Inference for Losers

By Isaiah Andrews, Dillon Bowen, Toru Kitagawa and Adam McCloskey*

Empirical researchers sometimes estimate parameters of interest (e.g. average outcomes or treatment effects) for many parallel units or treatments and report the result in "league tables" which present a ranked list of the units considered, together with their estimated effects.

Andrews, Kitagawa and McCloskey (2020, henceforth AKM) note that inference on the estimated "best" treatment leads to bias and undercoverage for conventional estimators and confidence intervals (CIs), respectively. Similar problems emerge for league tables, where e.g. the estimate reported in the third position may systematically overestimate the true coefficient for the (random) unit ranked third. This paper extends the results of AKM from inference on the single "best" treatment to more general ranking problems.

Suppose we are interested in a finite collection of treatments $\Theta = \{\theta_1 \ldots, \theta_K\}$. For each treatment $\theta \in \Theta$ we observe normally distributed estimates $(X(\theta), Y(\theta))^T \in \mathbb{R}^2$, where $X = (X(\theta_1), \ldots, X(\theta_K))^T$ and $Y = (Y(\theta_1), \ldots, Y(\theta_K))^T$ are jointly normal, $(X^T, Y^T)^T \sim \mathcal{N}(\mu, \Sigma)$, where

$$E\left[X(\theta), Y(\theta)\right] = \mu(\theta) = \left[\mu_X(\theta), \mu_Y(\theta)\right]$$

is unknown, while

$$\Sigma(\theta, \theta') = \begin{pmatrix} \Sigma_X(\theta, \theta') & \Sigma_{XY}(\theta, \theta') \\ \Sigma_{YX}(\theta, \theta') & \Sigma_Y(\theta, \theta') \end{pmatrix}$$

$$= Cov\left(\begin{pmatrix} X(\theta) \\ Y(\theta) \end{pmatrix}, \begin{pmatrix} X(\theta') \\ Y(\theta') \end{pmatrix}\right)$$

is known. We abbreviate $\Sigma(\theta, \theta)$ as $\Sigma(\theta)$, and assume $\Sigma_Y(\theta) > 0$ for all $\theta \in \Theta$. AKM show that their results for this finite-sample normal setting translate to uniform asymptotic results for procedures based on asymptotically normal (X, Y) and consistent estimates of Σ, and we expect that the same will be true for results in the present paper, but focus on the normal model for brevity.

To describe our inference problem, let $\hat{\theta}_{(1)}, \ldots, \hat{\theta}_{(K)}$ sort the elements of Θ based on $X(\theta)$, so $X(\hat{\theta}_{(1)}) \geq X(\hat{\theta}_{(2)}) \geq \ldots \geq X(\hat{\theta}_{(K)})$. AKM show that conventional estimators and CIs for $\mu(\hat{\theta}_{(1)})$ can perform poorly. Inference on $\mu(\hat{\theta}_{(k)})$ for any k raises similar issues. For instance, standard CIs may perform poorly for inference on the third-ranked treatment, $\mu_Y(\hat{\theta}_{(3)})$.[1]

For a fixed set of ranks $R \subseteq \{1, \ldots, K\}$, let $\widehat{R} = \left\{\hat{\theta}_{(k)} : k \in R\right\} \subseteq \Theta$ denote the set of treatments with estimated rank in R. We consider two possible problems: first, conditional inference, where we seek to conduct inference on a particular treatment conditional on it being ranked in R. For CIs, for instance, this requires that, for all μ and $Pr_\mu\{\cdot\}$ the probability under μ,[2]

$$(1) \quad Pr_\mu\left\{\mu_Y(\theta) \in CI | \theta \in \widehat{R}\right\} \geq 1 - \alpha.$$

In the special case with $R = \{3\}$, for instance, this requires that $\mu_Y(\theta)$ be covered with probability $1-\alpha$ conditional on θ being ranked third. Second, we consider unconditional inference, which for a set of ranks R controls the average, unconditional perfor-

* Andrews: Harvard University, Department of Economics, Littauer Center M-18, Cambridge, MA 02138, iandrews@fas.harvard.edu. Bowen: University of Pennsylvania, The Wharton School, 3730 Walnut St, Huntsman Hall, Office 532.3, Philadelphia, PA 19104, dsbowen@wharton.upenn.edu. Kitagawa: Brown University and University College London, Robinson Hall, 64 Waterman Street, Providence, RI 02912, toru_kitagawa@brown.edu. McCloskey: University of Colorado at Boulder, Department of Economics, 256 UCB, Boulder, CO 80309, adam.mccloskey@colorado.edu. Code implementing the procedures discussed in this paper is available at https://dsbowen-conditional-inference.readthedocs.io/.

[1]Note that our goal is to conduct inference given an approach to forming a ranking, rather than to form improved rankings as in e.g. Gu and Koenker (2020)

[2]Henceforth, all statements involving $Pr_\mu\{\cdot\}$ are taken to hold for all $\mu \in \mathbb{R}^{2K}$.

mance for the (random) set of treatments ranked in R.[3] For CIs, for instance, unconditional coverage requires that

$$(2)\quad \frac{1}{|R|}\sum_{k\in R} Pr_{\mu}\left\{\mu_Y(\hat{\theta}_{(k)})\in CI\right\}\geq 1-\alpha.$$

In the special case with $R=\{3\}$, this requires that the *random* target parameter $\mu_Y(\hat{\theta}_{(3)})$ be covered with probability at least $1-\alpha$. We refer the reader to AKM for further discussion and comparison of unconditional and conditional validity.

I. Conditional Inference

Following AKM, we seek an optimal estimator $\hat{\mu}_\alpha$ such that conditional α-quantile unbiasedness given $\theta\in\widehat{R}$ holds,

$$(3)\qquad Pr_{\mu}\left\{\tilde{\mu}_\alpha\geq\mu_Y(\theta)|\theta\in\widehat{R}\right\}=\alpha.$$

Our construction uses the distribution of (X,Y) given $X\in\mathcal{X}_R(\theta):=\{X:\theta\in\widehat{R}\}$. Define $Z_\theta=(Z_\theta(\theta_1),\ldots,Z_\theta(\theta_K))^T$ for

$$Z_\theta(\theta'):=X(\theta')-\Big(\Sigma_{XY}(\theta',\theta)/\Sigma_Y(\theta)\Big)Y(\theta),$$

and denote the set of possible values of $Y(\theta)$ such that $X\in\mathcal{X}_R(\theta)$ and $Z_\theta=z$ as $\mathcal{Y}_R(\theta,z)$. The set $\mathcal{Y}_R(\theta,z)$ is equal to

$$\left\{y:z+\Big(\Sigma_{XY}(\cdot,\theta)/\Sigma_Y(\theta)\Big)y\in\mathcal{X}_R(\theta)\right\}.$$

Conditional on $\{\theta\in\widehat{R},Z_\theta=z\}$, $Y(\theta)$ follows a truncated normal distribution:

$$(4)\qquad \xi|\xi\in\mathcal{Y}_R(\theta,z),$$

where $\xi\sim\mathcal{N}\big(\mu_Y(\theta),\Sigma_Y(\theta)\big)$ is the unconditional distribution of $Y(\theta)$.

To derive the quantile-unbiased estimator, let $F_{TN}(y;\mu_Y(\theta),R,z)$ be the cumulative distribution function for the truncated normal distribution from (4). Define $\hat{\mu}_\alpha$ as the unique solution to

$$F_{TN}(Y(\theta);\hat{\mu}_\alpha,R,Z_\theta)=1-\alpha.$$

[3] For another recent application of average coverage, see Armstrong, Kolesár and Plagborg-Møller (2021).

Proposition 7 of AKM establishes that $\hat{\mu}_\alpha$ is the optimal estimator of $\mu_Y(\theta)$ for which (3) holds, in the sense that any other estimator satisfying (3) must, on average, be less accurate than $\hat{\mu}_\alpha$. Hence, $\hat{\mu}_{\frac{1}{2}}$ is a median-unbiased estimator, while $[\hat{\mu}_{\alpha/2},\hat{\mu}_{1-\alpha/2}]$ is a level $1-\alpha$ CI with conditional coverage (1).

To compute $\hat{\mu}_\alpha$ we need to compute the set $\mathcal{Y}_R(\theta,Z_\theta)$, which takes a simple form in some cases. To illustrate, suppose $Y=X$, Σ_Y is diagonal, and θ was selected because $Y(\theta)$ was ranked as the k^{th} best. In this special case, $\mathcal{Y}_R(\theta,Z_\theta)=[Y(\hat{\theta}_{(k+1)}),Y(\hat{\theta}_{(k-1)})]$, so we truncate the normal distribution for $Y(\theta)$ based on the treatments ranked immediately above and below. In the general case with non-diagonal Σ_{XY} the truncation set $\mathcal{Y}_R(\theta,z)$ takes a more complex form that is nonetheless straightforward to compute.

PROPOSITION 1: *Define*

$$Q_\theta(\theta'):=\frac{\Sigma_Y(\theta)\Big(Z_\theta(\theta')-Z_\theta(\theta)\Big)}{\Sigma_{XY}(\theta)-\Sigma_{XY}(\theta',\theta)}$$

and $\Theta_{-\theta}:=\Theta\setminus\{\theta\}$. For any two disjoint subsets $\tau^{\mathcal{L}}\subseteq\Theta_{-\theta}$ and $\tau^{\mathcal{U}}\subseteq\Theta_{-\theta}$ define

$$S_\theta^{\mathcal{L}}(\tau^{\mathcal{L}},\tau^{\mathcal{U}}):=\Big\{\theta'\in\Theta_{-\theta}: \big(\theta'\in\tau^{\mathcal{L}}\text{ and }\Sigma_{XY}(\theta)>\Sigma_{XY}(\theta',\theta)\big)\text{ or }\big(\theta'\in\tau^{\mathcal{U}}\text{ and }\Sigma_{XY}(\theta)<\Sigma_{XY}(\theta',\theta)\big)\Big\},$$

$$\mathcal{L}(\theta,Z_\theta,\tau^{\mathcal{L}},\tau^{\mathcal{U}}):=\max_{\theta'\in S_\theta^{\mathcal{L}}(\tau^{\mathcal{L}},\tau^{\mathcal{U}})}Q_\theta(\theta'),$$

$$S_\theta^{\mathcal{U}}(\tau^{\mathcal{L}},\tau^{\mathcal{U}}):=\Big\{\theta'\in\Theta_{-\theta}: \big(\theta'\in\tau^{\mathcal{L}}\text{ and }\Sigma_{XY}(\theta)<\Sigma_{XY}(\theta',\theta)\big)\text{ or }\big(\theta'\in\tau^{\mathcal{U}}\text{ and }\Sigma_{XY}(\theta)>\Sigma_{XY}(\theta',\theta)\big)\Big\},$$

$$\mathcal{U}(\theta,Z_\theta,\tau^{\mathcal{L}},\tau^{\mathcal{U}}):=\min_{\theta'\in S_\theta^{\mathcal{U}}(\tau^{\mathcal{L}},\tau^{\mathcal{U}})}Q_\theta(\theta').$$

Also define

$$S_\theta^{\mathcal{V}}:=\Big\{\theta'\in\Theta:\Sigma_{XY}(\theta)=\Sigma_{XY}(\theta',\theta)\Big\},$$

$$\mathcal{V}(\theta,Z_\theta):=\min_{\theta'\in S_\theta^{\mathcal{V}}}-\Big(Z_\theta(\theta')-Z_\theta(\theta)\Big).$$

Now consider all possible arrangements of parameters θ such that $\theta = \hat{\theta}_{(k)}$,

$$\mathcal{S}_k(\theta) := \{(\tau^{\mathcal{L}} \subseteq \Theta_{-\theta}, \tau^{\mathcal{U}} \subseteq \Theta_{-\theta}) : |\tau^{\mathcal{L}}| = K - k, |\tau^{\mathcal{U}}| = k - 1\},$$

and a subset of arrangements such that $\theta \in \widehat{R}$,

$$\mathcal{S}_P(\theta) := \{(\tau^{\mathcal{L}}, \tau^{\mathcal{U}}) \in \bigcup_{k \in R} \mathcal{S}_k(\theta) : Q_\theta(\theta') \leq Q_\theta(\theta'') \; \forall \theta' \in S_\theta^{\mathcal{L}}(\tau^{\mathcal{L}}, \tau^{\mathcal{U}}), \theta'' \in S_\theta^{\mathcal{U}}(\tau^{\mathcal{L}}, \tau^{\mathcal{U}})\}.$$

If $\mathcal{V}(\theta, z) \geq 0$, then $\mathcal{Y}_R(\theta, z)$ is equal to

$$\bigcup_{(\tau^{\mathcal{L}}, \tau^{\mathcal{U}}) \in \mathcal{S}_P(\theta)} \left[\mathcal{L}(\theta, z, \tau^{\mathcal{L}}, \tau^{\mathcal{U}}), \mathcal{U}(\theta, z, \tau^{\mathcal{L}}, \tau^{\mathcal{U}})\right].$$

If $\mathcal{V}(\theta, z) < 0$, then $\mathcal{Y}_R(\theta, z) = \emptyset$.

Algorithm 1 in the Online Appendix provides an efficient means of computing $\mathcal{Y}_R(\theta, z)$, making implementation computationally feasible even for a large number of indices K. We also provide a user-friendly Python package (available at https://dsbowen.gitlab.io/conditional-inference/) to implement our approach.

II. Unconditional Inference

Conditional coverage (1) is a demanding requirement, and AKM note in their setting that even optimal conditional CIs can be quite long. To improve performance, we may weaken the requirement for quantile unbiasedness to bound only the average unconditional quantile bias across treatments ranked in R

$$\frac{1}{|R|} \left| \sum_{k \in R} \left(Pr_\mu \left\{ \hat{\mu}_{\alpha,k} \geq \mu_Y(\hat{\theta}_{(k)}) \right\} - \alpha \right) \right| \leq \beta \cdot \max\{\alpha, 1 - \alpha\} \tag{5}$$

for $\hat{\mu}_{\alpha,k}$ the estimator of $\mu_Y(\hat{\theta}_{(k)})$ and $\beta \in [0, \alpha)$ a user-selected constant.[4]

As in AKM, we control unconditional bias using a hybrid approach that builds upon our conditional results. Specifically, for $\beta \in [0, \alpha)$, let c_β denote the $1 - \beta$ quantile of $\max_{\theta \in \Theta} |\zeta(\theta)| / \sqrt{\Sigma_Y(\theta)}$ for $(\zeta(\theta_1), \ldots, \zeta(\theta_K)) \sim \mathcal{N}(0, \Sigma_Y)$ and define

$$\mathcal{Y}_R^H(\theta, \mu_Y(\theta), z) := \mathcal{Y}_R(\theta, z) \bigcap \left[\mu_Y(\theta) - c_\beta \sqrt{\Sigma_Y(\theta)}, \mu_Y(\theta) + c_\beta \sqrt{\Sigma_Y(\theta)}\right].$$

Define $\hat{\mu}_{\alpha,k}^H$ as the unique solution to

$$F_{TN}^H(Y(\hat{\theta}_{(k)}); \hat{\mu}_{\alpha,k}^H, R, Z_{\hat{\theta}_{(k)}}) = 1 - \alpha,$$

where $F_{TN}^H(y; \mu_Y(\theta), R, z)$ denotes the cumulative distribution function of $\xi | \xi \in \mathcal{Y}_R^H(\theta, \mu_Y(\theta), z)$ for $\xi \sim \mathcal{N}(\mu_Y(\theta), \Sigma_Y(\hat{\theta}_{(k)}))$.

PROPOSITION 2: *The hybrid estimator $\hat{\mu}_{\alpha,k}^H$ satisfies (5), while the hybrid interval $\left[\hat{\mu}^H_{\frac{\alpha-\beta}{2(1-\beta)},k}, \hat{\mu}^H_{1-\frac{\alpha-\beta}{2(1-\beta)},k}\right]$ satisfies (2).*

III. Movers Design and Opportunity Atlas Illustration

To illustrate our methods, we revisit recent work on the effects of neighborhoods using data from Chetty and Hendren (2018) and the Opportunity Atlas (OA, Chetty et al., 2018). Chetty and Hendren (2018) use a movers design to estimate the causal effect of growing up in a given area on a wide variety of outcomes, while Chetty et al. (2018) report observational estimates for the average income rank in adulthood for children growing up in different areas. Following Mogstad et al. (2021) we focus on average income rank in adulthood for children growing up in households at the 25th percentile of the income distribution across different commuting zones (CZs).

We first calibrate simulations to the movers results from Chetty and Hendren (2018) and Chetty et al. (2018), treating their point estimates as the true parameter values and adding normal noise with standard deviation equal to their standard errors to generate simulated estimates. For each simulation draw we sort the estimates and calculate (i) the estimation error for the CZ appearing in each rank when using

[4] We set $\beta = \frac{\alpha}{10}$ in our application and simulations.

the conventional estimator and (ii) whether the conventional 95% CI for the CZ appearing in each rank covers the true coefficient for that CZ. The upper left panel of Figure 1 reports the median estimation error across ten million simulation draws based on the Chetty and Hendren (2018) estimates, while the lower left panel reports simulation results based on the Chetty et al. (2018) estimates. In both cases, we report results in standard deviation units, normalizing by the standard deviation of true effect sizes. The right two panels report coverage based on the same designs.

In the simulations calibrated to the Chetty and Hendren (2018) estimates, we see that estimates at the top of the rankings are upward-biased, while those at the bottom are downward-biased. Conventional CIs undercover at the top and bottom of rankings, while overcovering in the middle. By contrast, the Chetty et al. (2018) estimates (based on large observational samples, rather than smaller samples of movers) are considerably more precise, and we see that median estimation errors for conventional estimates are uniformly close to zero. Coverage of conventional CIs, while not exactly 95%, never falls below 90%.

Motivated by the presence of these biases and coverage distortions, we next report corrected estimates. Table 1 presents estimates for the top and bottom 5 CZs based on the movers data, while Table 2 provides analagous results based on the OA data. Results for all 50 CZs may be found in Tables 1 and 2 of the Online Appendix. Considering first the results in Table 1, we see that the conditional approach frequently generates very large changes in the point estimates and CIs.[5] While the hybrid approach also generates substantial changes, these are less extreme than those from the conditional approach. Turning next to the results in Table 2, based on the OA estimates, we see that the adjustments from both the conditional and hybrid approaches are far smaller, consistent with the smaller degree of noise in this setting.

[5] Indeed, some of the estimates suggested by the conditional approach are impossible under the model in Chetty and Hendren (2018), which assumes a constant average effect on income rank in adulthood for each year spent in a given neighborhood until age 23, implying that no place can have an absolute effect larger than $100/23 \approx 4.35$. We could censor estimates and CIs to reflect this or other economic constraints, but here report uncensored estimates to highlight the underlying behavior of the conditional approach.

REFERENCES

Andrews, I., T. Kitagawa, and A. McCloskey. 2020. "Inference on Winners." Working Paper.

Armstrong, Timothy B., Michal Kolesár, and Mikkel Plagborg-Møller. 2021. "Robust Empirical Bayes Confidence Intervals." Working Paper.

Chetty, R., and N. Hendren. 2018. "The Impacts of Neighborhoods on Intergenerational Mobility ii: County-Level Estimates." *Quarterly Journal of Economics*, 133: 1163–1228.

Chetty, R., J.N. Friedman, N. Hendren, M.R. Jones, and S.R. Porter. 2018. "The Opportunity Atlas: Mapping Childhood Roots of Social Mobility." NBER Working Paper 25147.

Gu, Jiaying, and Roger Koenker. 2020. "Invidious Comparisons: Ranking and Selection as Compound Decisions." Working Paper.

Lee, J. D., D. L. Sun, Y. Sun, and J. E. Taylor. 2016. "Exact Post-Selection Inference, with Application to the LASSO." *Annals of Statistics*, 44: 907–927.

Mogstad, M., J.P. Romano, A.M. Shaikh, and D. Wilhelm. 2021. "Inference for Ranks with Applications to Mobility Across Neighborhoods and Acadmic Achievement Across Countries." Working Paper.

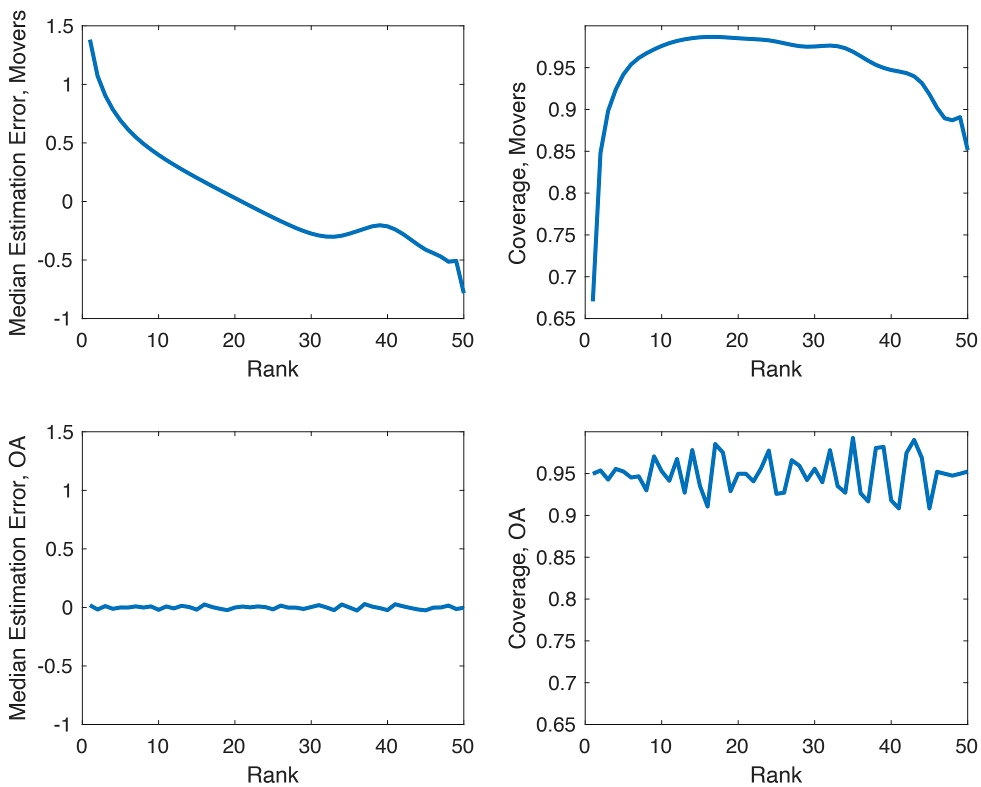

Figure 1. Median Bias and Coverage

Note: Median bias and coverage in simulations calibrated to movers results (from Chetty and Hendren, 2018) and OA results (from Chetty et al., 2018). Estimation errors are normalized by the standard deviation of true effects across CZs.

Rank	CZ	Naive Estimate	SE	$\hat{\mu}_{\frac{1}{2}}$	CI, L	CI, U	$\hat{\mu}^{H}_{\frac{1}{2}}$	CI^H, L	CI^H, U
1	Seattle	0.229	0.082	0.189	-0.163	0.385	0.189	-0.076	0.389
2	Washington DC	0.163	0.077	0.108	-0.407	0.502	0.108	-0.130	0.447
3	Cleveland	0.124	0.107	-2.528	-7.229	0.666	-0.284	-0.290	0.533
4	Fort Worth	0.121	0.090	1.175	-3.884	5.305	0.464	-0.230	0.470
5	Minneapolis	0.116	0.120	1.821	-3.341	5.677	0.571	-0.359	0.582
46	Charlotte	-0.248	0.096	-0.607	-2.499	0.498	-0.581	-0.632	0.162
47	Port St. Lucie	-0.263	0.090	-0.263	-2.245	1.719	-0.263	-0.624	0.098
48	Raleigh	-0.278	0.105	0.216	-0.605	2.436	0.093	-0.620	0.141
49	Fresno	-0.377	0.100	-0.858	-3.056	-0.085	-0.732	-0.775	-0.071
50	New Orleans	-0.391	0.111	0.206	-0.524	2.859	0.009	-0.532	0.050

Table 1—Estimates based on Chetty and Hendren (2018) data

Note: Naive and corrected estimates based on movers data. The column $\hat{\mu}_{\frac{1}{2}}$ reports the conditionally median-unbiased estimate for each CZ, conditioning on the CZ's rank, while columns CI, L and CI, U report the lower and upper endpoints for the conditional CI, respectively. Columns $\hat{\mu}^{H}_{\frac{1}{2}}$, CI^H, L, and CI^H, U similarly report estimates and CI endpoints based on the hybrid approach.

Rank	CZ	Naive Estimate	SE	$\hat{\mu}_{\frac{1}{2}}$	CI, L	CI, U	$\hat{\mu}^{H}_{\frac{1}{2}}$	CI^H, L	CI^H, U
1	San Francisco	0.456617	0.000816	0.444	0.422	0.457	0.454	0.453	0.457
2	Salt Lake City	0.456580	0.001297	0.488	0.457	0.527	0.462	0.457	0.462
3	Boston	0.453016	0.000907	0.453	0.450	0.455	0.453	0.450	0.455
4	Minneapolis	0.451732	0.001171	0.452	0.449	0.456	0.452	0.449	0.456
5	San Jose	0.448515	0.001040	0.449	0.446	0.451	0.449	0.446	0.451
46	Raleigh	0.368923	0.001296	0.369	0.366	0.372	0.369	0.366	0.372
47	Indianapolis	0.363872	0.001201	0.364	0.362	0.366	0.364	0.361	0.366
48	Jacksonville	0.358469	0.001047	0.354	0.332	0.360	0.355	0.354	0.360
49	Atlanta	0.358312	0.000684	0.360	0.357	0.369	0.360	0.357	0.361
50	Charlotte	0.355478	0.001087	0.355	0.353	0.358	0.355	0.353	0.358

Table 2—Estimates based on Chetty et al. (2018) data

Note: Naive and corrected estimates based on OA data. The column $\hat{\mu}_{\frac{1}{2}}$ reports the conditionally median-unbiased estimate for each CZ, conditioning on the CZ's rank, while columns CI, L and CI, U report the lower and upper endpoints for the conditional CI, respectively. Columns $\hat{\mu}^{H}_{\frac{1}{2}}$, CI^H, L, and CI^H, U similarly report estimates and CI endpoints based on the hybrid approach.

Online Appendix to the paper

Inference for Losers

Isaiah Andrews, Dillon Bowen, Toru Kitagawa, and Adam McCloskey

February 26, 2022

1. Proof of Proposition 1

Note that the event $\{\theta = \hat{\theta}_{(k)}\}$ is equivalent to the union of events

$$\bigcup_{(\tau^{\mathcal{L}},\tau^{\mathcal{U}})\in S_k(\theta)} \{M(\tau^{\mathcal{L}},\tau^{\mathcal{U}})X \leq 0\}, \tag{1}$$

where the $K \times K$ matrices $M(\tau^{\mathcal{L}},\tau^{\mathcal{U}})$ are defined as follows, supposing $\theta = \theta_i$:

1) for $j \neq i$ such that $\theta_j \in \tau^{\mathcal{U}}$, the j^{th} row of $M(\tau^{\mathcal{L}},\tau^{\mathcal{U}})$ is composed entirely of zeros except for a -1 in the j^{th} entry and a 1 in the i^{th} entry,
2) for $j \neq i$ such that $\theta_j \in \tau^{\mathcal{L}}$, the j^{th} row of $M(\tau^{\mathcal{L}},\tau^{\mathcal{U}})$ is composed entirely of zeros except for a 1 in the j^{th} entry and a -1 in the i^{th} entry, and
3) the i^{th} row of $M(\tau^{\mathcal{L}},\tau^{\mathcal{U}})$ is composed entirely of zeros.

For example, if $\tau^{\mathcal{L}} = \{\theta_j \in \Theta : j > i\}$ and $\tau^{\mathcal{U}} = \{\theta_j \in \Theta : j < i\}$, then

$$M(\tau^{\mathcal{L}},\tau^{\mathcal{U}}) = \begin{array}{c} \begin{array}{cccccccc} 1 & 2 & \dots & i & i+1 & i+2 & \dots & K \end{array} \\ \begin{pmatrix} -1 & 0 & \dots & 1 & 0 & 0 & \dots & 0 \\ 0 & -1 & \dots & 1 & 0 & 0 & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & 0 & 0 & 0 & \dots & 0 \\ 0 & 0 & \dots & -1 & 1 & 0 & \dots & 0 \\ 0 & 0 & \dots & -1 & 0 & 1 & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & -1 & 0 & 0 & \dots & 1 \end{pmatrix} \end{array} \begin{array}{c} \\ 1 \\ 2 \\ \vdots \\ i \\ i+1 \\ i+2 \\ \vdots \\ K. \end{array}$$

Define

$$A(\tau^{\mathcal{L}},\tau^{\mathcal{U}}) \coloneqq \begin{bmatrix} M(\tau^{\mathcal{L}},\tau^{\mathcal{U}}) & \mathbf{0} \end{bmatrix},$$

where $\mathbf{0}$ is a $K \times K$ matrix of zeros. Define

$$W \coloneqq (X^T, Y^T)^T$$

and note that the event $\{M(\tau^{\mathcal{L}},\tau^{\mathcal{U}})X \leq 0\}$ is equivalent to the event $\{A(\tau^{\mathcal{L}},\tau^{\mathcal{U}})W \leq 0\}$. Define

$$\tilde{Z}^*_\theta \coloneqq W - cY(\theta),$$

where $c = Cov(W, Y(\theta))/\Sigma_Y(\theta)$. By Lemma 5.1 of Lee et al. (2016), the event $\{A(\tau^{\mathcal{L}},\tau^{\mathcal{U}})W \leq 0\}$ is equivalent to the event

$$\left\{\mathcal{L}(\theta, \tilde{Z}^*_\theta, \tau^{\mathcal{L}},\tau^{\mathcal{U}}) \leq Y(\theta) \leq \mathcal{U}(\theta, \tilde{Z}^*_\theta, \tau^{\mathcal{L}},\tau^{\mathcal{U}}), 0 \leq \mathcal{V}(\theta, \tilde{Z}^*_\theta)\right\},$$

where

$$\mathcal{L}(\theta, z, \tau^{\mathcal{L}}, \tau^{\mathcal{U}}) \coloneqq \max_{j:(Ac)_j<0} \frac{-(Az)_j}{(Ac)_j},$$

$$\mathcal{U}(\theta, z, \tau^{\mathcal{L}}, \tau^{\mathcal{U}}) \coloneqq \min_{j:(Ac)_j>0} \frac{-(Az)_j}{(Ac)_j},$$

$$\mathcal{V}(\theta, z) \coloneqq \min_{j:(Ac)_j=0} -(Az)_j.$$

Note that

$$(A\tilde{Z}^*_{\theta_i})_j = \begin{cases} -(Z_\theta(\theta') - Z_\theta(\theta)) & \text{if } \theta' \in \tau^{\mathcal{U}} \\ Z_\theta(\theta') - Z_\theta(\theta) & \text{if } \theta' \in \tau^{\mathcal{L}} \\ 0 & \text{if } \theta' = \theta \end{cases} \tag{2}$$

and

$$(Ac)_j = \begin{cases} -\frac{\Sigma_{XY}(\theta',\theta) - \Sigma_{XY}(\theta)}{\Sigma_Y(\theta)} & \text{if } \theta' \in \tau^{\mathcal{U}} \\ \frac{\Sigma_{XY}(\theta',\theta) - \Sigma_{XY}(\theta)}{\Sigma_Y(\theta)} & \text{if } \theta' \in \tau^{\mathcal{L}} \\ 0 & \text{if } \theta' = \theta. \end{cases} \tag{3}$$

Thus,

$$\{\theta' : (Ac)_j < 0\} = \Big\{\theta' : \big(\theta' \in \tau^{\mathcal{L}} \text{ and } \Sigma_{XY}(\theta) > \Sigma_{XY}(\theta', \theta)\big) \text{ or } \big(\theta' \in \tau^{\mathcal{U}} \text{ and } \Sigma_{XY}(\theta) < \Sigma_{XY}(\theta', \theta)\big)\Big\}$$

and

$$\{\theta' : (Ac)_j > 0\} = \Big\{\theta' : \big(\theta' \in \tau^{\mathcal{L}} \text{ and } \Sigma_{XY}(\theta) < \Sigma_{XY}(\theta', \theta)\big) \text{ or } \big(\theta' \in \tau^{\mathcal{U}} \text{ and } \Sigma_{XY}(\theta) > \Sigma_{XY}(\theta', \theta)\big)\Big\}.$$

Define

$$S^{\mathcal{L}}_\theta(\tau^{\mathcal{L}}, \tau^{\mathcal{U}}) \coloneqq \{\theta' : (Ac)_j > 0\},$$

$$S^{\mathcal{U}}_\theta(\tau^{\mathcal{L}}, \tau^{\mathcal{U}}) \coloneqq \{\theta' : (Ac)_j < 0\},$$

$$S^{\mathcal{V}}_\theta \coloneqq \{\theta' : (Ac)_j) = 0\}$$

Equations (2) and (3) imply

$$\frac{-(A\tilde{Z}^*_{\theta_i})_j}{(Ac)_j} = Q_\theta(\theta').$$

This allows us to rewrite

$$\mathcal{L}(\theta, \tilde{Z}^*_{\theta_i}, \tau^{\mathcal{L}}, \tau^{\mathcal{U}}) \coloneqq \max_{\theta' \in S^{\mathcal{U}}_\theta(\tau^{\mathcal{L}}, \tau^{\mathcal{U}})} Q_\theta(\theta'),$$

$$\mathcal{U}(\theta, \tilde{Z}^*_{\theta_i}, \tau^{\mathcal{L}}, \tau^{\mathcal{U}}) \coloneqq \min_{\theta' \in S^{\mathcal{L}}_\theta(\tau^{\mathcal{L}}, \tau^{\mathcal{U}})} Q_\theta(\theta')$$

and

$$\mathcal{V}(\theta, \tilde{Z}^*_{\theta_i}) \coloneqq \min_{j \in S^{\mathcal{V}}_\theta} -\Big(Z_\theta(\theta') - Z_\theta(\theta)\Big).$$

Thus by (1), we have established that the event $\{\theta = \hat{\theta}_{(k)}\}$ is equivalent to the union of events

$$\bigcup_{(\tau^{\mathcal{L}},\tau^{\mathcal{U}})\in S_k(\theta)} \left\{\mathcal{L}(\theta,\tilde{Z}^*_\theta,\tau^{\mathcal{L}},\tau^{\mathcal{U}}) \le Y(\theta) \le \mathcal{U}(\theta,\tilde{Z}^*_\theta,\tau^{\mathcal{L}},\tau^{\mathcal{U}}), 0 \le \mathcal{V}(\theta,\tilde{Z}^*_\theta)\right\}.$$

Note that the event $\{\theta \in \widehat{R}\}$ is equivalent to the union of events $\cup_{k\in R}\{\theta = \hat{\theta}_{(k)}\}$ so that when $\mathcal{V}(\theta,z) \ge 0$,

$$\mathcal{Y}_R(\theta,z) = \bigcup_{(\tau^{\mathcal{L}},\tau^{\mathcal{U}})\in \mathcal{S}'_P(\theta)} \left[\mathcal{L}(\theta,z,\tau^{\mathcal{L}},\tau^{\mathcal{U}}), \mathcal{U}(\theta,z,\tau^{\mathcal{L}},\tau^{\mathcal{U}})\right],$$

where

$$\mathcal{S}'_P(\theta) := \bigcup_{k\in R} \mathcal{S}_k(\theta).$$

Finally, note that

$$\begin{aligned}&\bigcup_{(\tau^{\mathcal{L}},\tau^{\mathcal{U}})\in \mathcal{S}'_P(\theta)} \left[\mathcal{L}(\theta,z,\tau^{\mathcal{L}},\tau^{\mathcal{U}}), \mathcal{U}(\theta,z,\tau^{\mathcal{L}},\tau^{\mathcal{U}})\right] \\ &= \bigcup_{(\tau^{\mathcal{L}},\tau^{\mathcal{U}})\in \mathcal{S}_P(\theta)} \left[\mathcal{L}(\theta,z,\tau^{\mathcal{L}},\tau^{\mathcal{U}}), \mathcal{U}(\theta,z,\tau^{\mathcal{L}},\tau^{\mathcal{U}})\right],\end{aligned}$$

where

$$\begin{aligned}\mathcal{S}_P(\theta) := \{(\tau^{\mathcal{L}},\tau^{\mathcal{U}}) \in \bigcup_{k\in R} \mathcal{S}_k(\theta) : Q_\theta(\theta') \le Q_\theta(\theta'')\ \forall \theta' \in S^{\mathcal{L}}_\theta(\tau^{\mathcal{L}},\tau^{\mathcal{U}}), \\ \theta'' \in S^{\mathcal{U}}_\theta(\tau^{\mathcal{L}},\tau^{\mathcal{U}})\},\end{aligned}$$

since if $Q_\theta(\theta') > Q_\theta(\theta'')$ for some $\theta' \in S^{\mathcal{L}}_\theta(\tau^{\mathcal{L}},\tau^{\mathcal{U}}), \theta'' \in S^{\mathcal{U}}_\theta(\tau^{\mathcal{L}},\tau^{\mathcal{U}})$, then

$$\mathcal{L}(\theta,z,\tau^{\mathcal{L}},\tau^{\mathcal{U}}) \ge Q_\theta(\theta') > Q_\theta(\theta'') \ge \mathcal{U}(\theta,z,\tau^{\mathcal{L}},\tau^{\mathcal{U}}).$$

This yields the statement of the proposition. ■

2. Proof of Proposition 2

Note that for

$$CI^\beta_{P,\theta} = \left[Y(\theta) - c_\beta\sqrt{\Sigma_Y(\theta)}, Y(\theta) + c_\beta\sqrt{\Sigma_Y(\theta)}\right]$$

the level β projection interval for $\mu_Y(\theta)$,

$$\begin{aligned}&\sum_{k\in R} Pr_\mu\left\{\hat{\mu}^H_{\alpha,k} \ge \mu_Y\left(\hat{\theta}_{(k)}\right)\right\} \\ &= \sum_{k\in R}\sum_{\theta\in\Theta} Pr_\mu\left\{\theta = \hat{\theta}_{(k)}, \hat{\mu}^H_{\alpha,k} \ge \mu_Y(\theta)\right\} \\ &= \sum_{\theta\in\Theta} E_\mu\left[\sum_{k\in R} 1\left\{\theta = \hat{\theta}_{(k)}, \hat{\mu}^H_{\alpha,k} \ge \mu_Y(\theta)\right\}\right] \\ &= \sum_{\theta\in\Theta} E_\mu\left[1\left\{\theta \in \widehat{R}, \hat{\mu}^H_{\alpha,\theta} \ge \mu_Y(\theta)\right\}\right] \\ &= \sum_{\theta\in\Theta} Pr_\mu\left\{\theta \in \widehat{R}, \hat{\mu}^H_{\alpha,\theta} \ge \mu_Y(\theta)\right\}\end{aligned}$$

$$= \begin{array}{c} \sum_{\theta\in\Theta} Pr_{\mu}\left\{\hat{\mu}^{H}_{\alpha,\theta} \geq \mu_Y(\theta) \,|\theta \in \widehat{R}, \mu_Y(\theta) \in CI^{\beta}_{P,\theta}\right\} Pr_{\mu}\left\{\theta \in \widehat{R}, \mu_Y(\theta) \in CI^{\beta}_{P,\theta}\right\} + \\ \sum_{\theta\in\Theta} Pr_{\mu}\left\{\hat{\mu}^{H}_{\alpha,\theta} \geq \mu_Y(\theta) \,|\theta \in \widehat{R}, \mu_Y(\theta) \notin CI^{\beta}_{P,\theta}\right\} Pr_{\mu}\left\{\theta \in \widehat{R}, \mu_Y(\theta) \notin CI^{\beta}_{P,\theta}\right\}, \end{array}$$

where we write $\hat{\mu}^{H}_{\alpha,\theta}$ to denote the hybrid estimator of $\mu_Y(\theta)$ for $\theta \in \widehat{R}$,

$$\hat{\mu}^{H}_{\alpha,\theta} = \sum_{k\in R} 1\left\{\hat{\theta}_{(k)} = \theta\right\} \hat{\mu}^{H}_{\alpha,k}.$$

Proposition 7 of AKM implies, however, that

$$Pr_{\mu}\left\{\hat{\mu}^{H}_{\alpha,\theta} \geq \mu_Y(\theta) \,|\theta \in \widehat{R}, \mu_Y(\theta) \in CI^{\beta}_{P,\theta}\right\} = \alpha,$$

so

$$\left|\sum_{k\in R}\left(Pr_{\mu}\left\{\hat{\mu}^{H}_{\alpha,k} \geq \mu_Y\left(\hat{\theta}_{(k)}\right)\right\} - \alpha\right)\right|$$

$$= \left| \begin{array}{c} \sum_{\theta\in\Theta}\left(Pr_{\mu}\left\{\hat{\mu}^{H}_{\alpha,\theta} \geq \mu_Y(\theta) \,|\theta \in \widehat{R}, \mu_Y(\theta) \in CI^{\beta}_{P,\theta}\right\} - \alpha\right) Pr_{\mu}\left\{\theta \in \widehat{R}, \mu_Y(\theta) \in CI^{\beta}_{P,\theta}\right\} \\ +\sum_{\theta\in\Theta}\left(Pr_{\mu}\left\{\hat{\mu}^{H}_{\alpha,\theta} \geq \mu_Y(\theta) \,|\theta \in \widehat{R}, \mu_Y(\theta) \notin CI^{\beta}_{P,\theta}\right\} - \alpha\right) Pr_{\mu}\left\{\theta \in \widehat{R}, \mu_Y(\theta) \notin CI^{\beta}_{P,\theta}\right\} \end{array} \right|$$

$$= \left|\sum_{\theta\in\Theta}\left(Pr_{\mu}\left\{\hat{\mu}^{H}_{\alpha,\theta} \geq \mu_Y(\theta) \,|\theta \in \widehat{R}, \mu_Y(\theta) \notin CI^{\beta}_{P,\theta}\right\} - \alpha\right) Pr_{\mu}\left\{\theta \in \widehat{R}, \mu_Y(\theta) \notin CI^{\beta}_{P,\theta}\right\}\right|$$

$$\leq \sum_{\theta\in\Theta}\left|Pr_{\mu}\left\{\hat{\mu}^{H}_{\alpha,\theta} \geq \mu_Y(\theta) \,|\theta \in \widehat{R}, \mu_Y(\theta) \notin CI^{\beta}_{P,\theta}\right\} - \alpha\right| Pr_{\mu}\left\{\theta \in \widehat{R}, \mu_Y(\theta) \notin CI^{\beta}_{P,\theta}\right\}$$

$$\leq |R| \cdot \max\{\alpha, 1-\alpha\} \cdot \beta.$$

A parallel argument establishes the coverage statement. ■

3. Efficient Computation of the Truncation Set

The number of parameter arrangements that characterize the truncation set in Proposition 1 is $|\mathcal{S}_P(\theta)| = \sum_{k\in R} C_{k-1}(K-1)$ where C_k is the "choose k element" operator. Even assuming we can compute the truncation set for each arrangement $(\tau^{\mathcal{L}}, \tau^{\mathcal{U}}) \in \mathcal{S}_P(\theta)$ in constant time, exhaustive search will find the truncation set $\mathcal{Y}_R(\theta, z)$ in exponential time. This is infeasible when considering a large parameter space.

Here we describe an algorithm to compute the truncation set in $O(K \log K)$ time. We begin with the following observation.

OBSERVATION 1: *Suppose $\Sigma_{XY}(\theta', \theta) \neq \Sigma_{XY}(\theta)$ $\forall \theta' \in \Theta_{-\theta}$. Then for all $\tau^{\mathcal{L}}, \tau^{\mathcal{U}}$ and $\tilde{\tau}^{\mathcal{L}}, \tilde{\tau}^{\mathcal{U}}$*

$$S^{\mathcal{L}}_{\theta}(\tau^{\mathcal{L}}, \tau^{\mathcal{U}}) = S^{\mathcal{L}}_{\theta}(\tilde{\tau}^{\mathcal{L}}, \tilde{\tau}^{\mathcal{U}}) \iff (\tau^{\mathcal{L}}, \tau^{\mathcal{U}}) = (\tilde{\tau}^{\mathcal{L}}, \tilde{\tau}^{\mathcal{U}})$$

and

$$S^{\mathcal{U}}_{\theta}(\tau^{\mathcal{L}}, \tau^{\mathcal{U}}) = S^{\mathcal{U}}_{\theta}(\tilde{\tau}^{\mathcal{L}}, \tilde{\tau}^{\mathcal{U}}) \iff (\tau^{\mathcal{L}}, \tau^{\mathcal{U}}) = (\tilde{\tau}^{\mathcal{L}}, \tilde{\tau}^{\mathcal{U}}).$$

That is, we can construct a unique $\tau^{\mathcal{L}}$ and $\tau^{\mathcal{U}}$ given $S^{\mathcal{L}}_{\theta}(\tau^{\mathcal{L}}, \tau^{\mathcal{U}})$ and $S^{\mathcal{U}}_{\theta}(\tau^{\mathcal{L}}, \tau^{\mathcal{U}})$. For example, if $\theta' \in S^{\mathcal{L}}_{\theta}(\tau^{\mathcal{L}}, \tau^{\mathcal{U}})$ and $\Sigma_{XY}(\theta) > \Sigma_{XY}(\theta', \theta)$, then $\theta' \in \tau^{\mathcal{L}}$. Moreover, we can construct $(\tau^{\mathcal{L}}, \tau^{\mathcal{U}})$ in $O(K)$ (i.e., linear) time.

This observation suggests a $O(K \log K)$ algorithm for computing the truncation set as follows:

1) Begin by setting $\mathcal{Y}_R(\theta, z)$ to $\emptyset$ and sorting $\Theta_{-\theta} := \{\tilde{\theta}_1, \ldots, \tilde{\theta}_{K-1}\}$ such that $Q_\theta(\tilde{\theta}_1) \geq \ldots \geq Q_\theta(\tilde{\theta}_{K-1})$. Sorting takes $O(K \log K)$ time.

2) Assume $S_\theta^{\mathcal{L}}(\tau^{\mathcal{L}}, \tau^{\mathcal{U}}) = \Theta_{-\theta}$ and $S_\theta^{\mathcal{U}}(\tau^{\mathcal{L}}, \tau^{\mathcal{U}}) = \emptyset$ and construct $\tau^{\mathcal{L}}$ and $\tau^{\mathcal{U}}$ to satisfy this assumption. By Observation 1, these $\tau^{\mathcal{L}}$ and $\tau^{\mathcal{U}}$ are unique and can be constructed in linear time.

3) Step through $\tilde{\theta}_1, \ldots, \tilde{\theta}_{K-1}$, removing each parameter $\tilde{\theta}'$ from $S_\theta^{\mathcal{L}}(\tau^{\mathcal{L}}, \tau^{\mathcal{U}})$ and adding it to $S_\theta^{\mathcal{U}}(\tau^{\mathcal{L}}, \tau^{\mathcal{U}})$. Observation 1 implies that this step also switches the parameter from $\tau^{\mathcal{L}}$ to $\tau^{\mathcal{U}}$ or vice versa. This guarantees we consider all $S_\theta^{\mathcal{L}}(\tau^{\mathcal{L}}, \tau^{\mathcal{U}}) \subseteq \Theta_{-\theta}$ and $S_\theta^{\mathcal{U}}(\tau^{\mathcal{L}}, \tau^{\mathcal{U}}) \subseteq \Theta_{-\theta}$ satisfying

$$Q_\theta(\theta') \leq Q_\theta(\theta'') \; \forall \theta' \in S_\theta^{\mathcal{L}}(\tau^{\mathcal{L}}, \tau^{\mathcal{U}}), \theta'' \in S_\theta^{\mathcal{U}}(\tau^{\mathcal{L}}, \tau^{\mathcal{U}})$$

and

$$S_\theta^{\mathcal{L}}(\tau^{\mathcal{L}}, \tau^{\mathcal{U}}) \cap S_\theta^{\mathcal{U}}(\tau^{\mathcal{L}}, \tau^{\mathcal{U}}) = \emptyset.$$

Thus, $\mathcal{S}_P(\theta)$ is contained in the set of arrangements we consider in this stepwise procedure.

Because we step through $\tilde{\theta}_1, \ldots, \tilde{\theta}_{K-1}$, this step takes linear time.

4) After each step in (3), check if $\exists k \in R$ such that $K - |\tau^{\mathcal{L}}| \geq k \geq |\tau^{\mathcal{U}}| + 1$, implying $(\tau^{\mathcal{L}}, \tau^{\mathcal{U}}) \in \mathcal{S}_P(\theta)$. If so, add $[Q(\tilde{\theta}_{j+1}), Q(\tilde{\theta}_j)]$ to $\mathcal{Y}_R(\theta, z)$.

Assuming ties are rare (i.e., $|S_\theta^{\mathcal{V}}|$ is small), we can check if $k \in R$ in constant time by first creating an array $A = [\mathbf{1}\{1 \in R\}, \ldots, \mathbf{1}\{K-1 \in R\}]^T$ in linear time then using $k \in R \iff A[k]$.

After the initial sorting in $O(K \log K)$ time, the rest of the algorithm takes linear time. Therefore, the entire algorithm takes $O(K \log K)$ time.

We formalize the algorithm as follows.

Algorithm 1: Efficient computation of the truncation set

Result: The truncation set $\mathcal{Y}_R(\theta, z)$

$\mathcal{Y}_R(\theta, z) \leftarrow \emptyset$;

Sort $\Theta_{-\theta}$ such that $Q(\tilde{\theta}_1) \geq \ldots \geq Q(\tilde{\theta}_{K-1})$;

Construct $\tau^{\mathcal{L}}, \tau^{\mathcal{U}}$ such that $S_\theta^{\mathcal{L}}(\tau^{\mathcal{L}}, \tau^{\mathcal{U}}) = \Theta_{-\theta}$ and $S_\theta^{\mathcal{U}}(\tau^{\mathcal{L}}, \tau^{\mathcal{U}}) = \emptyset$;

if $\exists k \in R : |\tau^{\mathcal{L}}| = K - k, |\tau^{\mathcal{U}}| = k - 1$ **then**

  $\mathcal{Y}_R(\theta, z) \leftarrow \mathcal{Y}_R(\theta, z) \cup [Q_\theta(\tilde{\theta}_1), \infty)$;

end

for $\tilde{\theta}_j \in \Theta_{-\theta}$ **do**

  if $\tilde{\theta}_j \in \tau^{\mathcal{L}}$ **then**

    $\tau^{\mathcal{L}} \leftarrow \tau^{\mathcal{L}} \setminus \{\tilde{\theta}_j\}$;

    $\tau^{\mathcal{U}} \leftarrow \tau^{\mathcal{U}} \cup \{\tilde{\theta}_j\}$;

  else

    $\tau^{\mathcal{L}} \leftarrow \tau^{\mathcal{L}} \cup \{\tilde{\theta}_j\}$;

    $\tau^{\mathcal{U}} \leftarrow \tau^{\mathcal{U}} \setminus \{\tilde{\theta}_j\}$;

  end

  if $\exists k \in R : |\tau^{\mathcal{L}}| = K - k, |\tau^{\mathcal{U}}| = k - 1$ **then**

    if $j < K - 1$ **then**

      $\mathcal{Y}_R(\theta, z) \leftarrow \mathcal{Y}_R(\theta, z) \cup [Q_\theta(\tilde{\theta}_{j+1}), Q_\theta(\tilde{\theta}_j)]$;

    else

      $\mathcal{Y}_R(\theta, z) \leftarrow \mathcal{Y}_R(\theta, z) \cup (-\infty, Q_\theta(\tilde{\theta}_{K-1})]$;

    end

  end

end

4. Additional Simulation Results

Figure 1 reports simulation results for conventional estimators and CIs, along with our corrected procedures, in the simulations calibrated to Chetty and Hendren (2018) and Chetty et al. (2018), based on 5,000 simulation draws.

5. Full Empirical Results

Tables 1 and 2 report naive and corrected estimates and CIs for all 50 CZs.

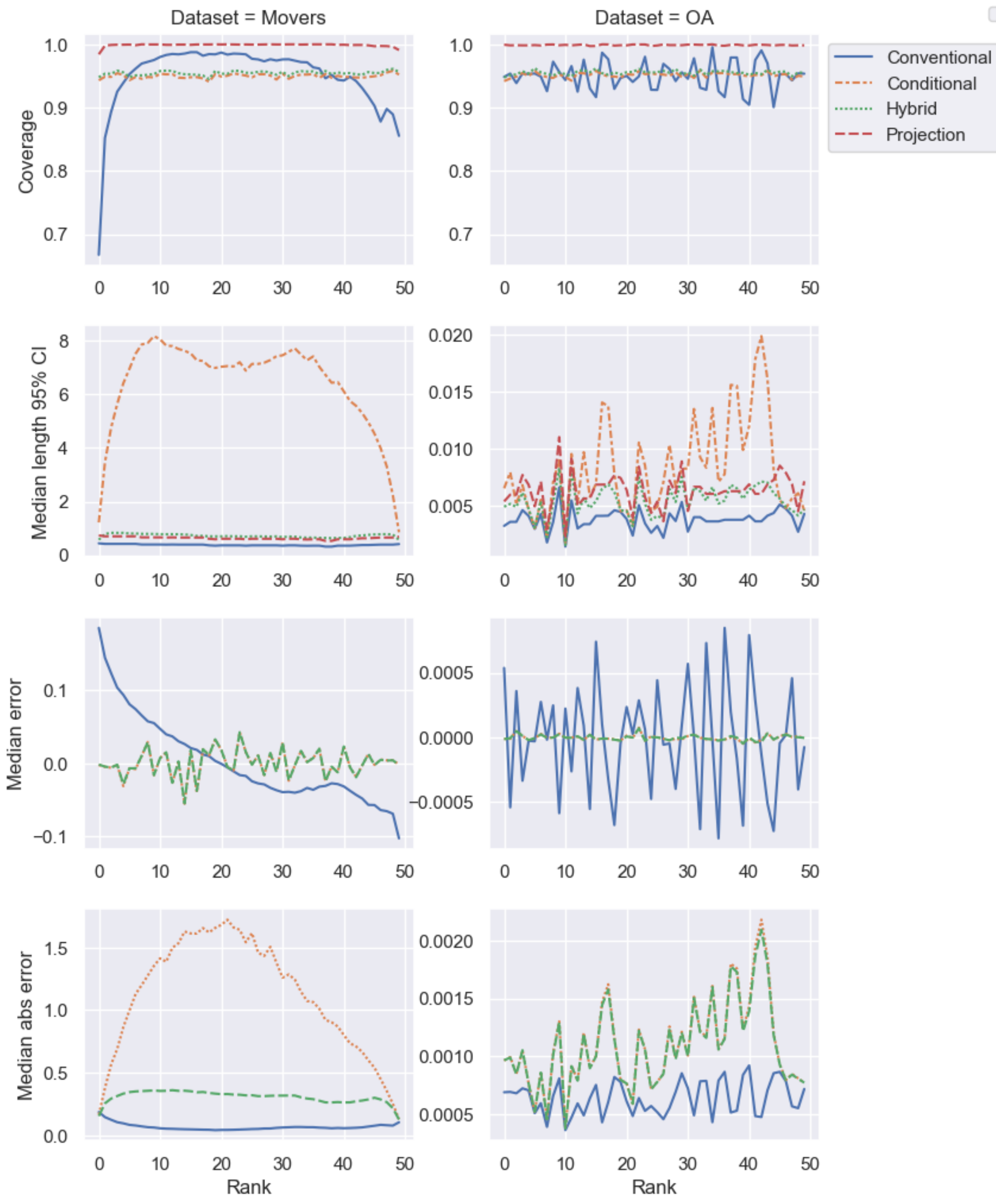

Figure 1. Simulated performance of estimators and CIs

Note: Estimator and CI performance for simulations calibrated to Chetty and Hendren, 2018 (first column) and Chetty et al., 2018 (second column). The first row shows the simulated coverage of nominal 95% CIs, while the second shows the median estimation error. The third row reports the median length of nominal 95% CIs, while the fourth row plots the median absolute estimation error of the estimators considered.

Rank	CZ	Naive Estimate	SE	$\hat{\mu}_{\frac{1}{2}}$	CI, L	CI, U	$\hat{\mu}^H_{\frac{1}{2}}$	CI^H, L	CI^H, U
1	Seattle	0.229	0.082	0.189	-0.163	0.385	0.189	-0.076	0.389
2	Washington DC	0.163	0.077	0.108	-0.407	0.502	0.108	-0.130	0.447
3	Cleveland	0.124	0.107	-2.528	-7.229	0.666	-0.284	-0.290	0.533
4	Fort Worth	0.121	0.090	1.175	-3.884	5.305	0.464	-0.230	0.470
5	Minneapolis	0.116	0.120	1.821	-3.341	5.677	0.571	-0.359	0.582
6	Portland	0.102	0.114	-1.979	-6.182	3.066	-0.332	-0.340	0.554
7	Dayton	0.098	0.152	3.340	-6.863	9.898	0.678	-0.496	0.686
8	Las Vegas	0.088	0.071	-0.623	-2.859	1.804	-0.180	-0.189	0.371
9	Buffalo	0.084	0.112	2.159	-2.180	6.154	0.510	-0.364	0.519
10	Kansas City	0.067	0.126	-2.575	-6.633	2.909	-0.413	-0.421	0.568
11	Salt Lake City	0.063	0.129	2.253	-5.547	6.363	0.554	-0.442	0.563
12	San Diego	0.054	0.080	-0.793	-3.029	2.510	-0.249	-0.258	0.371
13	Jacksonville	0.050	0.100	0.548	-4.874	4.825	0.430	-0.339	0.438
14	Milwaukee	0.045	0.135	-2.343	-11.619	7.979	-0.471	-0.477	0.569
15	Denver	0.042	0.101	0.884	-4.158	4.242	0.427	-0.350	0.433
16	Boston	0.038	0.089	0.988	-3.315	3.838	0.375	-0.312	0.384
17	Cincinnati	0.030	0.130	-0.669	-5.722	6.061	-0.463	-0.476	0.538
18	Grand Rapids	0.024	0.155	-5.905	-7.176	7.224	-0.570	-0.574	0.626
19	Philadelphia	0.022	0.071	1.436	-3.016	4.679	0.293	-0.256	0.297
20	San Francisco	0.017	0.083	0.612	-2.708	3.630	0.330	-0.311	0.341
21	Newark	0.008	0.064	-0.049	-1.872	1.671	-0.049	-0.246	0.263
22	Phoenix	0.000	0.065	0.142	-1.396	1.947	0.142	-0.261	0.258
23	San Jose	-0.011	0.101	-3.891	-7.711	1.863	-0.398	-0.400	0.388
24	Sacramento	-0.012	0.081	3.098	-3.216	3.588	0.298	-0.327	0.300
25	Columbus	-0.016	0.122	0.725	-4.816	4.784	0.449	-0.490	0.457
26	Miami	-0.021	0.063	-0.218	-3.041	2.401	-0.218	-0.267	0.226
27	St. Louis	-0.025	0.120	-3.504	-7.525	7.166	-0.480	-0.488	0.440
28	Dallas	-0.027	0.073	1.796	-1.248	4.517	0.251	-0.320	0.256
29	Houston	-0.039	0.069	-1.676	-3.939	1.044	-0.302	-0.306	0.238
30	Nashville	-0.041	0.129	4.901	-3.845	12.259	0.453	-0.549	0.457
31	Providence	-0.053	0.130	-3.767	-5.353	4.298	-0.550	-0.556	0.459
32	Pittsburgh	-0.056	0.125	-0.056	-5.656	5.544	-0.056	-0.540	0.428
33	Indianapolis	-0.059	0.129	3.315	-5.538	5.841	0.434	-0.564	0.440
34	Baltimore	-0.068	0.100	0.574	-1.486	3.778	0.303	-0.476	0.325
35	Detroit	-0.092	0.081	-3.197	-9.292	0.155	-0.402	-0.404	0.202
36	Atlanta	-0.093	0.055	1.998	-0.865	5.777	0.117	-0.313	0.119
37	Manchester	-0.102	0.139	0.550	-5.445	6.337	0.421	-0.649	0.441
38	Bridgeport	-0.115	0.093	0.164	-1.462	2.342	0.164	-0.496	0.256
39	Tampa	-0.138	0.068	-0.383	-1.832	0.561	-0.377	-0.409	0.146
40	New York	-0.148	0.047	-0.036	-0.502	0.669	-0.036	-0.339	0.043
41	Los Angeles	-0.170	0.040	-1.282	-3.215	-0.089	-0.323	-0.325	-0.076
42	Austin	-0.171	0.115	4.240	-3.533	16.929	0.270	-0.622	0.272
43	Chicago	-0.180	0.059	-0.180	-1.599	1.239	-0.180	-0.416	0.056
44	Orlando	-0.189	0.069	0.050	-1.186	1.764	0.050	-0.471	0.085
45	San Antonio	-0.206	0.098	0.104	-0.987	1.882	0.104	-0.625	0.188
46	Charlotte	-0.248	0.096	-0.607	-2.499	0.498	-0.581	-0.632	0.162
47	Port St. Lucie	-0.263	0.090	-0.263	-2.245	1.719	-0.263	-0.624	0.098
48	Raleigh	-0.278	0.105	0.216	-0.605	2.436	0.093	-0.620	0.141
49	Fresno	-0.377	0.100	-0.858	-3.056	-0.085	-0.732	-0.775	-0.071
50	New Orleans	-0.391	0.111	0.206	-0.524	2.859	0.009	-0.532	0.050

TABLE 1—ESTIMATES BASED ON CHETTY AND HENDREN (2018) DATA

Note: Naive and corrected estimates based on movers data. The column $\hat{\mu}_{\frac{1}{2}}$ reports the conditionally median-unbiased estimate for each CZ, conditioning on the CZ's rank, while columns CI, L and CI, U report the lower and upper endpoints for the conditional CI, respectively. Columns $\hat{\mu}^H_{\frac{1}{2}}$, CI^H, L, and CI^H, U similarly report estimates and CI endpoints based on the hybrid approach.

Rank	CZ	Naive Estimate	SE	$\hat{\mu}_{\frac{1}{2}}$	CI, L	CI, U	$\hat{\mu}^H_{\frac{1}{2}}$	CI^H, L	CI^H, U
1	San Francisco	0.456617	0.000816	0.444	0.422	0.457	0.454	0.453	0.457
2	Salt Lake City	0.456580	0.001297	0.488	0.457	0.527	0.462	0.457	0.462
3	Boston	0.453016	0.000907	0.453	0.450	0.455	0.453	0.450	0.455
4	Minneapolis	0.451732	0.001171	0.452	0.449	0.456	0.452	0.449	0.456
5	San Jose	0.448515	0.001040	0.449	0.446	0.451	0.449	0.446	0.451
6	Newark	0.445783	0.000750	0.446	0.444	0.447	0.446	0.444	0.447
7	Pittsburgh	0.441542	0.001082	0.441	0.438	0.444	0.441	0.438	0.444
8	New York	0.440292	0.000445	0.440	0.439	0.441	0.440	0.439	0.441
9	Seattle	0.438663	0.000903	0.438	0.435	0.441	0.438	0.436	0.441
10	Manchester	0.437685	0.001678	0.439	0.435	0.449	0.439	0.434	0.444
11	Los Angeles	0.430816	0.000352	0.431	0.430	0.432	0.431	0.430	0.432
12	Providence	0.430099	0.001387	0.432	0.427	0.440	0.432	0.427	0.436
13	Washington DC	0.427107	0.000758	0.427	0.425	0.429	0.427	0.425	0.429
14	Sacramento	0.426089	0.000925	0.424	0.416	0.429	0.424	0.422	0.429
15	San Diego	0.425785	0.000856	0.427	0.424	0.435	0.427	0.424	0.429
16	Houston	0.421016	0.000605	0.416	0.398	0.422	0.419	0.419	0.422
17	Denver	0.420967	0.001042	0.436	0.416	0.464	0.425	0.417	0.425
18	Bridgeport	0.420453	0.001045	0.421	0.417	0.428	0.421	0.417	0.424
19	Portland	0.419309	0.001166	0.420	0.417	0.424	0.420	0.417	0.424
20	Buffalo	0.415368	0.001131	0.415	0.413	0.418	0.415	0.413	0.418
21	Fort Worth	0.406522	0.000972	0.406	0.404	0.408	0.406	0.404	0.408
22	Miami	0.405049	0.000599	0.405	0.404	0.406	0.405	0.404	0.406
23	Austin	0.403423	0.001282	0.403	0.398	0.408	0.403	0.399	0.408
24	San Antonio	0.402263	0.000873	0.402	0.397	0.405	0.402	0.399	0.405
25	Philadelphia	0.401706	0.000661	0.402	0.400	0.405	0.402	0.400	0.404
26	Phoenix	0.397362	0.000814	0.397	0.392	0.399	0.397	0.394	0.399
27	Chicago	0.396876	0.000549	0.397	0.396	0.399	0.397	0.396	0.399
28	Kansas City	0.395848	0.001097	0.396	0.391	0.400	0.396	0.392	0.400
29	Fresno	0.394903	0.000925	0.395	0.393	0.398	0.395	0.393	0.398
30	Grand Rapids	0.392111	0.001353	0.392	0.389	0.395	0.392	0.389	0.395
31	Dallas	0.389610	0.000687	0.389	0.387	0.391	0.389	0.387	0.391
32	Milwaukee	0.388879	0.001242	0.364	0.321	0.392	0.384	0.384	0.392
33	Las Vegas	0.388836	0.001012	0.405	0.389	0.435	0.393	0.388	0.393
34	Orlando	0.385193	0.000903	0.379	0.353	0.386	0.382	0.382	0.386
35	Cleveland	0.385097	0.000917	0.382	0.350	0.417	0.382	0.381	0.389
36	Port St. Lucie	0.385023	0.001035	0.395	0.384	0.430	0.389	0.384	0.389
37	St. Louis	0.378417	0.000943	0.357	0.331	0.378	0.375	0.375	0.378
38	Nashville	0.378389	0.001202	0.413	0.378	0.473	0.383	0.377	0.383
39	New Orleans	0.377470	0.000954	0.374	0.356	0.381	0.374	0.374	0.381
40	Tampa	0.377310	0.000804	0.380	0.376	0.392	0.380	0.376	0.381
41	Dayton	0.374576	0.001244	0.370	0.350	0.377	0.371	0.370	0.377
42	Detroit	0.374336	0.000708	0.375	0.368	0.382	0.375	0.371	0.377
43	Cincinnati	0.374056	0.001046	0.375	0.365	0.389	0.375	0.370	0.378
44	Baltimore	0.373670	0.000918	0.375	0.371	0.382	0.375	0.371	0.378
45	Columbus	0.372527	0.001104	0.373	0.370	0.377	0.373	0.370	0.376
46	Raleigh	0.368923	0.001296	0.369	0.366	0.372	0.369	0.366	0.372
47	Indianapolis	0.363872	0.001201	0.364	0.362	0.366	0.364	0.361	0.366
48	Jacksonville	0.358469	0.001047	0.354	0.332	0.360	0.355	0.354	0.360
49	Atlanta	0.358312	0.000684	0.360	0.357	0.369	0.360	0.357	0.361
50	Charlotte	0.355478	0.001087	0.355	0.353	0.358	0.355	0.353	0.358

Table 2—Estimates based on Chetty et al. (2018) data

Note: Naive and corrected estimates based on OA data. The column $\hat{\mu}_{\frac{1}{2}}$ reports the conditionally median-unbiased estimate for each CZ, conditioning on the CZ's rank, while columns CI, L and CI, U report the lower and upper endpoints for the conditional CI, respectively. Columns $\hat{\mu}^H_{\frac{1}{2}}$, CI^H, L, and CI^H, U similarly report estimates and CI endpoints based on the hybrid approach.