CENTER FOR POLICY RESEARCH

THE MAXWELL SCHOOL

WORKING PAPER SERIES

Nonparametric Tests of Tail Behavior in Stochastic Frontier Models

William C. Horrace and Yulong Wang

Paper No. 230
June 2020

ISSN: 1252-3066

SYRACUSE UNIVERSITY
Maxwell School
Center for Policy Research
426 Eggers Hall
Syracuse University
Syracuse, NY 13244-1020
(315) 443-3114/email: ctrpol@syr.edu
http://www.maxwell.syr.edu/CPR_Working_Papers.aspx

CENTER FOR POLICY RESEARCH – Summer 2020

Leonard M. Lopoo, Director
Professor of Public Administration and International Affairs (PAIA)

Associate Directors

Margaret Austin
Associate Director, Budget and Administration

John Yinger
Trustee Professor of Economics (ECON) and Public Administration and International Affairs (PAIA)
Associate Director, Center for Policy Research

SENIOR RESEARCH ASSOCIATES

Badi Baltagi, ECON
Robert Bifulco, PAIA
Leonard Burman, PAIA
Carmen Carrión-Flores, ECON
Alfonso Flores-Lagunes, ECON
Sarah Hamersma, PAIA
Madonna Harrington Meyer, SOC
Colleen Heflin, PAIA
William Horrace, ECON
Yilin Hou, PAIA
Hugo Jales, ECON
Jeffrey Kubik, ECON
Yoonseok Lee, ECON
Amy Lutz, SOC
Yingyi Ma, SOC
Katherine Michelmore, PAIA
Jerry Miner, ECON
Shannon Monnat, SOC
Jan Ondrich, ECON
David Popp, PAIA
Stuart Rosenthal, ECON
Michah Rothbart, PAIA
Alexander Rothenberg, ECON
Rebecca Schewe, SOC
Amy Ellen Schwartz, PAIA/ECON
Ying Shi, PAIA
Saba Siddiki, PAIA
Perry Singleton, ECON
Yulong Wang, ECON
Peter Wilcoxen, PAIA
Maria Zhu, ECON

GRADUATE ASSOCIATES

Rhea Acuña, PAIA
Mariah Brennan, SOC. SCI.
Ziqiao Chen, PAIA
Yoon Jung Choi, PAIA
Dahae Choo, ECON
Stephanie Coffey, PAIA
Giuseppe Germinario, ECON
Myriam Gregoire-Zawilski, PAIA
Jeehee Han, PAIA
Mary Helander, SOC. SCI.
Hyoung Kwon, PAIA
Mattie Mackenzie-Liu, PAIA
Maeve Maloney, ECON
Austin McNeill Brown, SOC. SCI.
Qasim Mehdi, PAIA
Claire Pendergrast, SOC
Krushna Ranaware, SOC
Christopher Rick, PAIA
Huong Tran, ECON
Joaquin Urrego, ECON
Yao Wang, ECON
Yi Yang, ECON
Xiaoyan Zhang, Human Dev.
Bo Zheng, PAIA
Dongmei Zuo, SOC. SCI.

STAFF

Joseph Boskovski, Manager, Maxwell X Lab
Katrina Fiacchi, Administrative Specialist
Michelle Kincaid, Senior Associate, Maxwell X Lab
Emily Minnoe, Administrative Assistant
Candi Patterson, Computer Consultant
Samantha Trajkovski, Postdoctoral Scholar
Laura Walsh, Administrative Assistant

Abstract

This article studies tail behavior for the error components in the stochastic frontier model, where one component has bounded support on one side, and the other has unbounded support on both sides. Under weak assumptions on the error components, we derive nonparametric tests that the unbounded component distribution has thin tails and that the component tails are equivalent. The tests are useful diagnostic tools for stochastic frontier analysis and kernel deconvolution density estimation. A simulation study and an application to a stochastic cost frontier for 6,100 US banks from 1998 to 2005 are provided. The new tests reject the normal or Laplace distributional assumptions, which are commonly imposed in the existing literature.

JEL No.: C12, C21, D24

Keywords: Hypothesis Testing, Production, Inefficiency, Deconvolution, Extreme Value Theory

Authors: William C. Horrace, Department of Economics and Center for Policy Research, Syracuse University, Syracuse, NY 13244; Yulong Wang, Department of Economics and Center for Policy Research, Syracuse University, Syracuse, NY 13244, ywang402@maxwell.syr.edu

Acknowledgement

We thank Chris Parmeter for very helpful comments.

1 Introduction

Stochastic frontier analysis (SFA) has a vast literature, both methodological and empirical, and practitioners have applied the methods to myriad industries, most notably agriculture, banking, education, healthcare, and energy. A common practice in SFA is to impose parametric assumptions on the error components, but the set of statistical tools to investigate the validity of these assumptions is still limited. This paper expands this set of tools by drawing on recently developed techniques in extreme value (EV) theory and by developing new diagnostic tests. Comparing with the existing parametric methods that test for specific distributions,[1] our tests are designed nonparmatrically to test for a broad class of distributions that share a certain tail behavior.

In particular, the parametric stochastic frontier model for cross-sectional data (Meeusen and van den Broeck, 1977 and Aigner et al., 1997) is a leading case of the error component regression model but with the unique feature that one error component (U) is a non-negative random variable, while the other (V) is a random variable of unbounded support. To learn some features of U and V, a common assumption in the stochastic frontier literature is that V is drawn from a normal or Laplace distribution, both of which have thin tails (e.g., Aigner et al., 1977 and Horrace and Parmeter, 2018).[2] However, heavy-tailed distributions are now also being considered. For example, the findings of Wheat et al. (2019) suggest that a cost inefficiency model of highway maintenance costs in England has Student-t errors.[3] These parametric distributions, such as normal and Student-t, display similar patterns in the

[1] See, for example, Kopp and Mullahy (1990), Coelli (1995), Li (1996), Wang et al. (2011), and Papadopoulos and Parmeter (2020).

[2] For other parametric specifications of the model, see Li (1996), Carree (2002), Tsionas (2007), Kumbhakar et al. (2013), and Almanidis et al. (2014).

[3] There are semi-parametric estimators of the model that relax the distributional assumptions on one component and estimate the density of the other using kernel deconvolution techniques. See Hall and Simar (2002), Horrace and Parmeter (2011), Kneip et al. (2015), Simar et al. (2017), Cai et al. (2020), and Florens et al. (2020).

middle of their supports but exhibit substantially different tail behaviors. This observation motivates and plays an essential role in our diagnostic tests, which we believe are a timely and appropriate contribution to the literature.

The key idea of our test is as follows. Assuming independence of the error components, the largest order statistics of the composed error term ($Z = V - U$) approximately arise from the right tail of V, because U is non-negative. Also, assuming that V is in the domain of attraction (DOA) of extreme value distributions, the asymptotic distribution of the largest order statistics of V is the EV distribution, which may be fully characterized (after location and scale normalization) by a single parameter that captures its tail heaviness.[4] Then, likelihood ratio statistics for hypotheses on this single parameter can be derived based on the largest order statistics of Z and their limiting EV distribution.

To be specific, consider the right tail of V. If the DOA assumption is satisfied, then tail behavior may be entirely characterized by a *tail index*, $\xi \in \mathbb{R}$. If $\xi = 0$, then V has a thin tail. If $\xi > 0$, then V has a heavy tail. Otherwise, V has bounded support. Under very weak assumptions on the error components, we first derive a test that the right tail of V is thin ($H_0 : \xi = 0$), based on Z. We prove that this test is valid whether Z is observed or appended to a regression model. The former case applies to the kernel devoncolution estimation (e.g., Stefanski and Caroll, 1990 and Meister, 2006), and the latter case corresponds to the stochastic frontier model.

Second, if we assume that U is also in the DOA of extreme value distributions and that V is symmetric (a common assumption), we also derive a test that the (right) tail of U is thinner than the left tail of V. Finally, if we further assume that V is a member of the normal or Laplace family, then we may test the hypotheses that the tails of U and V are both thin. Given their potentially wide applicability, our nonparametric tests are therefore useful

[4]The assumption that V is in the DOA of extreme value distributions is not restrictive, as we shall see in Section 2.1

diagnostic tools to help empiricists make parametric choices on the distributions of both U and V. This is particularly important for the stochastic frontier model for cross-sectional data, where distributional assumptions on the components are typically necessary for the identification of the model's parameters.

The paper is organized as follows. The next section presents the tests. Section 3 provides a simulation study of their power and size properties. Section 4 applies the tests to a stochastic cost function for the US banks data, revealing that the tails of V are heavy. Therefore, a normal or Laplace assumption for V is not justified, and perhaps a Student-t assumption may be appropriate. Section 5 concludes.

2 Tests of tail behavior

In Section 2.1, we begin a review of the DOA assumption and present the test for the case where Z is directly observed. While the test is not applicable to SFA *per se*, it may be of interest to empiricists who use kernel deconvolution density estimators.[5] In Section 2.2, we move to the case where Z is appended to a regression model and has to be estimated, which covers the linear regression stochastic frontier model. Additional tests under different sets of weak assumptions are also presented.[6]

2.1 The case with no covariates

Consider a random sample of $Z_i = V_i - U_i$ for $i = 1, \ldots, n$, where $U_i \geq 0$ represents *inefficiency*, and $V_i \in \mathbb{R}$ is *noise* with unbounded support. We start with testing the *shape*

[5] These are kernel estimators of the density of U, which is known to be one-sided *ex ante*, based on parametric assumptions on V with Z observed. See, for example, Stefanski and Carroll (1990) and Meister (2006).

[6] While the analyses that follow are for cross-sectional data, they can easily be applied to panel data, as long as one is willing to assume independence in both the time and cross-sectional dimensions.

of the right tail of V_i in a nonparametric way.

The key assumption is that the distribution of V_i satisfies the DOA assumption. In particular, a cumulative distribution function (CDF) F is in the domain of attraction of EV distributions, denoted as $F \in \mathcal{D}(G_\xi)$, if there exist constants $a_n > 0$ and b_n such that for any $w \in \mathbb{R}$,

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

where G_ξ is the generalized EV distribution,

$$G_\xi(w) = \begin{cases} \exp[-(1+\xi w)^{-1/\xi}],\ 1+\xi w \geq 0, \text{ for } \xi \neq 0 \\ \exp[-e^{-w}],\ w \in \mathbb{R},\ \xi = 0 \end{cases} \tag{1}$$

and ξ is the tail index, measuring the decay rate of the tail.

The DOA condition is satisfied by a large range of commonly used distributions. If ξ is positive, this condition is equivalent to regularly varying at infinity, i.e.,

$$\lim_{t\to\infty} \frac{1-F(tw)}{1-F(t)} = w^{-1/\xi} \qquad \text{for } w > 0. \tag{2}$$

This covers Pareto, Student-t,[7] and F distributions, for example. The case with $\xi = 0$ covers the normal and the Laplace families, and the case with $\xi < 0$ corresponds to distributions with bounded support.[8] See Ch.1 in de Haan and Ferreira (2007) for a complete review.

Note that the above notation is for the right tail of V, which can be easily adapted to the left tail by considering $-V$. For expositional simplicity, we denote ξ_{V_-} and ξ_{V_+} as the tail indices for the left and right tails of V, respectively. The same notation applies to other variables (e.g., U and Z) introduced later.

[7] The tail index of the Student-t distribution with ν degrees of freedom is $\xi = 1/\nu$.

[8] The uniform distribution has $\xi = -1$, and the triangular distribution has $\xi = -1/2$.

Returning to SFA, a common assumption is that V_i is normal or Laplace, which implies that $\xi_{V_+} = 0$. So our hypothesis testing problem is as follows:

$$H_0 : \xi_{V_+} = 0 \text{ against } H_1 : \xi_{V_+} > 0. \tag{3}$$

If the null hypothesis is rejected, we would then argue that some heavy-tailed distribution should be used to model the noise and maybe the inefficiency as well.

To obtain a feasible test, we argue that, since U_i is bounded from below at zero, the largest order statistics of Z_i are approximately stemming from the right tail of V_i. This is formalized in Proposition 1, which requires the following conditions. Let $Z_{n:n} \geq, \ldots, \geq Z_{n:1}$ be the order statistics of $\{Z_i\}_{i=1}^n$ by descend sorting. Denote

$$\mathbf{Z}_+ = (Z_{n:n}, ..., Z_{n:n-k+1})^\intercal$$

as the k largest observations. From now on, we use bold letters to denote vectors. Denote F_V and $Q_V(p) = \inf\{y \in \mathbb{R} : p \leq F_V(y)\}$ as the CDF and the quantile function of V_i, respectively. Write $Q_V(1)$ as the right end-point of the support of V_i. For a generic column vector $\mathbf{X}$ and scalar c, the notation $\mathbf{X} - c$ means $\mathbf{X} - (c, \ldots, c)^\intercal$.

Assumption 1

(i) $(U_i, V_i)^\intercal$ is *i.i.d.*

(ii) U_i and V_i are independent.

(iii) $U_i \geq 0$ with $\mathbb{E}[|U_i|] < \infty$ and $V_i \in \mathbb{R}$ with $Q_V(1) = \infty$.

(iv) $F_V \in \mathcal{D}(G_{\xi_{V_+}})$ with $\xi_{V_+} \geq 0$. In addition, $F_V(\cdot)$ is twice continuously differentiable with bounded derivatives, and the density $f_V(\cdot)$ satisfies that $\partial f_V(t)/\partial t \nearrow 0$ as $t \to \infty$

on $[c, \infty)$ for some constant c.

Assumptions 1(i)-(iii) are common in the SFA literature (see Horrace and Parmeter, 2018 and the references therein). Assumption 1(iv) requires the tail of F_V to be within the domain of attraction of EV distributions with an infinite upper bound. Moreover, it requires that the density derivative monotonically increases to zero. This is again a mild assumption and is satisfied by many commonly used distributions. For example, the normal distribution is covered as seen by

$$\frac{\partial f_V(t)}{\partial t} \propto -t \exp(-\frac{t^2}{2}) \nearrow 0 \text{ as } t \to \infty,$$

and the Pareto distribution is covered as seen by

$$\frac{\partial f_V(t)}{\partial t} \propto (-\alpha - 1)t^{-\alpha-2} \nearrow 0 \text{ as } t \to \infty \text{ for some } \alpha > 0.$$

Under Assumption 1, the following proposition derives the asymptotic distribution of $\mathbf{Z}_+$.

Proposition 1 *Suppose Assumption 1 holds. Then, there exist sequences of constants* a_n *and* b_n *such that for any fixed* k

$$\frac{\mathbf{Z}_+ - b_n}{a_n} \xrightarrow{d} \mathbf{W}_+ = (W_1, ..., W_k)^\intercal \quad \text{as } n \to \infty,$$

where the joint density of $\mathbf{W}_+$ *is given by* $f_{\mathbf{W}_+|\xi_{V_+}}(w_1, ..., w_k) = G_{\xi_{V_+}}(w_k) \prod_{i=1}^k g_{\xi_{V_+}}(a_i)/G_{\xi_{V_+}}(w_i)$ *on* $w_k \leq w_{k-1} \leq \ldots \leq w_1$, *and* $g_{\xi_{V_+}}(w) = dG_{\xi_{V_+}}(w)/dw$.

The proof is in Appendix A. From now on, we reserve the letter $\mathbf{W}$ and its variants as the limiting observations.

Proposition 1 implies that the distributions of Z_i and V_i share the same (right) tail shape, which is entirely characterized by the tail index ξ_{V_+}. Such tail equivalence does not hold, however, for the left tails due to the existence of U. This is further studied in Section 2.3 under the additional assumption that V is symmetric.

If the constants a_n and b_n were known, $\mathbf{Z}_+$ is then approximately distributed as $\mathbf{W}_+$, and the limiting problem is reduced to the well-defined finite sample problem: constructing some inference method based on one draw $\mathbf{W}_+$ whose density $f_{\mathbf{W}_+|\xi_{V_+}}$ is known up to ξ_{V_+}. However, a_n and b_n depend on F_V and hence are unknown *a priori*.

To avoid the need for knowledge of a_n and b_n, we consider the following self-normalized statistic

$$\begin{aligned} \mathbf{Z}_+^* &= \frac{\mathbf{Z}_+ - Z_{n:n-k+1}}{Z_{n:n} - Z_{n:n-k+1}} \\ &= \left(1, \frac{Z_{n:n-1} - Z_{n:n-k+1}}{Z_{n:n} - Z_{n:n-k+1}}, ..., \frac{Z_{n:n-k+2} - Z_{n:n-k+1}}{Z_{n:n} - Z_{n:n-k+1}}, 0\right)^\intercal. \end{aligned} \tag{4}$$

It is easy to establish that $\mathbf{Z}_+^*$ is maximally invariant with respect to the group of location and scale transformations (e.g., Ch.6 in Lehmann and Romano, 2005). In words, the estimator constructed as a function of $\mathbf{Z}_+^*$ remains unchanged if data are shifted and multiplied by any non-zero constant. This makes senses since the tail shape should be preserved no matter how data are linearly transformed. This invariance property allows us to construct nonparametric tests for a stochastic frontier model that is otherwise not identified without parametric assumptions on U and V.[9] As such, our tests do not reveal anything about the location or the scale of the error components.

The continuous mapping theorem and Proposition 1 imply that for any fixed k, as $n \to \infty$,

$$\mathbf{Z}_+^* \xrightarrow{d} \mathbf{W}_+^* \equiv \frac{\mathbf{W}_+ - W_k}{W_1 - W_k}.$$

[9]In particular the non-zero expectation of U precludes identification of unknown parameter δ in the model $Z_i = \delta + V_i - U_i$.

The CDF of $\mathbf{W}_{+}^{*}$ can be calculated via change of variables as

$$f_{\mathbf{W}_{+}^{*}|\xi_{V_{+}}}\left(\mathbf{w}_{+}^{*}\right)=\Gamma\left(k\right)\int_{0}^{b_{0}\left(\xi_{V_{+}}\right)}t^{k-2}\exp\left(-(1+1/\xi_{V_{+}})\sum_{i=1}^{k}\log(1+\xi_{V_{+}}w_{i}^{*}t)\right)dt, \qquad (5)$$

where $\mathbf{w}_{+}^{*}=(w_{1}^{*},\ldots,w_{k}^{*})$, $b_{0}\left(\xi\right)=\infty$ if $\xi\geq 0$ and $-1/\xi$ otherwise, and $\Gamma\left(k\right)$ is the gamma function. Note that the invariance restriction costs two degrees of freedom since the first and last elements of $\mathbf{W}_{+}^{*}$ are always 1 and 0, respectively. We calculate this density by numerical quadrature.

Given $f_{\mathbf{W}_{+}^{*}|\xi_{V_{+}}}$, we can construct a likelihood-ratio type statistic for testing (3) if the alternative hypothesis is simple. To this end, we follow Andrews and Ploberger (1994) and Elliott et al. (2015) to consider the weighted average alternative

$$\int_{\Xi}f_{\mathbf{W}_{+}^{*}|\xi}(\cdot)h(\xi)d\xi,$$

where Ξ denotes the parameter space that includes all empirically relevant values of the tail index, and $h(\cdot)$ is a weighting function that reflects the importance of rejecting different alternative values.[10] Then our test is constructed as

$$\varphi(\mathbf{w}_{+}^{*})=\mathbf{1}\left[\frac{\int_{\Xi}f_{\mathbf{W}_{+}^{*}|\xi}(\mathbf{w}_{+}^{*})h(\xi)d\xi}{f_{\mathbf{W}_{+}^{*}|0}(\mathbf{w}_{+}^{*})}>\mathrm{cv}(k,\alpha)\right], \qquad (6)$$

where the critical value $\mathrm{cv}(k,\alpha)$ depends on k and the level of significance α. We can obtain it by simulation. By Proposition 1 and the continuous mapping theorem, this test controls size asymptotically as $\lim_{n\to\infty}\varphi(\mathbf{Z}_{+}^{*})=\alpha$.

We end this subsection by briefly discussing the choice of k, that is, the number of

[10]In later sections, we set $\Xi=(0,1)$ and $h(\cdot)$ to be the standard uniform distribution over $(0,1)$ for simplicity.

the largest order statistics used to approximate the EV distribution. On the one hand, larger k means including more mid-sample observations, which induces a larger finite sample bias in the EV approximation. On the other hand, smaller k provides a better asymptotic approximation but uses less sample information, leading to a lower power test. This trade-off leads to difficulty in theoretical justification of an optimal k in standard EV theory literature (cf., Müller and Wang, 2017). It is even more difficult, if at all possible, in our case, since we only observe Z, and not V. Nonetheless, our asymptotic arguments show that the test (6) controls size for any fixed k, as long as n is sufficiently large. Figure 1 depicts the *asymptotic* power of the test (6) with $\mathbf{W}_{+}^{*}$ generated from the density (5) based on 10,000 simulation draws. (More simulations about the finite sample performance are presented in Section 3.) The test controls size for all values of k by construction and has reasonably large power when k exceeds 20.

We now turn to the regression version of the test, with application to SFA.

2.2 The case with covariates: SFA

Now consider the linear regression with

$$Y_i = \mathbf{X}_i^{\intercal}\beta_0 + Z_i,$$

where $Z_i = -U_i + V_i$ is as in the previous section, and β_0 is some pseudo-true parameter in some compact parameter space. This could be a Cobb-Douglas production function (in logarithms), where Y is productive output and U is now called *technical efficiency*, which measures distance (U_i) from a stochastic frontier ($\mathbf{X}_i^{\intercal}\beta_0 + V_i$). The slopes ($\beta_0$) are marginal products of the productive inputs, $\mathbf{X}_i$. It could also be a stochastic cost function if we multiply U by -1. Suppose we have some estimator, $\hat{\beta}$ of β_0. The following assumption is imposed to construct our diagnostic test.

Assumption 2

(i) $(X_i, U_i, V_i)^\intercal$ is i.i.d.

(ii) U_i and V_i are independent.

(iii) $U_i \geq 0$ with $\mathbb{E}[|U_i|] < \infty$ and $V_i \in \mathbb{R}$ with $Q_V(1) = \infty$.

(iv) $F_V \in \mathcal{D}(G_{\xi_{V_+}})$ with $\xi_{V_+} \geq 0$. In addition, $F_V(\cdot)$ is twice continuously differentiable with bounded derivatives, and the density $f_V(\cdot)$ satisfies that $\partial f_V(t)/\partial t \nearrow 0$ as $t \to \infty$ on $[c, \infty)$ for some constant c.

(v) $\left|\left|\hat{\beta} - \beta_0\right|\right| \sup_i ||\mathbf{X}_i|| = o_p(n^{\xi_{V_+}})$, if $\xi_{V_+} > 0$. $\left|\left|\hat{\beta} - \beta_0\right|\right| \sup_i ||\mathbf{X}_i|| / f_V\left(Q_V(1 - 1/n)\right) = o_p(1)$, otherwise.

Assumption 2 is similar to Assumption 1 with additional restrictions on the covariate $\mathbf{X}$. In particular, Assumption 2(v) bounds the norm of $\hat{\beta}$ and $||\mathbf{X}_i||$. A sufficient condition when ξ_{V_+} is positive is that $\left|\left|\hat{\beta} - \beta_0\right|\right| = O_p(n^{-1/2})$ and $\sup_i ||\mathbf{X}_i|| = o_p(n^{1/2})$, which is easily satisfied in many applications.[11] When ξ_{V_+} is zero, we need slightly stronger bounds. Straightforward calculations show that the normal distribution satisfies Assumption 2(v) for the $\xi_{V_+} = 0$ case, if $\left|\left|\hat{\beta} - \beta_0\right|\right| = O_p\left(n^{-1/2}\right)$ and $\sup_i ||\mathbf{X}_i|| = O_p(n^{1/2-\varepsilon})$ for some $\varepsilon > 0$. This is seen by $1/f_V\left(Q_V\left(1 - 1/n\right)\right) \leq O(\log(n))$ (e.g., Example 1.1.7 in de Haan and Ferreira, 2007).

Denote $\hat{Z}_i$ as the OLS residuals and

$$\hat{\mathbf{Z}}_+ = \left(\hat{Z}_{n:n}, ..., \hat{Z}_{n:n-k+1}\right)^\intercal$$

[11]Even though $\mathbb{E}[|U_i|] \neq 0$, ordinary least squares (OLS) will typically suffice for $\hat{\beta}$, because our test is invariant to relocation.

the largest k order statistics. Then given Assumption 2, the following proposition derives the asymptotic distribution of $\hat{\mathbf{Z}}_+$

Proposition 2 *Suppose Assumption 2 holds. Then, there exist sequences of constants a_n and b_n such that for any fixed k*

$$\frac{\hat{\mathbf{Z}}_+ - b_n}{a_n} \xrightarrow{d} \mathbf{W}_+ \text{ as } n \to \infty,$$

where the joint density of $\mathbf{W}_+$ is the same as in Proposition 1.

The proof is in Appendix A. Proposition 2 implies that the largest order statistics of the regression residuals satisfy the same convergence as the no-covariate case. In other words, the estimation error from the OLS becomes negligible so that the largest order statistics are stemming from the right tail of V asymptotically. This validates the construction of the test (6) by replacing $\mathbf{Z}_+^*$ with $\hat{\mathbf{Z}}_+^*$, where

$$\hat{\mathbf{Z}}_+^* = \frac{\hat{\mathbf{Z}}_+ - \hat{Z}_{n:n-k+1}}{\hat{Z}_{n:n} - \hat{Z}_{n:n-k+1}}.$$

Then Proposition 2 and the continuous mapping theorem entail the asymptotic size control that $\lim_{n\to\infty} \varphi(\hat{\mathbf{Z}}_+^*) = \alpha$.

2.3 Symmetry of noise V

The previous analysis studies the right tail of V (and equivalently Z). Suppose we assume V has a symmetric distribution, then the tail indices of both tails of V become equivalent, and hence we can learn about the tail of U using the left tail index of Z. To this end, we make the following additional assumption.

Assumption 3

(i) V_i is symmetric at zero.

(ii) $F_U \in \mathcal{D}(G_{\xi_{U_+}})$ with $\xi_{U_+} \geq 0$.

Assumption 3(i) implies that $\xi_{V_-} = \xi_{V_+}$, and the condition that $U \geq 0$ implies its left tail index is negative. Therefore, in this subsection only, we simply denote ξ_U and ξ_V as the right tail indices of U and V, respectively. Now we can test if U has a thinner or equal right tail than V by specifying the following hypothesis testing problem,

$$H_0 : \xi_U \leq \xi_V \text{ against } H_1 : \xi_U > \xi_V. \tag{7}$$

Moreover, if V is in the normal or Laplace family ($\xi_V = 0$), since we limit the tail indices to be non-negative, the null hypothesis then reduces to $\xi_U = \xi_V = 0$.

Under the null hypothesis of (7), V is the leading term in Z in both the left and right tails. Then the DOA assumption for both V and U implies that $\xi_{Z_-} = \max\{\xi_U, \xi_V\}$, and Proposition 2 entails $\xi_{Z_+} = \xi_V$. Therefore, the above testing problem becomes equivalent to

$$H_0 : \xi_{Z_-} = \xi_{Z_+} \text{ against } H_1 : \xi_{Z_-} > \xi_{Z_+}. \tag{8}$$

We now construct a test for (8). Define $\hat{\mathbf{Z}}_-$ as the smallest k order statistics of the estimation residuals, that is,

$$\hat{\mathbf{Z}}_- = \left(\hat{Z}_{n:1}, \hat{Z}_{n,2}, \ldots, \hat{Z}_{n,k}\right)^\intercal$$

and its self-normalized analogue as

$$\hat{\mathbf{Z}}_-^* = \frac{\hat{\mathbf{Z}}_- - \hat{Z}_{n:k}}{\hat{Z}_{n:1} - \hat{Z}_{n:k}}.$$

The following proposition establishes that $\hat{\mathbf{Z}}_-^*$ asymptotically has the EV distribution with tail index ξ_{Z_-} and is independent from $\hat{\mathbf{Z}}_+^*$.

Proposition 3 *Suppose Assumptions 2 and 3 hold. Then, for any fixed k,*

$$\begin{pmatrix} \hat{\mathbf{Z}}_-^* \\ \hat{\mathbf{Z}}_+^* \end{pmatrix} \xrightarrow{d} \begin{pmatrix} \mathbf{W}_-^* \\ \mathbf{W}_+^* \end{pmatrix} \text{ as } n \to \infty,$$

where $\mathbf{W}_-^*$ and $\mathbf{W}_+^*$ are independent and both EV distributed with density (5) and tail indices ξ_{Z_-} and ξ_{Z_+}, respectively.

The proof is in Appendix A. Given the above proposition, we aim to construct a generalized likelihood ratio test for (8) as follows,

$$\varphi_{\pm}\left(\mathbf{w}_-^*, \mathbf{w}_+^*\right) = \mathbf{1}\left[\frac{\int_{\{(\xi_-,\xi_+)\in\Xi^2:\xi_+<\xi_-\}} f_{\mathbf{W}_-^*|\xi_-}(\mathbf{w}_-^*) f_{\mathbf{W}_+^*|\xi_+}\left(\mathbf{w}_+^*\right) h\left(\xi_-, \xi_+\right) d\xi_- d\xi_+}{\int_{\Xi} f_{\mathbf{W}_-^*|\xi}(\mathbf{w}_-^*) f_{\mathbf{W}_+^*|\xi}\left(\mathbf{w}_+^*\right) d\Lambda\left(\xi\right)} > \mathrm{cv}(k, \alpha)\right]. \tag{9}$$

Similarly as in (6), we denote Ξ as the parameter space of the tail index and $h(\cdot)$ to be the weigtht that transforms the composite alternative hypothesis into a simple one. The weight $\Lambda(\cdot)$ can be considered as the least favorable distribution, which we discuss more now.

Note that the null hypothesis of (8) is composite. We need to control size uniformly over all $\xi_{Z_-} = \xi_{Z_+} \in \Xi$. To that end, we can transform the composite null into a simple one by considering the weighted average density with respect to the weight Λ. Together with a suitably chosen critical value, this test (9) maintains the uniform size control. Now the problem reduces to determining an appropriate weight Λ. Elliott et al. (2015) study the generic hypothesis testing problem where a nuisance parameter exists in the null hypothesis. We tailor their argument for our test (9) and adopt their computational algorithm for implementation. In particular, $\Lambda(\cdot)$ and $\mathrm{cv}(k, \alpha)$ are numerically calculated and provided in tables along with the corresponding MATLAB code. Empiricists who use our test need only to construct the order statistics $\hat{\mathbf{Z}}_-^*$ and $\hat{\mathbf{Z}}_+^*$ and numerically evaluate the density. We provide more computational details in Appendix B. By the continuous mapping theorem and Proposition 3, for any fixed k, $\limsup_{n\to\infty} \mathbb{E}\left[\varphi_\pm\left(\hat{\mathbf{Z}}_-^*, \hat{\mathbf{Z}}_+^*\right)\right] \leq \alpha$ under the null hypothesis of (8).

As we discussed above, the hypothesis testing problem (8) simplifies to

$$H_0 : \xi_{Z_-} = \xi_{Z_+} = 0 \text{ against } H_1 : \xi_{Z_-} > \xi_{Z_+} = 0,$$

if V is assumed to be in the normal or Laplace family ($\xi_V = 0$). Proposition 3 implies $\hat{\mathbf{Z}}_-^*$ and $\hat{\mathbf{Z}}_+^*$ are asymptotically independent and both of them are EV distributed. Then accordingly, our test (9) reduces to

$$\mathbf{1}\left[\frac{\int_\Xi f_{\mathbf{W}_-^*|\xi_-}(\mathbf{w}_-^*)h(\xi_-)d\xi_-}{f_{\mathbf{W}_-^*|0}(\mathbf{w}_-^*)} > \mathrm{cv}(k, \alpha)\right],$$

which is identical to (6). This suggests that we can simply substitute $\hat{\mathbf{Z}}_-^*$ for $\mathbf{w}_-^*$ into (6) for implementation.

3 Simulation study

3.1 Hypothesis testing about noise V

We set $h(\cdot)$ to be the uniform weight on $[0, 1)$ to include all distributions with a finite mean and the level of significance to be 0.05. In Table 1, we report the small sample rejection probabilities of the test (6). We generate U_i from the right half-standard normal and the right half-Laplace(0,1) distributions and V_i from four distributions: standard normal, Laplace(0,1) (denoted La(0,1)), Student-t(2), Pareto(0.5), and F(4,4). The normal and Laplace distributions correspond to the null hypothesis, and the other three are alternative hypotheses. The results suggest that the test (6) has excellent performance in size and power. Note that when $k = 50$ and $n = 100$, we essentially include too many mid-sample observations so that the EV approximation is poor.

Now we consider the linear regression model that $Y_i = \mathbf{X}_i^\intercal \beta_0 + Z_i$ with $\mathbf{X}_i = (1, X_{2i})^\intercal$ and $\beta_0 = (1, 1)^\intercal$. We assume $X_{2i} \sim \mathcal{N}(0, 1)$ and independent from Z_i. Table 2 reports the rejection probabilities of our test (6). Findings are similar to those in Table 1.

3.2 Hypothesis testing about inefficiency U and noise V

Consider the hypothesis testing problem (8). We implement the test (9) with the same setup as above. Table 3 reports the rejection probabilities under the null and alternative hypotheses. We make the following observations. First, the test controls size well unless k is too large relative to n, as seen in the column with $n = 100$ and $k = 50$. This is again because we are using too many mid-sample observations to approximate the tail so that the EV convergence in Propositions 1-3 provides poor approximations. Second, the test has good power properties as seen from the last five rows. In particular, using only the largest 50 order statistics from 1,000 observations leads to the power of 0.94. Finally, the power decreases as the alternative hypothesis becomes closer to the null, as we move down along rows.

Now we consider the special case where V is in the normal or Laplace family. Then we implement (6) with $\hat{\mathbf{Z}}^*_-$ as the input. Table 4 contains the rejection probabilities under the null and alternative hypotheses. The rows with F_U being half-normal or Laplace correspond to the size under the null hypothesis, while other rows the power under the alternative hypothesis. The new test has excellent size and power properties.

4 Empirical illustration

We illustrate the new method using the US bank data collected by Feng and Serletis (2009). The data are a sample of US banks covering the period from 1998 to 2005 (inclusive). After deleting banks with negative or zero input prices, we are left with a balanced panel of 6,010 banks observed annually over the 8-year period. A more detailed description of the data may be found in Feng and Serletis (2009), who assume V is normal and U is half-normal. However, our test rejects such thin-tail assumption.

In particular, we specify a stochastic cost function, letting $Z = V + U$, so $U \geq 0$ is cost inefficiency, and more inefficient banks have higher total costs, Y. Since our tests are designed for cross-sectional data, we divide the original panel data into cross-sections (one for each year) and regress the logarithm of total bank cost on a constant and the logarithms of six control variables, including the wage rate for labor, the interest rate for borrowed funds, the price of physical capital, and the amounts of consumer loans, non-consumer loans, and securities.

Since the object of interest is the cost function, under the assumption that V is symmetric,[12] we multiply the OLS residuals by -1 and take the smallest and the largest $k \in \{25, 50, 75, 100\}$ order statistics, respectively, to implement the test (6). The p-values are reported in Table 5. These small p-values suggest that Z has heavy tails on both sides, so a Student-t assumption on V (e.g., Wheat et al, 2019) is more appropriate.

5 Concluding remarks

We derive several nonparametric tests of the tail behavior of the error components in the stochastic frontier model, which also apply to the kernel deconvolution density estimation when the target density of U is one-sided. The tests are easy to implement in MATLAB and are useful diagnostic tools for empiricists.

As the final remark, we provide a heuristic empirical guidance about some diagnostic methods, including ours, for SFA. Often a first-step diagnostic tool for SFA is to calculate the skewness of the OLS residuals to see if they are properly skewed. See Waldman (1982), Simar and Wilson (2010), and Horrace and Wright (2020). If they are positively skewed, the maximum likelihood estimator of the variance of inefficiency is zero, and OLS is the maximum likelihood estimator of β_0. If they are negatively skewed, then OLS is not a stationary point in the parameter space of the likelihood, and the stochastic frontier model is well-posed. After calculating negatively skewed OLS residuals, a useful second-step diagnostic tool is to implement our nonparametric tests to understand the tail behaviors of the error component distributions and to guide parametric choices subsequently.

[12]The symmetry assumption is reasonable here and is also imposed in Feng and Serletis (2009).

References

[1] Andrews, D. W. K. and W. Ploberger (1994). Optimal tests when a nuisance parameter is present only under the alternative, *Econometrica*, 62, 1383-1414.

[2] Aigner, D., C. A. K. Lovell, and P. Schmidt (1977). Formulation and estimation of stochastic production frontier models, *Journal of Econometrics*, 6, 21-37.

[3] Almanidis, P., J. Qian, and R. C. Sickles (2014). Stochastic frontiermodels with bounded inefficiency. In Sickles, R. C. and Horrace, W. C., eds. Festschrift in Honor of Peter Schmidt Econometric Methods and Applications, *New York: Springer*, 47–82.

[4] Arnold, B. C., N. Balakrishnan, and H. H. N. Nagaraja (1992). A First Course in Order Statistics, *Siam*.

[5] Cai, J., W. C. Horrace, and C. F. Parmeter (2020). Density deconvolution with Laplace errors and unknown variance, *Unpublished Manuscript, Syracuse University, Center for Policy Research*.

[6] Carree, M. A. (2002). Technological inefficiency and the skewness of the error component in stochastic frontier analysis, *Economics Letters* 77,101–107.

[7] Coelli T. (1995). Estimators and hypothesis tests for a stochastic frontier function: A Monte Carlo analysis. *Journal of Productivity Analysis* 6, 247–268.

[8] Elliott, G., U. K. Müller, and M. W. Watson (2015). Nearly optimal tests when a nuisance parameter is present under the null hypothesis, *Econometrica*, 83, 771-811.

[9] Feng, G., and A. Serletis (2009). Efficiency and productivity of the US banking in-dustry, 1998-2005: Evidence from the Fourier cost function satisfying global regularity conditions, *Journal of Applied Econometrics*, 24, 105-138.

[10] Florens, J. P., L. Simar, and I. Van Keilegom (2020). Estimation of the boundary of a variable observed with symmetric error, *Journal of the American Statistical Association*, 115:529, 425-441.

[11] Greene, W. H. (1990). A gamma-distributed stochastic frontier model, *Journal of Econometrics*, 46, 141-164.

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

[13] Hall, P. and L. Simar (2002). Estimating a change point, boundary, or frontier in the presence of observation error, *Journal of the American Statistical Association*, 97, 523-534.

[14] Horrace, W. C. and C. F. Parmeter (2011). Semiparametric deconvolution with unknown error variance, *Journal of Productivity Analysis*, 35, 129-141

[15] Horrace, W. C. and C. F. Parmeter (2018). A Laplace stochastic frontier model, *Econometric Reviews*, 37, 260-280.

[16] Horrace, W. C. and I. A. Wright (2020). Stationary points for parametric stochastic frontier models, *Journal of Business and Economic Statistics*, 38, 615-526

[17] Kneip, A., L. Simar, and I. Van Keilegom (2015). Frontier estimation in the presence of measurement error with unknown variance, *Journal of Econometrics*, 184, 379-393.

[18] Kopp, R. J., and J. Mullahy (1990). Moment-based estimation and testing of stochastic frontier models, *Journal of Econometrics*, 46, 165-83.

[19] Kumbhakar, S. C., C. F. Parmeter, and E. G. Tsionas (2013). A zero inefficiency stochastic frontier model, *Journal of Econometrics*, 172, 66–76.

[20] Li, Q. (1996). Estimating a stochastic production frontier when the adjusted error is symmetric, *Economics Letters* 52, 221–228.

[21] Meeusen, W., and J. van den Broeck (1977). Efficiency estimation from Cobb-Douglas production functions with composed error, *International Economic Review* 18, 435-444

[22] Müller, U. K. and Y. Wang (2017). Fixed-k asymptotic inference about tail properties, *Journal of the American Statistical Association*, 112, 1134-1143.

[23] Papadopoulos, A. and C. F. Parmeter (2020). Moment diagnostics and quasi-maximum likelihood estimation for the stochastic frontier model, *Unpublished Manuscript*.

[24] Simar, L., I, Van Keilegom, and V. Zelenyuk (2017). Nonparametric least squares methods for stochastic frontier models, *Journal of Productivity Analysis*, 47, 189-204.

[25] Simar, L., and P. W. Wilson (2010). Inference from cross-sectional stochastic frontier models, *Econometric Reviews*, 29, 62–98.

[26] Stefanski L. A. and R. J. Carroll (1990). Deconvolving kernel density estimators, *Statistics*, 21, 169-184,

[27] Tsionas, E. G. (2007). Effciency measurement with the Weibull stochastic frontier, *Oxford Bulletin of Economics and Statistics*, 69, 693–706.

[28] Waldman, D. M. (1982). A stationary point for the stochastic frontier likelihood, *Journal of Econometrics*, 18, 275–279.

[29] Wang, W. S., C. Amsler, and P. Schmidt (2011). Goodness of fit tests in stochastic frontier models. *Journal of Productivity Analysis* 35, 95–118.

[30] Wheat, P., A. D. Stead, and W. H. Greene (2019). Robust stochastic frontier analysis: a Student's t-half normal model with application to highway maintenance costs in England, *Journal of Productivity Analysis*, 51, 21–38.

Appendix

A Proofs

Proof of Proposition 1

Since only the right tail index of V shows up in this proof, we simply denote $\xi = \xi_{V_+}$ in this proof.

We prove the case with $k = 1$ first. By Corollary 1.2.4 and Remark 1.2.7 in de Haan and Ferreira (2007), the constants a_n and b_n can be chosen as follows. If $\xi > 0$, we choose $a_n = Q_V(1 - 1/n)$ and $b_n(\xi) = 0$. If $\xi = 0$, we choose $a_n = 1/(nf_V(b_n))$ and $b_n = Q_V(1 - 1/n)$. By construction, these constants satisfy that $1 - F_V(a_n v + b_n) = O(n^{-1})$ for any fixed $v > 0$ in both cases (e.g., Ch.1.1.2 in de Haan and Ferreira, 2007).

By Assumption 1(iv), we have that

$$
\begin{aligned}
&= \mathbb{P}\left(Z_{n:n} \leq a_n v + b_n\right) \\
&= \mathbb{P}\left(Z_i \leq a_n v + b_n\right)^n \\
&\equiv A_n(v) \cdot \left(1 + \frac{B_n(v)}{\mathbb{P}\left(V_i \leq a_n v + b_n\right)}\right)^n,
\end{aligned}
$$

where $A_n = \mathbb{P}\left(V_i \leq a_n v + b_n\right)^n$, and

$$
B_n(v) = \mathbb{P}\left(-U_i + V_i \leq a_n v + b_n\right) - \mathbb{P}\left(V_i \leq a_n v + b_n\right).
$$

By Assumption 1(iv), $A_n(v) \to G_\xi(v)$ for any constant $v > 0$. Then by the facts that $\mathbb{P}\left(V_i \leq a_n v + b_n\right) \to 1$ and $(1 + t/n)^n \to \exp(t)$, it suffices to show that $B_n(v) = o\left(n^{-1}\right)$. To this end, we have

$$
\begin{aligned}
& B_n(v) \\
=_{(1)} & \ \mathbb{E}\left[F_V(a_n v + b_n + U_i) - F_V(a_n v + b_n)\right] \\
\leq_{(2)} & \sup_{t \in [a_n v + b_n, \infty]} f_V(t) \cdot \mathbb{E}\left[|U_i|\right] \\
\leq_{(3)} & \ f_V(a_n v + b_n) \cdot \mathbb{E}\left[|U_i|\right] \\
=_{(4)} & \ o(n^{-1}),
\end{aligned}
$$

where eq.(1) is by Assumption 1(ii) (U_i is independent from V_i), ineq.(2) is by the intermediate value theorem, ineq.(3) follows from Assumption 1(iv) ($f_V(t)$ is non-increasing when $t > c$ for some constant c), and eq.(4) is seen by Assumption 1(iii) ($\mathbb{E}\left[|U_i|\right] < \infty$) and Assumption 1(iv). In particular, the fact that $nf_V(a_n v + b_n) = o(1)$ is implied by the von Mises condition. See, for example, Corollary 1.1.10 in de Haan and Ferreira (2007) with $t = Q_V(1 - 1/n)$.

Generalization to $k > 1$ is as follows. Consider $v_1 > v_2 > \cdots > v_k$. Chapter 8.4 in Arnold et al. (1992, p.219) gives that

$$
\begin{aligned}
& \mathbb{P}\left(Z_{n:n} \leq a_n v_1 + b_n, ..., Z_{n:n-k+1} \leq a_n v_k + b_n\right) \\
= \ & F_Z^{n-k}(a_n v_k + b_n) \prod_{r=1}^{k} (n - r + 1)\, a_n f_Z\left(a_n v_r + b_n\right) \\
= \ & \left[F_V^{n-k}\left(a_n v_k + b_n\right) \prod_{r=1}^{k} (n - r + 1)\, a_n f_V\left(a_n v_r + b_n\right)\right] \times \\
& \left[\left(\frac{F_Z\left(a_n v_k + b_n\right)}{F_V\left(a_n y_k + b_n\right)}\right)^{n-k} \prod_{r=1}^{k} \frac{f_Z\left(a_n v_r + b_n\right)}{f_V\left(a_n v_r + b_n\right)}\right] \\
\equiv \ & \tilde{A}_n \times \tilde{B}_n.
\end{aligned}
$$

The convergence that $\tilde{A}_n \to G_\xi(v_k)\prod_{r=1}^{k}\{g_\xi(v_r)/G_\xi(v_k)\}$ is established by Theorem 8.4.2 in Arnold et al. (1992). It now remains to show $\tilde{B}_n \to 1$. First, the fact that

$$(F_Z(a_n v_k + b_n)/F_V(a_n v_k + b_n))^{n-k} \to 1$$

is shown by the same argument as above in the $k = 1$ case. Second, for any v

$$\begin{aligned}
\frac{f_Z(v)}{f_V(v)} &= \frac{\frac{\partial \mathbb{E}[F_V(v+U_i)]}{\partial v}}{f_V(v)} \\
&= \frac{\frac{\partial}{\partial v}\int F_V(v+u) f_U(u)\, du}{f_V(v)} \\
&= \frac{\int \frac{\partial}{\partial v} F_V(v+u) f_U(u)\, du}{f_V(v)} \quad \text{(by Leibniz's rule)} \\
&= \frac{\mathbb{E}[f_V(v+U_i)]}{f_V(v)},
\end{aligned}$$

where applying Leibniz's rule is permitted by Assumption 1(iv), which implies that $f_V(v)$ is uniformly continuous in v. Then similarly as bounding B_n above, we use the mean value expansion and Assumptions 1(ii)-(iv) to derive that for any $r \in \{1, ..., k\}$ and some constant $0 < C < \infty$,

$$\begin{aligned}
& \left|\frac{f_Z(a_n v_r + b_n)}{f_V(a_n v_r + b_n)} - 1\right| \\
= & \left|\frac{\mathbb{E}[f_V(a_n v_r + b_n + U_i) - f_V(a_n v_r + b_n)]}{f_V(a_n v_r + b_n)}\right| \\
\leq & \sup_{t \in [a_n v_r + b_n, \infty]} \left|\frac{\partial f_V(t)/\partial t}{f_V(a_n v_r + b_n)}\right| \mathbb{E}[|U_i|] \\
\leq & \left|\frac{\partial f_V(a_n v_r + b_n)/\partial t}{f_V(a_n v_r + b_n)}\right| \mathbb{E}[|U_i|] \quad (\text{by} \frac{\partial f_V(t)}{\partial t} \nearrow 0) \\
\leq & C\left|\frac{f_V(a_n v_r + b_n)}{1 - F_V(a_n v_r + b_n)}\right| \mathbb{E}[|U_i|] \\
= & o(1),
\end{aligned}$$

where the last inequality follows from the fact that $\lim_{t\to\infty} \frac{\partial f_V(t)/\partial t(1-F_V(t))}{f_V(t)^2} \to -1-\xi$, which is implied by the von Mises condition (cf. Theorem 1.1.8 in de Haan and Ferreira (2007)), and the last equality follows from the facts that $n(1-F_V(a_n v_r + b_n)) = O(1)$ and $n f_V(a_n v_r + b_n) = o(1)$ (see again Corollary 1.1.10 in de Haan and Ferreira, 2007, with $t = Q_V(1-1/n)$). The proof is then complete. ■

Proof of Proposition 2

In this proof, we drop the subscript V_+ in ξ_{V_+} since it is the only tail index here.

Proposition 1 implies that

$$\frac{\mathbf{Z}_+ - b_n}{a_n} \xrightarrow{d} \mathbf{W}_+, \tag{10}$$

where $\mathbf{W}_+$ is jointly EV distributed with tail index ξ, and the constants a_n and b_n are chosen in the proof of Proposition 1.

Let $I = (I_1, \ldots, I_k) \in \{1, \ldots, T\}^k$ be the k random indices such that $Z_{n:n-j+1} = Z_{I_j}$, $j = 1, \ldots, k$, and let $\hat{I}$ be the corresponding indices such that $\hat{Z}_{n:n-j+1} = \hat{Z}_{\hat{I}_j}$. Then the convergence of $\hat{\mathbf{Z}}_+$ follows from (10) once we establish $|\hat{Z}_{\hat{I}_j} - Z_{I_j}| = o_p(a_n)$ for $j = 1, \ldots, k$. We consider $k = 1$ for simplicity and the argument for a general k is very similar. Denote $\varepsilon_i \equiv \hat{Z}_i - Z_i$.

Consider the case with $\xi > 0$. the part in Assumption 2(v) for $\xi > 0$ yields that

$$\begin{aligned} \sup_i |\varepsilon_i| &= \sup_i \left| \mathbf{X}_i \left(\beta_0 - \hat{\beta} \right) \right| \\ &\leq \left|\left| \beta_0 - \hat{\beta} \right|\right| \sup_i ||\mathbf{X}_i|| \\ &= o_p(1). \end{aligned}$$

Given this, we have that, on the one hand, $\hat{Z}_{\hat{I}} = \max_i \{Z_i + \varepsilon_i\} \leq Z_I + \sup_i |\varepsilon_i| = Z_I + o_p(1)$;

and on the other hand, $\hat{Z}_{\hat{I}} = \max_i\{Z_i + \varepsilon_i\} \geq \max_i\{Z_i + \min_i\{\varepsilon_i\}\} \geq Z_I + \min_i\{\varepsilon_i\} \geq Z_I - \sup_i |\varepsilon_i| = Z_I - o_p(1)$. Therefore, $|\hat{Z}_{\hat{I}} - Z_I| \leq o_p(1) = o_p(a_n)$ since $a_n \to \infty$.

Consider the case with $\xi = 0$. Corollary 1.2.4 in de Haan and Ferreira (2007) implies that $a_n = f_V\left(Q_V(1-1/n)\right)$. Thus, the part in Condition 2.3 for $\xi = 0$ implies that

$$\begin{aligned} \frac{1}{a_n}\sup_i |\varepsilon_i| &\leq \frac{\sup_i ||\mathbf{X}_i|| \cdot \left|\left|\beta_0 - \hat{\beta}\right|\right|}{f_W\left(Q_W(1-1/n)\right)} \\ &= o_p(1). \end{aligned}$$

Then the same argument as above yields that $\left|\hat{Z}_{\hat{I}} - Z_I\right| \leq O_p\left(\sup_i |\varepsilon_i|\right) = o_p(a_n)$. ■

Proof of Proposition 3

Let $\mathbf{Z}_-^*$ denote the k smallest order statistics of $\{Z_i\}$. Let $(a_n^+, b_n^+)^\intercal$ and $(a_n^-, b_n^-)^\intercal$ be the sequences of normalizing constants for the right and left tails of Z, respectively. Then by the same argument as in Proposition 2, we have $\hat{\mathbf{Z}}_- - \mathbf{Z}_- = o_p(a_n^-)$ and $\hat{\mathbf{Z}}_+ - \mathbf{Z}_+ = o_p(a_n^+)$. Therefore, it suffices to establish $\mathbf{Z}_+$ and $\mathbf{Z}_-$ jointly converge to $(\mathbf{W}_+^\intercal, \mathbf{W}_-^\intercal)^\intercal$ where $\mathbf{W}_+$ and $\mathbf{W}_-$ are independent and both EV distributed with indices ξ_{Z_+} and ξ_{Z_-}, respectively. To this end, note that the case with $k = 1$ is established as Theorem 8.4.3 in Arnold et al. (1992). We now generalize their argument for $k \geq 2$.

By elementary calculation and the i.i.d. assumption, the joint density of the order statistics $Z_{n:n}, \ldots, Z_{n:1}$ is $n! \prod_{i=1}^n f_Z(z_i)$ for $z_1 \leq z_2 \leq \ldots \leq z_n$. Then by a change of variables, the joint density of $(Z_{n:n} - b_n^+)/a_n^+, \ldots, (Z_{n:n-k+1} - b_n^+)/a_n^+, (Z_{n:k} - b_n^-)/a_n^-, \ldots, (Z_{n:1} - b_n^+)/a_n^+$ satisfies that for $v_1^- \leq v_2^- \leq \ldots \leq v_k^- \leq v_k^+ \leq \ldots v_1^+$,

$$
\begin{aligned}
& \mathbb{P}\left(\begin{array}{l} Z_{n:n} \leq a_n^+ v_1^+ + b_n^+, ..., Z_{n:n-k+1} \leq a_n^+ v_k^+ + b_n^+, \\ Z_{n:1} \geq a_n^- v_1^- + b_n^-, ..., Z_{n:k} \leq a_n^- v_k^- + b_n^- \end{array}\right) \\
= \;& \left(F_Z(a_n^+ v_k^+ + b_n^+) - F_Z(a_n^- v_k^- + b_n^-)\right)^{n-2k} \\
& \times \prod_{r=1}^{k} (n-r+1)\, a_n^- f_Z\left(a_n^- v_r^- + b_n^-\right) \\
& \times \prod_{r=1}^{k} (n-r+1)\, a_n^+ f_Z\left(a_n^+ v_r^+ + b_n^+\right) \\
\equiv \;& P_{1n} \times P_{2n} \times P_{3n}.
\end{aligned}
$$

By the DOA assumption for both the left and right tails and equations (8.3.1) and (8.4.9) in Arnold et al. (1992),

$$
P_{1n} \to G_{\xi_{Z_+}}\left(v_k^+\right)\left(1 - G_{\xi_{Z_-}}\left(v_k^-\right)\right).
$$

By (8.4.4) in Arnold et al. (1992) and the fact that k is fixed, $P_{2n} \to \prod_{r=1}^{k} g_{\xi_{Z_-}}\left(v_r^-\right)/G_{\xi_{Z_-}}(v_r^-)$ and $P_{3n} \to \prod_{r=1}^{k} g_{\xi_{Z_+}}\left(v_r^+\right)/\left(1 - G_{\xi_{Z_+}}(v_r^+)\right)$. The proof is then complete by combining P_{jn} for $j = 1, 2, 3$ and the continuous mapping theorem. ■

B Computational details

This section provides more details for constructing the test (9), which is based on the limiting observations $\mathbf{W}_-^*$ and $\mathbf{W}_+^*$. The density is given by (5), which is computed by Gaussian Quadrature. To construct the test (9), we specify the weight h to be uniform over the alternative space for expositional simplicity, which can be easily changed. Then, it remains to determine a suitable candidate for the weight Λ and the critical value $\mathrm{cv}(k, \alpha)$. We do this by the generic algorithm provided by Elliott et al. (2015) and Müller and Wang (2017).

The idea of identifying a suitable choice of Λ and $\mathrm{cv}(k, \alpha)$ is as follows. First, we can discretize Ξ into a grid Ξ_a and determine Λ accordingly as the point masses. Then we can simulate N random draws of $\mathbf{W}_-^*$ and $\mathbf{W}_+^*$ from $\xi \in \Xi_a$ and estimate the rejection probability $P_\xi(\varphi_\pm(\mathbf{W}_-^*, \mathbf{W}_+^*) = 1)$ by sample fractions. The subscript ξ emphasizes that the rejection probability depends on the value of ξ that generates the data. By iteratively increasing or decreasing the point masses as a function of whether the estimated $P_\xi(\varphi_\pm(\mathbf{W}_-^*, \mathbf{W}_+^*) = 1)$ is larger or smaller than the nominal level, we can find a candidate Λ together with $\mathrm{cv}(k, \alpha)$ that numerically satisfy the uniform size control. Note that we allow $P_\xi(\varphi_\pm(\mathbf{W}_-^*, \mathbf{W}_+^*) = 1)$ to be less than the nominal level for some ξ.

In practice, we implement the following steps. Let c be short for $\mathrm{cv}(k, \alpha)$.

1. Simulate $N = 10{,}000$ i.i.d. random draws from some proposal density with ξ drawn uniformly from Ξ_a, which is an equally spaced grid on $[0, 0.99]$ with 50 points.

2. Start with $\Lambda_{(0)} = \{1/50, 1/50, \ldots, 1/50\}$ and $c = 1$. Calculate the (estimated) coverage probabilities $P_{\xi_j}(\varphi_\pm(\mathbf{W}_-^*, \mathbf{W}_+^*) = 1)$ for every $\xi_j \in \Xi_a$ using importance sampling. Denote them by $P = (P_1, ..., P_{50})^\intercal$.

3. Update Λ and c by setting $c\Lambda_{(s+1)} = c\Lambda_{(s)} + \kappa(P - 0.05)$ with some step-length constant $\kappa > 0$, so that the j-th point mass in $c\Lambda$ is increased/decreased if the coverage probability for ξ_j is larger/smaller than the nominal level.

4. Integrate for 500 times. Then, the resulting $\Lambda_{(500)}$ and c are a valid candidate.

5. Numerically check if $\varphi_\pm$ with $\Lambda_{(500)}$ and c indeed controls the size uniformly by simulating the rejection probabilities over a much finer grid on Ξ. If not, go back to step 2 with a finer Ξ_a.

n	100			1000		
k	10	20	50	10	20	50
F_V	Rejection Prob. under half-normal U_i					
N(0,1)	0.01	0.00	0.00	0.03	0.02	0.01
La(0,1)	0.05	0.04	0.02	0.05	0.05	0.05
t(2)	0.31	0.45	0.35	0.30	0.49	0.76
±Pa(0.5)	0.31	0.52	0.12	0.30	0.48	0.78
±F(4,4)	0.32	0.50	0.50	0.29	0.50	0.80
F_V	Rejection Prob. under half-Laplace U_i					
N(0,1)	0.02	0.00	0.00	0.03	0.01	0.01
La(0,1)	0.05	0.04	0.050	0.06	0.06	0.04
t(2)	0.32	0.46	0.29	0.32	0.50	0.80
±Pa(0.5)	0.31	0.52	0.06	0.30	0.49	0.79
±F(4,4)	0.31	0.52	0.49	0.28	0.48	0.78

Table 1: Small sample rejection probabilities of test (6) based on the k largest order statistics of $Z = -U + V$. U_i is generated from half-standard normal or half-Laplace(0,1) and V_i is generated from standard normal, Laplace(0,1), Student-t(2), Pareto(0.5), and F(4,4). Based on 1,000 simulation draws. Significance level is 0.05.

n	100			1000		
k	10	20	50	10	20	50
F_V	Rejection Prob. under half-normal U_i					
N(0,1)	0.01	0.00	0.00	0.03	0.02	0.01
La(0,1)	0.05	0.04	0.02	0.05	0.05	0.05
t(2)	0.30	0.44	0.33	0.30	0.49	0.80
±Pa(0.5)	0.32	0.51	0.10	0.30	0.48	0.80
±F(4,4)	0.32	0.50	0.47	0.29	0.50	0.80
F_V	Rejection Prob. under half-Laplace U_i					
N(0,1)	0.01	0.00	0.00	0.03	0.01	0.01
La(0,1)	0.04	0.05	0.01	0.06	0.06	0.04
t(2)	0.32	0.46	0.28	0.31	0.50	0.80
±Pa(0.5)	0.32	0.50	0.07	0.30	0.49	0.80
±F(4,4)	0.31	0.51	0.47	0.28	0.48	0.78

Table 2: Small sample rejection probabilities of test (6) based on the k largest order statistics of the OLS residuals. U_i is generated from half-normal and V_i is generated from standard normal, Laplace(0,1), Student-t(2), Pareto(0.5), and F(4,4). Based on 1,000 simulation draws. Significance level is 0.05.

n		100			1000		
k		10	20	50	10	20	50
F_V	F_U	Rejection Prob. under H_0					
N(0,1)	half-N(0,1)	0.06	0.06	0.03	0.05	0.05	0.05
La(0,1)	half-La(0,1)	0.04	0.05	0.07	0.05	0.04	0.04
t(2)	half-t(2)	0.04	0.06	0.24	0.06	0.06	0.05
$\pm$Pa(0.5)	$-$Pa(0.5)	0.05	0.06	0.67	0.05	0.06	0.05
$\pm$F(4,4)	$-$F(4,4)	0.04	0.05	0.24	0.04	0.06	0.03
F_V	F_U	Rejection Prob. under H_1					
N(0,1)	$-$Pa(0.75)	0.28	0.68	0.99	0.25	0.55	0.94
Laplace(0,1)	$-$Pa(0.75)	0.25	0.46	0.89	0.21	0.44	0.82
t(2)	$-$Pa(0.75)	0.13	0.21	0.62	0.09	0.09	0.19
$\pm$Pa(0.5)	$-$Pa(0.75)	0.10	0.23	0.87	0.07	0.12	0.17
$\pm$F(4,4)	$-$Pa(0.75)	0.09	0.14	0.46	0.07	0.09	0.15

Table 3: Small sample rejection probabilities of test (9) based on the k largest and k smallest order statistics of the OLS residuals. U_i is generated from half-norma, half-Laplace, Student-t(2), Pareto(0.5), F(4,4), and Pareto(0.75) and V_i is generated from standard normal, Laplace(0,1), Student-t(2), Pareto(0.5), and F(4,4). Based on 1,000 simulation draws. Significance level is 0.05.

n	100			1000		
k	10	20	50	10	20	50
F_U	Rejection Prob. under Normal V_i					
half-N(0,1)	0.02	0.01	0.00	0.02	0.02	0.00
half-La(0,1)	0.05	0.03	0.00	0.06	0.05	0.04
half-t(2)	0.31	0.45	0.46	0.34	0.54	0.88
−Pa(0.5)	0.34	0.52	0.48	0.33	0.56	0.89
−F(4,4)	0.31	0.52	0.70	0.32	0.52	0.86
F_U	Rejection Prob. under Laplace V_i					
half-N(0,1)	0.05	0.03	0.01	0.05	0.05	0.05
half-La(0,1)	0.04	0.03	0.00	0.05	0.04	0.03
half-t(2)	0.28	0.37	0.38	0.35	0.58	0.89
−Pa(0.5)	0.30	0.40	0.44	0.33	0.58	0.89
−F(4,4)	0.30	0.49	0.59	0.33	0.52	0.86

Table 4: Rejection probabilities of test (6), based on the smallest k order statistics of the OLS residuals. U_i is generated from various distributions and V_i is generated from standard normal or Laplace(0,1). Based on 1,000 simulation draws. Significance level is 0.05.

		left tail				right tail			
year	$k =$	25	50	75	100	25	50	75	100
1998		> 0.1	0.03	0.00	0.00	0.00	0.00	0.00	0.00
1999		0.00	0.00	0.00	0.00	0.05	0.00	0.00	0.00
2000		> 0.1	0.03	0.00	0.00	0.00	0.00	0.00	0.00
2001		0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
2002		> 0.1	0.00	0.00	0.00	> 0.1	0.00	0.00	0.00
2003		> 0.1	0.05	0.00	0.00	> 0.1	0.00	0.00	0.00
2004		0.04	0.00	0.00	0.00	> 0.1	0.00	0.00	0.00
2005		0.09	0.00	0.00	0.00	> 0.1	0.00	0.00	0.00

Table 5: P-values of the test (6) for the US Banks data collected by Feng and Serletis (2009).

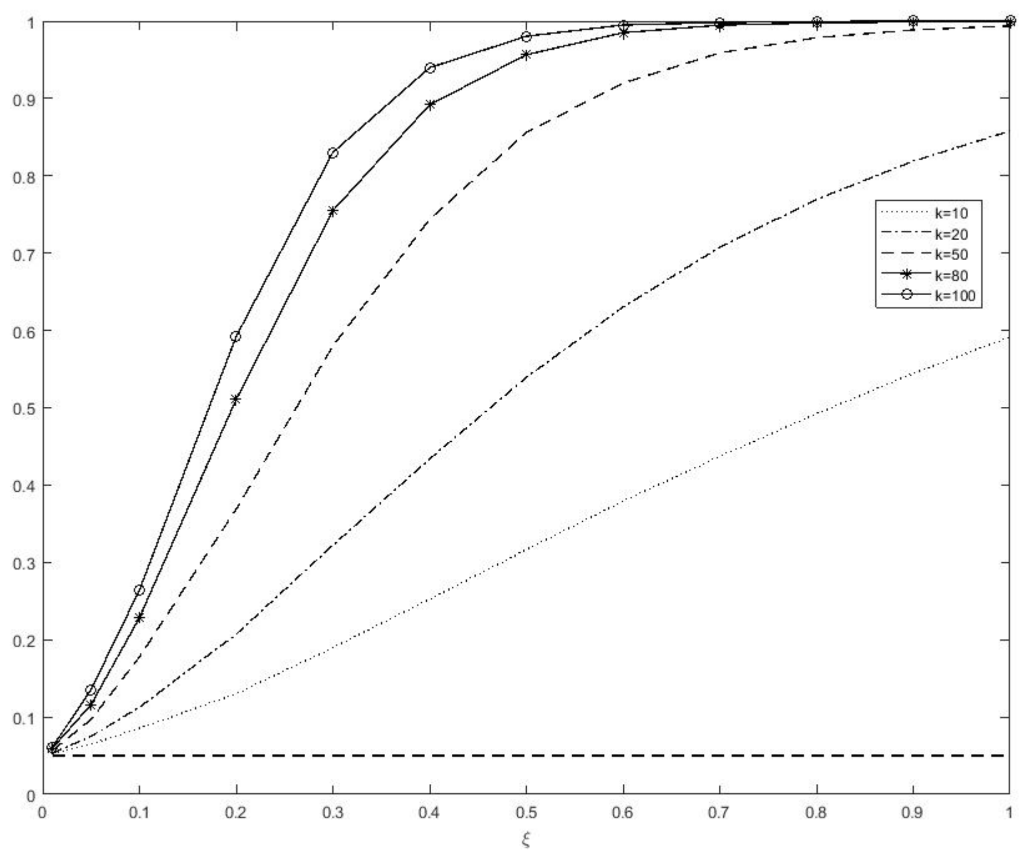

Figure 1: Asymptotic rejection probabilities of the test (6) with $\mathbf{W}_{+}^{*}$ generated from the joint extreme value distribution (5) and the nominal size of 0.05. The plots are based on numerical simulations with 10,000 random draws.