Online Appendix:
Binary Outcome Models with Extreme Covariates: Estimation and Prediction

Laura Liu Yulong Wang

February 21, 2025

This online appendix is organized as follows. In Appendix A, we provide proofs for all propositions and theorems in the main paper, and establish the asymptotic properties of the proposed tail estimator in both cross-sectional and panel data setups. In Appendix B, we present additional figures and tables that supplement the simulation and empirical results in the main text.

A Proofs

A.1 Proofs for Section 2

We first state some useful properties of RV functions in the following lemma.

Lemma A.1 (Properties of RV functions) *Let $f \in RV_{-\alpha}$ and $g \in RV_{-\beta}$ with $\alpha, \beta > 0$. We have the following:*

(a) $f(x)g(x) \in RV_{-\alpha-\beta}$.

(b) If $f'(x)$ is non-increasing, $\frac{xf'(x)}{f(x)} \to -\alpha$ *as* $x \to \infty$, *and hence* $f'(x) \in RV_{-\alpha-1}$.

(c) If $\alpha > 1$, $\int_x^\infty f(s)ds \in RV_{-\alpha+1}$.

Proof. See Section B in de Haan and Ferreira (2006). ■

Proof of Proposition 2.1.

Recall that

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

Let $R = \frac{\mathbb{P}(Y=0)}{\mathbb{P}(Y=1)}$. Then, $0 < R < \infty$ according to the non-degeneracy condition (c). Define the term in the denominator

$$\pi^*(x) = \frac{1}{\pi(x)} - 1 = \frac{\mathbb{P}(Y=0) f_{X|Y}(x|0)}{\mathbb{P}(Y=1) f_{X|Y}(x|1)} = R \cdot \frac{f_{X|Y}(x|0)}{f_{X|Y}(x|1)}.$$

Given the RV condition (a) and the non-increasing condition (b), we have $f_{X|Y}(x|y) \in RV_{-\alpha^{(y)}-1}$ and

$$\frac{x f'_{X|Y}(x|y)}{f_{X|Y}(x|y)} \to -\alpha^{(y)} - 1,$$

as $x \to \infty$, by Lemma A.1(b). Then the elasticity of $\pi^*(x)$ is given by

$$\begin{aligned} \delta^*(x) = \frac{\partial \pi^*(x)}{\partial x} \frac{x}{\pi^*(x)} &= \frac{x f'_{X|Y}(x|0)}{f_{X|Y}(x|0)} - \frac{x f'_{X|Y}(x|1)}{f_{X|Y}(x|1)} \\ &\to \left(-\alpha^{(0)} - 1\right) - \left(-\alpha^{(1)} - 1\right) = \alpha^{(1)} - \alpha^{(0)}. \end{aligned} \tag{A.1}$$

As in (10), the extreme elasticity can be decomposed as

$$\delta(x) = \delta_\pi(x) + \delta_{1-\pi}(x).$$

And each component can be calculated via the chain rule

$$\begin{aligned} \delta_\pi(x) &= -\delta^*(x) \frac{\pi^*(x)}{1 + \pi^*(x)} = -\delta^*(x) \left(1 - \pi(x)\right), \\ \delta_{1-\pi}(x) &= \delta^*(x) \frac{1}{1 + \pi^*(x)} = \delta^*(x) \pi(x), \end{aligned}$$

where $\delta^*(x)$ is given by (A.1).

- If $\alpha^{(0)} > \alpha^{(1)}$, $\pi(x) \to 1$ as $x \to \infty$, and hence $\delta(x) \sim \delta^*(x) \to \alpha^{(1)} - \alpha^{(0)}$.
- If $\alpha^{(0)} < \alpha^{(1)}$, $\pi(x) \to 0$ as $x \to \infty$, and hence $\delta(x) \sim -\delta^*(x) \to \alpha^{(0)} - \alpha^{(1)}$.
- If $\alpha^{(0)} = \alpha^{(1)}$, $\delta^*(x) \to 0$ and $\pi(x) \to \frac{1}{1+R}$, as $x \to \infty$, so we have

$$\delta(x) = \delta_\pi(x) + \delta_{1-\pi}(x) \sim \delta^*(x) \frac{1-R}{1+R}.$$

As $0 < R < \infty$, it follow that $\left|\frac{1-R}{1+R}\right| < 1$, and thus $\delta(x) \to 0$.

Combining all cases, we have that, as $x \to \infty$,

$$\delta(x) \to -|\alpha^{(1)} - \alpha^{(0)}| = -|\alpha^*|.$$

■

Proof of Proposition 2.2.

For $Y = 1$,

$$\begin{aligned} 1 - F_{X|Y}(x|1) = \mathbb{P}(X > x|Y = 1) &= \frac{\mathbb{P}(X > x, Y = 1)}{\mathbb{P}(Y = 1)} \\ &= \frac{\int_x^\infty \mathbb{P}(Y = 1|X = s) f_X(s) \, ds}{\mathbb{P}(Y = 1)} \\ &= \frac{\int_x^\infty F_\varepsilon(s) f_X(s) \, ds}{\mathbb{P}(Y = 1)}, \end{aligned}$$

where the third equality is by Bayes' rule, and the fourth equality is by the independence condition (c). For the numerator, we have that $F_\varepsilon(x) \to 1$ as $x \to \infty$, and thus $\int_x^\infty F_\varepsilon(s) f_X(s) \, ds \to \int_x^\infty f_X(s) \, ds = 1 - F_X(x) \in RV_{-\alpha_X}$ by the RV condition (a). For the denominator, the RV condition (a) implies that X and ε have overlapping support, at least in the right tail, so $0 < \mathbb{P}(Y = 1) < 1$. Then,

$$1 - F_{X|Y}(x|1) \in RV_{-\alpha_X} \text{ and } \alpha^{(1)} = \alpha_X. \tag{A.2}$$

For $Y = 0$, a similar argument yields that

$$\begin{aligned} 1 - F_{X|Y}(x|0) = \mathbb{P}(X > x|Y = 0) &= \frac{\mathbb{P}(X > x, Y = 0)}{\mathbb{P}(Y = 0)} \\ &= \frac{\int_x^\infty \mathbb{P}(Y = 0|X = s) f_X(s) \, ds}{\mathbb{P}(Y = 0)} \\ &= \frac{\int_x^\infty (1 - F_\varepsilon(s)) f_X(s) \, ds}{\mathbb{P}(Y = 0)}. \end{aligned}$$

For the numerator, by the RV condition (a), we have that $1 - F_\varepsilon(x) \in RV_{-\alpha_\varepsilon}$, and $f_X(x) \in RV_{-\alpha_X - 1}$ based on Lemma A.1(b), and thus $\int_x^\infty (1 - F_\varepsilon(s)) f_X(s) \, ds \in RV_{-\alpha_X - \alpha_\varepsilon}$ by Lemma A.1(a,c). For the denominator, again, the overlapping support implies that $0 < \mathbb{P}(Y = 0) <$

1. Then,

$$1 - F_{X|Y}(x|0) \in RV_{-\alpha_X - \alpha_\varepsilon} \text{ and } \alpha^{(0)} = \alpha_X + \alpha_\varepsilon. \tag{A.3}$$

Combining (A.2) and (A.3), we further obtain that $\alpha_\varepsilon = \alpha^{(0)} - \alpha^{(1)}$. ■

Proposition A.1 (Cross-sec. data: existence of tail average of partial effects) *Under conditions (a) and (b) of Proposition 2.1, assume further that $\pi(x)$ and $f_{X|Y}(x|y)$ are continuously differentiable for $x \geq \underline{x}$, for some $\underline{x} > 0$. Then, the tail average of the partial effects $\mathbb{E}[\pi'(X)|X \geq \underline{x}]$ exists, and converges to 0 as $\underline{x} \to \infty$.*

Proof. Let the unconditional pdf be $f(x) = f_{X|Y}(x|1)\,\mathbb{P}(Y=1) + f_{X|Y}(x|0)\,\mathbb{P}(Y=0)$, and $F(x)$ is the corresponding unconditional cdf. Then, the tail average of the partial effects is given by

$$\mathbb{E}[\pi'(X)|X \geq \underline{x}] = \frac{\int_{\underline{x}}^{\infty} \pi'(x) f(x) dx}{1 - F(\underline{x})}.$$

For the numerator,

$$\begin{aligned}
\left| \int_{\underline{x}}^{\infty} \pi'(x) f(x) dx \right| &= \left| \pi(x) f(x)|_{\underline{x}}^{\infty} - \int_{\underline{x}}^{\infty} \pi(x) f'(x) dx \right| \\
&\leq \pi(\underline{x}) f(\underline{x}) + \int_{\underline{x}}^{\infty} \pi(x) \left| f'(x) \right| dx \\
&\leq f(\underline{x}) + \int_{\underline{x}}^{\infty} \left| f'(x) \right| dx \\
&= 2 f(\underline{x}).
\end{aligned}$$

The first line is by integration by parts, given that $\pi(x)$ and $f_{X|Y}(x|y)$ are continuously differentiable for $x \geq \underline{x}$. The second and third lines are by the fact that $0 \leq \pi(x) \leq 1$ and that $f(x) \to 0$ as $x \to \infty$. The last line follows from that $f'(x)$ is non-increasing in $x \geq \underline{x}$.

Combining with the denominator, we have that

$$\begin{aligned}
\left| \mathbb{E}[\pi'(X)|X \geq \underline{x}] \right| &\leq \frac{2 f(\underline{x})}{1 - F(\underline{x})} \\
&= 2 \frac{f_{X|Y}(\underline{x}|1)\,\mathbb{P}(Y=1) + f_{X|Y}(\underline{x}|0)\,\mathbb{P}(Y=0)}{\left(1 - F_{X|Y}(\underline{x}|1)\right)\mathbb{P}(Y=1) + \left(1 - F_{X|Y}(\underline{x}|0)\right)\mathbb{P}(Y=0)} \\
&\leq 2 \max\left\{ \frac{f_{X|Y}(\underline{x}|1)}{1 - F_{X|Y}(\underline{x}|1)}, \frac{f_{X|Y}(\underline{x}|0)}{1 - F_{X|Y}(\underline{x}|0)} \right\}
\end{aligned}$$

$$= \frac{2}{\underline{x}} \max\{\alpha^{(1)}, \alpha^{(0)}\}(1 + o(1)),$$

as $\underline{x} \to \infty$. The second line plugs in the expressions of $f(x)$ and $F(x)$. The third line is by the properties of weighted averages. The last line is by the RV approximation of the pdf in (6). Therefore, the tail average of the partial effects exists, and converges to 0 as $\underline{x} \to \infty$. ■

Proof of Theorem 2.1.

This proof builds on Wang and Tsai (2009). Recall that the log likelihood function is given by

$$\ell_N^{(y)}\left(\theta^{(y)}\right) = \sum_{i=1}^{N} \left(\log Z_i'\theta^{(y)} - Z_i'\theta^{(y)} \log \frac{X_i}{\underline{x}_N^{(y)}} \right) \mathbf{1}\left\{\Xi_{N,i}^{(y)}\right\},$$

and its first and second derivatives, i.e., the score and Hessian, are

$$S_N^{(y)}\left(\theta^{(y)}\right) = \frac{\partial \ell_N^{(y)}\left(\theta^{(y)}\right)}{\partial \theta^{(y)}} = \sum_{i=1}^{N} \left(\frac{1}{Z_i'\theta^{(y)}} - \log \frac{X_i}{\underline{x}_N^{(y)}} \right) Z_i \mathbf{1}\left\{\Xi_{N,i}^{(y)}\right\}, \tag{A.4}$$

$$H_N^{(y)}\left(\theta^{(y)}\right) = \frac{\partial^2 \ell_N^{(y)}\left(\theta^{(y)}\right)}{\partial \theta^{(y)} \partial \theta^{(y)\prime}} = -\sum_{i=1}^{N} \frac{Z_i Z_i'}{\left(Z_i'\theta^{(y)}\right)^2} \mathbf{1}\left\{\Xi_{N,i}^{(y)}\right\}. \tag{A.5}$$

Also note that as $0 < \xi_N < 1$, Assumption 2.3(b) implies that $\underline{x}_N^{(y)} \to \infty$ as $N \to \infty$. Combined with the tail conditions in Assumption 2.2(a,b), this further leads to that for $x \geq \underline{x}_N^{(y)}$,

$$C^{(y)}\left(Z_i\right)(x)^{-\alpha^{(y)}(Z_i)} \left(1 + D^{(y)}\left(Z_i\right)(x)^{-\beta^{(y)}(Z_i)} + r_i^{(y)}\left(x, Z_i\right)\right) \tag{A.6}$$

$$= C^{(y)}\left(Z_i\right)(x)^{-\alpha^{(y)}(Z_i)}\left(1 + o(1)\right),$$

$$D^{(y)}\left(Z_i\right)(x)^{-\beta^{(y)}(Z_i)} + r_i^{(y)}\left(x, Z_i\right) = D^{(y)}\left(Z_i\right)(x)^{-\beta^{(y)}(Z_i)}\left(1 + o(1)\right), \tag{A.7}$$

as $N \to \infty$, almost surely.

Note that $N^{(y)}$ is a random variable, so we introduce $\xi_N^{(y)}$ in (13), a non-random sequence representing the asymptotic proportion of tail observations for $Y_i = y$. Following from (A.6),

$$\mathbb{P}\left(\Xi_{N,i}^{(y)}\right) = \mathbb{P}(X_i > \underline{x}_N^{(y)} | Y_i = y)\mathbb{P}(Y_i = y) \tag{A.8}$$

$$= \mathbb{E}\left[C^{(y)}(Z_i)\left(\underline{x}_N^{(y)}\right)^{-\alpha^{(y)}(Z_i)}\right](1 + o(1)) \cdot \mathbb{P}(Y_i = y)$$

$$= \xi_N^{(y)} \left(1 + o(1)\right).$$

Part 1. Let

$$S_{N,i}^{(y)} = \frac{1}{\sqrt{\xi_N^{(y)}}} \left(H_{N0}^{(y)}\right)^{-1/2} \left(\frac{1}{Z_i'\theta_0^{(y)}} - \log \frac{X_i}{\underline{x}_N^{(y)}}\right) Z_i \mathbf{1}\left\{\Xi_{N,i}^{(y)}\right\}, \tag{A.9}$$

which is i.i.d. across i by Assumption 2.1(a). Then, according to the score defined in (A.4),

$$\frac{1}{\sqrt{\xi_N^{(y)}}} \left(H_{N0}^{(y)}\right)^{-1/2} S_N^{(y)}\left(\theta_0^{(y)}\right) = \sum_{i=1}^{N} S_{N,i}^{(y)},$$

and we can apply the central limit theorem (CLT) to obtain its asymptotic normality.

First, for the mean,

$$\mathbb{E}\left[S_{N,i}^{(y)}\right] = \frac{1}{\sqrt{\xi_N^{(y)}}} \left(H_{N0}^{(y)}\right)^{-1/2} \mathbb{E}\left[\left(\frac{1}{Z_i'\theta_0^{(y)}} - \log \frac{X_i}{\underline{x}_N^{(y)}}\right) Z_i \mathbf{1}\left\{\Xi_{N,i}^{(y)}\right\}\right].$$

For the first term in the expectation,

$$\begin{aligned}
\mathbb{E}\left[\frac{1}{Z_i'\theta_0^{(y)}} Z_i \mathbf{1}\left\{\Xi_{N,i}^{(y)}\right\}\right] &= \mathbb{E}\left[\frac{1}{Z_i'\theta_0^{(y)}} Z_i \mathbb{P}\left(\Xi_{N,i}^{(y)} \middle| Z_i\right)\right] \\
&= \mathbb{E}\left[\frac{1}{Z_i'\theta_0^{(y)}} Z_i C^{(y)}\left(Z_i\right) \left(\underline{x}_N^{(y)}\right)^{-\alpha^{(y)}(Z_i)}\right] \\
&\quad + \mathbb{E}\left[\frac{1}{Z_i'\theta_0^{(y)}} Z_i C^{(y)}\left(Z_i\right) D^{(y)}\left(Z_i\right) \left(\underline{x}_N^{(y)}\right)^{-\alpha^{(y)}(Z_i)-\beta^{(y)}(Z_i)}\right] \left(1 + o(1)\right),
\end{aligned}$$

where we substitute the tail approximation in Assumption 2.2(a) and apply the supremum condition of the remainder term in Assumption 2.2(b) as in equation (A.7). For the second term in the expectation,

$$\begin{aligned}
&\mathbb{E}\left[\log \frac{X_i}{\underline{x}_N^{(y)}} Z_i \mathbf{1}\left\{\Xi_{N,i}^{(y)}\right\}\right] \\
&= \mathbb{E}\left[Z_i \int_0^\infty \mathbb{P}\left(\log \frac{X_i}{\underline{x}_N^{(y)}} > s \middle| Z_i\right) ds\right]
\end{aligned}$$

$$
\begin{aligned}
&= \mathbb{E}\left[Z_i \int_0^\infty \mathbb{P}\left(X_i > \underline{x}_N^{(y)} e^s \,\middle|\, Z_i\right) ds\right] \\
&= \mathbb{E}\left[Z_i C^{(y)}\left(Z_i\right)\left(\underline{x}_N^{(y)}\right)^{-\alpha^{(y)}(Z_i)} \int_0^\infty e^{-s\alpha^{(y)}(Z_i)} ds\right] \\
&\quad + \mathbb{E}\left[Z_i C^{(y)}\left(Z_i\right) D^{(y)}\left(Z_i\right)\left(\underline{x}_N^{(y)}\right)^{-\alpha^{(y)}(Z_i)-\beta^{(y)}(Z_i)} \int_0^\infty e^{-s\left(\alpha^{(y)}(Z_i)+\beta^{(y)}(Z_i)\right)} ds\right](1+o(1)) \\
&= \mathbb{E}\left[Z_i C^{(y)}\left(Z_i\right)\left(\underline{x}_N^{(y)}\right)^{-\alpha^{(y)}(Z_i)} \cdot \frac{1}{\alpha^{(y)}\left(Z_i\right)}\right] \\
&\quad + \mathbb{E}\left[Z_i C^{(y)}\left(Z_i\right) D^{(y)}\left(Z_i\right)\left(\underline{x}_N^{(y)}\right)^{-\alpha^{(y)}(Z_i)-\beta^{(y)}(Z_i)} \cdot \frac{1}{\alpha^{(y)}\left(Z_i\right)+\beta^{(y)}\left(Z_i\right)}\right](1+o(1)),
\end{aligned}
$$

The first equality follows from that for a generic random variable $X > 0$, $\mathbb{E}[X] = \int_0^\infty \mathbb{P}(X > s)ds$, which is given by integration by parts. The third equality is again by the tail approximation in Assumption 2.2(a,b) and equation (A.7). For both terms, we re-derive and fix minor typos regarding the $o(1)$ terms in equations (B.1) and (B.2) in Wang and Tsai (2009). Plugging both terms back into the mean expression, and noting that $\alpha^{(y)}\left(Z_i\right) = Z_i'\theta_0^{(y)}$, we have that as $N \to \infty$,

$$
\begin{aligned}
&\mathbb{E}\left[S_{N,i}^{(y)}\right] \\
&= \frac{1}{\sqrt{\xi_N^{(y)}}}\left(H_{N0}^{(y)}\right)^{-1/2} \mathbb{E}\left[Z_i \frac{\beta^{(y)}\left(Z_i\right) C^{(y)}\left(Z_i\right) D^{(y)}\left(Z_i\right)}{\alpha^{(y)}\left(Z_i\right)\left(\alpha^{(y)}\left(Z_i\right)+\beta^{(y)}\left(Z_i\right)\right)}\left(\underline{x}_N^{(y)}\right)^{-\alpha^{(y)}(Z_i)-\beta^{(y)}(Z_i)}\right](1+o(1)) \\
&= o\left(\frac{1}{\sqrt{N}}\right),
\end{aligned}
\tag{A.10}
$$

where the last line is by Assumption 2.3(b).

Second, for the variance,

$$
\begin{aligned}
\mathbb{V}\left[S_{N,i}^{(y)}\right] &= \mathbb{E}\left[S_{N,i}^{(y)} S_{N,i}^{(y)\prime}\right] - \mathbb{E}\left[S_{N,i}^{(y)}\right]\mathbb{E}\left[S_{N,i}^{(y)}\right]' \\
&= \frac{1}{\xi_N^{(y)}}\left(H_{N0}^{(y)}\right)^{-1/2} \mathbb{E}\left[\left(\frac{1}{Z_i'\theta_0^{(y)}} - \log\frac{X_i}{\underline{x}_N^{(y)}}\right)^2 Z_i Z_i' \mathbf{1}\left\{\Xi_{N,i}^{(y)}\right\}\right]\left(H_{N0}^{(y)}\right)^{-1/2} + o\left(\frac{1}{N}\right).
\end{aligned}
$$

In the second equality, the first term is by the definition of $S_{N,i}^{(y)}$ in (A.9), and the second term is by $\mathbb{E}\left[S_{N,i}^{(y)}\right] = o\left(\frac{1}{\sqrt{N}}\right)$ in (A.10). For the expectation term, by the law of total expectation,

$$\mathbb{E}\left[\left(\frac{1}{Z_i'\theta_0^{(y)}}-\log\frac{X_i}{\underline{x}_N^{(y)}}\right)^2 Z_iZ_i'\mathbf{1}\left\{\Xi_{N,i}^{(y)}\right\}\right]$$
$$=\mathbb{E}\left[\mathbb{E}\left[\left(1-Z_i'\theta_0^{(y)}\log\frac{X_i}{\underline{x}_N^{(y)}}\right)^2\middle|Z_i,\Xi_{N,i}^{(y)}\right]\frac{1}{\left(Z_i'\theta_0^{(y)}\right)^2}Z_iZ_i'\mathbf{1}\left\{\Xi_{N,i}^{(y)}\right\}\right].$$

Let us first consider the inner expectation $\mathbb{E}\left[\left(1-Z_i'\theta_0^{(y)}\log\frac{X_i}{\underline{x}_N^{(y)}}\right)^2\middle|Z_i,\Xi_{N,i}^{(y)}\right]$. Denote $\tilde{X}_i = Z_i'\theta_0^{(y)}\log\frac{X_i}{\underline{x}_N^{(y)}}$. Then, by the tail approximation in Assumption 2.2(a,b) and equation (A.6), we have that $1-F_{\tilde{X}_i|Z_i,\Xi_{N,i}^{(y)}}\left(\tilde{x}\,\middle|Z_i,\Xi_{N,i}^{(y)}\right)=\exp(-\tilde{x})\left(1+o(1)\right)$ for $\tilde{x}>0$, almost surely. Intuitively, $\tilde{X}_i$ approximately follows a standard exponential distribution given $\left\{Z_i,\Xi_{N,i}^{(y)}\right\}$. Then, following from integration by parts,

$$\begin{aligned}
\mathbb{E}\left[\tilde{X}_i\,\middle|Z_i,\Xi_{N,i}^{(y)}\right] &= \int_0^\infty 1-F_{\tilde{X}_i|Z_i,\Xi_{N,i}^{(y)}}\left(s\,\middle|Z_i,\Xi_{N,i}^{(y)}\right)ds\\
&= \int_0^\infty \exp(-s)\left(1+o(1)\right)ds = 1+o(1),\\
\mathbb{E}\left[\tilde{X}_i^2\,\middle|Z_i,\Xi_{N,i}^{(y)}\right] &= \int_0^\infty 2s\left[1-F_{\tilde{X}_i|Z_i,\Xi_{N,i}^{(y)}}\left(s\,\middle|Z_i,\Xi_{N,i}^{(y)}\right)\right]ds\\
&= \int_0^\infty 2s\exp(-s)\left(1+o(1)\right)ds = 2\left(1+o(1)\right).
\end{aligned}$$

Therefore, we have that

$$\mathbb{E}\left[\left(1-Z_i'\theta_0^{(y)}\log\frac{X_i}{\underline{x}_N^{(y)}}\right)^2\middle|Z_i,\Xi_{N,i}^{(y)}\right]=\mathbb{E}\left[\left(1-\tilde{X}_i\right)^2\middle|Z_i,\Xi_{N,i}^{(y)}\right]=1+o(1),$$

and thus

$$\begin{aligned}
\mathbb{E}\left[\left(\frac{1}{Z_i'\theta_0^{(y)}}-\log\frac{X_i}{\underline{x}_N^{(y)}}\right)^2 Z_iZ_i'\mathbf{1}\left\{\Xi_{N,i}^{(y)}\right\}\right] &= \mathbb{E}\left[\frac{1}{\left(Z_i'\theta_0^{(y)}\right)^2}Z_iZ_i'\mathbf{1}\left\{\Xi_{N,i}^{(y)}\right\}\right]\left(1+o(1)\right)\\
&= \mathbb{E}\left[\frac{1}{\left(Z_i'\theta_0^{(y)}\right)^2}Z_iZ_i'\middle|\,\Xi_{N,i}^{(y)}\right]\mathbb{P}\left(\Xi_{N,i}^{(y)}\right)\left(1+o(1)\right)
\end{aligned}$$

$$= H_{N0}^{(y)} \xi_N^{(y)} \left(1 + o_p(1)\right),$$

where the second equality is by the definition of $H_{N0}^{(y)}$ in Assumption 2.3(c) and the approximation to $\mathbb{P}\left(\Xi_{N,i}^{(y)}\right)$ in (A.8). Substituting this back to the expression of $\mathbb{V}\left[S_{N,i}^{(y)}\right]$, we have that as $N \to \infty$,

$$\mathbb{V}\left[S_{N,i}^{(y)}\right] = \mathcal{I}_{d_Z}\left(1 + o(1)\right) + o\left(\frac{1}{N}\right) \to \mathcal{I}_{d_Z}. \tag{A.11}$$

Therefore, as $\theta_0^{(y)} \in \operatorname{int}\left(\Theta^{(y)}\right)$, by the CLT, we have that as $N \to \infty$,

$$\frac{1}{\sqrt{N\xi_N^{(y)}}}\left(H_{N0}^{(y)}\right)^{-1/2} S_N^{(y)}\left(\theta_0^{(y)}\right) = \frac{1}{\sqrt{N}}\sum_{i=1}^{N} S_{N,i}^{(y)} \xrightarrow{d} \mathcal{N}\left(0, \mathcal{I}_{d_Z}\right), \tag{A.12}$$

where the mean and variance of $S_{N,i}^{(y)}$ are given in (A.10) and (A.11), respectively.

Part 2. Similarly, for the Hessian matrix,

$$\begin{aligned}
&\frac{1}{N\xi_N^{(y)}}\left(H_{N0}^{(y)}\right)^{-1/2} \cdot H_N^{(y)}\left(\theta_0^{(y)}\right) \cdot \left(H_{N0}^{(y)}\right)^{-1/2} \qquad\qquad (\text{A.13})\\
&= \frac{1}{N\xi_N^{(y)}}\left(H_{N0}^{(y)}\right)^{-1/2} \cdot \left[-\sum_{i=1}^{N} \frac{Z_i Z_i'}{\left(Z_i'\theta_0^{(y)}\right)^2}\mathbf{1}\left\{\Xi_{N,i}^{(y)}\right\}\right] \cdot \left(H_{N0}^{(y)}\right)^{-1/2}\\
&= -\frac{1}{\xi_N^{(y)}}\left(H_{N0}^{(y)}\right)^{-1/2} \cdot \mathbb{E}\left[\frac{Z_i Z_i'}{\left(Z_i'\theta_0^{(y)}\right)^2}\mathbf{1}\left\{\Xi_{N,i}^{(y)}\right\}\right] \cdot \left(H_{N0}^{(y)}\right)^{-1/2} \cdot \left(1 + o_p(1)\right)\\
&= -\frac{1}{\xi_N^{(y)}}\left(H_{N0}^{(y)}\right)^{-1/2} \cdot \mathbb{E}\left[\frac{Z_i Z_i'}{\left(Z_i'\theta_0^{(y)}\right)^2}\middle| \Xi_{N,i}^{(y)}\right]\mathbb{P}\left(\Xi_{N,i}^{(y)}\right) \cdot \left(H_{N0}^{(y)}\right)^{-1/2} \cdot \left(1 + o_p(1)\right)\\
&= -\frac{1}{\xi_N^{(y)}}\left(H_{N0}^{(y)}\right)^{-1/2} \cdot H_{N0}^{(y)}\xi_N^{(y)} \cdot \left(H_{N0}^{(y)}\right)^{-1/2} \cdot \left(1 + o_p(1)\right)\\
&\xrightarrow{p} -\mathcal{I}_{d_Z}.
\end{aligned}$$

The first equality is by the definition of the Hessian matrix (A.5). The second equality follows from the law of large numbers (LLN), where the finite mean is given by Assumption 2.3(c). The fourth equality is by the definition of $H_{N0}^{(y)}$ in Assumption 2.3(c) and the approximation to $\mathbb{P}\left(\Xi_{N,i}^{(y)}\right)$ in (A.8).

Part 3. For $y \in \{0,1\}$, let $\zeta_N = \left(H_{N0}^{(y)}\right)^{1/2}\left(\theta^{(y)} - \theta_0^{(y)}\right)$, $\zeta_{N0} = \left(H_{N0}^{(y)}\right)^{1/2}\theta_0^{(y)}$, and $W_i = \left(H_{N0}^{(y)}\right)^{-1/2} Z_i$. Then, the log likelihood function can be rewritten as

$$\tilde{\ell}_N^{(y)}\left(\zeta_N\right) = \sum_{i=1}^{N}\left(\log\left(W_i'\left(\zeta_N + \zeta_{N0}\right)\right) - W_i'\left(\zeta_N + \zeta_{N0}\right)\log\frac{X_i}{\underline{x}_N^{(y)}}\right)\mathbf{1}\left\{\Xi_{N,i}^{(y)}\right\},$$

the corresponding score and Hessian are denoted by $\tilde{S}_N^{(y)}\left(\zeta_N\right)$ and $\tilde{H}_N^{(y)}\left(\zeta_N\right)$, respectively, and the MLE estimate is denoted by $\hat{\zeta}_N$.

First, the new Hessian matrix is given by

$$\tilde{H}_N^{(y)}\left(\zeta_N\right) = -\sum_{i=1}^{N}\frac{W_i W_i'}{\left(W_i'\left(\zeta_N + \zeta_{N0}\right)\right)^2}\mathbf{1}\left\{\Xi_{N,i}^{(y)}\right\}.$$

It is positive definite for all $\zeta_N = \left(H_{N0}^{(y)}\right)^{1/2}\left(\theta^{(y)} - \theta_0^{(y)}\right)$ with $\theta^{(y)} \in \Theta^{(y)}$, as $\Theta^{(y)}$ is a convex cone such that for all $\theta^{(y)} \in \Theta^{(y)}$, $Z_i'\theta^{(y)} = W_i'\left(\zeta_N + \zeta_{N0}\right) > 0$ almost surely: see the discussion after Assumption 2.2. Then, the log likelihood function is $\tilde{\ell}_N^{(y)}\left(\zeta_N\right)$ is strictly concave over its domain, and the MLE estimate $\hat{\zeta}_N$ is unique.

Second, let

$$\mathcal{U}_{N,C}^{(y)} = \left\{u \in \mathbb{R}^{d_Z} : \frac{1}{\sqrt{N\xi_N^{(y)}}}\left(H_{N0}^{(y)}\right)^{-1/2} u + \theta_0^{(y)} \in \Theta^{(y)} \text{ and } \|u\| = C\right\}.$$

Note that $\mathcal{U}_{N,C}^{(y)} \neq \emptyset$ for any $C > 0$. This follows from three facts: first, $\Theta^{(y)}$ is a convex cone; second, $\operatorname{int}\left(\Theta^{(y)}\right) \neq \emptyset$ as $\theta_0^{(y)} \in \operatorname{int}\left(\Theta^{(y)}\right)$; and third, $H_{N0}^{(y)}$ is finite and full rank by Assumption 2.3(c). Then, let u be an arbitrary non-random vector in $\mathcal{U}_{N,C}^{(y)}$. Applying the second-order Taylor expansion of $\tilde{\ell}_N^{(y)}\left(\frac{u}{\sqrt{N\xi_N^{(y)}}}\right)$ around $\zeta_N = 0$, we have that

$$\tilde{\ell}_N^{(y)}\left(\frac{u}{\sqrt{N\xi_N^{(y)}}}\right) - \tilde{\ell}_N^{(y)}(0) = \frac{1}{\sqrt{N\xi_N^{(y)}}}u'\tilde{S}_N^{(y)}(0) + \frac{1}{2N\xi_N^{(y)}}u'\tilde{H}_N^{(y)}(0)\,u + o_p(1)$$

From (A.12) and (A.13), we have that as $N \to \infty$,

$$\frac{1}{\sqrt{N\xi_N^{(y)}}}\tilde{S}_N^{(y)}(0) = \frac{1}{\sqrt{N\xi_N^{(y)}}}S_N^{(y)}\left(\theta_0^{(y)}\right) \xrightarrow{d} \mathcal{N}\left(0, \mathcal{I}_{d_Z}\right), \tag{A.14}$$

$$\frac{1}{N\xi_N^{(y)}}\tilde{H}_N^{(y)}(0) = \frac{1}{N\xi_N^{(y)}}\left(H_{N0}^{(y)}\right)^{-1/2} \cdot H_N^{(y)}\left(\theta_0^{(y)}\right) \cdot \left(H_{N0}^{(y)}\right)^{-1/2} \xrightarrow{p} -\mathcal{I}_{d_Z}. \tag{A.15}$$

This implies that when C is large enough, the quadratic term dominates the linear one with an arbitrarily large probability. That is, for any $\varepsilon > 0$, there exists a $C > 0$ such that

$$\limsup_N \mathbb{P}\left(\sup_{u \in \mathcal{U}_{N,C}^{(y)}} \tilde{\ell}_N^{(y)}\left(\frac{u}{\sqrt{N\xi_N^{(y)}}}\right) < \tilde{\ell}_N^{(y)}(0)\right) > 1 - \varepsilon.$$

Therefore, $\tilde{\ell}_N^{(y)}(\cdot)$ must have at least one local maximizer, which is of order $O_p\left(\sqrt{N\xi_N^{(y)}}\right)$. And by the uniqueness of the MLE, this local maximizer is the global maximizer, and $\hat{\zeta}_N = O_p\left(\sqrt{N\xi_N^{(y)}}\right)$.

Finally, by the first order condition, $\tilde{S}_N^{(y)}\left(\hat{\zeta}_N\right) = 0$. Applying the first-order Taylor expansion around $\zeta_N = 0$, we have that

$$0 = \tilde{S}_N^{(y)}\left(\hat{\zeta}_N\right) = \tilde{S}_N^{(y)}(0) + \tilde{H}_N^{(y)}(0)\,\hat{\zeta}_N + o_p(1),$$

which implies that

$$\sqrt{N\xi_N^{(y)}}\hat{\zeta}_N = -\left(\frac{1}{N\xi_N^{(y)}}\tilde{H}_N^{(y)}(0)\right)^{-1} \cdot \frac{1}{\sqrt{N\xi_N^{(y)}}}\left(\tilde{S}_N^{(y)}(0) + o_p(1)\right) \xrightarrow{d} \mathcal{N}\left(0, \mathcal{I}_{d_Z}\right),$$

following from (A.14) and (A.15). This is equivalent to

$$\sqrt{N\xi_N^{(y)}}\left(H_{N0}^{(y)}\right)^{1/2}\left(\hat{\theta}^{(y)} - \theta_0^{(y)}\right) \xrightarrow{d} \mathcal{N}\left(0, \mathcal{I}_{d_Z}\right).$$

In addition, the asymptotic independence between $\hat{\theta}^{(1)}$ and $\hat{\theta}^{(0)}$ follows from i.i.d. observations across i in Assumption 2.1(a). This completes the proof of the theorem. ■

Proposition A.2 (Cross-sectional data: proportion of tail observations) *Suppose As-*

sumptions 2.1–2.3 hold. For $y \in \{0,1\}$, as $N \to \infty$,

$$\frac{N^{(y)}}{N} = \xi_N^{(y)} \left(1 + o_p(1)\right).$$

Proof. First, following from Assumption 2.2(a,b) and equation (A.8), we have that

$$\mathbb{E}\left[\frac{N^{(y)}}{N}\right] = \mathbb{P}\left(\Xi_{N,i}^{(y)}\right) = \xi_N^{(y)} \left(1 + o(1)\right). \tag{A.16}$$

Second, the variance of $N^{(y)}/N$ is bounded by

$$\begin{aligned}\mathbb{V}\left[\frac{N^{(y)}}{N}\right] &= \frac{1}{N}\mathbb{P}\left(\Xi_{N,i}^{(y)}\right)\left(1 - \mathbb{P}\left(\Xi_{N,i}^{(y)}\right)\right) \\ &\leq \frac{1}{N}\mathbb{P}\left(\Xi_{N,i}^{(y)}\right) = \frac{1}{N}\xi_N^{(y)} \left(1 + o(1)\right).\end{aligned} \tag{A.17}$$

Combining both, we obtain that as $N \to \infty$,

$$\begin{aligned}\mathbb{E}\left[\left(\frac{N^{(y)}}{N\xi_N^{(y)}} - 1\right)^2\right] &= \mathbb{V}\left[\frac{N^{(y)}}{N\xi_N^{(y)}}\right] + o(1) = \frac{1}{\left(\xi_N^{(y)}\right)^2}\mathbb{V}\left[\frac{N^{(y)}}{N}\right] + o(1) \\ &\leq \frac{1}{N\xi_N^{(y)}}\left(1 + o(1)\right) + o(1) \to 0,\end{aligned}$$

where the first line is by the expression for the mean (A.16), the second line is by the bound on the variance (A.17), and the last convergence is by $N\xi_N^{(y)} \to \infty$ in Assumption 2.3(a). This part of the proof is slightly different from Wang and Tsai (2009): first, we explicitly account for the additive $o(1)$ term; second, as $N^{(y)}$ is a random variable, we let $N\xi_N^{(y)} \to \infty$ instead.

By Chebyshev's inequality, convergence in the second moment implies convergence in probability. Then, we have that as $N \to \infty$, $\frac{N^{(y)}}{N\xi_N^{(y)}} \xrightarrow{p} 1$, and equivalently, $\frac{N^{(y)}}{N} = \xi_N^{(y)}\{1 + o_p(1)\}$.

■

Remark A.1 Combining Theorem 2.1 with Proposition A.2, we have that as $N \to \infty$,

$$\sqrt{N^{(y)}}\left(H_{N0}^{(y)}\right)^{1/2}\left(\hat{\theta}^{(y)} - \theta_0^{(y)}\right) \xrightarrow{d} \mathcal{N}\left(0, \mathcal{I}_{d_Z}\right).$$

A.2 Proofs for Section 3

We first present two lemmas for the LLN and the CLT with i.n.i.d. observations under rare events, and then provide the proof of our main result in Section 3: Theorem 3.1.

Let us consider a generic setup with i.n.i.d. observations under rare events. Let $\{X_{N,i} : 1 \leq i \leq N, N \geq 1\}$ be a triangular array of i.n.i.d. d_X-dimensional random vectors, and $\{\Xi_{N,i} : 1 \leq i \leq N, N \geq 1\}$ be a triangular array of events with $\mathbb{P}_i(\Xi_{N,i}) = p_{N,i} \to 0$ as $N \to \infty$. Subsript i denotes quantities given potential individual-specific parameters. Also define $p_N = \frac{1}{N}\sum_{i=1}^N p_{N,i}$.

Lemma A.2 (Rare event LLN for i.n.i.d. observations) *Suppose that:*

(a) $Np_N \to \infty$, *as* $N \to \infty$.

(b) $\mathbb{E}_i\left[\|X_{N,i}\|^2 | \Xi_{N,i}\right] \leq M < \infty$, *for sufficiently large* N *and all* i.

Then, $\mathbb{E}_i\left[X_{N,i}|\Xi_{N,i}\right] = \mu_{N,i}$ *exists, and as* $N \to \infty$,

$$\frac{1}{Np_N}\sum_{i=1}^N X_{N,i}\mathbf{1}\{\Xi_{N,i}\} - \frac{1}{Np_N}\sum_{i=1}^N p_{N,i}\mu_{N,i} \xrightarrow{p} 0.$$

Proof. Let $S_N = \sum_{i=1}^N \left(X_{N,i}\mathbf{1}\{\Xi_{N,i}\} - p_{N,i}\mu_{N,i}\right)$. We want to show that $S_N/(Np_N) \xrightarrow{p} 0$.

By the multivariate Chebyshev's inequality, for any $\varepsilon > 0$,

$$\mathbb{P}\left(\frac{\|S_N\|}{Np_N} > \varepsilon\right) \leq \frac{\operatorname{tr}\left(\mathbb{V}\left[S_N\right]\right)}{\varepsilon^2\left(Np_N\right)^2}.$$

Now, let's compute the trace of $\mathbb{V}\left[S_N\right]$:

$$\begin{aligned}
\operatorname{tr}\left(\mathbb{V}\left[S_N\right]\right) &= \sum_{i=1}^N \operatorname{tr}\left(\mathbb{V}_i\left[X_{N,i}\mathbf{1}\{\Xi_{N,i}\}\right]\right) \\
&\leq \sum_{i=1}^N \mathbb{E}_i\left[\|X_{N,i}\|^2\mathbf{1}\{\Xi_{N,i}\}\right] = \sum_{i=1}^N \mathbb{E}_i\left[\|X_{N,i}\|^2 \middle| \Xi_{N,i}\right]\mathbb{P}\left(\Xi_{N,i}\right) \\
&\leq M\sum_{i=1}^N p_{N,i} = MNp_N,
\end{aligned}$$

where the first line is by independence, the second line is by the definition of variance, and the third line is by condition (b). Substituting this back into the multivariate Chebyshev's inequality:

$$\mathbb{P}\left(\frac{\|S_N\|}{Np_N} > \varepsilon\right) \leq \frac{M}{\varepsilon^2 Np_N} \to 0,$$

as $N \to \infty$, following from condition (a).

Therefore, $S_N/(Np_N) \xrightarrow{p} 0$, which is equivalent to

$$\frac{1}{Np_N}\sum_{i=1}^{N} X_{N,i}\mathbf{1}\{\Xi_{N,i}\} - \frac{1}{Np_N}\sum_{i=1}^{N} p_{N,i}\mu_{N,i} \xrightarrow{p} 0.$$

■

Lemma A.3 (Rare event CLT for i.n.i.d. observations) *Suppose that:*

(a) $Np_N \to \infty$*, as* $N \to \infty$*.*

(b) $\mathbb{E}_i\left[\|X_{N,i}\|^{2+\kappa}|\,\Xi_{N,i}\right] \leq M < \infty$ *for some* $\kappa > 0$*, for sufficiently large* N *and all* i*.*

(c) Let $\mathbb{V}_i\left[X_{N,i}|\,\Xi_{N,i}\right] = \Sigma_{N,i}$*. Its smallest eigenvalue* $\lambda_{\min}\left(\Sigma_{N,i}\right) \geq \underline{\sigma}_N^2$*, where* $\underline{\sigma}_N^2\left(Np_N\right)^{\kappa/(2+\kappa)} \to \infty$ *as* $N \to \infty$*, for all* i*.*

Let $\mathbb{E}_i\left[X_{N,i}|\Xi_{N,i}\right] = \mu_{N,i}$*,* $S_N = \sum_{i=1}^{N}\left(X_{N,i}\mathbf{1}\{\Xi_{N,i}\} - p_{N,i}\mu_{N,i}\right)$*, and* $\Sigma_N = \sum_{i=1}^{N} p_{N,i}\Sigma_{N,i}$*. Then, as* $N \to \infty$*,*

$$\Sigma_N^{-1/2} S_N \xrightarrow{d} N\left(0, \mathcal{I}_{d_X}\right).$$

Proof. Let $X_{N,i}^* = X_{N,i}\mathbf{1}\{\Xi_{N,i}\} - p_{N,i}\mu_{N,i}$. So we have $\mathbb{E}_i\left[X_{N,i}^*\right] = 0$ and $\mathbb{V}_i\left[X_{N,i}^*\right] = p_{N,i}\Sigma_{N,i}$. We'll use the Lindeberg-Feller CLT for triangular arrays. We need to verify two conditions:

1. Lindeberg condition: for any $\varepsilon > 0$, as $N \to \infty$,

$$\frac{1}{\operatorname{tr}\left(\Sigma_N\right)}\sum_{i=1}^{N}\mathbb{E}_i\left[\left\|X_{N,i}^*\right\|^2\mathbf{1}\left\{\left\|X_{N,i}^*\right\| > \varepsilon\sqrt{\operatorname{tr}\left(\Sigma_N\right)}\right\}\right] \to 0.$$

2. Variance condition: as $N \to \infty$,

$$\frac{\max_{1\leq i\leq N}\operatorname{tr}\left(\mathbb{V}_i\left[X_{N,i}^*\right]\right)}{\operatorname{tr}\left(\Sigma_N\right)} \to 0.$$

First, for the Lindeberg condition,

$$\begin{aligned}\mathbb{E}_i\left[\left\|X_{N,i}^*\right\|^2\mathbf{1}\left\{\left\|X_{N,i}^*\right\|>\varepsilon\sqrt{\operatorname{tr}\left(\Sigma_N\right)}\right\}\right] &\leq \mathbb{E}_i\left[\left\|X_{N,i}\right\|^2\mathbf{1}\{\Xi_{N,i}\}\mathbf{1}\left\{\left\|X_{N,i}^*\right\|>\varepsilon\sqrt{\operatorname{tr}\left(\Sigma_N\right)}\right\}\right]\\ &\leq \frac{\mathbb{E}_i\left[\left\|X_{N,i}\right\|^{2+\kappa}\mathbf{1}\{\Xi_{N,i}\}\right]}{\left(\varepsilon\sqrt{\operatorname{tr}\left(\Sigma_N\right)}\right)^{\kappa}}\\ &\leq \frac{Mp_{N,i}}{\left(\varepsilon\sqrt{\operatorname{tr}\left(\Sigma_N\right)}\right)^{\kappa}},\end{aligned}$$

where the first line is by the definition of $X_{N,i}^*$, the second line is by the indicator function on $\left\|X_{N,i}^*\right\|>\varepsilon\sqrt{\operatorname{tr}\left(\Sigma_N\right)}$, and the third line is by condition (b). Also, we can obtain a lower bound of $\operatorname{tr}\left(\Sigma_N\right)$ in the denominator based on condition (c):

$$\operatorname{tr}\left(\Sigma_N\right)=\sum_{i=1}^N p_{N,i}\operatorname{tr}\left(\Sigma_{N,i}\right)\geq\sum_{i=1}^N p_{N,i}\cdot\min_{1\leq i\leq N}\operatorname{tr}\left(\Sigma_{N,i}\right)=Np_N\cdot d_X\underline{\sigma}_N^2. \tag{A.18}$$

Therefore, we have that

$$\begin{aligned}&\frac{1}{\operatorname{tr}\left(\Sigma_N\right)}\sum_{i=1}^N\mathbb{E}_i\left[\left\|X_{N,i}^*\right\|^2\mathbf{1}\left\{\left\|X_{N,i}^*\right\|>\varepsilon\sqrt{\operatorname{tr}\left(\Sigma_N\right)}\right\}\right]\\ &\leq\frac{MNp_N}{\varepsilon^\kappa\left(\operatorname{tr}\left(\Sigma_N\right)\right)^{1+\kappa/2}}\leq\frac{MNp_N}{\varepsilon^\kappa\left(Np_N\cdot d_X\underline{\sigma}_N^2\right)^{1+\kappa/2}}\\ &=\frac{MNp_N}{\varepsilon^\kappa d_X^{1+\kappa/2}\left(Np_N\right)^{\kappa/2}\left(\underline{\sigma}_N^2\right)^{1+\kappa/2}}\to 0,\end{aligned}$$

where the second inequality is by (A.18) and the convergence to 0 is by condition (c).

Second, for the variance condition,

$$\begin{aligned}\frac{\max_{1\leq i\leq N}\operatorname{tr}\left(\mathbb{V}_i\left[X_{N,i}^*\right]\right)}{\operatorname{tr}\left(\Sigma_N\right)} &= \frac{\max_{1\leq i\leq N}p_{N,i}\operatorname{tr}\left(\Sigma_{N,i}\right)}{\sum_{i=1}^N p_{N,i}\operatorname{tr}\left(\Sigma_{N,i}\right)}\\ &\leq\frac{\max_{1\leq i\leq N}p_{N,i}\cdot\max_{1\leq i\leq N}\operatorname{tr}\left(\Sigma_{N,i}\right)}{Np_N\min_{1\leq i\leq N}\operatorname{tr}\left(\Sigma_{N,i}\right)}\\ &\leq\frac{M^{2/(2+\kappa)}}{Np_N\cdot d_X\underline{\sigma}_N^2}\to 0.\end{aligned}$$

By conditions (b) and (c), $d_X\underline{\sigma}_N^2\leq\operatorname{tr}(\Sigma_{N,i})\leq M^{2/(2+\kappa)}$ for sufficiently large N and all i, so we obtain the inequality in the last line. Further by conditions (a) and (b), we have that in the denominator, $\underline{\sigma}_N^2 Np_N=\underline{\sigma}_N^2\left(Np_N\right)^{\kappa/(2+\kappa)}\cdot\left(Np_N\right)^{2/(2+\kappa)}\to\infty$, so the whole term converges to 0.

With both conditions satisfied, we can apply the Lindeberg-Feller CLT and obtain that as $N \to \infty$,

$$\Sigma_N^{-1/2} S_N \xrightarrow{d} N(0, \mathcal{I}_{d_X}).$$

■

Proof of Theorem 3.1.

The structure of the proof is similar to the one for Theorem 2.1 and further incorporates both panel data and i.n.i.d. observations. Recall that the log likelihood function is given by

$$\ell_N(\theta^*) = \sum_{i=1}^{N} \left[Z_i'\theta^*(1-Y_{i1}) \log \frac{X_{i1}}{X_{i2}} - \log\left(1 + \left(\frac{X_{i1}}{X_{i2}}\right)^{Z_i'\theta^*}\right)\right] \mathbf{1}\{\Xi_{N,i}\},$$

and its first and second derivatives, i.e., the score and Hessian, are

$$S_N(\theta^*) = \frac{\partial \ell_N(\theta^*)}{\partial \theta^*} = \sum_{i=1}^{N} \left(\frac{1}{1 + \left(\frac{X_{i1}}{X_{i2}}\right)^{Z_i'\theta^*}} - Y_{i1} \right) Z_i \log \frac{X_{i1}}{X_{i2}} \mathbf{1}\left\{\Xi_{N,i}\right\}, \tag{A.19}$$

$$H_N(\theta^*) = \frac{\partial^2 \ell_N(\theta^*)}{\partial \theta^* \partial \theta^{*\prime}} = -\sum_{i=1}^{N} \frac{\left(\frac{X_{i1}}{X_{i2}}\right)^{Z_i'\theta^*}}{\left(1 + \left(\frac{X_{i1}}{X_{i2}}\right)^{Z_i'\theta^*}\right)^2} Z_i Z_i' \left(\log \frac{X_{i1}}{X_{i2}}\right)^2 \mathbf{1}\left\{\Xi_{N,i}\right\}. \tag{A.20}$$

Also note that Assumption 3.3(a) implies that $\underline{x}_N \to \infty$ as $N \to \infty$. Combined with the tail conditions in Assumption 3.2(a,b), this further leads to that for $x \geq \underline{x}_N$,

$$\begin{aligned} &C_i^{(y)}(Z_i)(x)^{-\tilde{\alpha}_i^{(y)}(Z_i)} \left(1 + D_i^{(y)}(Z_i)(x)^{-\beta_i^{(y)}(Z_i)} + r_i^{(y)}(x, Z_i)\right) \\ &= C_i^{(y)}(Z_i)(x)^{-\tilde{\alpha}_i^{(y)}(Z_i)}(1 + o(1)), \end{aligned} \tag{A.21}$$

as $N \to \infty$, almost surely in Z_i and for all i.

Part 1. Note that N_Ξ is a random variable, so we introduce $\xi_{N,i}$ and ξ_N in (24), non-random sequences representing the asymptotic proportions of tail observations with switching outcome values. Let us first link $\xi_{N,i}$ and ξ_N to $p_{N,i} = \mathbb{P}_i(\Xi_{N,i})$ and $p_N = \frac{1}{N}\sum_{i=1}^{N} p_{N,i}$ in the framework of rare event with i.n.i.d. observations.

For each $i = 1, \cdots, N$, by Assumption 3.1 on model specification,

$$\begin{aligned}
\mathbb{P}_i\left(\Xi_{N,i} \middle| Z_i\right) &= \mathbb{P}_i\left(Y_{i1} + Y_{i2} = 1, X_{i1} \geq \underline{x}_N, X_{i2} \geq \underline{x}_N \middle| Z_i\right) \\
&= \mathbb{P}_i\left(Y_{i1} = 1, Y_{i2} = 0, X_{i1} \geq \underline{x}_N, X_{i2} \geq \underline{x}_N | Z_i\right) \\
&\quad + \mathbb{P}_i\left(Y_{i1} = 0, Y_{i2} = 1, X_{i1} \geq \underline{x}_N, X_{i2} \geq \underline{x}_N | Z_i\right) \\
&= \mathbb{P}_i\left(Y_{i1} = 1, X_{i1} \geq \underline{x}_N | Z_i\right) \mathbb{P}_i\left(Y_{i2} = 0, X_{i2} \geq \underline{x}_N | Z_i\right) \\
&\quad + \mathbb{P}_i\left(Y_{i1} = 0, X_{i1} \geq \underline{x}_N | Z_i\right) \mathbb{P}_i\left(Y_{i2} = 1, X_{i2} \geq \underline{x}_N | Z_i\right). \qquad \text{(A.22)}
\end{aligned}$$

By Assumption 3.2(a,b) and equation (A.21), the first term in (A.22) is given by

$$\begin{aligned}
\mathbb{P}_i\left(Y_{i1} = 1, X_{i1} \geq \underline{x}_N | Z_i\right) &= \mathbb{P}_i\left(X_{i1} \geq \underline{x}_N | Y_{i1} = 1, Z_i\right) \mathbb{P}_i\left(Y_{i1} = 1 | Z_i\right) \\
&= \underline{x}_N^{-\tilde{\alpha}_i^{(1)}(Z_i)} C_i^{(1)}\left(Z_i\right)\left(1 + o(1)\right) \cdot \mathbb{P}_i\left(Y_{i1} = 1 | Z_i\right),
\end{aligned}$$

as $\underline{x}_N \to \infty$, for all i. Similar expressions hold for the other three terms in (A.22), and it follows that as $N \to \infty$,

$$\begin{aligned}
p_{N,i} &= \mathbb{P}_i\left(\Xi_{N,i}\right) = \mathbb{E}_i\left[\mathbb{P}_i\left(\Xi_{N,i} \middle| Z_i\right)\right] \qquad \text{(A.23)} \\
&= 2\mathbb{E}_i\left[\prod_{y \in \{0,1\}} \underline{x}_N^{-\tilde{\alpha}_i^{(y)}(Z_i)} C_i^{(y)}\left(Z_i\right) \tilde{\mathcal{L}}_i^{(y)}\left(\underline{x}_N, Z_i\right) \mathbb{P}_i(Y_{it} = y | Z_i)\right] \\
&= 2\mathbb{E}_i\left[\prod_{y \in \{0,1\}} \underline{x}_N^{-\tilde{\alpha}_i^{(y)}(Z_i)} C_i^{(y)}\left(Z_i\right) \mathbb{P}_i(Y_{it} = y | Z_i)\right]\left(1 + o(1)\right) \\
&= \xi_{N,i}\left(1 + o(1)\right).
\end{aligned}$$

Accordingly, by the definition of ξ_N in (24),

$$p_N = \frac{1}{N}\sum_{i=1}^{N} p_{N,i} = \frac{1}{N}\sum_{i=1}^{N} \xi_{N,i}\left(1 + o(1)\right) = \xi_N\left(1 + o(1)\right). \qquad \text{(A.24)}$$

Therefore, $Np_N \to \infty$ as $N \to \infty$ by Assumption 3.3(a), so condition (a) in Lemmas A.2 and A.3 (rare event LLN and CLT for i.n.i.d. observations) is satisfied.

Part 2. In this part, we derive the asymptotic expression for

$$\mathbb{P}_i\left(Y_{i1}=1 \middle| X_{i1}=x_1, X_{i2}=x_2, Z_i=z, \Xi_{N,i}\right).$$

First, by the tail approximation in Assumption 3.2(a,b), for $i=1,\cdots,N$, for $y\in\{0,1\}$, and for $\mathbb{P}_{Z_i}$-almost all $z\in supp(Z_i)$, the pdf is given by

$$\begin{aligned} f_{i,X_{it}|Y_{it},Z_i}(x|y,z) &= \frac{\partial\left(1-F_{i,X_{it}|Y_{it},Z_i}(x|y,z)\right)}{\partial x} \\ &= C_i^{(y)}(z)\,\tilde{\alpha}_i^{(y)}(z)\,x^{-\tilde{\alpha}_i^{(y)}(z)-1}\left(1+B_i^{(y)}(x,z)\right), \end{aligned} \tag{A.25}$$

where

$$\begin{aligned} &B_i^{(y)}(x,z) \\ &= x^{-\beta_i^{(y)}(z)}\left[D_i^{(y)}(z)\left(1+\frac{\beta_i^{(y)}(z)}{\tilde{\alpha}_i^{(y)}(z)}\right)+x^{\beta_i^{(y)}(z)}r_i^{(y)}(x,z)-\frac{1}{\tilde{\alpha}_i^{(y)}(z)}x^{\beta_i^{(y)}(z)+1}\frac{\partial r_i^{(y)}(x,z)}{\partial x}\right] \\ &= O\left(x^{-\beta_i^{(y)}(z)}\right). \end{aligned} \tag{A.26}$$

Second, by the Bayes' theorem representation, for $x\geq \underline{x}_N$,

$$\begin{aligned} &\mathbb{P}_i\left(Y_{it}=1 \middle| X_{it}=x, Z_i=z, X_{it}\geq \underline{x}_N\right)=\mathbb{P}_i\left(Y_{it}=1 \middle| X_{it}=x, Z_i=z\right) \\ &= \frac{f_{i,X_{it}|Y_{it},Z_i}(x|1,z)\,\mathbb{P}_i\left(Y_{it}=1 \middle| Z_i=z\right)}{\sum_{y\in\{0,1\}} f_{i,X_{it}|Y_{it},Z_i}(x|y,z)\,\mathbb{P}_i\left(Y_{it}=y \middle| Z_i=z\right)} \\ &= \frac{1}{1+\frac{f_{i,X_{it}|Y_{it},Z_i}(x|0,z)\mathbb{P}_i(Y_{it}=0|Z_i=z)}{f_{i,X_{it}|Y_{it},Z_i}(x|1,z)\mathbb{P}_i(Y_{it}=1|Z_i=z)}} \\ &= \frac{1}{1+\frac{C_i^{(0)}(z)\tilde{\alpha}_i^{(0)}(z)x^{-\tilde{\alpha}_i^{(0)}(z)-1}\left(1+B_i^{(0)}(x,z)\right)\mathbb{P}_i(Y_{it}=0|Z_i=z)}{C_i^{(1)}(z)\tilde{\alpha}_i^{(1)}(z)x^{-\tilde{\alpha}_i^{(1)}(z)-1}\left(1+B_i^{(1)}(x,z)\right)\mathbb{P}_i(Y_{it}=1|Z_i=z)}} \\ &= \frac{1}{1+A_i(z)x^{\tilde{\alpha}_i^{(1)}(z)-\tilde{\alpha}_i^{(0)}(z)}}\left(1+O\left(x^{-\beta_i^{(1)}(z)}\right)+O\left(x^{-\beta_i^{(0)}(z)}\right)\right) \\ &= \frac{1}{1+A_i(z)x^{z'\theta_0^*}}\left(1+O\left(x^{-\beta_i^{\min}(z)}\right)\right), \end{aligned} \tag{A.27}$$

where $A_i(z)=\frac{C_i^{(0)}(z)\tilde{\alpha}_i^{(0)}(z)\mathbb{P}_i(Y_{it}=0|Z_i=z)}{C_i^{(1)}(z)\tilde{\alpha}_i^{(1)}(z)\mathbb{P}_i(Y_{it}=1|Z_i=z)}$ and $\beta_i^{\min}(z)=\min\left\{\beta_i^{(1)}(z),\beta_i^{(0)}(z)\right\}$. The third and fourth lines in (A.27) follow from the tail approximation in Assumption 3.2(a,b) and equations (A.25) and (A.26), and the last line is by additive form of the tail index in (20).

Finally, we cancel out $A_i(z)$ by conditioning on event $Y_{i1}+Y_{i2}=1$. By the model specification in Assumption 3.1,

$$
\begin{aligned}
&\mathbb{P}_i\left(Y_{i1}=1\middle|\, X_{i1}=x_1, X_{i2}=x_2, Z_i=z, \Xi_{N,i}\right) && \text{(A.28)}\\
&= \frac{\mathbb{P}_i\left(Y_{i1}=1\middle|\, X_{i1}=x, Z_i=z, X_{i1}\geq \underline{x}_N\right)\mathbb{P}_i\left(Y_{i2}=0\middle|\, X_{i2}=x, Z_i=z, X_{i2}\geq \underline{x}_N\right)}{\sum_{y\in\{0,1\}}\mathbb{P}_i\left(Y_{i1}=y\middle|\, X_{i1}=x, Z_i=z, X_{i1}\geq \underline{x}_N\right)\mathbb{P}_i\left(Y_{i2}=1-y\middle|\, X_{i2}=x, Z_i=z, X_{i2}\geq \underline{x}_N\right)}\\
&= \frac{1}{1+\frac{\mathbb{P}_i(Y_{i1}=0|X_{i1}=x,Z_i=z,X_{i1}\geq\underline{x}_N)\mathbb{P}_i(Y_{i2}=1|X_{i2}=x,Z_i=z,X_{i2}\geq\underline{x}_N)}{\mathbb{P}_i(Y_{i1}=1|X_{i1}=x,Z_i=z,X_{i1}\geq\underline{x}_N)\mathbb{P}_i(Y_{i2}=0|X_{i2}=x,Z_i=z,X_{i2}\geq\underline{x}_N)}}\\
&= \frac{1}{1+\left(\frac{x_1}{x_2}\right)^{z'\theta_0^*}}\left(1+O\left(\max\{x_1,x_2\}^{-\beta_i^{\min}(z)}\right)\right),
\end{aligned}
$$

where we plug in (A.27) to obtain the last equality. Similarly,

$$
\begin{aligned}
&\mathbb{P}_i\left(Y_{i1}=0\middle|\, X_{i1}=x_1, X_{i2}=x_2, Z_i=z, \Xi_{N,i}\right) && \text{(A.29)}\\
&= \frac{\left(\frac{x_1}{x_2}\right)^{z'\theta_0^*}}{1+\left(\frac{x_1}{x_2}\right)^{z'\theta_0^*}}\left(1+O\left(\max\{x_1,x_2\}^{-\beta_i^{\min}(z)}\right)\right).
\end{aligned}
$$

Part 3. Let

$$
S_{N,i} = \frac{1}{\sqrt{N\xi_N}}\left(H_{N0}\right)^{-1/2}\left(\frac{1}{1+\left(\frac{X_{i1}}{X_{i2}}\right)^{Z_i'\theta_0^*}} - Y_{i1}\right) Z_i \log\frac{X_{i1}}{X_{i2}}, \qquad \text{(A.30)}
$$

which is i.n.i.d. across i by Assumption 3.1(a). Then, according to the score defined in (A.19),

$$
\frac{1}{\sqrt{N\xi_N}}\left(H_{N0}\right)^{-1/2} S_N\left(\theta_0^*\right) = \sum_{i=1}^{N} S_{N,i}\mathbf{1}\left\{\Xi_{N,i}\right\},
$$

and we can apply Lemma A.3 (rare event CLT for i.n.i.d. observations) to obtain its asymptotic normality.

First, for the mean,

$$
\mathbb{E}_i\left[S_{N,i}\middle|\,\Xi_{N,i}\right] = \frac{1}{\sqrt{N\xi_N}}\left(H_{N0}\right)^{-1/2}\mathbb{E}_i\left[\left(\frac{1}{1+\left(\frac{X_{i1}}{X_{i2}}\right)^{Z_i'\theta_0^*}} - Y_{i1}\right) Z_i \log\frac{X_{i1}}{X_{i2}}\middle|\,\Xi_{N,i}\right]
$$

$$= \frac{1}{\sqrt{N\xi_N}} \left(H_{N0}\right)^{-1/2} \mathbb{E}_i \left[\left(\frac{1}{1+\left(\frac{X_{i1}}{X_{i2}}\right)^{Z_i'\theta_0^*}} - \mathbb{E}_i \left[Y_{i1} \middle| X_i, Z_i, \Xi_{N,i} \right] \right) Z_i \log \frac{X_{i1}}{X_{i2}} \middle| \Xi_{N,i} \right]$$
$$= \frac{1}{\sqrt{N\xi_N}} \left(H_{N0}\right)^{-1/2} \mathbb{E}_i \left[O\left(\max\{X_{i1}, X_{i2}\}^{-\beta_i^{\min}(Z_i)} \right) Z_i \log \frac{X_{i1}}{X_{i2}} \middle| \Xi_{N,i} \right]$$
$$= \frac{1}{\sqrt{N\xi_N}} \left(H_{N0}\right)^{-1/2} o\left(\underline{x}_N{}^{-\underline{\beta}} \right) \mathbb{E}_i \left[Z_i \log \frac{X_{i1}}{X_{i2}} \middle| \Xi_{N,i} \right],$$

where the third equality is by (A.28) and $0 < \frac{1}{1+\left(\frac{X_{i1}}{X_{i2}}\right)^{Z_i'\theta_0^*}} < 1$, and the last equality is by the lower bound of $\beta_i^{(y)}(Z_i)$ in Assumption 3.2(a). Note that from Assumption 3.3(a), $\underline{x}_N{}^{-\underline{\beta}} = o\left((N\xi_N)^{-\frac{1+\kappa}{2+\kappa}} \right)$; from Assumption 3.3(b,c), H_{N0} is finite and its smallest eigenvalue is $O\left((N\xi_N)^{-\frac{\kappa}{2+\kappa}} \right)$, so the largest eigenvalue of $\left(H_{N0}\right)^{-1/2}$ is $o\left((N\xi_N)^{\frac{\kappa}{2(2+\kappa)}} \right)$; from Assumption 3.3(c), $\mathbb{E}_i \left[Z_i \log \frac{X_{i1}}{X_{i2}} \middle| \Xi_{N,i} \right]$ is finite. Combining the bounds on all three terms, we have that

$$\mathbb{E}_i \left[S_{N,i} \middle| \Xi_{N,i} \right] = o\left(\frac{1}{N\xi_N} \right). \tag{A.31}$$

Therefore, as $p_{N,i} = \xi_{N,i}\left(1 + o(1)\right)$ in (A.23), we obtain that

$$\sum_{i=1}^{N} p_{N,i} \mathbb{E}_i \left[S_{N,i} \middle| \Xi_{N,i} \right] = \sum_{i=1}^{N} \xi_{N,i} \cdot \left(1 + o(1)\right) \cdot o\left(\frac{1}{N\xi_N} \right) = o\left(1\right). \tag{A.32}$$

Second, for the variance,

$$\mathbb{V}_i \left[S_{N,i} \middle| \Xi_{N,i} \right]$$
$$= \mathbb{E}_i \left[S_{N,i} S_{N,i}' \middle| \Xi_{N,i} \right] - \mathbb{E}_i \left[S_{N,i} \middle| \Xi_{N,i} \right] \mathbb{E}_i \left[S_{N,i} \middle| \Xi_{N,i} \right]'$$
$$= \frac{1}{N\xi_N} \left(H_{N0}\right)^{-1/2} \mathbb{E}_i \left[\left(\frac{1}{1+\left(\frac{X_{i1}}{X_{i2}}\right)^{Z_i'\theta_0^*}} - Y_{i1} \right)^2 Z_i Z_i' \left(\log \frac{X_{i1}}{X_{i2}} \right)^2 \middle| \Xi_{N,i} \right] \left(H_{N0}\right)^{-1/2}$$
$$+ o\left(\frac{1}{\left(N\xi_N\right)^2} \right).$$

In the second equality, the first term is by the definition of $S_{N,i}$ in (A.30), and the second term is by (A.31). For the expectation term, by the law of total expectation,

$$\begin{aligned}
&\mathbb{E}_i\left[\left.\left(\frac{1}{1+\left(\frac{X_{i1}}{X_{i2}}\right)^{Z_i'\theta_0^*}}-Y_{i1}\right)^2 Z_iZ_i'\left(\log\frac{X_{i1}}{X_{i2}}\right)^2\right|\Xi_{N,i}\right]\\
&=\mathbb{E}_i\left[\mathbb{E}_i\left[\left.\left(\frac{1}{1+\left(\frac{X_{i1}}{X_{i2}}\right)^{Z_i'\theta_0^*}}-Y_{i1}\right)^2\right|X_i,Z_i,\Xi_{N,i}\right]Z_iZ_i'\left(\log\frac{X_{i1}}{X_{i2}}\right)^2\right|\Xi_{N,i}\Bigg].
\end{aligned}$$

Note that the conditional expectation in the inner term is given by

$$\begin{aligned}
&\mathbb{E}_i\left[\left.\left(\frac{1}{1+\left(\frac{X_{i1}}{X_{i2}}\right)^{Z_i'\theta_0^*}}-Y_{i1}\right)^2\right|X_i,Z_i,\Xi_{N,i}\right]\\
&=\mathbb{E}_i\left[\left.\left(\left(Y_{i1}-\mathbb{E}_i\left[Y_{i1}|X_i,Z_i,\Xi_{N,i}\right]\right)+\left(\mathbb{E}_i\left[Y_{i1}|X_i,Z_i,\Xi_{N,i}\right]-\frac{1}{1+\left(\frac{X_{i1}}{X_{i2}}\right)^{Z_i'\theta_0^*}}\right)\right)^2\right|X_i,Z_i,\Xi_{N,i}\right]\\
&=\mathbb{V}_i\left[Y_{i1}|X_i,Z_i,\Xi_{N,i}\right]+\mathbb{E}_i\left[\left.\left(\mathbb{E}_i\left[Y_{i1}|X_i,Z_i,\Xi_{N,i}\right]-\frac{1}{1+\left(\frac{X_{i1}}{X_{i2}}\right)^{Z_i'\theta_0^*}}\right)^2\right|X_i,Z_i,\Xi_{N,i}\right]\\
&=\mathbb{P}_i\left(Y_{i1}=1|X_i,Z_i,\Xi_{N,i}\right)\mathbb{P}_i\left(Y_{i1}=0|X_i,Z_i,\Xi_{N,i}\right)+o\left(\underline{x}_N{}^{-2\underline{\beta}}\right)\\
&=\frac{\left(\frac{X_{i1}}{X_{i2}}\right)^{Z_i'\theta_0^*}}{\left(1+\left(\frac{X_{i1}}{X_{i2}}\right)^{Z_i'\theta_0^*}\right)^2}\left(1+o\left(\underline{x}_N{}^{-\underline{\beta}}\right)\right)+o\left(\underline{x}_N{}^{-2\underline{\beta}}\right),
\end{aligned}$$

where the last three lines are by (A.28), (A.29), and $0<\frac{1}{1+\left(\frac{X_{i1}}{X_{i2}}\right)^{Z_i'\theta_0^*}}<1$. Therefore, we have that

$$\begin{aligned}
&\mathbb{E}_i\left[\left.\left(\frac{1}{1+\left(\frac{X_{i1}}{X_{i2}}\right)^{Z_i'\theta_0^*}}-Y_{i1}\right)^2 Z_iZ_i'\left(\log\frac{X_{i1}}{X_{i2}}\right)^2\right|\Xi_{N,i}\right]\\
&=\mathbb{E}_i\left[\left.\frac{\left(\frac{X_{i1}}{X_{i2}}\right)^{Z_i'\theta_0^*}}{\left(1+\left(\frac{X_{i1}}{X_{i2}}\right)^{Z_i'\theta_0^*}\right)^2}Z_iZ_i'\left(\log\frac{X_{i1}}{X_{i2}}\right)^2\right|\Xi_{N,i}\right]\left(1+o\left(\underline{x}_N{}^{-\underline{\beta}}\right)\right)
\end{aligned}$$

$$+\mathbb{E}_i\left[Z_iZ_i'\left(\log\frac{X_{i1}}{X_{i2}}\right)^2\middle|\,\Xi_{N,i}\right]o\left(\underline{x}_N{}^{-\underline{\beta}}\right)$$
$$=H_{N0,i}\left(1+o\left(1\right)\right)+o\left(\underline{x}_N{}^{-2\underline{\beta}}\right),$$

where the second equality is by the definition of $H_{N0,i}$ and the fact that $\mathbb{E}_i\left[Z_iZ_i'\left(\log\frac{X_{i1}}{X_{i2}}\right)^2\middle|\,\Xi_{N,i}\right]$ is finite by Assumption 3.3(c). Substituting this back to the expression of $\mathbb{V}\left[S_{N,i}\right]$, we have that as $N\to\infty$,

$$\begin{aligned}
&\mathbb{V}_i\left[S_{N,i}\middle|\,\Xi_{N,i}\right]\\
&=\frac{1}{N\xi_N}\left[\left(H_{N0}\right)^{-1/2}H_{N0,i}\left(H_{N0}\right)^{-1/2}\left(1+o\left(1\right)\right)+\left(H_{N0}\right)^{-1}o\left(\underline{x}_N{}^{-\underline{\beta}}\right)\right]+o\left(\frac{1}{\left(N\xi_N\right)^2}\right)\\
&=\frac{1}{N\xi_N}\left(H_{N0}\right)^{-1/2}H_{N0,i}\left(H_{N0}\right)^{-1/2}\left(1+o\left(1\right)\right)+o\left(\frac{1}{\left(N\xi_N\right)^2}\right)
\end{aligned}$$

The argument for the second line is similar to that for (A.31): $\underline{x}_N{}^{-2\underline{\beta}}=o\left(\left(N\xi_N\right)^{-\frac{2(1+\kappa)}{2+\kappa}}\right)$ by Assumption 3.3(a), and the largest eigenvalue of $\left(H_{N0}\right)^{-1}$ is $o\left(\left(N\xi_N\right)^{\frac{\kappa}{2+\kappa}}\right)$ by Assumption 3.3(b,c), so $\left(H_{N0}\right)^{-1}o\left(\underline{x}_N{}^{-\underline{\beta}}\right)=o\left(\frac{1}{N\xi_N}\right)$. Then, as $p_{N,i}=\xi_{N,i}\left(1+o(1)\right)$ in (A.23), we obtain that

$$\begin{aligned}
&\sum_{i=1}^N p_{N,i}\mathbb{V}_i\left[S_{N,i}\middle|\,\Xi_{N,i}\right] &\text{(A.33)}\\
&=\sum_{i=1}^N\xi_{N,i}\left(1+o\left(1\right)\right)\cdot\left[\frac{1}{N\xi_N}\left(H_{N0}\right)^{-1/2}H_{N0,i}\left(H_{N0}\right)^{-1/2}\left(1+o\left(1\right)\right)+o\left(\frac{1}{\left(N\xi_N\right)^2}\right)\right]\\
&=\left(H_{N0}\right)^{-1/2}\left(H_{N0}\right)\left(H_{N0}\right)^{-1/2}\left(1+o\left(1\right)\right)+o\left(1\right)\to\mathcal{I}_{d_Z}.
\end{aligned}$$

For Lemma A.3, condition (a) is satisfied by (A.24) in Part 1, condition (b) is by Assumption 3.3(c), and condition (c) is by Assumption 3.3(b). Therefore, as $\theta_0^*\in\text{int}\left(\theta^*\right)$, we have that as $N\to\infty$,

$$\frac{1}{\sqrt{N\xi_N}}\left(H_{N0}\right)^{-1/2}S_N\left(\theta_0^*\right)=\sum_{i=1}^N S_{N,i}\mathcal{I}\left\{\Xi_{N,i}\right\}\xrightarrow{d}\mathcal{N}\left(0,\mathcal{I}_{d_Z}\right),\tag{A.34}$$

where $\sum_{i=1}^N p_{N,i}\mathbb{E}_i\left[S_{N,i}\middle|\,\Xi_{N,i}\right]$ and $\sum_{i=1}^N p_{N,i}\mathbb{V}_i\left[S_{N,i}\middle|\,\Xi_{N,i}\right]$ are given in (A.32) and (A.33), respectively.

Part 4. Similarly, for the Hessian matrix,

$$
\begin{aligned}
&\frac{1}{N\xi_N}\left(H_{N0}\right)^{-1/2}\cdot H_N\left(\theta_0^*\right)\cdot\left(H_{N0}\right)^{-1/2} \qquad \text{(A.35)}\\
&=\frac{1}{N\xi_N}\left(H_{N0}\right)^{-1/2}\cdot\left[-\sum_{i=1}^{N}\frac{\left(\frac{X_{i1}}{X_{i2}}\right)^{Z_i'\theta_0^*}}{\left(1+\left(\frac{X_{i1}}{X_{i2}}\right)^{Z_i'\theta_0^*}\right)^2}Z_iZ_i'\left(\log\frac{X_{i1}}{X_{i2}}\right)^2\mathbf{1}\left\{\Xi_{N,i}^{(y)}\right\}\right]\cdot\left(H_{N0}\right)^{-1/2}\\
&=-\left(H_{N0}\right)^{-1/2}\cdot\frac{1}{N\xi_N}\sum_{i=1}^{N}p_{N,i}\mathbb{E}_i\left[\left.\frac{\left(\frac{X_{i1}}{X_{i2}}\right)^{Z_i'\theta_0^*}}{\left(1+\left(\frac{X_{i1}}{X_{i2}}\right)^{Z_i'\theta_0^*}\right)^2}Z_iZ_i'\left(\log\frac{X_{i1}}{X_{i2}}\right)^2\right|\Xi_{N,i}\right]\cdot\left(H_{N0}\right)^{-1/2}\cdot\left(1+o_p(1)\right)\\
&=-\left(H_{N0}\right)^{-1/2}\cdot\frac{1}{N\xi_N}\sum_{i=1}^{N}\xi_{N,i}\mathbb{E}_i\left[\left.\frac{\left(\frac{X_{i1}}{X_{i2}}\right)^{Z_i'\theta_0^*}}{\left(1+\left(\frac{X_{i1}}{X_{i2}}\right)^{Z_i'\theta_0^*}\right)^2}Z_iZ_i'\left(\log\frac{X_{i1}}{X_{i2}}\right)^2\right|\Xi_{N,i}\right]\cdot\left(H_{N0}\right)^{-1/2}\cdot\left(1+o_p(1)\right)\\
&=-\left(H_{N0}\right)^{-1/2}\cdot H_{N0}\cdot\left(H_{N0}\right)^{-1/2}\cdot\left(1+o_p(1)\right)\\
&\xrightarrow{p}-\mathcal{I}_{d_Z}.
\end{aligned}
$$

The first equality is by the definition of the Hessian matrix (A.20). The second equality follows from Lemma A.2 (rare event LLN for i.n.i.d. observations), where condition (a) is satisfied by (A.24) in Part 1, and condition (b) is by Assumption 3.3(c). The third equality is by $p_{N,i}=\xi_{N,i}\left(1+o(1)\right)$ in (A.23). The fourth equality is by the definition of H_{N0}.

Part 5. Let $\zeta_N=\left(H_{N0}\right)^{1/2}\left(\theta^*-\theta_0^*\right)$, $\zeta_{N0}=\left(H_{N0}\right)^{1/2}\theta_0^*$, and $W_i=\left(H_{N0}\right)^{-1/2}Z_i$. Then, the log likelihood function can be rewritten as

$$
\tilde{\ell}_N\left(\zeta_N\right)=\sum_{i=1}^{N}\left[W_i'\left(\zeta_N+\zeta_{N0}\right)\left(1-Y_{i1}\right)\log\frac{X_{i1}}{X_{i2}}-\log\left(1+\left(\frac{X_{i1}}{X_{i2}}\right)^{W_i'\left(\zeta_N+\zeta_{N0}\right)}\right)\right]\mathbf{1}\{\Xi_{N,i}\},
$$

the corresponding score and Hessian are denoted by $\tilde{S}_N\left(\zeta_N\right)$ and $\tilde{H}_N\left(\zeta_N\right)$, respectively, and the MLE estimate is denoted by $\hat{\zeta}_N$.

First, the Hessian matrix is given by

$$
\tilde{H}_N\left(\zeta_N\right)=-\sum_{i=1}^{N}\frac{\left(\frac{X_{i1}}{X_{i2}}\right)^{W_i'\left(\zeta_N+\zeta_{N0}\right)}}{\left(1+\left(\frac{X_{i1}}{X_{i2}}\right)^{W_i'\left(\zeta_N+\zeta_{N0}\right)}\right)^2}W_iW_i'\left(\log\frac{X_{i1}}{X_{i2}}\right)^2\mathbf{1}\left\{\Xi_{N,i}^{(y)}\right\}.
$$

It is positive definite for all $\zeta_N = (H_{N0})^{1/2}(\theta^* - \theta_0^*)$ with $\theta^* \in \Theta^*$. The reason is that Assumption 3.3(b,c) implies that both H_{N0} and $\mathbb{E}_i[Z_i Z_i' | \Xi_{N,i}]$ are finite and positive definite, and thus $\mathbb{E}_i[W_i W_i' | \Xi_{N,i}]$ is finite and positive definite. Then, the log likelihood function is $\tilde{\ell}_N(\zeta_N)$ is strictly concave over its domain, and the MLE estimate $\hat{\zeta}_N$ is unique.

Second, let

$$\mathcal{U}_{N,C} = \left\{ u \in \mathbb{R}^{d_Z} : \frac{1}{\sqrt{N\xi_N}} (H_{N0})^{-1/2} u + \theta_0^* \in \theta^* \text{ and } \|u\| = C \right\}.$$

Note that $\mathcal{U}_{N,C} \neq \emptyset$ for any $C > 0$. This follows from three facts: first, θ^* is a convex cone; second, $\operatorname{int}(\theta^*) \neq \emptyset$ as $\theta_0^* \in \operatorname{int}(\theta^*)$; and third, H_{N0} is finite and full rank by Assumption 3.3(b,c). Then, let u be an arbitrary non-random vector in $\mathcal{U}_{N,C}$. Applying the second-order Taylor expansion of $\tilde{\ell}_N\left(\frac{u}{\sqrt{N\xi_N}}\right)$ around $\zeta_N = 0$, we have that

$$\tilde{\ell}_N\left(\frac{u}{\sqrt{N\xi_N}}\right) - \tilde{\ell}_N(0) = \frac{1}{\sqrt{N\xi_N}} u' \tilde{S}_N(0) + \frac{1}{2N\xi_N} u' \tilde{H}_N(0) u + o_p(1)$$

From (A.34) and (A.35), we have that as $N \to \infty$,

$$\frac{1}{\sqrt{N\xi_N}} \tilde{S}_N(0) = \frac{1}{\sqrt{N\xi_N}} S_N(\theta_0^*) \xrightarrow{d} \mathcal{N}(0, \mathcal{I}_{d_Z}), \tag{A.36}$$

$$\frac{1}{N\xi_N} \tilde{H}_N(0) = \frac{1}{N\xi_N} (H_{N0})^{-1/2} \cdot H_N(\theta_0^*) \cdot (H_{N0})^{-1/2} \xrightarrow{p} -\mathcal{I}_{d_Z}. \tag{A.37}$$

This implies that when C is large enough, the quadratic term dominates the linear one with an arbitrarily large probability. That is, for any $\varepsilon > 0$, there exists a $C > 0$ such that

$$\limsup_N \mathbb{P}\left(\sup_{u \in \mathcal{U}_{N,C}} \tilde{\ell}_N\left(\frac{u}{\sqrt{N\xi_N}}\right) < \tilde{\ell}_N(0) \right) > 1 - \varepsilon.$$

Therefore, $\tilde{\ell}_N(\cdot)$ must have at least one local maximizer, which is of order $O_p\left(\sqrt{N\xi_N}\right)$. And by the uniqueness of the MLE, this local maximizer is the global maximizer, and $\hat{\zeta}_N = O_p\left(\sqrt{N\xi_N}\right)$.

Finally, by the first order condition, $\tilde{S}_N\left(\hat{\zeta}_N\right) = 0$. Applying the first-order Taylor expansion around $\zeta_N = 0$, we have that

$$0 = \tilde{S}_N\left(\hat{\zeta}_N\right) = \tilde{S}_N(0) + \tilde{H}_N(0) \hat{\zeta}_N + o_p(1),$$

which implies that

$$\sqrt{N\xi_N}\hat{\zeta}_N = -\left(\frac{1}{N\xi_N}\tilde{H}_N(0)\right)^{-1}\cdot\frac{1}{\sqrt{N\xi_N}}\left(\tilde{S}_N(0)+o_p(1)\right)\xrightarrow{d}\mathcal{N}(0,\mathcal{I}_{d_Z}),$$

following from (A.36) and (A.37). This is equivalent to

$$\sqrt{N\xi_N}\,(H_{N0})^{1/2}\left(\hat{\theta}^*-\theta_0^*\right)\xrightarrow{d}\mathcal{N}(0,\mathcal{I}_{d_Z}).$$

■

Proposition A.3 (Panel data: proportion of tail switchers) *Suppose Assumptions 3.1–3.3 hold. as $N\to\infty$,*

$$\frac{N_\Xi}{N}=\xi_N(1+o_p(1)).$$

Proof. First, for the mean of N_Ξ/N,

$$\mathbb{E}\left[\frac{N_\Xi}{N}\right]=\mathbb{E}\left[\frac{1}{N}\sum_{i=1}^{N}\mathbf{1}\{\Xi_{N,i}\}\right]=\frac{1}{N}\sum_{i=1}^{N}\mathbb{P}_i(\Xi_{N,i})=\xi_N(1+o(1)),$$

where the second equality follows from the independence across i in Assumption 3.1(a), and the last equality follows from (A.24). Second, let $\mathbb{V}_i$ be the variance given $\{\lambda_i,\mathcal{C}_i\}$. Similarly, based on the independence across i and (A.24),

$$\begin{aligned}\mathbb{V}\left[\frac{N_\Xi}{N}\right]&=\mathbb{V}\left[\frac{1}{N}\sum_{i=1}^{N}\mathbf{1}\{\Xi_{N,i}\}\right]=\frac{1}{N^2}\sum_{i=1}^{N}\mathbb{V}_i\left[\mathbf{1}\{\Xi_{N,i}\}\right]=\frac{1}{N^2}\sum_{i=1}^{N}\mathbb{P}_i(\Xi_{N,i})\left(1-\mathbb{P}_i(\Xi_{N,i})\right),\\&\le\frac{1}{N^2}\sum_{i=1}^{N}\mathbb{P}_i(\Xi_{N,i})=\frac{\xi_N}{N}(1+o(1)).\end{aligned}$$

Together, we have that as $N\to\infty$,

$$\begin{aligned}\mathbb{E}\left[\left(\frac{N_\Xi}{N\xi_N}-1\right)^2\right]&=\mathbb{V}\left[\frac{N_\Xi}{N\xi_N}\right]+o(1)=\frac{1}{\xi_n^2}\mathbb{V}\left[\frac{N_\Xi}{N}\right]+o(1)\\&\le\frac{1}{N\xi_N}(1+o(1))+o(1)\to 0,\end{aligned}$$

where the convergence to 0 follows from Assumption 3.3(a). As convergence in the second moment implies convergence in probability, we obtain that as $N \to \infty$,

$$\frac{N_\Xi}{N} \to \xi_N \left(1 + o_p(1)\right).$$

■

Remark A.2 Combining Theorem 3.1 with Proposition A.3, we have that as $N \to \infty$,

$$\sqrt{N_\Xi} \left(H_{N0}\right)^{1/2} \left(\hat{\theta}^* - \theta_0^*\right) \xrightarrow{d} \mathcal{N}\left(0, \mathcal{I}_{d_Z}\right).$$

B Additional Tables and Figures

Tables A.1–A.4 provide detailed results, including bias, standard deviation, and RMSE, for various estimands across all model specifications in Monte Carlo Experiment 1. See Table A.1 for tail parameters, Table A.2 for extreme elasticity, Table A.3 for conditional probability, and Table A.4 for partial effects. Figure A.1 shows the scatter plots of the LPS from all Monte Carlo repetitions in Experiment 2 with $\alpha_X = 1$ and $\alpha_\varepsilon = 1$.

Figure A.2 and Tables A.5–A.7 present more details for the empirical example on housing prices and bank riskiness. Figure A.2 depicts the histograms for distributions of $X_{it}|Y_{it} = y$ for the baseline sample. Table A.5 summarizes the descriptive statistics for all samples. Table A.6 reports a forecasting comparison across estimators for all samples, along with the parameter and APE estimates based on the tail estimator. Lastly, Table A.7 shows a robustness check across different values of level c for the baseline sample.

Table A.1: Parameter estimation - Experiment 1

		$\alpha_X = 0.5$			$\alpha_X = 1$			$\alpha_X = 1.5$			$\alpha_X = 2$		
		Bias	SD	RMSE	Bias	SD	RMSE	Bias	SD	RMSE	Bias	SD	RMSE
$\alpha_\varepsilon = 0.5$	$\hat{\alpha}^{(0)}$	0.002	0.131	0.131	-0.020	0.180	0.181	-0.025	0.237	0.238	-0.107	0.289	0.309
	$\hat{\alpha}^{(1)}$	0.003	0.062	0.062	-0.005	0.118	0.118	-0.036	0.174	0.178	-0.075	0.231	0.243
$\alpha_\varepsilon = 1$	$\hat{\alpha}^{(0)}$	0.026	0.186	0.188	-0.023	0.253	0.254	-0.081	0.279	0.290	-0.155	0.325	0.360
	$\hat{\alpha}^{(1)}$	0.004	0.063	0.063	0.010	0.124	0.124	-0.005	0.182	0.182	-0.033	0.238	0.241
$\alpha_\varepsilon = 1.5$	$\hat{\alpha}^{(0)}$	0.067	0.269	0.277	-0.043	0.301	0.304	-0.144	0.327	0.357	-0.277	0.367	0.460
	$\hat{\alpha}^{(1)}$	0.004	0.063	0.063	0.011	0.124	0.125	0.003	0.185	0.185	-0.017	0.243	0.244
$\alpha_\varepsilon = 2$	$\hat{\alpha}^{(0)}$	0.105	0.326	0.342	-0.095	0.353	0.365	-0.267	0.373	0.458	-0.427	0.387	0.576
	$\hat{\alpha}^{(1)}$	0.004	0.063	0.063	0.011	0.124	0.125	0.004	0.185	0.185	-0.013	0.244	0.244

Notes: Bias, SD, and RMSE are calculated based on $N_{sim} = 1000$ repetitions and with respect to the theoretical values as the tails, $\alpha^{(0)} = \alpha_X + \alpha_\varepsilon$ and $\alpha^{(1)} = \alpha_X$.

Table A.2: Extreme elasticity estimation - Experiment 1

		$\alpha_X = 0.5$			$\alpha_X = 1$			$\alpha_X = 1.5$			$\alpha_X = 2$		
		Bias	SD	RMSE	Bias	SD	RMSE	Bias	SD	RMSE	Bias	SD	RMSE
$\alpha_\varepsilon = 0.5$	Tail	0.00	0.15	0.15	0.01	0.21	0.21	-0.01	0.30	0.30	0.03	0.36	0.37
	Logit, tail X	0.47	0.04	0.47	0.23	0.08	0.25	0.07	0.10	0.12	-0.02	0.11	0.11
	Logit, all X	-17.49	10.26	20.27	-2.21	0.81	2.35	-1.21	0.29	1.25	-0.84	0.17	0.86
	Local Logit	2091.55	29194.28	29215.53	-39.20	1033.30	1033.48	-2.70	42.08	42.14	-0.39	9.44	9.44
	Local linear	0.47	0.07	0.47	-3.88	10.72	11.40	-0.30	1.17	1.21	-0.11	1.77	1.77
$\alpha_\varepsilon = 1$	Tail	-0.02	0.20	0.20	0.03	0.29	0.29	0.08	0.34	0.35	0.12	0.40	0.42
	Logit, tail X	1.00	0.01	1.00	0.87	0.19	0.89	0.50	0.29	0.58	0.31	0.26	0.40
	Logit, all X	-14.24	11.21	17.85	-16.82	1.59	16.90	-6.49	0.46	6.51	-4.11	0.26	4.11
	Local Logit	202.17	1870.36	1877.34	-129.29	1661.59	1664.55	-8.35	363.05	362.97	-1.92	30.73	30.78
	Local linear	0.96	0.08	0.97	4.28	46.29	46.47	-3.25	56.96	57.03	-0.49	3.34	3.37
$\alpha_\varepsilon = 1.5$	Tail	-0.06	0.28	0.28	0.05	0.33	0.33	0.15	0.38	0.40	0.26	0.45	0.52
	Logit, tail X	1.50	0.01	1.50	1.44	0.51	1.53	1.19	0.42	1.26	0.92	0.48	1.04
	Logit, all X	-12.09	12.31	16.81	-30.68	1.76	30.73	-10.91	0.53	10.92	-6.47	0.31	6.48
	Local Logit	1.50	0.00	1.50	-860.37	7495.52	7512.01	-100.10	1580.91	1583.14	-11.44	96.72	97.35
	Local linear	1.46	0.08	1.47	3.06	159.59	159.54	2.16	81.37	81.36	-0.95	7.16	7.21
$\alpha_\varepsilon = 2$	Tail	-0.10	0.33	0.35	0.11	0.38	0.39	0.27	0.41	0.49	0.41	0.47	0.63
	Logit, tail X	2.00	0.01	2.00	1.81	1.10	2.12	1.84	0.40	1.89	1.60	0.59	1.70
	Logit, all X	-11.67	12.40	16.57	-29.01	8.22	30.12	-13.84	0.56	13.85	-7.93	0.31	7.94
	Local Logit	2.00	0.00	2.00	14.23	1434.99	1410.94	-137.52	610.54	625.27	-34.39	216.59	219.19
	Local linear	1.96	0.08	1.97	7.36	53.36	53.84	-13.43	151.57	152.08	-3.60	18.53	18.87

Notes: Bias, SD, and RMSE are calculated based on $N_{sim} = 1000$ repetitions and with respect to the theoretical values as the tails, $-|\alpha^{(1)} - \alpha^{(0)}| = -\alpha_\varepsilon$. Evaluated at the 97.5th percentile of the distribution of X.

Table A.3: $\hat{\mathbb{P}}(Y=1|X=x)$ - Experiment 1

			$\alpha_X = 0.5$			$\alpha_X = 1$			$\alpha_X = 1.5$			$\alpha_X = 2$		
			Bias	SD	RMSE	Bias	SD	RMSE	Bias	SD	RMSE	Bias	SD	RMSE
$\alpha_\varepsilon = 0.5$	$x: 90\%$	Tail	0.01	0.03	0.03	0.03	0.07	0.07	0.02	0.09	0.09	0.03	0.09	0.09
		Logit, tail X	-0.88	0.05	0.88	-0.68	0.08	0.68	-0.60	0.07	0.61	-0.55	0.07	0.56
		Logit, all X	-0.18	0.13	0.22	-0.13	0.03	0.13	-0.07	0.02	0.08	-0.05	0.01	0.05
		Local Logit	0.01	0.05	0.06	0.00	0.04	0.04	0.00	0.03	0.03	-0.00	0.03	0.03
		Local linear	-0.41	0.02	0.41	-0.12	0.06	0.13	-0.01	0.02	0.02	-0.00	0.02	0.02
	$x: 95\%$	Tail	0.00	0.02	0.02	0.01	0.04	0.04	0.02	0.06	0.06	0.02	0.06	0.06
		Logit, tail X	-0.93	0.06	0.93	-0.73	0.09	0.73	-0.62	0.08	0.63	-0.56	0.08	0.56
		Logit, all X	-0.02	0.13	0.13	-0.05	0.06	0.08	-0.03	0.03	0.04	-0.01	0.02	0.02
		Local Logit	0.01	0.12	0.12	0.01	0.07	0.07	0.00	0.05	0.05	-0.00	0.04	0.04
		Local linear	-0.45	0.02	0.45	-0.06	0.09	0.11	-0.00	0.02	0.02	-0.00	0.02	0.02
	$x: 97.5\%$	Tail	-0.00	0.01	0.01	0.00	0.03	0.03	0.01	0.03	0.04	0.02	0.04	0.04
		Logit, tail X	-0.94	0.07	0.94	-0.70	0.09	0.70	-0.57	0.07	0.58	-0.50	0.06	0.51
		Logit, all X	0.00	0.08	0.08	0.05	0.06	0.08	0.05	0.03	0.06	0.04	0.02	0.05
		Local Logit	-0.11	0.25	0.27	0.02	0.12	0.12	0.01	0.09	0.09	0.01	0.07	0.07
		Local linear	-0.47	0.03	0.47	0.00	0.09	0.09	0.00	0.03	0.03	0.00	0.04	0.04
	$x: 99\%$	Tail	-0.00	0.01	0.01	-0.00	0.02	0.02	0.00	0.03	0.03	0.00	0.03	0.03
		Logit, tail X	-0.03	0.10	0.11	-0.17	0.13	0.21	-0.23	0.09	0.25	-0.22	0.07	0.23
		Logit, all X	0.00	0.05	0.05	0.07	0.04	0.08	0.11	0.02	0.12	0.11	0.02	0.11
		Local Logit	-0.48	0.08	0.48	-0.01	0.27	0.27	0.02	0.19	0.19	0.01	0.14	0.14
		Local linear	-0.41	0.11	0.43	0.03	0.08	0.08	0.01	0.06	0.06	-0.00	0.08	0.08
$\alpha_\varepsilon = 1$	$x: 90\%$	Tail	-0.00	0.01	0.01	0.01	0.03	0.03	0.02	0.05	0.06	0.04	0.06	0.07
		Logit, tail X	-0.98	0.02	0.98	-0.90	0.01	0.90	-0.83	0.02	0.83	-0.79	0.01	0.79
		Logit, all X	0.01	0.04	0.04	0.06	0.01	0.06	0.03	0.02	0.03	0.02	0.01	0.02
		Local Logit	0.00	0.02	0.02	0.00	0.03	0.03	-0.00	0.02	0.02	-0.00	0.02	0.02
		Local linear	-0.51	0.02	0.51	-0.20	0.10	0.23	-0.01	0.03	0.04	-0.00	0.01	0.01
	$x: 95\%$	Tail	-0.00	0.00	0.00	0.00	0.02	0.02	0.01	0.03	0.03	0.02	0.03	0.04
		Logit, tail X	-1.00	0.02	1.00	-0.94	0.02	0.94	-0.88	0.03	0.88	-0.83	0.03	0.83
		Logit, all X	0.00	0.03	0.03	0.05	0.00	0.05	0.08	0.01	0.08	0.07	0.01	0.07
		Local Logit	0.00	0.04	0.04	0.00	0.04	0.04	0.00	0.04	0.04	-0.00	0.03	0.03
		Local linear	-0.52	0.02	0.52	-0.07	0.14	0.16	0.00	0.03	0.03	-0.00	0.02	0.02
	$x: 97.5\%$	Tail	-0.00	0.00	0.00	-0.00	0.01	0.01	0.00	0.02	0.02	0.01	0.02	0.02
		Logit, tail X	-1.00	0.02	1.00	-0.94	0.06	0.94	-0.83	0.07	0.84	-0.78	0.06	0.78
		Logit, all X	-0.00	0.02	0.02	0.02	0.00	0.02	0.06	0.00	0.06	0.08	0.00	0.08
		Local Logit	-0.12	0.21	0.24	0.01	0.05	0.05	0.01	0.05	0.05	0.01	0.05	0.05
		Local linear	-0.51	0.03	0.51	0.04	0.13	0.14	0.01	0.03	0.03	0.00	0.03	0.03
	$x: 99\%$	Tail	-0.00	0.00	0.00	-0.00	0.01	0.01	-0.00	0.01	0.01	0.00	0.02	0.02
		Logit, tail X	-0.00	0.02	0.02	-0.01	0.04	0.04	-0.05	0.07	0.09	-0.07	0.08	0.11
		Logit, all X	-0.00	0.02	0.02	0.01	0.00	0.01	0.04	0.00	0.04	0.06	0.00	0.06
		Local Logit	-0.49	0.08	0.49	0.00	0.09	0.09	0.01	0.08	0.08	0.02	0.07	0.08
		Local linear	-0.44	0.12	0.45	0.06	0.11	0.12	0.01	0.05	0.05	0.00	0.04	0.04

Notes: Bias, SD, and RMSE are calculated based on $N_{sim} = 1000$ repetitions w.r.t. the theoretical values.

Table A.3: $\hat{\mathbb{P}}(Y = 1|X = x)$ - Experiment 1 (cont.)

			$\alpha_X = 0.5$			$\alpha_X = 1$			$\alpha_X = 1.5$			$\alpha_X = 2$	
		Bias	SD	RMSE	Bias	SD	RMSE	Bias	SD	RMSE	Bias	SD	RMSE
$\alpha_\varepsilon = 1.5$	$x : 90\%$												
	Tail	-0.00	0.00	0.00	0.00	0.02	0.02	0.02	0.03	0.04	0.03	0.04	0.05
	Logit, tail X	-1.00	0.02	1.00	-0.95	0.00	0.95	-0.90	0.01	0.90	-0.86	0.01	0.86
	Logit, all X	-0.00	0.04	0.04	0.05	0.00	0.05	0.06	0.01	0.06	0.05	0.01	0.05
	Local Logit	0.00	0.01	0.01	0.00	0.02	0.02	0.00	0.02	0.02	-0.00	0.02	0.02
	Local linear	-0.52	0.02	0.52	-0.23	0.12	0.26	-0.01	0.04	0.04	-0.00	0.01	0.01
	$x : 95\%$												
	Tail	-0.00	0.00	0.00	-0.00	0.01	0.01	0.00	0.02	0.02	0.02	0.02	0.02
	Logit, tail X	-1.00	0.02	1.00	-0.98	0.01	0.98	-0.94	0.01	0.94	-0.91	0.02	0.91
	Logit, all X	-0.00	0.03	0.03	0.02	0.00	0.02	0.05	0.00	0.05	0.06	0.00	0.06
	Local Logit	0.00	0.01	0.01	0.00	0.02	0.02	0.00	0.03	0.03	0.00	0.02	0.02
	Local linear	-0.52	0.02	0.52	-0.06	0.16	0.17	0.01	0.03	0.03	-0.00	0.02	0.02
	$x : 97.5\%$												
	Tail	-0.00	0.00	0.00	-0.00	0.00	0.00	0.00	0.01	0.01	0.01	0.01	0.01
	Logit, tail X	-1.00	0.02	1.00	-0.99	0.03	0.99	-0.93	0.07	0.93	-0.87	0.07	0.88
	Logit, all X	-0.00	0.02	0.02	0.01	0.00	0.01	0.02	0.00	0.02	0.05	0.00	0.05
	Local Logit	-0.12	0.21	0.24	0.00	0.04	0.04	0.01	0.03	0.03	0.01	0.03	0.03
	Local linear	-0.50	0.03	0.50	0.05	0.14	0.15	0.01	0.04	0.04	0.00	0.02	0.02
	$x : 99\%$												
	Tail	-0.00	0.00	0.00	-0.00	0.00	0.00	-0.00	0.01	0.01	-0.00	0.01	0.01
	Logit, tail X	-0.00	0.02	0.02	-0.00	0.01	0.01	-0.01	0.03	0.03	-0.02	0.05	0.05
	Logit, all X	-0.00	0.02	0.02	0.00	0.00	0.00	0.01	0.00	0.01	0.02	0.00	0.02
	Local Logit	-0.49	0.08	0.49	-0.00	0.06	0.06	0.00	0.03	0.03	0.01	0.04	0.04
	Local linear	-0.43	0.12	0.45	0.06	0.12	0.14	0.01	0.05	0.05	0.00	0.03	0.03
$\alpha_\varepsilon = 2$	$x : 90\%$												
	Tail	-0.00	0.00	0.00	0.00	0.01	0.01	0.01	0.02	0.03	0.03	0.03	0.04
	Logit, tail X	-1.00	0.02	1.00	-0.97	0.00	0.97	-0.93	0.00	0.93	-0.90	0.00	0.90
	Logit, all X	-0.00	0.04	0.04	0.03	0.00	0.03	0.05	0.00	0.05	0.05	0.00	0.06
	Local Logit	0.00	0.00	0.00	0.00	0.01	0.01	0.00	0.02	0.02	0.00	0.02	0.02
	Local linear	-0.51	0.02	0.51	-0.24	0.13	0.27	-0.01	0.04	0.05	-0.00	0.01	0.01
	$x : 95\%$												
	Tail	-0.00	0.00	0.00	-0.00	0.00	0.00	0.00	0.01	0.01	0.01	0.01	0.02
	Logit, tail X	-1.00	0.02	1.00	-0.99	0.00	0.99	-0.97	0.01	0.97	-0.94	0.01	0.94
	Logit, all X	-0.00	0.03	0.03	0.01	0.00	0.01	0.03	0.00	0.03	0.04	0.00	0.04
	Local Logit	0.00	0.00	0.00	0.00	0.01	0.01	0.00	0.02	0.02	0.00	0.02	0.02
	Local linear	-0.51	0.02	0.51	-0.06	0.17	0.18	0.01	0.04	0.04	-0.00	0.01	0.01
	$x : 97.5\%$												
	Tail	-0.00	0.00	0.00	-0.00	0.00	0.00	-0.00	0.01	0.01	0.00	0.01	0.01
	Logit, tail X	-1.00	0.02	1.00	-0.99	0.03	0.99	-0.97	0.06	0.97	-0.93	0.08	0.93
	Logit, all X	-0.00	0.02	0.02	0.00	0.00	0.00	0.01	0.00	0.01	0.02	0.00	0.02
	Local Logit	-0.12	0.21	0.24	0.00	0.01	0.01	0.00	0.02	0.02	0.00	0.03	0.03
	Local linear	-0.50	0.03	0.50	0.06	0.14	0.16	0.01	0.04	0.04	-0.00	0.02	0.02
	$x : 99\%$												
	Tail	-0.00	0.00	0.00	-0.00	0.00	0.00	-0.00	0.00	0.00	-0.00	0.01	0.01
	Logit, tail X	-0.00	0.02	0.02	0.00	0.00	0.00	-0.00	0.01	0.01	-0.00	0.02	0.02
	Logit, all X	-0.00	0.02	0.02	0.00	0.00	0.00	0.00	0.00	0.00	0.01	0.00	0.01
	Local Logit	-0.49	0.08	0.49	-0.00	0.02	0.02	0.00	0.02	0.02	0.00	0.03	0.03
	Local linear	-0.43	0.12	0.44	0.07	0.12	0.14	0.01	0.05	0.05	0.00	0.02	0.02

Notes: Bias, SD, and RMSE are calculated based on $N_{sim} = 1000$ repetitions w.r.t. the theoretical values.

Table A.4: Partial effects estimation - Experiment 1

			$\alpha_X = 0.5$			$\alpha_X = 1$			$\alpha_X = 1.5$			$\alpha_X = 2$	
			Bias	SD	RMSE	Bias	SD	RMSE	Bias	SD	RMSE	Bias	SD	RMSE
$\alpha_\varepsilon = 0.5$	$x : 90\%$	Tail	0.01	0.03	0.03	0.03	0.07	0.07	0.02	0.09	0.09	0.03	0.09	0.09
		Logit, tail X	-0.88	0.05	0.88	-0.68	0.08	0.68	-0.60	0.07	0.61	-0.55	0.07	0.56
		Logit, all X	-0.18	0.13	0.22	-0.13	0.03	0.13	-0.07	0.02	0.08	-0.05	0.01	0.05
		Local Logit	0.01	0.05	0.06	0.00	0.04	0.04	0.00	0.03	0.03	-0.00	0.03	0.03
		Local linear	-0.41	0.02	0.41	-0.12	0.06	0.13	-0.01	0.02	0.02	-0.00	0.02	0.02
	$x : 95\%$	Tail	0.00	0.02	0.02	0.01	0.04	0.04	0.02	0.06	0.06	0.02	0.06	0.06
		Logit, tail X	-0.93	0.06	0.93	-0.73	0.09	0.73	-0.62	0.08	0.63	-0.56	0.08	0.56
		Logit, all X	-0.02	0.13	0.13	-0.05	0.06	0.08	-0.03	0.03	0.04	-0.01	0.02	0.02
		Local Logit	0.01	0.12	0.12	0.01	0.07	0.07	0.00	0.05	0.05	-0.00	0.04	0.04
		Local linear	-0.45	0.02	0.45	-0.06	0.09	0.11	-0.00	0.02	0.02	-0.00	0.02	0.02
	$x : 97.5\%$	Tail	-0.00	0.01	0.01	0.00	0.03	0.03	0.01	0.03	0.04	0.02	0.04	0.04
		Logit, tail X	-0.94	0.07	0.94	-0.70	0.09	0.70	-0.57	0.07	0.58	-0.50	0.06	0.51
		Logit, all X	0.00	0.08	0.08	0.05	0.06	0.08	0.05	0.03	0.06	0.04	0.02	0.05
		Local Logit	-0.11	0.25	0.27	0.02	0.12	0.12	0.01	0.09	0.09	0.01	0.07	0.07
		Local linear	-0.47	0.03	0.47	0.00	0.09	0.09	0.00	0.03	0.03	0.00	0.04	0.04
	$x : 99\%$	Tail	-0.00	0.01	0.01	-0.00	0.02	0.02	0.00	0.03	0.03	0.00	0.03	0.03
		Logit, tail X	-0.03	0.10	0.11	-0.17	0.13	0.21	-0.23	0.09	0.25	-0.22	0.07	0.23
		Logit, all X	0.00	0.05	0.05	0.07	0.04	0.08	0.11	0.02	0.12	0.11	0.02	0.11
		Local Logit	-0.48	0.08	0.48	-0.01	0.27	0.27	0.02	0.19	0.19	0.01	0.14	0.14
		Local linear	-0.41	0.11	0.43	0.03	0.08	0.08	0.01	0.06	0.06	-0.00	0.08	0.08
$\alpha_\varepsilon = 1$	$x : 90\%$	Tail	-0.00	0.01	0.01	0.01	0.03	0.03	0.02	0.05	0.06	0.04	0.06	0.07
		Logit, tail X	-0.98	0.02	0.98	-0.90	0.01	0.90	-0.83	0.02	0.83	-0.79	0.01	0.79
		Logit, all X	0.01	0.04	0.04	0.06	0.01	0.06	0.03	0.02	0.03	0.02	0.01	0.02
		Local Logit	0.00	0.02	0.02	0.00	0.03	0.03	-0.00	0.02	0.02	-0.00	0.02	0.02
		Local linear	-0.51	0.02	0.51	-0.20	0.10	0.23	-0.01	0.03	0.04	-0.00	0.01	0.01
	$x : 95\%$	Tail	-0.00	0.00	0.00	0.00	0.02	0.02	0.01	0.03	0.03	0.02	0.03	0.04
		Logit, tail X	-1.00	0.02	1.00	-0.94	0.02	0.94	-0.88	0.03	0.88	-0.83	0.03	0.83
		Logit, all X	0.00	0.03	0.03	0.05	0.00	0.05	0.08	0.01	0.08	0.07	0.01	0.07
		Local Logit	0.00	0.04	0.04	0.00	0.04	0.04	0.00	0.04	0.04	-0.00	0.03	0.03
		Local linear	-0.52	0.02	0.52	-0.07	0.14	0.16	0.00	0.03	0.03	-0.00	0.02	0.02
	$x : 97.5\%$	Tail	-0.00	0.00	0.00	-0.00	0.01	0.01	0.00	0.02	0.02	0.01	0.02	0.02
		Logit, tail X	-1.00	0.02	1.00	-0.94	0.06	0.94	-0.83	0.07	0.84	-0.78	0.06	0.78
		Logit, all X	-0.00	0.02	0.02	0.02	0.00	0.02	0.06	0.00	0.06	0.08	0.00	0.08
		Local Logit	-0.12	0.21	0.24	0.01	0.05	0.05	0.01	0.05	0.05	0.01	0.05	0.05
		Local linear	-0.51	0.03	0.51	0.04	0.13	0.14	0.01	0.03	0.03	0.00	0.03	0.03
	$x : 99\%$	Tail	-0.00	0.00	0.00	-0.00	0.01	0.01	-0.00	0.01	0.01	0.00	0.02	0.02
		Logit, tail X	-0.00	0.02	0.02	-0.01	0.04	0.04	-0.05	0.07	0.09	-0.07	0.08	0.11
		Logit, all X	-0.00	0.02	0.02	0.01	0.00	0.01	0.04	0.00	0.04	0.06	0.00	0.06
		Local Logit	-0.49	0.08	0.49	0.00	0.09	0.09	0.01	0.08	0.08	0.02	0.07	0.08
		Local linear	-0.44	0.12	0.45	0.06	0.11	0.12	0.01	0.05	0.05	0.00	0.04	0.04

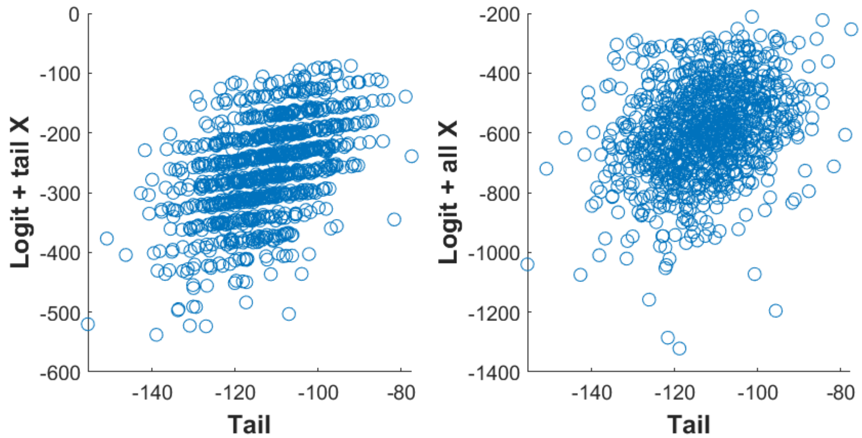
Notes: Bias, SD, and RMSE are calculated based on $N_{sim} = 1000$ repetitions w.r.t. the theoretical values.

Table A.4: Partial effects estimation - Experiment 1 (cont.)

			$\alpha_X = 0.5$			$\alpha_X = 1$			$\alpha_X = 1.5$			$\alpha_X = 2$		
			Bias	SD	RMSE	Bias	SD	RMSE	Bias	SD	RMSE	Bias	SD	RMSE
$\alpha_\varepsilon = 1.5$	$x: 90\%$	Tail	-0.00	0.00	0.00	0.00	0.02	0.02	0.02	0.03	0.04	0.03	0.04	0.05
		Logit, tail X	-1.00	0.02	1.00	-0.95	0.00	0.95	-0.90	0.01	0.90	-0.86	0.01	0.86
		Logit, all X	-0.00	0.04	0.04	0.05	0.00	0.05	0.06	0.01	0.06	0.05	0.01	0.05
		Local Logit	0.00	0.01	0.01	0.00	0.02	0.02	0.00	0.02	0.02	-0.00	0.02	0.02
		Local linear	-0.52	0.02	0.52	-0.23	0.12	0.26	-0.01	0.04	0.04	-0.00	0.01	0.01
	$x: 95\%$	Tail	-0.00	0.00	0.00	-0.00	0.01	0.01	0.00	0.02	0.02	0.02	0.02	0.02
		Logit, tail X	-1.00	0.02	1.00	-0.98	0.01	0.98	-0.94	0.01	0.94	-0.91	0.02	0.91
		Logit, all X	-0.00	0.03	0.03	0.02	0.00	0.02	0.05	0.00	0.05	0.06	0.00	0.06
		Local Logit	0.00	0.01	0.01	0.00	0.02	0.02	0.00	0.03	0.03	0.00	0.02	0.02
		Local linear	-0.52	0.02	0.52	-0.06	0.16	0.17	0.01	0.03	0.03	-0.00	0.02	0.02
	$x: 97.5\%$	Tail	-0.00	0.00	0.00	-0.00	0.00	0.00	0.00	0.01	0.01	0.01	0.01	0.01
		Logit, tail X	-1.00	0.02	1.00	-0.99	0.03	0.99	-0.93	0.07	0.93	-0.87	0.07	0.88
		Logit, all X	-0.00	0.02	0.02	0.01	0.00	0.01	0.02	0.00	0.02	0.05	0.00	0.05
		Local Logit	-0.12	0.21	0.24	0.00	0.04	0.04	0.01	0.03	0.03	0.01	0.03	0.03
		Local linear	-0.50	0.03	0.50	0.05	0.14	0.15	0.01	0.04	0.04	0.00	0.02	0.02
	$x: 99\%$	Tail	-0.00	0.00	0.00	-0.00	0.00	0.00	-0.00	0.01	0.01	-0.00	0.01	0.01
		Logit, tail X	-0.00	0.02	0.02	-0.00	0.01	0.01	-0.01	0.03	0.03	-0.02	0.05	0.05
		Logit, all X	-0.00	0.02	0.02	0.00	0.00	0.00	0.01	0.00	0.01	0.02	0.00	0.02
		Local Logit	-0.49	0.08	0.49	-0.00	0.06	0.06	0.00	0.03	0.03	0.01	0.04	0.04
		Local linear	-0.43	0.12	0.45	0.06	0.12	0.14	0.01	0.05	0.05	0.00	0.03	0.03
$\alpha_\varepsilon = 2$	$x: 90\%$	Tail	-0.00	0.00	0.00	0.00	0.01	0.01	0.01	0.02	0.03	0.03	0.03	0.04
		Logit, tail X	-1.00	0.02	1.00	-0.97	0.00	0.97	-0.93	0.00	0.93	-0.90	0.00	0.90
		Logit, all X	-0.00	0.04	0.04	0.03	0.00	0.03	0.05	0.00	0.05	0.05	0.00	0.06
		Local Logit	0.00	0.00	0.00	0.00	0.01	0.01	0.00	0.02	0.02	0.00	0.02	0.02
		Local linear	-0.51	0.02	0.51	-0.24	0.13	0.27	-0.01	0.04	0.05	-0.00	0.01	0.01
	$x: 95\%$	Tail	-0.00	0.00	0.00	-0.00	0.00	0.00	0.00	0.01	0.01	0.01	0.01	0.02
		Logit, tail X	-1.00	0.02	1.00	-0.99	0.00	0.99	-0.97	0.01	0.97	-0.94	0.01	0.94
		Logit, all X	-0.00	0.03	0.03	0.01	0.00	0.01	0.03	0.00	0.03	0.04	0.00	0.04
		Local Logit	0.00	0.00	0.00	0.00	0.01	0.01	0.00	0.02	0.02	0.00	0.02	0.02
		Local linear	-0.51	0.02	0.51	-0.06	0.17	0.18	0.01	0.04	0.04	-0.00	0.01	0.01
	$x: 97.5\%$	Tail	-0.00	0.00	0.00	-0.00	0.00	0.00	-0.00	0.01	0.01	0.00	0.01	0.01
		Logit, tail X	-1.00	0.02	1.00	-0.99	0.03	0.99	-0.97	0.06	0.97	-0.93	0.08	0.93
		Logit, all X	-0.00	0.02	0.02	0.00	0.00	0.00	0.01	0.00	0.01	0.02	0.00	0.02
		Local Logit	-0.12	0.21	0.24	0.00	0.01	0.01	0.00	0.02	0.02	0.00	0.03	0.03
		Local linear	-0.50	0.03	0.50	0.06	0.14	0.16	0.01	0.04	0.04	-0.00	0.02	0.02
	$x: 99\%$	Tail	-0.00	0.00	0.00	-0.00	0.00	0.00	-0.00	0.00	0.00	-0.00	0.01	0.01
		Logit, tail X	-0.00	0.02	0.02	0.00	0.00	0.00	-0.00	0.01	0.01	-0.00	0.02	0.02
		Logit, all X	-0.00	0.02	0.02	0.00	0.00	0.00	0.00	0.00	0.00	0.01	0.00	0.01
		Local Logit	-0.49	0.08	0.49	-0.00	0.02	0.02	0.00	0.02	0.02	0.00	0.03	0.03
		Local linear	-0.43	0.12	0.44	0.07	0.12	0.14	0.01	0.05	0.05	0.00	0.02	0.02

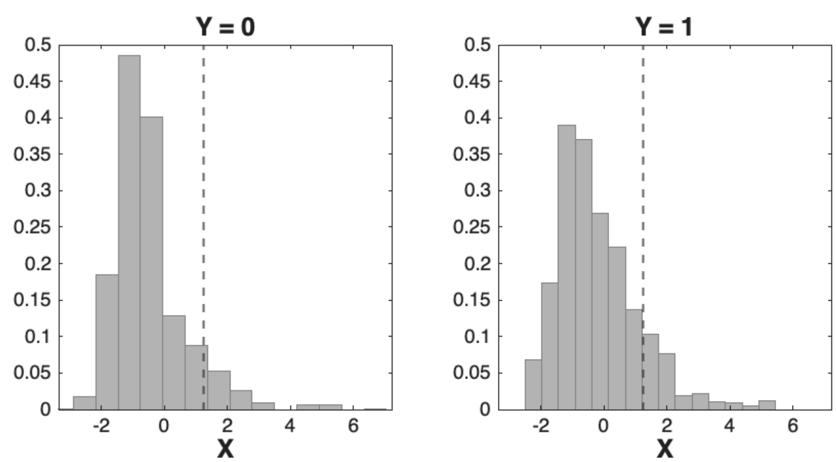
Notes: Bias, SD, and RMSE are calculated based on $N_{sim} = 1000$ repetitions w.r.t. the theoretical values.

Figure A.1: Log predictive score - Experiment 2, $\alpha_X = 1$, $\alpha_\varepsilon = 1$



Notes: Each circle represents one Monte Carlo repetition. Note that the x- and y-scales are substantially different.

Figure A.2: Histogram - banking application, baseline sample



Notes: Black dashed lines mark the 90th percentile of the X_{it} distribution and indicate the tail region.

Table A.5: Sample Statistics for Charge-Off Rates

				Sample Sizes			Sample Statistics			
Loan	t_{start}	t_{fcst}	T	N	$N_e^\dagger$	$N_f^\dagger$	$\%(Y=1)$	SD_X	$Skew_X$	$Kurt_X$
CRE	2001Q3	2009Q3	32	7882	1286	1086	3.27	0.94	0.70	12.26
CRE	2000Q3	2009Q3	36	8224	1408	1194	2.90	0.95	0.73	12.37
CRE	1999Q3	2009Q3	40	8588	1451	1212	2.60	0.99	0.67	11.96
CRE	1998Q3	2009Q3	44	9047	1506	1260	2.36	1.00	0.66	12.28
CRE	1997Q3	2009Q3	48	9495	1546	1303	2.14	1.02	0.63	12.28
CRE	2001Q4	2009Q4	32	7798	1602	1367	3.91	0.92	0.73	10.87
CRE	2000Q4	2009Q4	36	8128	1726	1476	3.46	0.93	0.79	11.18
CRE	1999Q4	2009Q4	40	8510	1806	1533	3.10	0.95	0.73	10.96
CRE	1998Q4	2009Q4	44	8899	1909	1607	2.81	0.97	0.74	11.28
CRE	1997Q4	2009Q4	48	9382	1959	1654	2.56	0.99	0.72	11.43
RRE	2001Q3	2009Q3	32	7909	2104	1693	18.11	0.94	0.68	12.16
RRE	2000Q3	2009Q3	36	8244	2348	1868	16.03	0.95	0.70	12.26
RRE	1999Q3	2009Q3	40	8614	2497	1958	14.32	0.98	0.65	11.84
RRE	1998Q3	2009Q3	44	9085	2604	2038	12.94	1.00	0.64	12.15
RRE	1997Q3	2009Q3	48	9531	2719	2143	11.74	1.02	0.60	12.14
RRE	2001Q4	2009Q4	32	7824	2248	1813	19.24	0.92	0.71	10.80
RRE	2000Q4	2009Q4	36	8148	2470	1986	17.01	0.92	0.77	11.10
RRE	1999Q4	2009Q4	40	8538	2642	2098	15.18	0.95	0.71	10.87
RRE	1998Q4	2009Q4	44	8936	2849	2264	13.74	0.97	0.72	11.19
RRE	1997Q4	2009Q4	48	9418	2974	2366	12.47	0.98	0.70	11.32

Notes: N is the total number of banks in each sample. $N_e^\dagger$ is the number of banks in the tail and contributing to the likelihood, characterized by X_{it} being above the 90th percentile of the estimation sample and Y_{it} switching values across t for these tail observations. $N_f^\dagger$ is the number of banks that meet the criteria for $N_e^\dagger$ and, additionally, have $X_{i,T+1}$ in the tail during the forecasting period. Sample statistics are calculated for each N-by-T panel. The panels are unbalanced due to missing values.

Table A.6: Forecast evaluation and parameter estimation - banking application, all samples

		$T = 32$	$T = 36$	$T = 40$	$T = 44$	$T = 48$
RRE, prediction period = 2009Q3						
$\text{LPS} \cdot N_f^{\dagger}$	*Tail*	*-1112.60*	*-1217.91*	*-1291.65*	*-1355.10*	*-1443.19*
	Logit, tail X	-7.65 ***	-10.01 ***	-10.75 ***	-12.46 ***	-12.84 ***
	Logit, all X	-87.83 ***	-85.66 ***	-132.49 ***	-170.68 ***	-207.60 ***
Tail	$\hat{\theta}^*_{\log X}$	0.96 ***	1.00 ***	0.97 ***	1.00 ***	1.01 ***
	$\hat{\theta}^*_{Z \log X}$	0.89 ***	0.61 ***	0.78 ***	0.67 ***	0.58 ***
	APE	0.13 ***	0.13 ***	0.12 ***	0.12 ***	0.12 ***
	$\hat{V}(\tilde{A}_i)$	0.99	1.00	0.98	0.98	0.98
RRE, prediction period = 2009Q4						
$\text{LPS} \cdot N_f^{\dagger}$	*Tail*	*-1232.46*	*-1347.38*	*-1433.72*	*-1559.47*	*-1649.76*
	Logit, tail X	-14.12 ***	-17.95 ***	-18.66 ***	-21.85 ***	-22.37 ***
	Logit, all X	-154.94 ***	-144.03 ***	-178.91 ***	-206.57 ***	-232.10 ***
Tail	$\hat{\theta}^*_{\log X}$	1.11 ***	1.14 ***	1.09 ***	1.11 ***	1.12 ***
	$\hat{\theta}^*_{Z \log X}$	0.52 **	0.54 **	0.63 ***	0.53 ***	0.42 **
	APE	0.16 ***	0.16 ***	0.15 ***	0.15 ***	0.14 ***
	$\hat{V}(\tilde{A}_i)$	1.06	1.06	1.04	1.03	1.03
CRE, prediction period = 2009Q3						
$\text{LPS} \cdot N_f^{\dagger}$	*Tail*	*-777.51*	*-853.64*	*-878.00*	*-910.85*	*-953.34*
	Logit, tail X	-8.93 ***	-9.51 ***	-10.38 ***	-11.71 ***	-11.83 ***
	Logit, all X	-15.95 ***	-12.40 ***	-21.61 ***	-29.59 ***	-45.66 ***
Tail	$\hat{\theta}^*_{\log X}$	1.20 ***	1.25 ***	1.26 ***	1.29 ***	1.29 ***
	$\hat{\theta}^*_{Z \log X}$	0.46	0.30	0.30	0.26	0.23
	APE	0.11 ***	0.11 ***	0.11 ***	0.10 ***	0.10 ***
	$\hat{V}(\tilde{A}_i)$	1.00	1.00	1.01	1.03	1.04
CRE, prediction period = 2009Q4						
$\text{LPS} \cdot N_f^{\dagger}$	*Tail*	*-1017.50*	*-1088.70*	*-1134.48*	*-1189.45*	*-1237.07*
	Logit, tail X	-17.18 ***	-22.72 ***	-23.49 ***	-28.68 ***	-29.12 ***
	Logit, all X	-50.76 ***	-68.91 ***	-77.75 ***	-95.20 ***	-106.86 ***
Tail	$\hat{\theta}^*_{\log X}$	1.31 ***	1.30 ***	1.31 ***	1.37 ***	1.42 ***
	$\hat{\theta}^*_{Z \log X}$	0.07	0.20	0.04	-0.02	-0.10
	APE	0.14 ***	0.14 ***	0.13 ***	0.13 ***	0.13 ***
	$\hat{V}(\tilde{A}_i)$	1.03	1.05	1.06	1.08	1.08

Notes: For the tail estimator, the table reports the exact values of $\text{LPS} \cdot N_f^{\dagger}$. For other estimators, the table reports their differences from the tail estimator. The tests compare other estimators with the tail estimator. The last four rows in each panel report the estimated parameters and APEs as well as the sample variances of the estimated $\tilde{A}_i$'s using the tail estimator. All significance levels are indicated by *: 10%, **: 5%, and ***: 1%.

Table A.7: Forecast evaluation and parameter estimation - banking application, varying c

		$c=0$	$c=0.01$	$c=0.02$	$c=0.05$
LPS $\cdot N_f^{\dagger}$	*Tail*	*-1433.72*	*-1453.03*	*-1462.00*	*-1447.46*
	Logit, tail X	-18.66 ***	-18.71 ***	-18.21 ***	-18.63 ***
	Logit, all X	-178.91 ***	-174.92 ***	-175.49 ***	-180.03 ***
Tail	$\hat{\theta}^*_{\log X}$	1.09 ***	1.08 ***	1.08 ***	1.09 ***
	$\hat{\theta}^*_{Z \log X}$	0.63 ***	0.61 ***	0.60 ***	0.58 ***
	APE	0.15 ***	0.15 ***	0.15 ***	0.15 ***
	$\hat{V}(\tilde{A}_i)$	1.06	1.05	1.04	1.04
		$c=0.1$	$c=0.2$	$c=0.5$	$c=1$
LPS $\cdot N_f^{\dagger}$	*Tail*	*-1411.56*	*-1343.85*	*-1131.54*	*-833.62*
	Logit, tail X	-18.40 ***	-15.77 ***	-11.74 ***	-9.10 ***
	Logit, all X	-178.44 ***	-180.10 ***	-166.46 ***	-95.35 ***
Tail	$\hat{\theta}^*_{\log X}$	1.14 ***	1.12 ***	1.11 ***	1.22 ***
	$\hat{\theta}^*_{Z \log X}$	0.52 **	0.52 **	0.44 *	0.16
	APE	0.15 ***	0.14 ***	0.11 ***	0.10 ***
	$\hat{V}(\tilde{A}_i)$	1.04	1.01	0.99	0.96

Notes: Baseline sample: RRE, forecasting period = 2009Q4, $T = 40$. 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. The last four rows in each panel report the estimated parameters and APEs as well as the sample variances of the estimated $\tilde{A}_i$'s using the tail estimator. All significance levels are indicated by *: 10%, **: 5%, and ***: 1%.