large T, bias correction: estimate unit-specific heterogeneity
  - Establish consistency and asymptotic normality for i.n.i.d. tail data
- Extensions
  - Additional non-extreme covariates Z
  - Dynamic panel data models
- Many potential micro/macro applications, esp. given recent extreme events

Thank you!