A Simple and Powerful Diagnostic Test for Binary Choice Models*

Ting Ji
Central University of Finance and Economics

Laura Liu
University of Pittsburgh

Yulong Wang
Lehigh University

Jiahe Xing
Syracuse University

First version: November 3, 2025
This version: March 29, 2026

Abstract

This paper proposes a specification test for the conventional distributional assumptions of error terms in binary choice models, focusing on its tail properties. Based on extreme value theory, we first establish that the tail index of the unobserved error can be recovered by that of the observed covariates. The null hypothesis of the index being zero essentially covers the widely used probit and logit models. We then construct a simple and powerful statistical test for both cross-sectional and panel data, requiring no model estimation and no parametric assumptions. Monte Carlo simulations demonstrate that our test performs well in size and power, and applications to three empirical examples on firm export and innovation decisions and female labor force participation illustrate its general applicability.

Keywords: hypothesis testing, binary choice, extreme value theory, tail index, panel data.

JEL code: C12, C23, C25, C52

*Ting Ji: Central University of Finance and Economics, email: `jiting@cufe.edu.cn`. Laura Liu: University of Pittsburgh, email: `laura.liu@pitt.edu`. Yulong Wang: Lehigh University, email: `yuw925@lehigh.edu`. Jiahe Xing: Syracuse University, email: `jxing04@syr.edu`. We thank Xu Cheng, Wayne Gao, and Bo Honoré for helpful discussions. All remaining errors are ours.

1 Introduction

Binary choice models are workhorses of empirical economics. They are widely used to study decisions such as labor force participation, educational choice, export decision, and innovation behavior. Since the pioneering work of McFadden (1974) and Chamberlain (1980), probit and logit models have become standard tools in empirical work. Their appeal comes from a simple index structure, where the binary outcome depends on whether the index crosses a threshold, and from the resulting tractability of estimation. Yet their credibility depends not only on the specification of the observable index, but also on the assumed distribution of the latent error term. In particular, probit and logit impose thin-tailed normal or logistic errors. In many economic settings, however, rare but consequential shocks may be important for the decision itself, so whether the latent error is thin-tailed is a crucial empirical question rather than a harmless assumption.

This distinction between thin-tailed and heavy-tailed latent errors matters for both estimation and inference. If the true latent error is heavier-tailed than the normal or logistic distribution, standard binary choice models can understate the probability of extreme unobserved shocks and hence misrepresent rare outcomes. Moreover, Khan and Tamer (2010) show that tail behavior can determine whether standard $\sqrt{n}$-asymptotically normal inference is valid in binary choice models, with heavier tails leading to irregular estimation, slower convergence rates, and non-Gaussian limiting distributions. For this reason, the thin-tail assumption underlying probit and logit should be treated as a substantive restriction that deserves direct testing.

To see the idea, consider the simple binary choice model

$$Y_i = I(X_i - \varepsilon_i \geq 0), \tag{1}$$

where $I(\cdot)$ denotes the indicator function, X_i is an observed covariate, and ε_i is an unobserved error. The key observation is that the tail of the latent error leaves a trace in the tail of the observed covariate once we condition on the outcome. In particular, among observations with $Y_i = 0$, large values of X_i must be matched by even larger realizations of ε_i. As a result, the right tail of $X_i \mid Y_i = 0$ reflects the right tail of the latent error. Symmetrically, the left tail of $X_i \mid Y_i = 1$ reflects the left tail of the error. With one heavy-tailed covariate, this link allows the tail behavior of the latent error to be inferred from observable conditional extremes.

We formalize this idea using extreme value theory. Under mild domain-of-attraction conditions, if the heavy-tailed covariate X_i has tail index $\lambda > 0$ and the latent error ε_i has tail index $\xi \geq 0$, then the conditional distribution $X_i \mid Y_i = 0$ has tail index $\gamma = \xi\lambda/(\xi + \lambda)$ in the right tail. This mapping is central to our diagnostic because it implies that testing whether the latent error is thin-tailed, namely whether $\xi = 0$, is equivalent to testing whether the corresponding observable conditional tail is thin, namely whether $\gamma = 0$. Thus, the tail property of an unobserved error term becomes testable using only observables.

Based on this equivalence, we construct a simple likelihood-ratio test from the normalized largest order statistics of a single heavy-tailed covariate in the outcome-specific subsamples $X_i \mid Y_i = 0$. The test compares the observed spacing pattern of these conditional extremes with the pattern implied under thin-tailed errors, and systematic departures provide evidence of heavier tails. This yields a direct test of whether the latent error distribution is thin-tailed, as assumed in standard probit and logit models.

The proposed diagnostic is attractive in practice. It requires neither model estimation nor a parametric specification of the link function, which makes it computationally light and easy to implement. It is also easy to present graphically and can be used as a pre-estimation diagnostic before committing to a logit or probit specification.

The framework extends to richer settings with additional covariates: when one observed covariate dominates the tail of the index, the diagnostic can be implemented using that covariate alone; alternatively, a plug-in version allows testing based on the estimated full index. The same idea also extends to static and dynamic panel data with unobserved individual effects in short-T settings. Because the procedure does not require estimating common coefficients or individual effects, it avoids the incidental parameters problem altogether. In this way, the test provides a simple and broadly applicable diagnostic that remains useful across a wide range of data structures in binary choice models.

Our diagnostic is also conceptually distinct from existing approaches. Conventional specification tests typically assess overall fit, the shape of the link function, or index linearity, while semiparametric single-index estimators relax the functional form of the link without directly testing the tail behavior of the latent error. In contrast, our procedure isolates the thin-tail restriction itself and does so without estimating the binary choice model.

Monte Carlo simulations show that the proposed diagnostic performs well in finite samples. In both cross-sectional and panel settings, it controls size under thin-tailed errors and has good power against heavy-tailed alternatives. Power generally rises with the number of tail observations, although using too many can weaken the tail approximation by bringing in mid-sample observations.

We further illustrate our proposed diagnostic in three empirical examples: firm export participation, firm innovation decisions, and female labor force participation. In the two firm examples, asset size serves as the heavy-tailed covariate, while in the labor supply example we use husband's income. Across these examples, the data provide evidence against the thin-tail assumption underlying conventional probit or logit models, suggesting that rare and unusually large unobserved shocks play an important role in shaping binary decisions. These empirical examples highlight that the proposed test is easy to implement in practice and can detect the tail misspecification that conventional diagnostics often overlook.

Related literature. The econometric analysis of binary outcomes has a long tradition; see earlier work by McFadden (1974) and Chamberlain (1980), as well as textbook treatments such as Wooldridge (2002). A substantial literature studies their specification and goodness-of-fit, typically asking whether the chosen link is appropriate, whether the parametric form is correctly specified, or whether the implied choice probabilities match the data; see, among others, Pregibon (1980), Bierens (1990), Horowitz and Härdle (1994), and Xia et al. (2004). Recently, Ota and Otsu (2026) develop a specification test for parametric binary choice models based on comparing maximum likelihood and maximum score estimators. This literature, however, does not directly test the tail behavior of the latent error. Our paper fills this gap by providing a simple diagnostic for the thin-tail assumption maintained by standard logit and probit models.

A related strand of work develops semiparametric estimators for binary choice models, beginning with the maximum score estimator of Manski (1975). This is followed by the maximum rank correlation estimator of Han (1987), the smoothed maximum score estimator of Horowitz (1992), the semiparametric least squares estimator of Ichimura (1993), and the semiparametric maximum likelihood estimator of Klein and Spady (1993). These methods relax the parametric form of the link function and greatly expand modeling flexibility, but they are estimation procedures rather than direct specification diagnostics. Importantly, Khan and Tamer (2010) show that in discrete choice and related limited dependent variable models, heavy tails can undermine standard asymptotic inference, rendering the thin-tail assumption both economically and econometrically consequential. Our approach is complementary: instead of estimating the binary choice model under a flexible link, we use extreme observations of a heavy-tailed covariate to test whether the latent error belongs to the thin-tailed class assumed by standard binary choice specifications. A rejection of the null would indicate that the conventional logit and probit specifications may not be adequate, and motivate the use of more flexible semiparametric estimators, including those discussed above.

In addition, our paper is related to the literature on panel binary choice models with unobserved individual effects, including settings with predetermined covariates. In short-T panels, fixed effects generate the well-known incidental parameters problem (Neyman and Scott, 1948). Moreover, in dynamic models, this problem is compounded by lagged dependence and the treatment of initial conditions (Heckman, 1981). A large literature has been developed to address these issues within parametric specifications; for example, Fernández-Val (2009) develops a bias-correction procedure for panel probit models. While these methods improve estimation, they generally retain the thin-tailed structure implicit in probit or logit. Semiparametric approaches relax the functional form restriction; for example, Manski (1987) provides a maximum score estimator for static panel binary choice models, and Honoré and Kyriazidou (2000) develop a pairwise differencing estimator for dynamic panel discrete choice models. We complement this literature by directly testing the thin-tail assumption without model estimation, thereby sidestepping the incidental parameters problem.

Finally, our paper also relates to the work that applies extreme value theory to econometric analysis. Classical references such as de Haan and Ferreira (2007) show that tail behavior determines the limiting laws of extremes. Recent methodological advances have further improved statistical tools for tail inference. For example, Müller and Wang (2017) develop a fixed-k asymptotic framework that enables inference based on only a finite number of extreme order statistics, and Einmahl and He (2023) provide modern results on tail index estimation under heterogeneity. Our diagnostic draws on these insights in a different way. Standard extreme value methods study the tail of an observed variable directly, whereas we use the conditional extremes of an observed heavy-tailed covariate to infer the tail class of an unobserved latent error. In contrast to Liu and Wang (2025), which considers estimation and forecasting with binary outcomes and extreme covariates, the contribution here focuses on testing and accommodates the generalized extreme value distribution with $\xi = 0$, including the thin-tailed null.

The remainder of the paper is organized as follows. Section 2 reviews the extreme value theory and introduces our test statistic in the baseline model. Section 3 extends our method to allow for multiple covariates and panel data. Section 4 evaluates the finite sample performance of our proposed method via Monte Carlo simulations. Section 5 applies our method to three widely studied empirical examples, and Section 6 concludes with some remarks. All the proofs are provided in the Appendix.

2 Main Result

In this section, we first provide an introduction to extreme value theory and show how to recover the tail heaviness of the latent error from the covariate. We then develop the test procedure and derive test statistics in the baseline model.

2.1 Recover Tail Heaviness of Latent Error

To characterize the tail properties of the latent error, we consider the tail behavior of X_i through extreme value theory. The relevant limit distributions form the *generalized extreme value* (GEV) family G_ξ, defined by

$$G_\xi(x) = \begin{cases} \exp(-(1+\xi x)^{-1/\xi}), & \text{if } \xi \neq 0, \\ \exp(-\exp(-x)), & \text{if } \xi = 0. \end{cases}$$

When $\xi = 0$, G_0 is the *Gumbel* distribution, also known as the Type I extreme value distribution. When $\xi > 0$, G_ξ is the *Fréchet* extreme value distribution with *Pareto* tails. The parameter ξ is referred to as the *tail index*, characterizing the tail heaviness of the underlying distribution. A distribution F is classified within the *domain of attraction* of the extreme value distribution $G_\xi(\cdot)$ if there exists a sequence of constants $a_n > 0$ and b_n such that

$$\lim_{n\to\infty} (F(a_n x + b_n))^n = G_\xi(x)$$

for all x in the support of G_ξ. We denote this condition by $F \in \mathcal{D}(G_\xi)$. Our key assumptions require that the distributions of our covariate and the latent error reside within the domain of attraction of extreme value distributions.

Moreover, if a distribution F has positive first and second order derivatives in a neighborhood of the right endpoint, the von Mises' condition provides a sufficient condition for belonging to a domain of attraction of G_ξ, that is

$$\lim_{x \to x^*} \frac{(1 - F(x))F''(x)}{(F'(x))^2} = -\xi - 1, \tag{2}$$

see for example, de Haan and Ferreira (2007, Theorem 1.1.8). Here x^* refers to the right endpoint of this distribution, which equals infinity when $\xi > 0$ and can be finite or infinite when $\xi = 0$.

Von Mises' condition is mild and satisfied by almost all common textbook continuous distributions. The case $\xi > 0$ covers heavy-tailed distributions such as Pareto, Student-t, Cauchy, and F, whose tails are well-approximated by Pareto distributions, i.e., power law. The case $\xi = 0$ covers thin-tailed ones such as normal, logistic, and exponential, with exponentially decaying tails. See de Haan and Ferreira (2007) for a comprehensive review.

Back to the binary choice model, two commonly assumed distributions are the Gaussian and logistic distributions, both satisfying (2) with $\xi = 0$. So, our hypothesis testing problem is

$$H_0 : \xi = 0 \text{ against } H_1 : \xi > 0. \tag{3}$$

If the null is rejected, we recommend avoiding logit or probit and using more flexible methods such as the (smoothed) maximum score and maximum rank correlation estimators, which remain consistent under heavy-tailed covariates and errors.

It is not straightforward to directly test (3) since ε_i is latent. Our next step is to show the test (3) is equivalent to imposing a test on the tail index of the observed covariate X_i in a subsample with $Y_i = y$ for any $y \in \{0, 1\}$. We focus on the right tail first for illustration, and the same argument applies to the left. Specifically, we establish that if the covariate X_i is distributed with a heavy tail, the tail behavior of the covariate in the subsample with $Y_i = 0$ will be determined by the tail of the error term ε_i.

To illustrate the key idea, let us first consider a simple example in which X follows a Pareto distribution and ε_i follows either a Pareto or an exponential distribution. By Bayes' rule, the conditional density is

$$f_{X|Y=0}(x) = \frac{\mathbb{P}(Y_i = 0 \mid X_i = x) f_X(x)}{\mathbb{P}(Y_i = 0)} = \frac{(1 - F_\varepsilon(x)) f_X(x)}{\mathbb{P}(Y_i = 0)}, \tag{4}$$

where f_Z denotes the PDF of a generic random variable Z. Suppose X_i follows a heavy-tailed distribution, such as a Pareto with tail index $\lambda > 0$, so that $f_X(x) \propto x^{-1/\lambda-1}$.

When ε_i also follows a heavy-tailed distribution, such as a Pareto with tail index $\xi > 0$, we have $1 - F_\varepsilon(x) \propto x^{-1/\xi}$, and

$$f_{X|Y=0}(x) \propto x^{-1/\lambda-1/\xi-1},$$

so $F_{X|Y=0} \in \mathcal{D}(G_{\xi\lambda/(\xi+\lambda)})$. When the error term has a thin tail with $\xi = 0$, for example, the exponential distribution with rate parameter η, we have $1 - F_\varepsilon(x) = \exp(-\eta x)$, and

$$f_{X|Y=0}(x) \propto x^{-1/\lambda-1} \exp(-\eta x).$$

Since the tail decays exponentially, $F_{X|Y=0} \in \mathcal{D}(G_0)$. Combining both cases, we obtain $F_{X|Y=0} \in \mathcal{D}(G_\gamma)$ with $\gamma = \xi\lambda/(\xi + \lambda)$, so $\xi = 0$ if and only if $\gamma = 0$.

To generalize beyond the Pareto and exponential cases, we adopt the following assumptions. Let $A \sim B$ denote $\lim_{x\to\infty} A/B = 1$.

Assumption 1. (i). Baseline model (1) holds, $\{(X_i, \varepsilon_i)\}_{i=1}^n$ are i.i.d., and $\mathbb{P}(Y_i = 0) > 0$.

(ii). $F_X \in \mathcal{D}(G_\lambda)$ with tail index $\lambda > 0$ and satisfies the von Mises condition in (2).

(iii). ε_i is independent from X_i. $F_\varepsilon \in \mathcal{D}(G_\xi)$ with tail index $\xi \geq 0$ and satisfies the von Mises condition in (2). Furthermore, if $\xi = 0$, $f_\varepsilon(x) \sim C_1 x^{C_2} \exp\{-x^{C_3}/C_4\}$ for some positive constants C_1, C_2, C_3, C_4.

In Assumption 1.(i), the model condition implies that X_i affects Y_i and thus has a nonzero coefficient. Without loss of generality, we take its coefficient to be positive by redefining X_i as $-X_i$ when necessary, and set it to one for normalization. The i.i.d. condition is standard in the binary choice literature. Moreover, $\mathbb{P}(Y_i = 0) > 0$ ensures that the denominator in (4) is well-defined and that the subsample $\{X_i : Y_i = 0\}$ is nondegenerate. Assumption 1.(ii) requires F_X to have a heavy tail, which is necessary to generate the identifying power. Assumption 1.(iii) requires that F_ε is within the domain of attraction of the extreme value distribution, which is quite general since most common distributions such as Gaussian, logistic, Pareto, Student-t, and F distributions meet this assumption.

Remark 1. We emphasize that Assumptions 1.(ii) and 1.(iii) are satisfied regardless of any constant shift. This is because $f_X(t + c)/f_X(t) \to 1$ as $t \to \infty$ for any constant c. This feature is very convenient for our test, especially when we further add discrete or bounded covariates. See for instance Examples 1 and 2 in Section 3.1.

Proposition 1. Suppose Assumption 1 holds. Then,

$$F_{X|Y=0} \in \mathcal{D}\left(G_{\gamma}\right) \text{ with } \gamma = \frac{\xi\lambda}{\xi + \lambda}.$$

Proposition 1 shows that when X_i has a heavy right tail with $\lambda > 0$, the tail behavior of $X_i \mid Y_i = 0$ is governed by that of the latent error ε_i. Under H_0, $\xi = 0$ so $\gamma = 0$, and then $F_{X|Y=0}$ lies in the Gumbel domain and inherits the thin tail behavior from the error. Under H_1, $\xi > 0$ so $\gamma > 0$, and then $F_{X|Y=0}$ again reflects the heavy tail behavior from the error. Hence, testing (3) is *equivalent* to testing whether $\gamma = 0$, which can be implemented using only the subsample $\{X_i : Y_i = 0\}$.

$$H_0 : \gamma = 0 \text{ against } H_1 : \gamma > 0. \tag{5}$$

The identifying power of our diagnostic arises from the extreme observations of the covariate X_i. Since $Y_i = 0$ implies $X_i < \varepsilon_i$, large values of X_i in this subsample correspond to extreme realizations of the latent error. A heavy-tailed regressor thus acts as a "carrier" of tail information: its extreme order statistics encode the tail domain of the unobserved error. Conversely, if X_i were thin-tailed, the sample would rarely generate sufficiently large observations to reveal tail behavior of the error, and the test would lack power. Therefore, requiring one heavy-tailed covariate is not merely technical but essential to distinguish between thin and heavy-tailed errors.

2.2 Test Statistic

We again illustrate the test procedure using the right tail. Proposition 1 directly implies the implementation of the test: under H_0, the normalized spacings of the largest k order statistics of $X_i \mid Y_i = 0$ follow a fixed-k Gumbel limit, whereas systematic deviations indicate $\xi > 0$. Let $n_0 = \sum_i I(Y_i = 0)$ be the number of subsample observations.

Remark 2. Note that by Assumption 1.(i), $n_0/n \to \mathbb{P}(Y_i = 0) > 0$ almost surely as $n \to \infty$. Moreover, for any fixed k, $\mathbb{P}(n_0 \geq k) \to 1$ as $n \to \infty$, so the top-k order statistics in the $Y_i = 0$ subsample are well-defined with probability approaching 1.

Let $\{X_{i_0}^{(0)}, i_0 = 1, \cdots, n_0\}$ denote the subsample with $Y_i = 0$. Sort this subsample descendingly into

$$X_{n_0:n_0}^{(0)} \geq X_{n_0:n_0-1}^{(0)} \geq \cdots \geq X_{n_0:1}^{(0)},$$

where $n_0 : n_0 - j + 1$ denotes the j-th largest observation in the subsample. Consider the largest k order statistics

$$\mathbf{X}^{(0)} = \left(X_{n_0:n_0}^{(0)}, X_{n_0:n_0-1}^{(0)}, \cdots, X_{n_0:n_0-k+1}^{(0)}\right)'.$$

Extreme value theory implies that for any fixed k, there exist sequences of constants a_{n_0} and b_{n_0} such that

$$\begin{aligned}\frac{\mathbf{X}^{(0)} - b_{n_0}}{a_{n_0}} &= \left(\frac{X_{n_0:n_0}^{(0)} - b_{n_0}}{a_{n_0}}, \frac{X_{n_0:n_0-1}^{(0)} - b_{n_0}}{a_{n_0}}, \cdots, \frac{X_{n_0:n_0-k+1}^{(0)} - b_{n_0}}{a_{n_0}}\right)' \\ &\overset{d}{\to} (V_1, \cdots, V_k)' = \mathbf{V},\end{aligned}$$

where $(V_1, \cdots, V_k)'$ is jointly extreme value distributed with PDF

$$f_{\mathbf{V}|\gamma}(v_1, \cdots, v_k) = G_\gamma(v_k) \prod_{j=1}^{k} \frac{g_\gamma(v_j)}{G_\gamma(v_j)} \tag{6}$$

on $(v_k \leq v_{k-1} \leq \cdots \leq v_1)$ with $g_\gamma(v) = \partial G_\gamma(v)/\partial v$, and zero otherwise.

Inference would be straightforward if the constants a_{n_0} and b_{n_0} were known. Unfortunately, they depend on the underlying distribution $F_{X|Y=0}$ and are difficult to estimate accurately. To sidestep this issue, we follow Müller and Wang (2017) to consider the self-normalized statistics

$$\mathbf{X}^{*(0)} = \frac{\mathbf{X}^{(0)} - X_{n_0:n_0-k+1}^{(0)}}{X_{n_0:n_0}^{(0)} - X_{n_0:n_0-k+1}^{(0)}} = \left(1, \frac{X_{n_0:n_0-1}^{(0)} - X_{n_0:n_0-k+1}^{(0)}}{X_{n_0:n_0}^{(0)} - X_{n_0:n_0-k+1}^{(0)}}, \cdots, 0\right)', \tag{7}$$

which is now invariant to location and scale shift. Such normalization is maximum-invariant (e.g., Lehmann and Romano, 2005, Chapter 6) and equivalent to other normalizations in terms of testing power.

By the continuous mapping theorem and the extreme value theorem, we have that

$$\mathbf{X}^{*(0)} \xrightarrow{d} \left(1, \frac{V_2 - V_k}{V_1 - V_k}, \cdots, 0\right)' = \mathbf{V}^*.$$

Some calculation yields that

$$f_{\mathbf{V}^*|\gamma}\left(\mathbf{v}^*\right) = \Gamma\left(k\right) \int_0^\infty u^{k-2} \exp\left(-\left(1 + \frac{1}{\gamma}\right)\left(\sum_{j=1}^k \log\left(1 + \gamma v_j^* u\right)\right)\right) du,$$

where $\Gamma\left(\cdot\right)$ denotes the gamma function. This density is now uniquely characterized by γ, and hence can be used to construct the likelihood ratio test.

Recall the hypotheses in (5), where the null is simple with $\gamma = 0$. If the alternative were also simple, say $\gamma = \gamma_1$ for some $\gamma_1 > 0$, the Neyman-Pearson lemma suggests that the optimal test in the limiting problem rejects if the likelihood ratio is greater than the critical value

$$\varphi(\mathbf{V}^*) = I\left(\frac{f_{\mathbf{V}^*|\gamma_1}\left(\mathbf{V}^*\right)}{f_{\mathbf{V}^*|0}\left(\mathbf{V}^*\right)} > \mathrm{cv}(\alpha, k)\right),$$

where the critical value $\mathrm{cv}(\alpha, k)$ depends on the significance level α and the tail size k.

Since the alternative hypothesis is composite, we follow Andrews and Ploberger (1994) and adopt a weighted average over alternatives, with $w(\gamma) > 0$ and $\int_{\mathbb{R}^+} w(\gamma)d\gamma = 1$. The likelihood ratio test is then given by

$$\varphi\left(\mathbf{V}^*\right) = I\left(\frac{\int_{\mathbb{R}^+} f_{\mathbf{V}^*|\gamma}\left(\mathbf{V}^*\right) w\left(\gamma\right) d\gamma}{f_{\mathbf{V}^*|\gamma=0}\left(\mathbf{V}^*\right)} > \mathrm{cv}\left(\alpha, k\right)\right), \tag{8}$$

In practice, we substitute $\mathbf{V}^*$ by its finite sample analog $\mathbf{X}^{*(0)}$ as in (7). By the continuous mapping theorem, our test controls size asymptotically.

Theorem 1. *Suppose Assumption 1 holds. Then, for any finite k, our test* (8) *controls size asymptotically, i.e.,* $\lim_{n\to\infty} \mathbb{P}(\varphi(\mathbf{X}^*) = 1 \mid H_0) = \alpha$.

Several remarks clarify the practical implementation and theoretical interpretation of this result. First, the weighting function $w(\gamma)$ determines how the test integrates information across alternative tail indices. We employ a uniform weight $w(\gamma) = 1$ for $\gamma \in [0, 1]$, which mirrors the exponential average tests of Andrews and Ploberger (1994) and yields balanced power against a broad class of heavy-tailed alternatives while maintaining a simple imple-

$k\backslash\alpha$	0.10	0.09	0.08	0.07	0.06	0.05	0.04	0.03	0.02	0.01
10	1.48	1.58	1.70	1.82	2.00	2.22	2.52	2.98	3.72	5.65
25	1.19	1.30	1.42	1.57	1.74	2.00	2.45	3.05	3.96	6.96
50	0.79	0.87	0.95	1.08	1.22	1.45	1.70	2.13	2.95	4.98
70	0.65	0.72	0.79	0.87	0.99	1.15	1.44	1.78	2.44	4.19
100	0.54	0.58	0.64	0.71	0.80	0.92	1.09	1.38	1.99	3.73

Table 1: Critical values of the likelihood ratio test (8) with different k and significance level α. Based on 10,000 simulation draws.

mentation. Other weighting schemes, such as exponential or truncated normal densities, may emphasize specific regions of γ, but simulations indicate that the uniform weight achieves a near-optimal trade-off between robustness and sensitivity. Critical values can be obtained by simulation. Table 1 presents the critical values for various combinations of α and k.

Second, following Müller (2011), our likelihood ratio tests based on fixed-k normalized spacings of the largest order statistics are asymptotically most powerful among all tests invariant to affine transformations. The self-normalization in (7) eliminates the nuisance parameters (a_{n_0}, b_{n_0}), leaving the joint law of $\mathbf{X}^{*(0)}$ dependent only on the tail index γ. Consequently, the test remains valid under arbitrary rescaling or shifts of the covariate X_i. This invariance ensures empirical robustness when covariates are measured in different units.

Third, many thin-tailed error distributions under the null, such as the Gaussian or logistic, are symmetric, so we test both tails of the conditional distribution: the right-tail statistic uses $X_i \mid Y_i = 0$ and the left-tail statistic uses $X_i \mid Y_i = 1$. To control size when testing both sides, we use a Bonferroni correction and reject the global null of thin-tailed errors if either tail rejects at level $\alpha/2$. Formally, letting T_R and T_L denote the right and left tail test statistics, respectively, we define

$$\varphi_{\text{two-sided}} = I\left(T_R > \text{cv}(\alpha/2, k) \text{ or } T_L > \text{cv}(\alpha/2, k)\right),$$

where we use the same k for both sides for simple implementation. Equivalently, the combined p-value can be written as

$$p_{\text{two-sided}} = 2\,\min\{p_R,\, p_L\},$$

where p_L and p_R denote the p-values for the left and right tail, respectively. Then the null hypothesis is rejected whenever $p_{\text{two-sided}} < \alpha$.

Fourth, the test is computationally light since it relies solely on tail observations and requires no model estimation, making it a convenient diagnostic for evaluating thin tail assumptions in binary choice models.

Taken together, the implementation features uniform weighting, fixed-k asymptotics, scale invariance, and bilateral Bonferroni correction. It yields a diagnostic that controls size asymptotically and is straightforward to apply.

3 Extensions

We now consider two important extensions. Section 3.1 discusses the case with multiple covariates and Section 3.2 analyzes panel data models.

3.1 Multiple Covariates

Consider the binary choice model

$$Y_i = I(X_i + X_i^{a\prime}\beta - \varepsilon_i \geq 0), \tag{9}$$

where X_i is the scalar heavy-tailed *dominating* regressor and X_i^a collects the *auxiliary* variables. The coefficient on the dominating regressor is again normalized to one without loss of generality, provided that X_i affects Y_i conditional on X_i^a. We impose following conditions.

Assumption 2. (i). Model (9) with multiple covariates holds, $\{(X_i, X_i^{a\prime}, \varepsilon_i)'\}_{i=1}^n$ are i.i.d., and $\mathbb{P}(Y_i = 0) > 0$.

(ii). $F_X \in \mathcal{D}(G_\lambda)$ with tail index $\lambda > 0$ and satisfies the von Mises condition in (2).

(iii). ε_i is independent from $(X_i, X_i^{a\prime})'$. $F_\varepsilon \in \mathcal{D}(G_\xi)$ with tail index $\xi \geq 0$ and satisfies the von Mises condition in (2). Furthermore, if $\xi = 0$, $f_\varepsilon(x) \sim C_1 x^{C_2} \exp\{-x^{C_3}/C_4\}$ for some positive constants C_1, C_2, C_3, C_4.

(iv). $f_{X+X^{a\prime}\beta}(x)/f_X(x) \to 1$ as $x \to \infty$.

Assumptions 2.(i)–(iii) are analogous to Assumptions 1.(i)–(iii). Assumption 2.(iv) implies that X_i dominates the tail so that $X_i^{a\prime}\beta$ has no asymptotic effect. This condition is satisfied if all components of X_i^a are discrete or bounded variables, as illustrated in the two empirical examples below. Also, if some components of X_i^a have unbounded support, we can estimate the tail heaviness of each component and check if they are thinner than the dominating regressor X_i.

Example 1 (Firm export)**.** One empirical example is the heterogeneous firm trade models following Melitz (2003). In these models, it is usually assumed that firms draw productivity from a heavy-tailed Pareto distribution (e.g., Chaney, 2008), and export participation is determined by whether productivity exceeds a fixed export cost threshold. Let $Y_i = 1$ if firm exports. Firm size, as a measure of productivity, serves as the dominating regressor X_i, while the auxiliary block X_i^a includes ownership indicators, policy variables, and industry or region fixed effects, each discrete or bounded.

Example 2 (Female labor force participation)**.** Another example is the female labor force participation, such as Fernández-Val (2009). Let $Y_i = 1$ if the woman participates in the labor force. Husband's income serves as the dominating regressor X_i, while the auxiliary block X_i^a includes demographic variables such as age, education, and the numbers of children in different age groups, each discrete or bounded.

Given Assumption 2, we establish the following proposition.

Proposition 2. Suppose Assumption 2 holds. Then,

$$F_{X|Y=0} \in \mathcal{D}\left(G_\gamma\right) \text{ with } \gamma = \frac{\xi\lambda}{\xi+\lambda}.$$

Proposition 2 shows that in the presence of multiple covariates, the tail behavior of the latent error can still be recovered from a single heavy-tailed regressor. When one covariate dominates the right tail of the linear index, the conditional distribution $X_i \mid Y_i = 0$ remains in the same extreme value domain of attraction as in the univariate case, yielding the tail index $\gamma = \xi\lambda/(\xi+\lambda)$. Consequently, our diagnostic can be implemented using only the regressor with the heaviest tail, without estimating β or forming the full linear index $X_i + X_i^{a\prime}\beta$.

Remark 3. We also provide a plug-in approach that does not require knowing which regressor dominates the tail in Appendix A. Write the linear index as $W_i = \widetilde{X}_i'\delta$ for a full vector of regressors $\widetilde{X}_i \in \mathbb{R}^p$, with δ identified up to scale. One may estimate δ by a distribution-free method, such as the maximum score estimator, and form the fitted index $\widehat{W}_i = \widetilde{X}_i'\widehat{\delta}$. Under Assumption 4 in Appendix A, the extreme order statistics of $\widehat{W}_i$ conditional on $Y_i = 0$ have the same limiting distribution as those of the true index W_i, so by Proposition 4 the tail test applied to $\widehat{W}$ has the same asymptotic distribution and size as if δ were known.

As both the dominating variable and the plug-in approaches are conceptually similar, for simplicity, we focus on the dominating variable one in the rest of the paper.

3.2 Panel Data

We extend our framework to short-T panel binary choice models with unobserved individual effects and possibly predetermined covariates, such as lagged dependent variables. Note that one advantage of our approach is that we do not need to integrate out or difference out the unobserved individual effects. More specifically, consider the model

$$Y_{it} = I\left(X_{it} + X_{it}^{a\prime}\beta + \rho Y_{i,t-1} + \alpha_i - \varepsilon_{it} \geq 0\right), \tag{10}$$

where α_i is the unobserved individual effect that is constant over t, ρ captures the persistence of outcome Y_{it}, and ε_{it} denotes the idiosyncratic error term.

Assumption 3. (i). Panel model (10) holds, with T fixed. $\left\{\{X_{it}, X_{it}^{a}, \varepsilon_{it}\}_{t=1}^{T}, Y_{i0}, \alpha_i\right\}_{i=1}^{n}$ are all i.i.d. across i. Also assume $\mathbb{P}(Y_{it} = 0) > 0$.

For each t:

(ii). $F_{X_t} \in \mathcal{D}(G_{\lambda_t})$ with tail index $\lambda_t > 0$ and satisfies the von Mises condition in (2). The subscript i is suppressed as the densities are written for the generic distribution at time t.

(iii). ε_{it} is independent from $(X_{it}, X_{it}^{a\prime}, Y_{i,t-1}, \alpha_i)'$. $F_{\varepsilon_t} \in \mathcal{D}(G_{\xi_t})$ with tail index $\xi_t \geq 0$ and satisfies the von Mises condition in (2). Furthermore, if $\xi_t = 0$, $f_{\varepsilon_t}(x) \sim C_{1t}x^{C_{2t}}\exp\{-x^{C_{3t}}/C_{4t}\}$ for some positive constants $C_{1t}, C_{2t}, C_{3t}, C_{4t}$.

(iv). $f_{X_t + X_t^{a\prime}\beta + \rho Y_{t-1} + \alpha}(x)/f_{X_t}(x) \to 1$ as $x \to \infty$,

Assumption 3 is analogous to Assumption 2. Since we work with short-T panels, we do not require strict stationarity across t. Given Assumption 3.(iv), the tail of $X_{it}^{a\prime}\beta + \rho Y_{i,t-1} + \alpha_i$ is dominated entirely by that of X_{it}. In particular, the lagged outcome is binary and thus enters the linear index only as a bounded shift. As a result, the tail behavior that underlies our diagnostic test is identical to that in the cross-sectional case period by period, and it suffices to focus on the dominating coordinate of X_{it} that exhibits the heaviest tail without estimating $(\beta, \rho')'$ and $\{\alpha_i\}_{i=1}^n$.

Proposition 3. Suppose Assumption 3 holds. Then, for each t,

$$F_{X_t|Y_t=0} \in \mathcal{D}(G_{\gamma_t}) \text{ with } \gamma_t = \frac{\xi_t \lambda_t}{\xi_t + \lambda_t}.$$

Proposition 3 shows that our tail-based diagnostic extends naturally to panel binary choice models with unobserved individual effects and predetermined covariates, so that a likelihood-ratio statistic based on the top-k order statistics provides a valid test of $H_0 : \xi = 0$.

Implementation is similar to the cross-sectional procedure. We can conduct the test separately for each period and combine the period-specific results using a Bonferroni adjustment. As in Proposition 1, asymptotic size control holds for each period, though the Bonferroni may be mildly conservative. If desired, we can also combine the two tails using the same adjustment. Unlike conventional nonlinear panel estimators, our method requires neither differencing nor integrating out the unobserved individual effects and remains feasible even when T is small and with lagged dependent variables.

4 Monte Carlo Simulations

We assess finite sample size and power of the proposed test (8) in both cross-sectional and panel designs in Section 4.1 and 4.2, respectively.

4.1 Cross-Sectional Model

We start with the cross-sectional case and generate data from the following model:

$$Y_i = I\left(X_i + X_{1,i}^a + X_{2,i}^a - \varepsilon_i \geq 0\right),$$

where the error term ε_i is drawn from four distributions: standard normal, logistic, and two Student-t distributions with different degrees of freedom. The normal and logistic distributions represent the null hypothesis, while the Student-t distributions serve as heavy-tailed alternatives. The dominating heavy-tailed covariate X_i is drawn from a Student-$t(2)$ distribution. The other two components, $X_{1,i}^a$ and $X_{2,i}^a$, are thin-tailed and drawn from a

Dist. of ε	n	$k=10$			$k=25$			$k=50$			$k=70$		
		Left	Right	Both	Left	Right	Both	Left	Right	Both	Left	Right	Both
Normal	1000	0.03	0.03	0.02	0.02	0.02	0.02	0.01	0.01	0.01	0.01	0.00	0.00
	2000	0.03	0.03	0.03	0.02	0.03	0.02	0.01	0.02	0.01	0.01	0.01	0.01
	5000	0.02	0.02	0.03	0.03	0.02	0.02	0.02	0.02	0.01	0.01	0.01	0.01
Logistic	1000	0.04	0.04	0.05	0.04	0.04	0.05	0.04	0.04	0.03	0.02	0.03	0.03
	2000	0.05	0.06	0.05	0.05	0.05	0.05	0.05	0.05	0.06	0.06	0.06	0.06
	5000	0.05	0.05	0.05	0.05	0.06	0.06	0.06	0.05	0.06	0.06	0.06	0.06
$t(2)$	1000	0.09	0.10	0.12	0.14	0.13	0.18	0.14	0.15	0.17	0.21	0.21	0.21
	2000	0.12	0.11	0.14	0.17	0.17	0.24	0.21	0.21	0.30	0.30	0.29	0.31
	5000	0.13	0.15	0.19	0.21	0.22	0.30	0.32	0.31	0.47	0.47	0.47	0.47
$t(1)$	1000	0.17	0.17	0.23	0.31	0.29	0.41	0.37	0.38	0.51	0.51	0.51	0.51
	2000	0.18	0.18	0.23	0.33	0.34	0.46	0.46	0.48	0.71	0.71	0.71	0.71
	5000	0.18	0.18	0.24	0.37	0.35	0.50	0.56	0.56	0.73	0.65	0.66	0.83

Table 2: Simulation results for cross-sectional data. Entries are the rejection rates of our proposed test (8). The normal and logistic distributions correspond to the null hypothesis, and the Student-t distributions correspond to the alternative. Significance level is 5%. Based on 2,000 simulation draws.

standard normal distribution and a binary distribution, respectively. We consider sample sizes of $n \in \{1000, 2000, 5000\}$.

We apply the test using only the dominating covariate X_i, so no model estimation is required, which in turn reduces potential misspecification issues. We report rejection rates for the left and right tails of ε_i separately in columns labeled Left and Right in Table 2, and for the joint null that both tails are thin using Bonferroni critical values in columns labeled Both. Results are shown for various numbers of tail observations $k \in \{10, 25, 50, 70\}$.

We summarize the findings in Table 2 as follows. First, size is well controlled under the null for both normal and logistic errors, lying in the domain of attraction with $\xi = 0$. The test yields slightly more conservative rejection rates under normal than under logistic errors, since the normal tail decays as $\exp(-cx^2)$ and the logistic as $\exp(-cx)$. Moreover, mild size distortions arise when k is large relative to n, such as $k = 70$ with relatively small n, as too many mid-sample order statistics then enter the tail approximation and the extreme value approximation becomes rough.

Second, the test exhibits strong power against heavy-tailed alternatives. The test retains decent power even for small k, such as $k = 10$. Power increases with k and with tail heaviness, with heavier tails such as $t(1)$ being detected more easily than $t(2)$. Joint testing of both tails yields higher power, as expected. Note that power increases mainly with k rather than n, since larger n improves the extreme-value approximation but does not directly raise power.

4.2 Panel Data Models

We examine both static and dynamic panel models with unobserved individual heterogeneity. In the static case, we generate data from

$$Y_{it} = I(X_{it} + X^a_{1,it} + X^a_{2,it} + \alpha_i - \varepsilon_{it} \geq 0), \tag{11}$$

where $\{X_{it}, X^a_{1,it}, X^a_{2,it}, \varepsilon_{it}\}$ are i.i.d. across i and t with the same marginal distributions as in the cross-sectional design. The individual effect α_i is drawn from the standard uniform distribution $U(0,1)$. We consider sample sizes $n \in \{1000, 2000, 5000\}$ and $T = 2$. In the dynamic case, the data generating process is

$$Y_{it} = I(X_{it} + Y_{i,t-1} + X^a_{1,it} + X^a_{2,it} + \alpha_i - \varepsilon_{it} \geq 0),$$

with the initial period Y_{i0} generated from (11). All other specifications match the static panel design.

We implement the test period by period and combine the period-specific results with a Bonferroni adjustment. The results in Tables 3 for the static panel and 4 for the dynamic panel show patterns similar to the cross-sectional case. The test maintains good size control, and power increases with k and tail heaviness. As expected, the Bonferroni correction is slightly conservative, but the test remains powerful. Overall, size and power behave similarly across cross-sectional, static panel, and dynamic panel settings, indicating that our test is well-suited to a wide range of data structures.

5 Empirical Examples

We apply the proposed diagnostic test in three empirical examples. Sections 5.1 and 5.2 examine firm-level exporting and innovation, illustrating the cross-sectional case; Section 5.3 studies female labor force participation, demonstrating the panel data case.

In each example, we use a single dominating heavy-tailed covariate X_i, such as firm asset size or husband's income, and the test infers the error tail from the conditional extremes of X_i without estimating the binary choice model, so our test is simple to implement and robust to potential model misspecification. Though the test does not require specifying the auxiliary regressors X^a_i, we discuss them in the examples for completeness as well as for

Dist. of ε	n		$k=10$			$k=25$			$k=50$			$k=70$	
		Left	Right	Both	Left	Right	Both	Left	Right	Both	Left	Right	Both
Normal	1000	0.04	0.03	0.04	0.02	0.02	0.01	0.01	0.01	0.01	0.00	0.00	0.00
	2000	0.03	0.03	0.03	0.01	0.02	0.02	0.01	0.01	0.01	0.01	0.01	0.01
	5000	0.03	0.03	0.03	0.02	0.03	0.02	0.01	0.01	0.01	0.01	0.02	0.01
Logistic	1000	0.04	0.05	0.05	0.06	0.05	0.06	0.05	0.05	0.05	0.03	0.03	0.03
	2000	0.03	0.05	0.05	0.05	0.05	0.05	0.06	0.05	0.07	0.06	0.05	0.06
	5000	0.04	0.04	0.04	0.06	0.06	0.05	0.08	0.06	0.07	0.06	0.07	0.08
$t(2)$	1000	0.13	0.13	0.16	0.16	0.17	0.22	0.19	0.20	0.27	0.16	0.16	0.22
	2000	0.14	0.15	0.18	0.23	0.23	0.29	0.30	0.29	0.39	0.30	0.30	0.41
	5000	0.15	0.18	0.21	0.30	0.31	0.40	0.40	0.41	0.54	0.48	0.45	0.61
$t(1)$	1000	0.22	0.22	0.29	0.40	0.40	0.55	0.50	0.49	0.68	0.52	0.52	0.69
	2000	0.24	0.24	0.32	0.46	0.46	0.60	0.64	0.64	0.81	0.71	0.70	0.86
	5000	0.25	0.25	0.34	0.47	0.47	0.64	0.73	0.73	0.88	0.83	0.83	0.95

Table 3: Simulation results for static panel data with $T = 2$. Entries are the rejection rates of our proposed test (8). The normal and logistic distributions correspond to the null hypothesis, and the Student-t distributions correspond to the alternative. Significance level is 5%. Based on 2,000 simulation draws.

Dist. of ε	n		$k=10$			$k=25$			$k=50$			$k=70$	
		Left	Right	Both	Left	Right	Both	Left	Right	Both	Left	Right	Both
Normal	1000	0.03	0.03	0.03	0.02	0.02	0.01	0.00	0.01	0.01	0.00	0.01	0.01
	2000	0.03	0.03	0.03	0.02	0.02	0.02	0.01	0.01	0.01	0.00	0.01	0.01
	5000	0.03	0.03	0.02	0.03	0.02	0.02	0.02	0.01	0.01	0.01	0.01	0.01
Logistic	1000	0.05	0.04	0.05	0.05	0.04	0.04	0.02	0.05	0.04	0.01	0.05	0.03
	2000	0.05	0.04	0.05	0.06	0.05	0.06	0.05	0.04	0.05	0.04	0.06	0.05
	5000	0.05	0.04	0.05	0.06	0.04	0.06	0.07	0.05	0.07	0.08	0.05	0.07
$t(2)$	1000	0.11	0.10	0.13	0.17	0.15	0.21	0.13	0.14	0.20	0.10	0.14	0.16
	2000	0.16	0.13	0.17	0.25	0.19	0.29	0.27	0.22	0.33	0.26	0.22	0.34
	5000	0.18	0.16	0.23	0.29	0.26	0.36	0.41	0.32	0.50	0.45	0.35	0.56
$t(1)$	1000	0.25	0.22	0.31	0.38	0.35	0.47	0.44	0.43	0.60	0.45	0.45	0.63
	2000	0.24	0.26	0.34	0.43	0.45	0.58	0.61	0.60	0.78	0.71	0.62	0.84
	5000	0.26	0.25	0.33	0.47	0.51	0.64	0.73	0.73	0.88	0.83	0.82	0.95

Table 4: Simulation results for dynamic panel data with $T = 2$. Entries are the rejection rates of our proposed test (8). The normal and logistic distributions correspond to the null hypothesis, and the Student-t distributions correspond to the alternative. Significance level is 5%. Based on 2,000 simulation draws.

verifying Assumption 2.(iv) that X_i dominates X_i^a in the tail.

5.1 Export and Firm Size

We first examine firms' decision to export, a classic binary choice problem in international trade. According to heterogeneous firm trade models (Melitz, 2003; Bernard et al., 2007), larger and more productive firms are more likely to enter export markets because they can better absorb fixed trade costs. Empirically, firm size is well known to display a heavy-tailed distribution: see for example, Axtell (2001), Gabaix (2009), and di Giovanni and Levchenko (2012).

Following the cross-sectional setup with multiple covariates in (9), let $Y_i = 1$ if firm i exports in a given year and $Y_i = 0$ otherwise. We focus on firm size as the key covariate, which can be measured in several ways, including employment, sales, and assets, and here we use total assets as our measure of firm size. The dominating regressor X_i is therefore the firm's total asset size, which proxies for productivity and financial capacity and satisfies the heavy-tailedness in Assumption 1.(ii). Ownership indicators, policy variables, and industry or region fixed effects enter as auxiliary regressors X_i^a. Since these variables are discrete or bounded, the dominance condition in Assumption 2.(iv) holds.

We use the National Tax Survey Database (NTSD) from China, which is jointly administered by the Chinese State Administration of Taxation and the Ministry of Finance. The NTSD is a stratified random sample of taxpayers drawn each year and broadly regarded as the most reliable source of firm-level tax and performance information in China. It contains detailed tax data and firm characteristics, including sales, employment, asset size, and export value. The dataset has been widely used in prior work such as Chen et al. (2023) and Liu and Mao (2019).

We construct our sample from the NTSD using observations for which both the export decision and asset size are available. The resulting sample is large, with $n = 429{,}032$ in 2014 and $n = 394{,}814$ in 2015. Figure 1 shows a Pareto fit to the top 0.1% of the asset distribution among non-exporters ($X_i \mid Y_i = 0$). The fitted Pareto CDF overlaps closely with the empirical one, suggesting that the domain of attraction assumption is plausibly satisfied. The Hill estimator yields $\widehat{\gamma} = 0.80$ for 2014 and $\widehat{\gamma} = 0.78$ for 2015 for the conditional distribution $X_i \mid Y_i = 0$, so the conditional distributions indeed exhibit heavy tails and even lack finite moments.

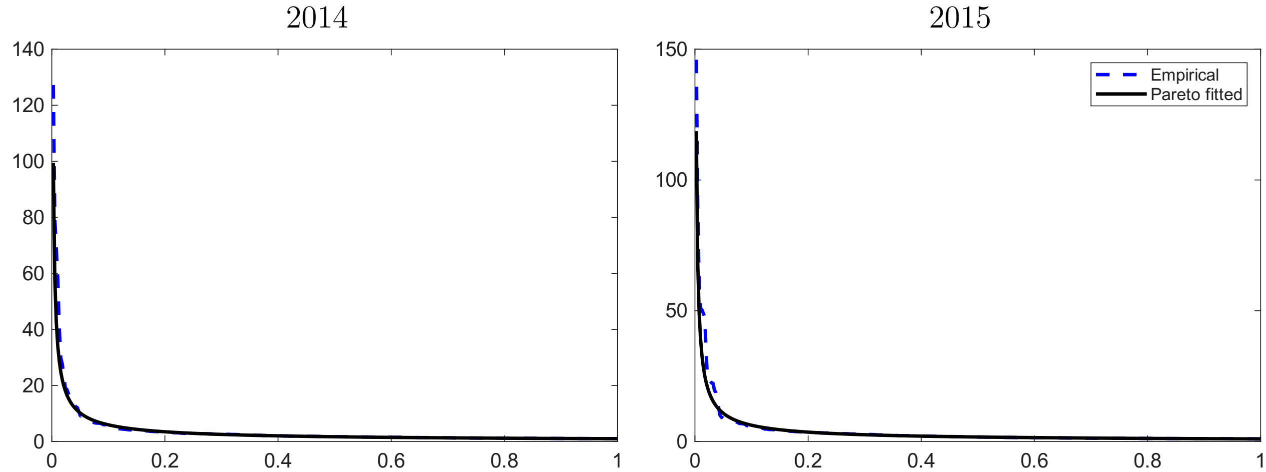

Figure 1: Pareto fit of firm assets among non-exporters ($X_i \mid Y_i = 0$): empirical (blue dashed) and fitted Pareto (black solid), top 0.1%.

H_0 : Right tail of ε is thin				
Sample	$k = 25$	50	70	100
2014	0.09	0.00	0.00	0.00
2015	0.09	0.00	0.00	0.00

Table 5: The p-values of our test (8) in firm export dataset. Y_i indicates whether firm exports or not, and X_i is the firm's asset size.

Given limited power at $k = 10$ in our simulations and the large sample sizes in the current example, we omit the $k = 10$ case and compute the likelihood ratio statistic (8) for $k \in \{25, 50, 70, 100\}$. Table 5 reports the resulting p-values. At $k = 25$, the p-value is 0.09, and it declines rapidly as k increases and reaches nearly zero for $k \geq 50$. The conditional asset distribution among non-exporters exhibits a heavy upper tail, so we reject the thin-tail null for the latent error. Economically, this suggests that export participation is likely influenced by rare but sizable unobserved shocks, such as changes in market access, policy shifts, or exchange rate movements, that affect export decisions more strongly than standard logit or probit models would allow. This interpretation is also consistent with the well-documented fact that a small share of firms export (Bernard et al., 2007).

5.2 Innovation and Firm Size

We next analyze firms' binary decision to innovate. In the Schumpeterian growth literature, innovation and firm size are closely related, because innovation generates rents and market

H_0 : Right tail of ε is thin				
Sample	$k = 25$	50	70	100
2014	0.09	0.00	0.00	0.00
2015	0.09	0.00	0.00	0.00

Table 6: The p-values of our test (8) in firm innovation dataset. Y_i indicates whether firm files at least one patent or not, and X_i is the firm's asset size.

leadership, and firm scale affects the returns to R&D and the ability to appropriate those rents (Cohen, 2010; Aghion et al., 2014). Compared to the exporting example in Section 5.1, exporting and innovation are often empirically linked (Aw et al., 2011; Lileeva and Trefler, 2010), though the shock processes governing export and innovation decisions can exhibit different tail behaviors.

Similarly, we maintain a setup analogous to the export example. The binary outcome Y_i is defined as an indicator for whether firm i files at least one patent in a given year, so that $Y_i = 1$ if the firm patents and $Y_i = 0$ otherwise. The dominating regressor X_i is again the firm's asset size, which proxies for the resources and scale economies in R&D. The potential auxiliary regressors X_i^a are the same as in the export example.

We merge the NTSD with patent filing data from China's State Intellectual Property Office (SIPO). Since all patent applications in China must be submitted to SIPO, these data provide comprehensive information on patent filings and grants. The publicly available records include the filing date, applicant name and address, firm name, and patent type. Several previous studies have used the SIPO data, such as Liu and Qiu (2016) and Liu and Ma (2020).

The samples are similar to the export example. We compute the test statistic (8) and report the p-values in Table 6. The thin-tailed null is rejected again, with p-values near zero for $k \geq 50$. Therefore, the latent error of the innovation decision is also heavy-tailed, and conventional probit or logit may understate the probability of extreme innovative responses.

5.3 Female Labor Force Participation and Husband's Income

Our final example revisits the classic female labor force participation setting using the PSID data. For example, Fernández-Val (2009) estimate a fixed effects panel probit with bias correction in this setting. The sample spans 1980–1988 with $T = 9$ years and consists of $n = 1{,}461$ married women who were aged 18–60 as of 1985 and whose husbands were

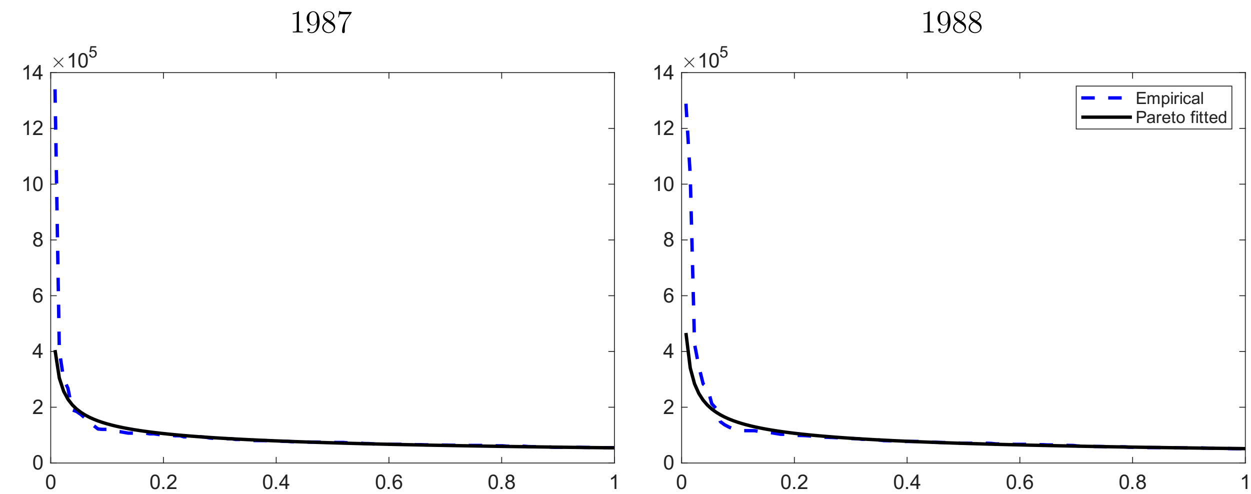

Figure 2: Pareto fit of husband's income among non-participants ($X_{it} \mid Y_{it} = 0$): empirical (blue dashed) and fitted Pareto (black solid), top 1%.

continuously employed.

We specify the dynamic panel data model as (10), similar to Eq. (5.1) in Fernández-Val (2009). The binary outcome Y_{it} equals one if the woman participates in the labor force and zero otherwise. The dominating regressor X_{it} is husband's income, which has heavy tails and satisfies Assumption 1.(ii). The auxiliary regressors X_{it}^a include log husband's income, which has thin tails, as well as the numbers of children aged 0–2, 3–5, and 6–17, age, and age squared, which are discrete or bounded, so the dominance condition in Assumption 2.(iv) applies. The individual effect α_i captures unobserved heterogeneity such as capability or willingness to work. Note that the coefficient on husband's income remains significant when the log of husband's income is also included, justifying the use of husband's income as the dominating regressor for the diagnostic test.

We now assess the normal error assumption in the conventional panel probit model. Figure 2 plots a Pareto fit to the upper 1% of the husband's income distribution among non-participants ($X_{it} \mid Y_{it} = 0$) for the latest two years 1987 and 1988. The plots for other years are very similar and hence omitted. For the conditional distribution $X_{it} \mid Y_{it} = 0$, the estimated tail indices range from $\widehat{\gamma}_t = 0.36 - 0.45$ across t, supporting that husband's income is heavy-tailed and thus provides sufficient identifying power for the test.

We apply the likelihood ratio statistic (8) to the right tail for $k \in \{25, 50, 70, 100\}$. Table 7 reports the year-specific p-values, as well as the p-values for the full panel via a Bonferroni correction. In the early years (1980–1982), the evidence against thin tails is

H_0 : Right tail of ε is thin									
Sample	$k=25$	50	70	100	Sample	$k=25$	50	70	100
1980	0.62	0.42	0.07	0.00	1985	0.00	0.00	0.00	0.00
1981	0.97	0.55	0.21	0.02	1986	0.00	0.00	0.00	0.00
1982	0.84	0.08	0.02	0.00	1987	0.00	0.00	0.00	0.00
1983	0.04	0.00	0.00	0.00	1988	0.00	0.00	0.00	0.00
1984	0.00	0.00	0.00	0.00	Panel	0.00	0.00	0.00	0.00

Table 7: The p-values of our test (8) in the female labor force participation data. Y_{it} indicates whether the individual participates in the labor force or not, and X_{it} is the husband's income.

mixed at small k and becomes stronger as k increases, 1983 already shows strong rejection, and from 1984 onward the p-values are essentially zero across all k. This pattern suggests possible nonstationarity across time. Notably, however, our test remains valid in the presence of such nonstationarity, as shown in Proposition 3. For the full panel, all p-values are close to zero, so we reject the thin-tailed null. This rejection indicates that unobserved shocks in female labor supply may be more dispersed than under standard probit or logit. For example, health shocks, childcare disruptions, or temporary wage opportunities may exert unusually large effects on labor force participation decisions.

From a modeling standpoint, our diagnostic complements the bias correction of Fernández-Val (2009) by targeting a different source of misspecification. While his correction mitigates the incidental parameters bias under normal errors, our test assesses whether the normality assumption is empirically plausible. Rejection of H_0 therefore suggests adopting a heavier-tailed or semiparametric link for the participation process.

Across three cross-sectional and panel examples, we see that heavy-tailed latent errors are empirically relevant in binary choice models, and our diagnostic effectively differentiates between when conventional probit/logit remains adequate versus when tail-robust alternatives are needed.

6 Conclusion

This paper develops a simple yet powerful diagnostic test for the distributional assumptions underlying binary choice models. It exploits the connection between the tail behavior of observed covariates and that of the latent error, and requires no model estimation or parametric assumption beyond the domain of attraction assumption. The test uses the normalized spacings of the covariate's largest order statistics given $Y_i = y$ for $y = 0, 1$, yielding a likelihood ratio statistic that is invariant to scale and location. The null corresponds to the conventional logit or probit specification, while rejection indicates a heavy-tailed alternative such as Student-t or Pareto. Because the procedure uses only the observable covariate and the binary outcome, it can serve as a pre-estimation diagnostic for the credibility of the assumed link function.

We also extend the diagnostic to settings with multiple covariates and to panel data with individual effects and dynamic specifications. The test relies only on the most heavy-tailed covariate, thereby preserving the simplicity of a one-dimensional procedure and avoiding the incidental parameters problem in nonlinear panel estimation.

Monte Carlo evidence shows good size control and high power against heavy-tailed alternatives, even with modest sample sizes. Applications to firm export and innovation and to female labor force participation illustrate the diagnostic. Rejections of the null hypothesis of thin tails suggest that rare and large shocks may play a significant role in shaping binary choices in these contexts, and that standard logit or probit specifications may be inadequate.

Moving forward, several extensions are promising, such as data-adaptive choices of the tail sample size k and multinomial discrete choice models. In general, our tail-based diagnostics complement conventional specification tests by focusing on the behavior of extremes, a crucial yet under-examined aspect of economic data. We hope this framework encourages broader use of extreme value theory in econometric analysis, especially for discrete choice models.

A Details for the Plug-in Approach

This appendix gives conditions under which the extreme order statistics of the *fitted* linear index have the same extreme value limit as the unknown true linear index. This justifies implementing the diagnostic on a fitted index without knowing which coordinate is tail dominating.

Let $\widetilde{X}_i \in \mathbb{R}^p$ and consider

$$Y_i = I(W_i - \varepsilon_i \geq 0), \quad W_i = \widetilde{X}_i'\delta, \tag{12}$$

where δ is identified up to scale. We use $\widetilde{X}_i$ and δ here to emphasize that we need not know which coordinate is tail dominating, nor normalize the coefficient on the dominating variable to one. Let $\widehat{\delta}$ be an estimator of δ under certain normalization, and define the fitted linear index $\widehat{W}_i = \widetilde{X}_i'\widehat{\delta}$.

Assumption 4. (i) Linear index model (12) holds, $(Y_i, \widetilde{X}_i, \varepsilon_i)'$ are i.i.d., and $\mathbb{P}(Y_i = 0) > 0$.

(ii) $F_W \in \mathcal{D}(G_\lambda)$ with tail index $\lambda > 0$ and satisfies the von Mises condition in (2).

(iii) ε_i is independent from $\widetilde{X}_i$. $F_\varepsilon \in \mathcal{D}(G_\xi)$ with tail index $\xi \geq 0$ and satisfies the von Mises condition in (2). Furthermore, if $\xi = 0$, $f_\varepsilon(x) \sim C_1 x^{C_2} \exp\{-x^{C_3}/C_4\}$ for some positive constants C_1, C_2, C_3, C_4.

(iv) $\limsup_{x\to\infty} \left(1 - F_{\|\widetilde{X}\| \,|Y=0}(x)\right) / \left(1 - F_{W|Y=0}(x)\right) < \infty$.

(v) $\|\widehat{\delta} - \delta\| = o_p(n^{-\eta})$ for some $\eta > 0$.

Assumptions 4.(i)–(iii) are analogous to Assumption 1.(i)–(iii). For Assumption 4.(ii), a special case is that there exists one dominating regressor $\widetilde{X}_{j^\star}$ with nonzero $\delta_{j^\star}$, such that the right tail of $\text{sign}(\delta_{j^\star})\widetilde{X}_{j^\star}$ is strictly heavier than all other $\text{sign}(\delta_j)\widetilde{X}_j$ for $j \neq j^\star$, so the right tail of W is asymptotically equivalent to that of $\delta_{j^\star}\widetilde{X}_{j^\star}$. Assumption 4.(iv) accounts for the interaction of estimation error in $\widehat{\delta}$ with $\widetilde{X}$ in both tails, and rules out cancellation and degenerate cases. Assumption 4.(v) holds, for example, for the maximum score estimator. This condition can be relaxed to require only that $\widehat{\delta}$ be consistent for δ when $\xi > 0$, and that $\|\widehat{\delta} - \delta\| = o_p(1/\log n)$ when $\xi = 0$: see the proof of Lemma 2.

Let a_{n_0} and b_{n_0} be the extreme value normalizing sequence for the right tail of $W_i \mid Y_i = 0$.

Proposition 4. Suppose Assumption 4 holds. Then, for any fixed $k \geq 1$, as $n \to \infty$,

$$\left(\frac{\widehat{W}^{(0)}_{n_0:n_0} - b_{n_0}}{a_{n_0}}, \cdots, \frac{\widehat{W}^{(0)}_{n_0:n_0-k+1} - b_{n_0}}{a_{n_0}} \right)' \xrightarrow{d} (V_1, \cdots, V_k)',$$

whose joint PDF is as in (6).

B Proofs

B.1 Proof of Proposition 1

Proof. By Assumption 1.(ii), we have $1 - F_X(x) \sim C_1 x^{-1/\lambda}$ as $x \to \infty$ for some constant $C_1 > 0$, which implies $f_X(x) \sim (C_1/\lambda) x^{-1/\lambda - 1}$ as $x \to \infty$. Below we consider the cases $\xi > 0$ and $\xi = 0$ for the tail of error term ε separately.

Case 1: $\xi > 0$. By Assumption 1.(iii), since $F_\varepsilon \in \mathcal{D}(G_\xi)$ with $\xi > 0$, we have $1 - F_\varepsilon(x) \sim C_2 x^{-1/\xi}$ as $x \to \infty$ for some constant $C_2 > 0$.

Applying Bayes' rule as in (4), we obtain

$$f_{X|Y=0}(x) = \frac{(1 - F_\varepsilon(x)) f_X(x)}{\mathbb{P}(Y_i = 0)} \sim \frac{C_1 C_2 x^{-1/\lambda - 1/\xi - 1}}{\lambda \mathbb{P}(Y_i = 0)}.$$

It follows that $1 - F_{X|Y=0}(x) \sim x^{-1/\lambda - 1/\xi}$ for large x. By the von Mises theorem (e.g., de Haan and Ferreira, 2007, Theorem 1.1.8), this implies $F_{X|Y=0} \in \mathcal{D}(G_{\lambda\xi/(\lambda+\xi)})$.

Case 2: $\xi = 0$. By Assumption 1.(iii), since $\xi = 0$, we have $f_\varepsilon(x) \sim C_1 x^{C_2} e^{-x^{C_3}/C_4}$ for some positive constants C_1, C_2, C_3, C_4. Define

$$L = \lim_{x \to \infty} \frac{x^{C_3 - 1} (1 - F_\varepsilon(x))}{f_\varepsilon(x)}.$$

Applying l'Hôpital's rule, we obtain

$$L = \lim_{x \to \infty} \frac{\frac{d}{dx} \left[x^{C_3 - 1} (1 - F_\varepsilon(x)) \right]}{\frac{d}{dx} f_\varepsilon(x)} = \lim_{x \to \infty} \frac{(C_3 - 1) x^{C_3 - 2} (1 - F_\varepsilon(x)) - x^{C_3 - 1} f_\varepsilon(x)}{C_2 \, x^{-1} f_\varepsilon(x) - \frac{C_3}{C_4} \, x^{C_3 - 1} f_\varepsilon(x)} = \frac{C_4}{C_3},$$

which yields

$$1 - F_\varepsilon(x) \sim \frac{C_4}{C_3} \frac{f_\varepsilon(x)}{x^{C_3 - 1}}.$$

By Bayes' rule as in (4), we have

$$f_{X|Y=0}(x) = \frac{(1 - F_\varepsilon(x))f_X(x)}{\mathbb{P}(Y_i = 0)} \sim \frac{C_4 C_1}{C_3 \lambda \mathbb{P}(Y_i = 0)} x^{C_2 - 1/\lambda - C_3} \exp\{-x^{C_3}/C_4\}.$$

Since $f_{X|Y=0}(x)$ has an exponentially decaying tail, so does $1 - F_{X|Y=0}(x)$. By the von Mises theorem, this implies $F_{X|Y=0} \in \mathcal{D}(G_0)$.

Combining both cases, we conclude that $F_{X|Y=0} \in \mathcal{D}(G_\gamma)$ with $\gamma = \xi\lambda/(\lambda + \xi)$. □

B.2 Proof of Proposition 2

Proof. Assumption 2.(iii) and Bayes' rule imply that

$$\mathbb{P}(X_i + X_i^{a\prime}\beta \le x \mid Y_i = 0) = \frac{\int_{-\infty}^{x} \mathbb{P}(Y_i = 0 \mid X_i + X_i^{a\prime}\beta = t) f_{X + X^{a\prime}\beta}(t) dt}{\mathbb{P}(Y_i = 0)}.$$

Taking derivative w.r.t. x on both sides of the equation to obtain

$$f_{X + X^{a\prime}\beta|Y=0}(x) = \frac{1 - F_\varepsilon(x)}{\mathbb{P}\left(Y_i = 0\right)} f_{X + X^{a\prime}\beta}\left(x\right).$$

Given Assumption 2.(iv),

$$\frac{f_{X + X^{a\prime}\beta}(x)}{f_X(x)} \to 1 \text{ as } x \to \infty,$$

and the rest of the argument follows analogously as in the proof of Proposition 1. □

B.3 Proof of Proposition 3

Proof. For each t, denote $Z_{it} = X_{it}^{a\prime}\beta + \rho Y_{i,t-1} + \alpha_i$. Assumption 3.(iii) and Bayes' rule imply that

$$\mathbb{P}(X_{it} + Z_{it} \le x \mid Y_{it} = 0) = \frac{\int_{-\infty}^{x} \mathbb{P}(Y_{it} = 0 \mid X_{it} + Z_{it} = u) f_{X_t + Z_t}(u) du}{\mathbb{P}(Y_{it} = 0)}.$$

Taking derivative w.r.t. x on both sides of the equation to obtain

$$f_{X_t + Z_t|Y_t=0}(x) = \frac{1 - F_{\varepsilon_t}(x)}{\mathbb{P}\left(Y_{it} = 0\right)} f_{X_t + Z_t}\left(x\right). \tag{13}$$

Given Assumption 3.(iv),

$$\frac{f_{X_t+Z_t}(x)}{f_{X_t}(x)} \to 1 \text{ as } x \to \infty,$$

and the rest of the argument follows analogously as in the proof of Proposition 1. □

B.4 Proof of Proposition 4

Lemma 1. *Under Assumption 4.(i)–(iii), we have*

$$F_{W|Y=0} \in \mathcal{D}(G_\gamma) \text{ with } \gamma = \frac{\xi\lambda}{\xi+\lambda}.$$

Then, for any fixed $k \geq 1$, as $n \to \infty$,

$$\left(\frac{W^{(0)}_{n_0:n_0} - b_{n_0}}{a_{n_0}}, \cdots, \frac{W^{(0)}_{n_0:n_0-k+1} - b_{n_0}}{a_{n_0}} \right)' \xrightarrow{d} (V_1, \cdots, V_k)',$$

with joint density (6)*.*

Proof. Applying Proposition 1 with X replaced by W yields the results. □

Lemma 2. *Under Assumption 4, as $n \to \infty$,*

$$\max_{1 \leq i_0 \leq n_0} \frac{\left| \widehat{W}^{(0)}_{i_0} - W^{(0)}_{i_0} \right|}{a_{n_0}} \xrightarrow{p} 0.$$

Proof. Conditional on $I^{(0)} = \{i : Y_i = 0\}$, relabel the observations as $\{\widetilde{X}^{(0)}_{i_0} : i_0 = 1, \cdots, n_0\}$, which are i.i.d. with distribution $\widetilde{X} \mid Y = 0$.

Write $\widehat{W}^{(0)}_{i_0} = W^{(0)}_{i_0} + R^{(0)}_{i_0}$ with $R^{(0)}_{i_0} = \widetilde{X}^{(0)\prime}_{i_0}(\widehat{\delta} - \delta)$. A standard perturbation bound and Cauchy-Schwarz give

$$\max_{1 \leq i_0 \leq n_0} \frac{\left| \widehat{W}^{(0)}_{i_0} - W^{(0)}_{i_0} \right|}{a_{n_0}} \leq \|\widehat{\delta} - \delta\| \cdot \frac{\max_{1 \leq i_0 \leq n_0} \|\widetilde{X}^{(0)}_{i_0}\|}{a_{n_0}}.$$

Then, it suffices to find the rate of $\max_{1 \leq i_0 \leq n_0} \|\widetilde{X}^{(0)}_{i_0}\|$. Fix $M > 0$ and set $x_{n_0}(M) = b_{n_0} + a_{n_0} M$. By the union bound,

$$\mathbb{P}\left(\max_{1 \leq i_0 \leq n_0} \|\widetilde{X}^{(0)}_{i_0}\| > x_{n_0}(M) \mid I^{(0)} \right) \leq n_0 \mathbb{P}\left(\|\widetilde{X}^{(0)}_{i_0}\| > x_{n_0}(M) \mid Y_{i_0} = 0 \right).$$

By Assumption 4.(iv), for some $C < \infty$ and $\underline{x}$, $\mathbb{P}(\|\widetilde{X}_{i_0}^{(0)}\| > x \mid Y_{i_0} = 0) \leq C\mathbb{P}(W_{i_0}^{(0)} > x \mid Y_{i_0} = 0)$ for all $x \geq \underline{x}$. For large n_0 so that $x_{n_0}(M) \geq \underline{x}$, we have

$$\mathbb{P}\left(\|\widetilde{X}_{i_0}^{(0)}\| > x_{n_0}(M) \mid Y_{i_0} = 0\right) \leq C\mathbb{P}\left(W_{i_0}^{(0)} > x_{n_0}(M) \mid Y_{i_0} = 0\right).$$

By Lemma 1, $F_{W|Y=0} \in \mathcal{D}(G_\gamma)$ with normalizing sequence (a_{n_0}, b_{n_0}), so

$$n_0\left(1 - F_{W|Y=0}(b_{n_0} + a_{n_0}M)\right) \to -\log G_\gamma(M) \text{ a.s.},$$

and therefore the right-hand side above is bounded for each fixed M. Letting $M \to \infty$ implies $\max_{1\leq i_0\leq n_0} \|\widetilde{X}_{i_0}^{(0)}\| = O_p(b_{n_0} + a_{n_0})$.

Therefore,

$$\max_{1\leq i_0\leq n_0} \frac{\left|\widehat{W}_{i_0}^{(0)} - W_{i_0}^{(0)}\right|}{a_{n_0}} \leq \|\widehat{\delta} - \delta\| \cdot \frac{\max_{1\leq i_0\leq n_0} \|\widetilde{X}_{i_0}^{(0)}\|}{a_{n_0}} = \|\widehat{\delta} - \delta\| \cdot O_p\left(\frac{b_{n_0} + a_{n_0}}{a_{n_0}}\right).$$

Under Assumption 4.(v), $\|\widehat{\delta} - \delta\| = o_p(n^{-\eta})$. Moreover, $\frac{b_{n_0}+a_{n_0}}{a_{n_0}} = O(1)$, if $\gamma > 0$; and $\frac{b_{n_0}+a_{n_0}}{a_{n_0}} = O(\log n_0) = O(\log n)$, if $\gamma = 0$ and under the tail distribution in Assumption 4.(iii). Hence, their product is $o_p(1)$, which proves the claim. □

Proof of Proposition 4. By Lemma 1,

$$\left(\frac{W_{n_0:n_0}^{(0)} - b_{n_0}}{a_{n_0}}, \cdots, \frac{W_{n_0:n_0-k+1}^{(0)} - b_{n_0}}{a_{n_0}}\right)' \xrightarrow{d} (V_1, \cdots, V_k)'.$$

Let $\Delta_{n_0} = \max_{1\leq i_0\leq n_0} |\widehat{W}_{i_0}^{(0)} - W_{i_0}^{(0)}|$ over the unsorted subsample. A deterministic property of order statistics implies that over the sorted subsample, for each $1 \leq j \leq n_0$,

$$\left|\widehat{W}_{n_0:n_0-j+1}^{(0)} - W_{n_0:n_0-j+1}^{(0)}\right| \leq \Delta_{n_0},$$

and hence

$$\max_{1\leq j\leq k} \left|\frac{\widehat{W}_{n_0:n_0-j+1}^{(0)} - b_{n_0}}{a_{n_0}} - \frac{W_{n_0:n_0-j+1}^{(0)} - b_{n_0}}{a_{n_0}}\right| \leq \frac{\Delta_{n_0}}{a_{n_0}}.$$

By Lemma 2, $\Delta_{n_0}/a_{n_0} \xrightarrow{p} 0$, so

$$\max_{1\le j\le k}\left|\frac{\widehat{W}^{(0)}_{n_0:n_0-j+1}-b_{n_0}}{a_{n_0}}-\frac{W^{(0)}_{n_0:n_0-j+1}-b_{n_0}}{a_{n_0}}\right| \xrightarrow{p} 0.$$

An application of Slutsky's theorem then implies that the vector of normalized top-k order statistics of $\widehat{W}$ converges in distribution to $(V_1,\cdots,V_k)'$ as well. □

References

Aghion, P., Akcigit, U., and Howitt, P. (2014). What do we learn from schumpeterian growth theory? In *Handbook of economic growth*, volume 2, pages 515–563. Elsevier.

Andrews, D. W. and Ploberger, W. (1994). Optimal tests when a nuisance parameter is present only under the alternative. *Econometrica: Journal of the Econometric Society*, pages 1383–1414.

Aw, B. Y., Roberts, M. J., and Xu, D. Y. (2011). R&d investment, exporting, and productivity dynamics. *American Economic Review*, 101(4):1312–1344.

Axtell, R. L. (2001). Zipf distribution of U.S. firm sizes. *Science*, 293(5536):1818–1820.

Bernard, A. B., Jensen, J. B., Redding, S. J., and Schott, P. K. (2007). Firms in international trade. *Journal of Economic Perspectives*, 21(3):105–130.

Bierens, H. J. (1990). A consistent conditional moment test of functional form. *Econometrica*, 58(6):1443–1458.

Chamberlain, G. (1980). Analysis of covariance with qualitative data. *Review of Economic Studies*, 47(1):225–238.

Chaney, T. (2008). Distorted gravity: The intensive and extensive margins of international trade. *American Economic Review*, 98(4):1707–1721.

Chen, Z., Jiang, X., Liu, Z., Serrato, J. C. S., and Xu, D. Y. (2023). Tax policy and lumpy investment behaviour: Evidence from china's VAT reform. *Review of Economic Studies*, 90(2):634–674.

Cohen, W. M. (2010). Fifty years of empirical studies of innovative activity and performance. *Handbook of the Economics of Innovation*, 1:129–213.

de Haan, L. and Ferreira, A. (2007). *Extreme Value Theory: An Introduction.* Springer Science and Business Media, New York.

di Giovanni, J. and Levchenko, A. A. (2012). Country size, international trade, and aggregate fluctuations. *Journal of Political Economy*, 120(6):1083–1132.

Einmahl, J. H. and He, Y. (2023). Extreme value inference for heterogeneous power law data. *The Annals of Statistics*, 51(3):1331–1356.

Fernández-Val, I. (2009). Fixed effects estimation of structural parameters and marginal effects in panel probit models. *Journal of Econometrics*, 150(1):71–85.

Gabaix, X. (2009). Power laws in economics and finance. *Annu. Rev. Econ.*, 1(1):255–294.

Han, A. K. (1987). Non-parametric analysis of a generalized regression model: The maximum rank correlation estimator. *Journal of Econometrics*, 35(2–3):303–316.

Heckman, J. J. (1981). The incidental parameters problem and the problem of initial conditions in estimating a discrete time–discrete data stochastic process. In Manski, C. and McFadden, D., editors, *Structural Analysis of Discrete Data with Econometric Applications*, pages 179–195. MIT Press, Cambridge, MA.

Honoré, B. E. and Kyriazidou, E. (2000). Panel data discrete choice models with lagged dependent variables. *Econometrica*, 68(4):839–874.

Horowitz, J. L. (1992). A smoothed maximum score estimator for the binary response model. *Econometrica*, 60(3):505–531.

Horowitz, J. L. and Härdle, W. (1994). Testing a parametric model against a semiparametric alternative. *Econometric Theory*, 10(5):821–848.

Ichimura, H. (1993). Semiparametric least squares (sls) and weighted sls estimation of single-index models. *Journal of Econometrics*, 58(1–2):71–120.

Khan, S. and Tamer, E. (2010). Irregular identification, support conditions, and inverse weight estimation. *Econometrica*, 78(6):2021–2042.

Klein, R. W. and Spady, R. H. (1993). An efficient semiparametric estimator for binary response models. *Econometrica*, 61(2):387–421.

Lehmann, E. L. and Romano, J. P. (2005). *Testing statistical hypotheses.* Springer.

Lileeva, A. and Trefler, D. (2010). Improved access to foreign markets raises plant-level productivity...for some plants. *Quarterly Journal of Economics*, 125(3):1051–1099.

Liu, L. and Wang, Y. (2025). Binary outcome models with extreme covariates: Estimation and prediction. *arXiv preprint arXiv:2502.16041.*

Liu, Q. and Ma, H. (2020). Trade policy uncertainty and innovation: Firm level evidence from China's WTO accession. *Journal of International Economics*, 127:103387.

Liu, Q. and Qiu, L. D. (2016). Intermediate input imports and innovations: Evidence from Chinese firms' patent filings. *Journal of International Economics*, 103:166–183.

Liu, Y. and Mao, J. (2019). How Do Tax Incentives Affect Investment and Productivity? Firm-Level Evidence from China. *American Economic Journal: Economic Policy*, 11(3):261–291.

Manski, C. F. (1975). Maximum score estimation of the stochastic utility model of choice. *Journal of Econometrics*, 3(3):205–228.

Manski, C. F. (1987). Semiparametric analysis of random effects linear models from binary panel data. *Econometrica*, 55(2):357–362.

McFadden, D. (1974). Conditional logit analysis of qualitative choice behavior. In Zarembka, P., editor, *Frontiers in Econometrics*, pages 105–142. Academic Press, New York.

Melitz, M. J. (2003). The impact of trade on intra-industry reallocations and aggregate industry productivity. *Econometrica*, 71(6):1695–1725. Publisher: Wiley Online Library.

Müller, U. K. (2011). Efficient tests under a weak convergence assumption. *Econometrica*, 79(2):395–435.

Müller, U. K. and Wang, Y. (2017). Fixed-k asymptotic inference about tail properties. *Journal of the American Statistical Association*, 112(519):1334–1343.

Neyman, J. and Scott, E. L. (1948). Consistent estimates based on partially consistent observations. *Econometrica*, 16(1):1–32.

Ota, Y. and Otsu, T. (2026). Specification testing for binary choice model via maximum score. Technical Report 194, Keio Economic Observatory, Keio University.

Pregibon, D. (1980). Goodness of link tests for generalized linear models. *Journal of the Royal Statistical Society. Series C (Applied Statistics)*, 29(1):15–24.

Wooldridge, J. M. (2002). *Econometric Analysis of Cross Section and Panel Data*. MIT Press, Cambridge, MA.

Xia, Y., Li, W. K., Tong, H., and Zhang, D. (2004). A goodness-of-fit test for single-index models. *Statistica Sinica*, 14(1):1–39.