Manuscript submitted to *The Econometrics Journal*, pp. S1–S7.

Online supplement: Robust IV Inference with Clustering Dependenc

Jianfei Cao[†]

[†]*Department of Economics, Northeastern University, Boston, MA 02115, United States.*
E-mail: j.cao@northeastern.edu

S1. MONTE CARLO SIMULATION

In this section, we study the finite-sample performance of the proposed estimator. In all the following settings, the data generating process follows the linear IV model (3.1), where $n = 900$ and $G = 30$ such that $n/G^2 = 1$, which deviates from the usual asymptotics. The null hypothesis is $H_0 : \beta = 0$. For each setting, 1,000 replications are conducted to calculate the empirical rejection rate.

For each setting, we observe $\{(X_i, Y_i, Z_i)\}_{i=1}^n$ and a partition $\{I_g\}_{g=1}^G$ of $\{i\}_{i=1}^n$. Let consecutive observations belong to the same group; that is, $I_1 = \{1, 2, \ldots, |I_1|\}$, $I_2 = \{|I_1| + 1, \ldots, |I_1| + |I_2|\}$, etc., where $|\cdot|$ is cardinality. The data are drawn according to the following process:

$$Y_i = Z_i\pi\beta + U_i$$

$$X_i = Z_i\pi + V_i$$

$$\begin{pmatrix} U_i \\ V_i \end{pmatrix} \sim N\left(0, \begin{bmatrix} 1 & 0.5 \\ 0.5 & 1 \end{bmatrix}\right), \text{ if } i = 1 + \sum_{h=1}^{g} |I_h| \text{ for some } g = 0, 1, \ldots, G-1$$

$$\begin{pmatrix} U_i \\ V_i \end{pmatrix} = 0.5 \begin{pmatrix} U_{i-1} \\ V_{i-1} \end{pmatrix} + \sqrt{1 - 0.5^2} \begin{pmatrix} \varepsilon_i^U \\ \varepsilon_i^V \end{pmatrix}, \text{ if } i \neq 1 + \sum_{h=1}^{g} |I_h| \text{ for any } g = 0, 1, \ldots, G-1$$

$$\begin{pmatrix} \varepsilon_i^U \\ \varepsilon_i^V \end{pmatrix} \sim N\left(0, \begin{bmatrix} 1 & 0.5 \\ 0.5 & 1 \end{bmatrix}\right) \text{ and is i.i.d. across } i.$$

Also, each dimension of the k-dimensional instruments Z_i takes one independent draw from the distribution of $\{U_i\}_{i=1}^n$ and is fixed across replications. Thus, (U_i, V_i) within each group follows an AR(1) process and is independent across different groups. The parameters $(\beta, \pi, \{I_g\}_{g=1}^G, k)$ vary accordingly across settings.

S1.1. Debiasing and truncation

We first investigate three Fama-MacBeth-type inferential procedures to show the necessity of debiasing and truncation. We consider (i) the t-test on group-level 2SLS estimators (FM), (ii) the t-test on group-level unbiased IV estimators (FMU), and (iii) the proposed method by Fama-MacBeth truncated unbiased IV estimators (FMTU), with truncation parameter selected as suggested in the implementation section. The full-sample 2SLS with CCE estimates of standard errors is also reported for comparison.

In this experiment, we have five instruments ($k = 5$) and one endogenous variable. Groups are imbalanced in sizes, with five groups of 90 observations and 25 groups of 18 observations. For each group, the observations follow an AR(1) process as previously described. The first-stage coefficient is $\pi = (0.1, 0.1, 0.1, 0.1, 0.1)'/\sqrt{5}$ such that $\|\pi\|_2 = 0.1$.

The power curves are reported in Figure 1. Estimators used in CCE and FM are both biased. FM has greater bias between the two, because it uses group-level 2SLS estimators with much larger finite-sample bias than the full-sample estimator. FMU is less powerful than FMTU, because the unbiased IV estimator does not have a bounded second moment, such that the resulting t-statistic has a tail that is too large.

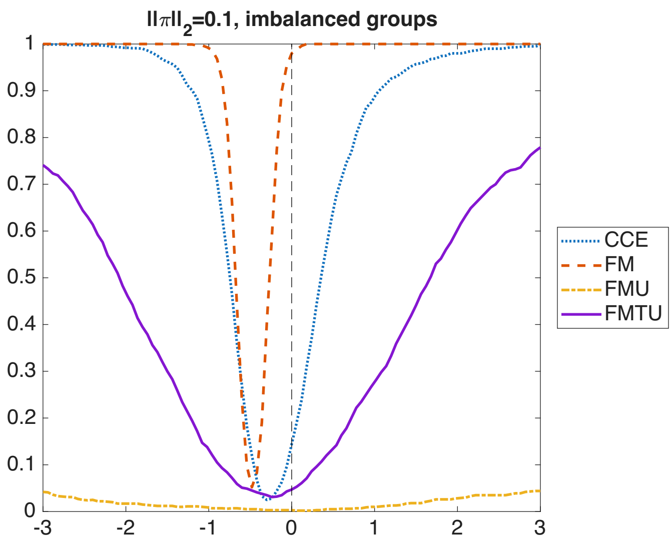

Figure S1. Power comparison among Fama-MacBeth procedures ($\alpha = 0.05$)

S1.2. Comparison with other inferential methods

In this section, we compare the proposed method with the existing inferential procedures. We consider the "clustered standard error" approach (CCE) and the natural extension of Anderson-Rubin test to our settings (AR-CCE). To implement the AR-CCE method, we apply CCE to the regression of $Y - X\beta_0$ on Z, where β_0 is the hypothesized value as in $H_0 : \beta = \beta_0$. In our case, we test $H_0 : \beta = 0$, so AR-CCE is equivalent to performing CCE to test the hypothesis $H_0 : \gamma = 0$ in the regression $Y = Z\gamma + U$.

We look at several configurations. The number of instruments k varies in the set $\{1, 5, 10\}$. The first-stage strength is chosen such that $\|\pi\|_2 \in \{0.1, 0.5\}$, with $\pi = \|\pi\|_2 \iota_k/\sqrt{k}$ and ι_k being a k-vector of 1's. For example, in the case of $\|\pi\|_2 = 0.1$ and $k = 5$, we have $\pi = (0.1, 0.1, 0.1, 0.1, 0.1)'/\sqrt{5}$. We also consider both balanced and imbalanced groups. In the balanced-group case, we have 30 groups of 30 observations; in the imbalanced-group case, we have 5 groups of 90 observations and 25 groups of 18 observations.

The sizes are reported in Table S1 and the power curves are in Figures S2, S3, and S4. Among all methods, only FMTU consistently provides valid inference results under the null across all settings. CCE displays a noticeable bias under small $\|\pi\|_2$ and over-identification. AR-CCE is robust to weak instruments, but not to group imbalance.

Table S1. Summary ($\alpha = 0.05$)

			Balanced Groups			Imbalanced Groups		
			Median	MAD	Size	Median	MAD	Size
$k = 1$	$\|\pi\|_2 = 0.5$	CCE	0.000	0.049	0.050	0.002	0.048	0.060
		AR-CCE	-	-	0.048	-	-	0.056
		FMTU	-0.019	0.059	0.042	-0.033	0.088	0.044
	$\|\pi\|_2 = 0.1$	CCE	0.002	0.240	0.048	0.011	0.243	0.048
		AR-CCE	-	-	0.048	-	-	0.056
		FMTU	0.010	0.306	0.047	-0.018	0.330	0.047
$k = 5$	$\|\pi\|_2 = 0.5$	CCE	0.012	0.052	0.063	0.012	0.052	0.076
		AR-CCE	-	-	0.042	-	-	0.087
		FMTU	-0.065	0.104	0.043	-0.103	0.151	0.037
	$\|\pi\|_2 = 0.1$	CCE	0.183	0.243	0.127	0.190	0.245	0.144
		AR-CCE	-	-	0.042	-	-	0.087
		FMTU	0.074	0.278	0.065	0.007	0.329	0.047
$k = 10$	$\|\pi\|_2 = 0.5$	CCE	0.029	0.052	0.066	0.029	0.052	0.073
		AR-CCE	-	-	0.052	-	-	0.127
		FMTU	-0.048	0.105	0.039	-0.085	0.130	0.033
	$\|\pi\|_2 = 0.1$	CCE	0.281	0.287	0.251	0.280	0.286	0.272
		AR-CCE	-	-	0.052	-	-	0.127
		FMTU	0.084	0.218	0.071	0.055	0.282	0.040

Notes: The "Median" columns show the median bias of the simulated estimates. The "MAD" columns show the median absolute deviation of the simulated estimates. The "Size" columns show the empirical rejection rates of the inference methods. AR-CCE does not report Median and MAD because itself does not produce an estimator for the parameter of interest.

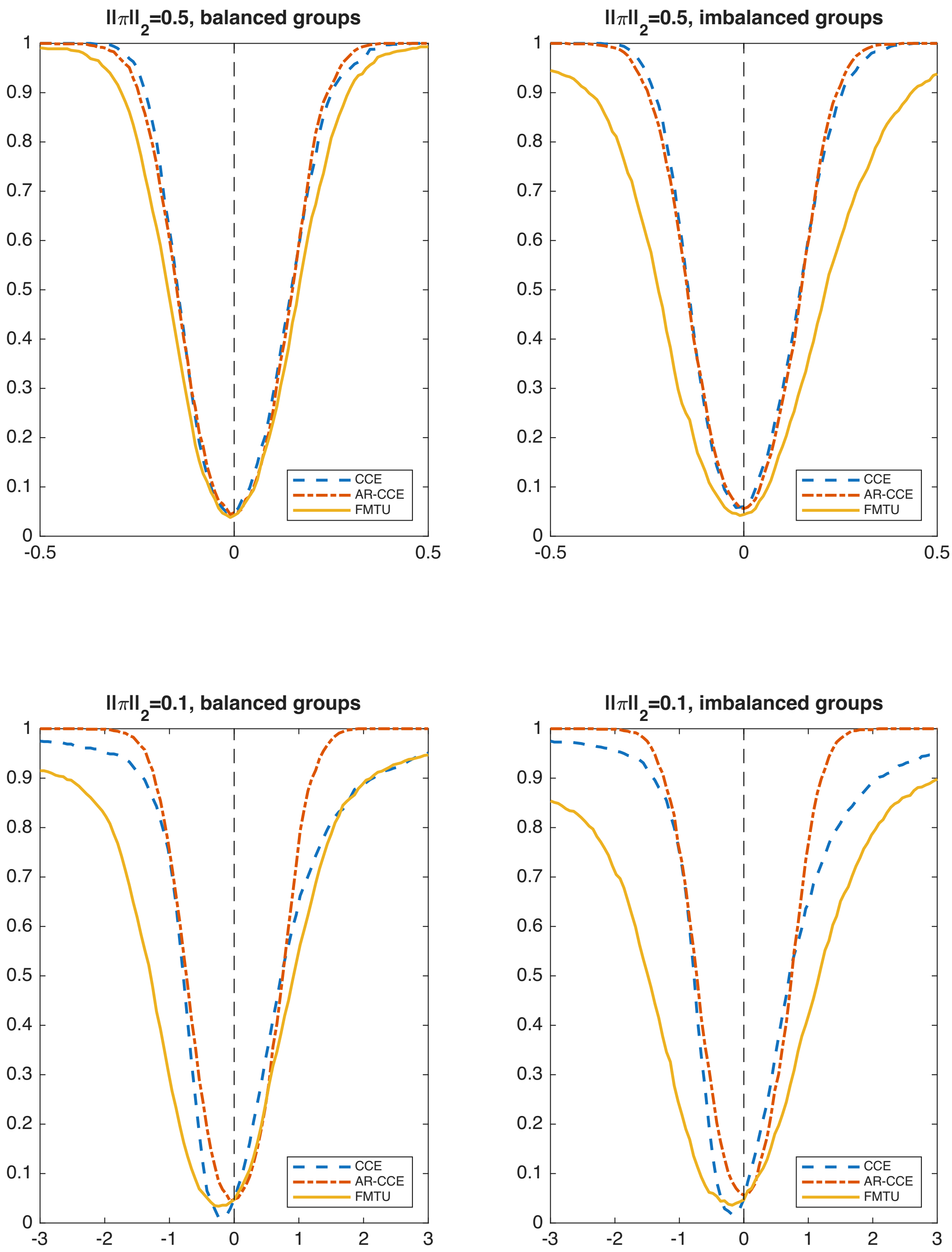

Figure S2. Power curves with nominal size $\alpha = 0.05$ and $k = 1$.

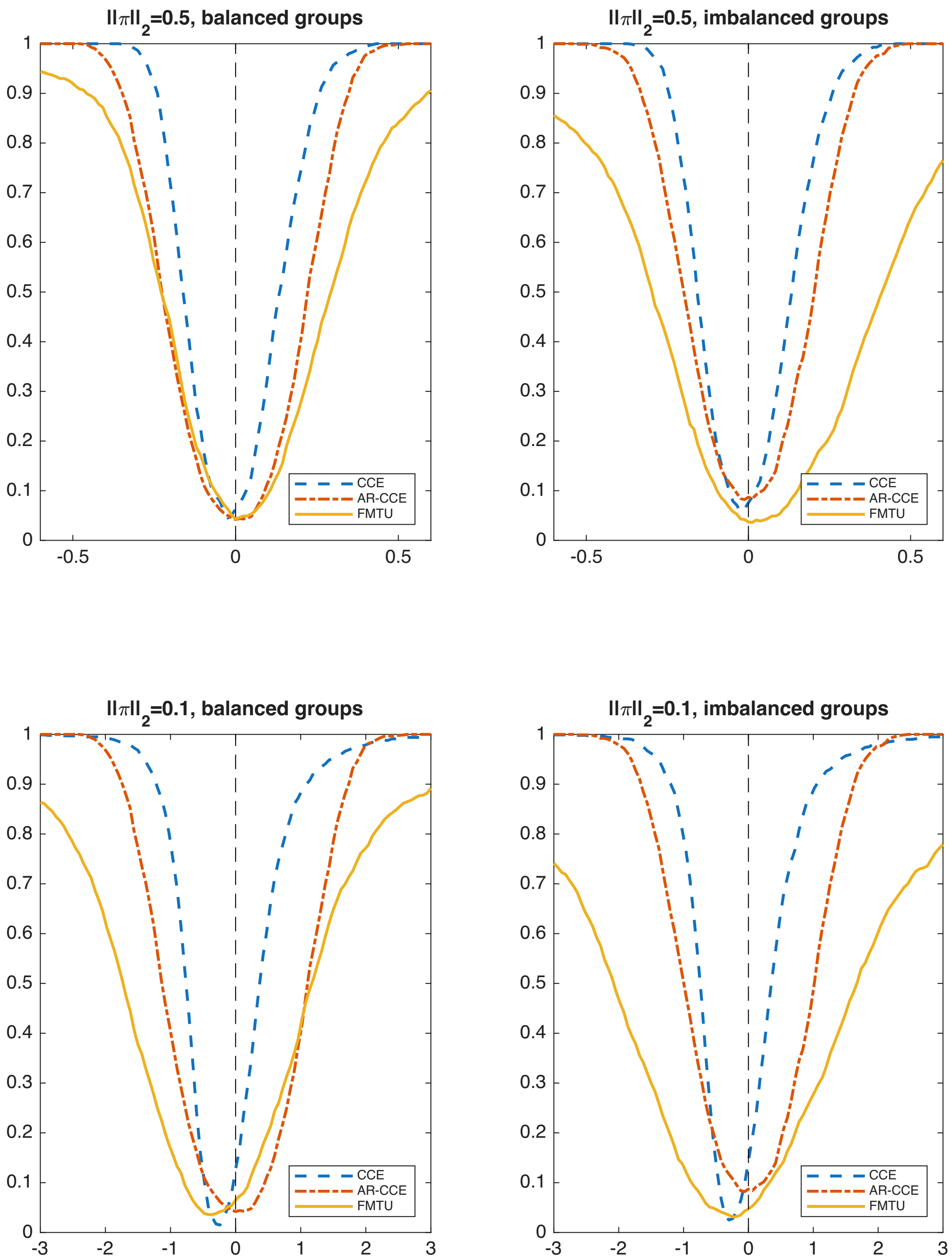

Figure S3. Power curves with nominal size $\alpha = 0.05$ and $k = 5$.

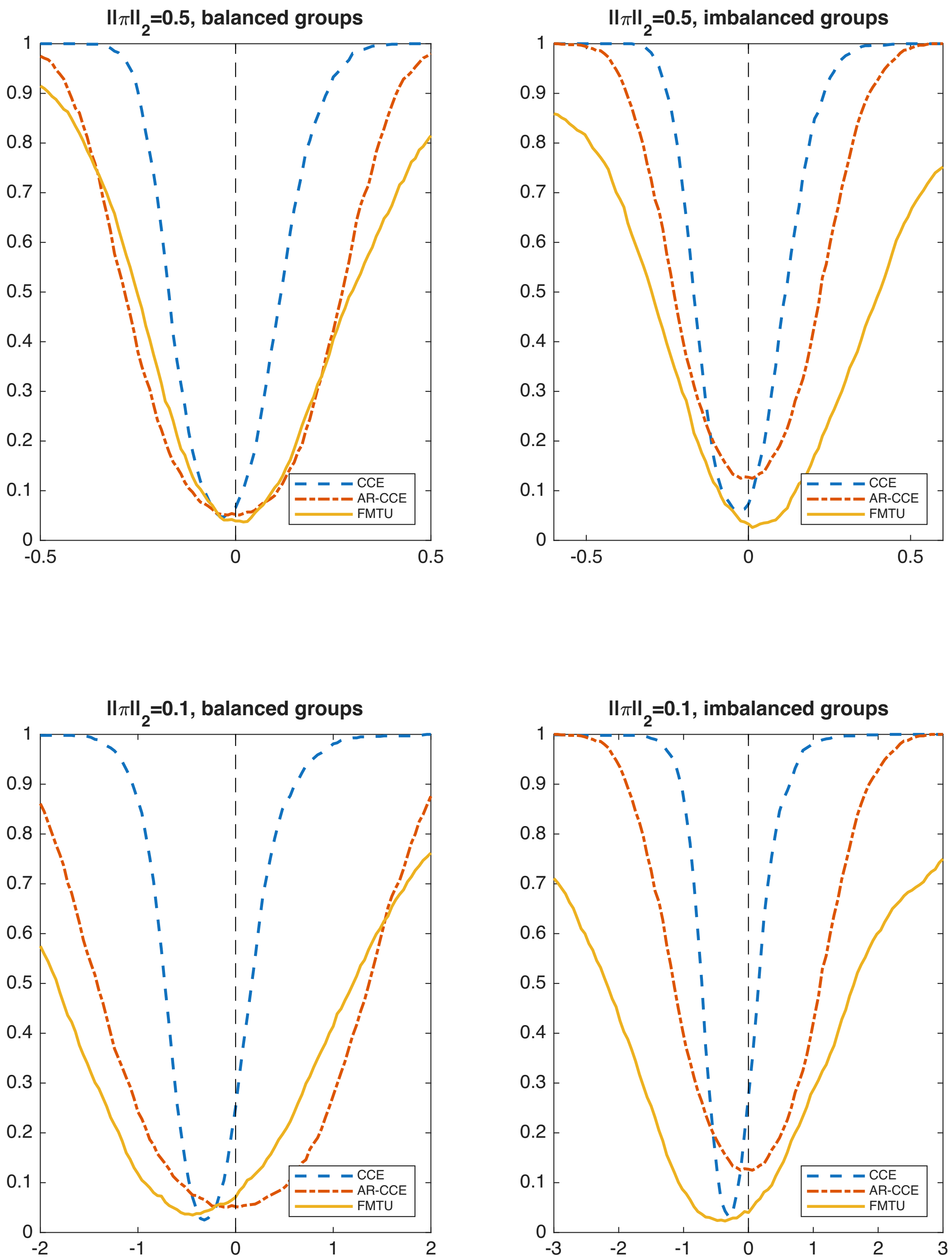

Figure S4. Power curves with nominal size $\alpha = 0.05$ and $k = 10$.

S1.3. Truncation parameter choices

In this section, we investigate the impact of the choice of the truncation parameter. The data regenerating process is the same as in section S1.1. FMTU methods with three different values of the truncation parameters π^* are reported ($\pi^* = -0.15, -0.2, -0.25$). The CCE method is also reported for comparison. The power curves are shown in Figure S5. Generally, the proposed method is quite robust to the choice of the truncation parameter in terms of null rejection rate. Moreover, Figure S5 exhibits a "bias-variance" tradeoff. That is, a higher π^* corresponds to high power but leads to more bias.

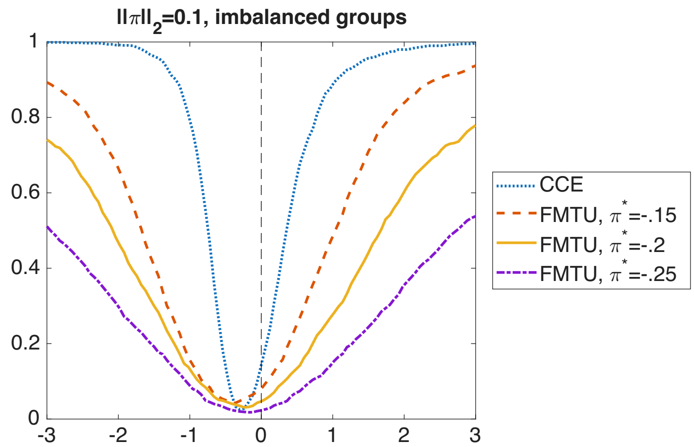

Figure S5. Power comparison among truncation-parameter choices ($\alpha = 0.05$)