Online Supplementary Material to "Relevant moment selection under mixed identification strength*"

Prosper Dovonon[†], Firmin Doko Tchatoka[‡], and Michael Aguessy[§]
Concordia University and University of Adelaide

October 2, 2022

This supplementary appendix presents further simulation results. Section S1 replicates the same experiment as in Section 5 in the main paper but with data generating processes (DGP) that account for nonzero correlation between the relevant instruments z_1 and z_2. Section S2 presents a cross-validation algorithm to determine the best choice of value for the tuning α appearing in the modified relevant moment selection criterion (mRMSC) - see Equation (24) in the main paper. Simulation results of this algorithm are also presented in this section.

Appendix S1: Additional Simulations

We perform additional simulations to investigate the performance of our proposed selection criterion and the properties of the post-selection GMM estimator (bias, mean squared error, coverage rate of confidence sets) when the instruments are not independent. We consider the same data generating process (DGP) of Sections 3 and 5 in the main paper where the instruments z_1 and z_2 can be arbitrarily correlated. Specifically, we consider the following DGP:

$$Y = X\theta + U, \quad X = z_1\pi_{1T} + z_2\pi_{2T} + V, \quad \pi_{iT} = \frac{c_i}{T^{\delta_i}}, \quad i = 1, 2,$$

*This research is supported by the Social Sciences and Humanities Research Council of Canada and by the Australian Research Council grant DP200101498.

[†]Economics Department, Concordia University, 1455 de Maisonneuve Blvd. West, H-1155, Montreal, Quebec, Canada, H3G 1M8. Email: prosper.dovonon@concordia.ca (Corresponding author).

[‡]School of Economics, The University of Adelaide, 10 Pulteney Street, Adelaide, Australia, SA 5005. Email: firmin.dokotchatoka@adelaide.edu.au.

[§]Economics Department, Concordia University, 1455 de Maisonneuve Blvd. West, H-1155, Montreal, Quebec, Canada, H3G 1M8. Email: michaelaguessy@gmail.com.

where Z is independent of the errors U and V. More specifically, for $t = 1, \ldots, T$, $(z_{1t}, z_{2t})' \sim \text{NID}(0, (1, \rho_z, 1))$ and $(u_t, v_t)' \sim \text{NID}(0, (1, \rho, 1))$, with $\rho_z, \rho \in (-1, 1)$.

We consider many combinations of ρ_z and ρ in our simulation experiments but, as the results do not change qualitatively, we only report the case $\rho_z = \rho = 0.5$. We also focus on the case of two endogenous variables ($p = 2$) as this enables us to evaluate the performance of our proposed selection method and the properties of the post-selection GMM estimator when the weaker of the two instruments is relevant and can improve the efficiency of GMM estimation. As in Sections 3 and 5 of the main paper, we analyze the cases where the instruments z_1 and z_2 are of equal strength - with $\delta_1 = \delta_2 = 0, 0.2, 0.3, 0.4$ - and of mixed strength - with $(\delta_1, \delta_2) = (0, 0.4), (0.1, 0.4), (0.2, 0.4), (0.3, 0.4)$. We set $\theta_0 = (0.1, 0.1)'$, $c_1 = (1.48, 0)$ and $c_2 = (0, 1.48)$, and include 10 extra instruments - $z_3, \ldots, z_{12}$ - which are independent of one another and of z_1, z_2, U and V, and with common distribution $N(0, I_T)$.

Figure S1.1 contains the plots of the hit rates of the selection procedures (mRMSC, RMSC and DN), while Figures S1.2-S1.4 present the plots of the bias, MSE, and coverage rate of joint confidence sets of the post-selection GMM estimator with the selection criteria (mRMSC, RMSC and DN) and the naive GMM estimator that uses all available instruments. The figures are plotted across the sample size T and for various identification strengths ($\delta_1 < \delta_2$ and $\delta_1 = \delta_2$). As seen in all cases, the results are similar to those reported in Section 5 in the main paper (Figures 5.1-(b), 5.2-(b), 5.3-(b) and 5.4-(b)). Moreover, Table S1.1 summarizes the empirical selection probabilities of all models with the various selection criteria (mRMSC, RMSC, DN). The results are similar to those shown in Table 5.1 in the main paper, and confirm the dominance of our criterion (mRMSC) in almost all cases. Table S1.2 contains the bias, MSE, and coverage rate of the joint confidence sets plotted in Figures S1.1-S1.4 for sample sizes $T = 100; 500; 1,000; 5,000; 100,000$. The gray columns of the table indicate the performance of our criterion (mRMSC) for the more weakly identified component θ_2 when $\delta_1 < \delta_2$. With the exception of the weak identification case ($\delta_1 = \delta_2 = 0.5$), GMM-mRMSC has an edge over GMM-RMSC, GMM-DN, and the naive GMM that uses all available instruments. These findings are in line with those in Section 5.

To enable comparison between the DGP in which z_1 and z_2 are uncorrelated (Section 5 in the main) and the one where z_1 and z_2 are correlated (Table S1.2), we have also reported in Table S1.3 the bias, MSE, and coverage rate of the joint confidence sets when $\rho_z = 0$. We clearly see from this table that the results are quite similar to those reported in Table S1.2.

Figure S1.1: Hit rate of mRMSC, RMSC and DN – Sample size $T = 100$; 200; 500; 1,000; 5,000; 10,000; 20,000; 50,000; 100,000; number of replications: 5,000. Model with $p = 2, \rho_z = 0.5$.

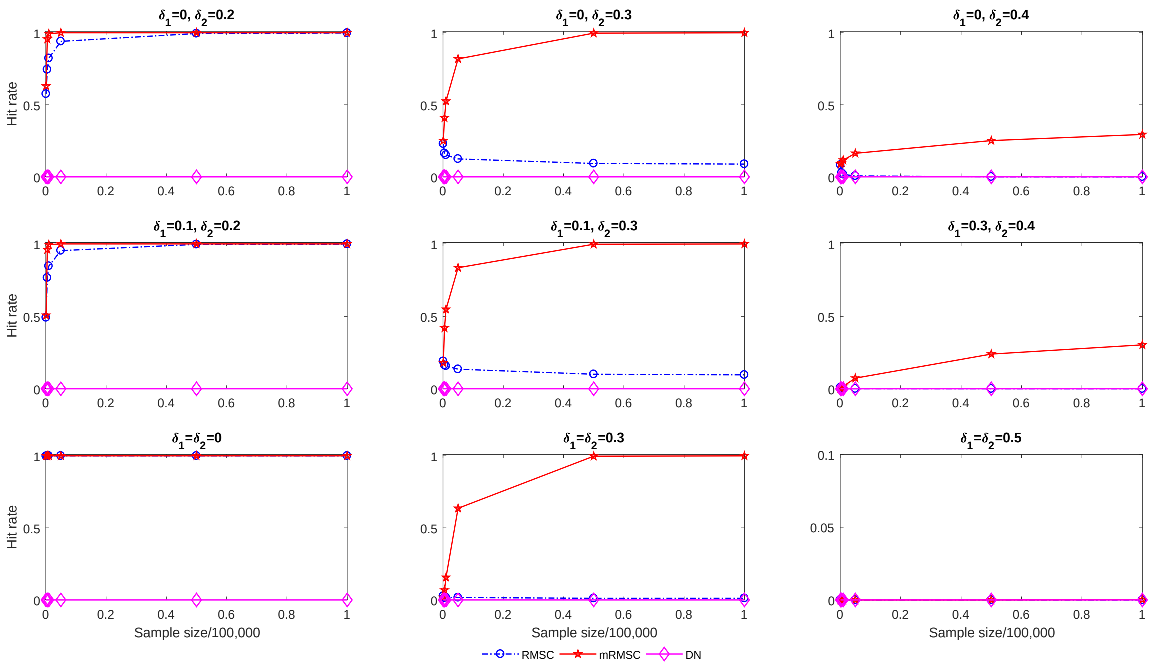
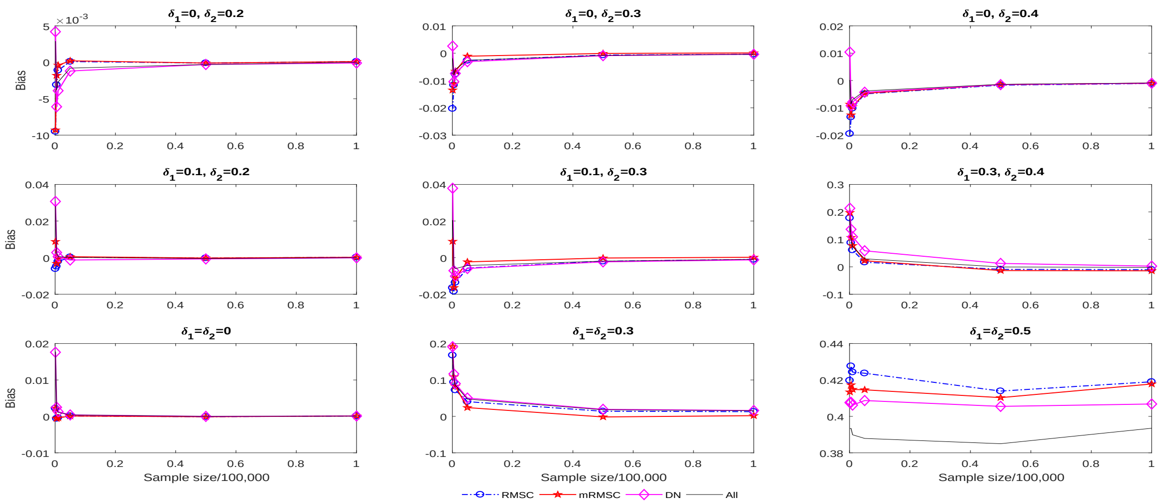
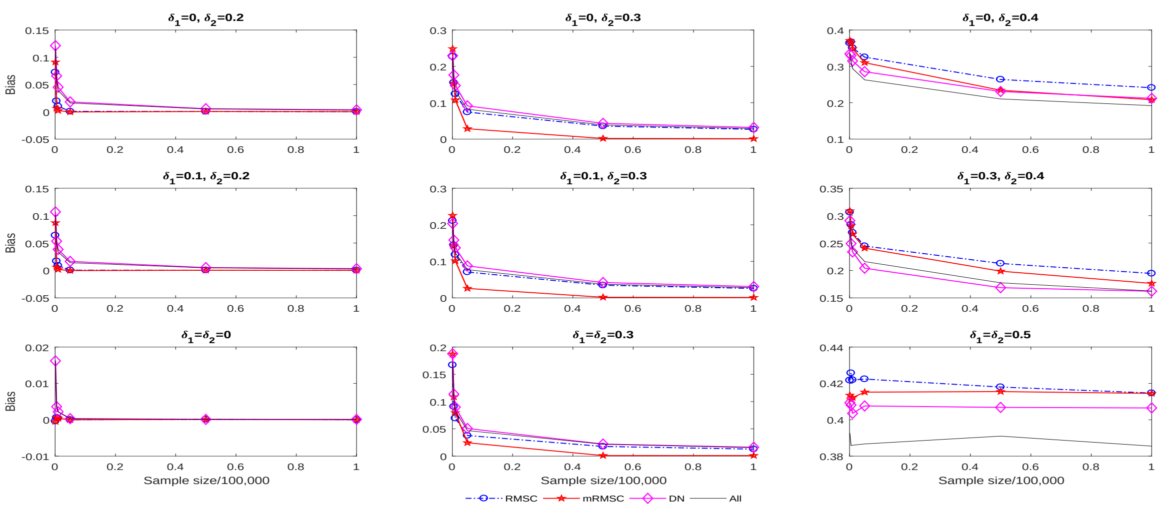


Figure S1.2: Bias of GMM post selection with mRMSC, RMSC and DN, and GMM with full set of IVs – Sample size T = 100; 200; 500; 1,000; 5,000; 10,000; 20,000; 50,000; 100,000; number of replications: 5,000.

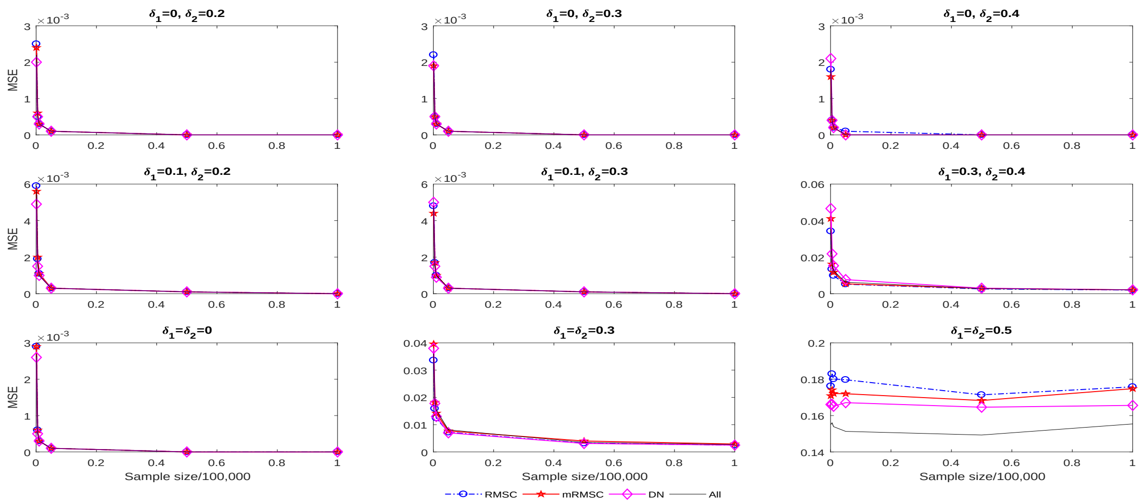


(a) Model with $p = 2, \rho_z = 0.5$: bias of $\hat{\theta}_1$

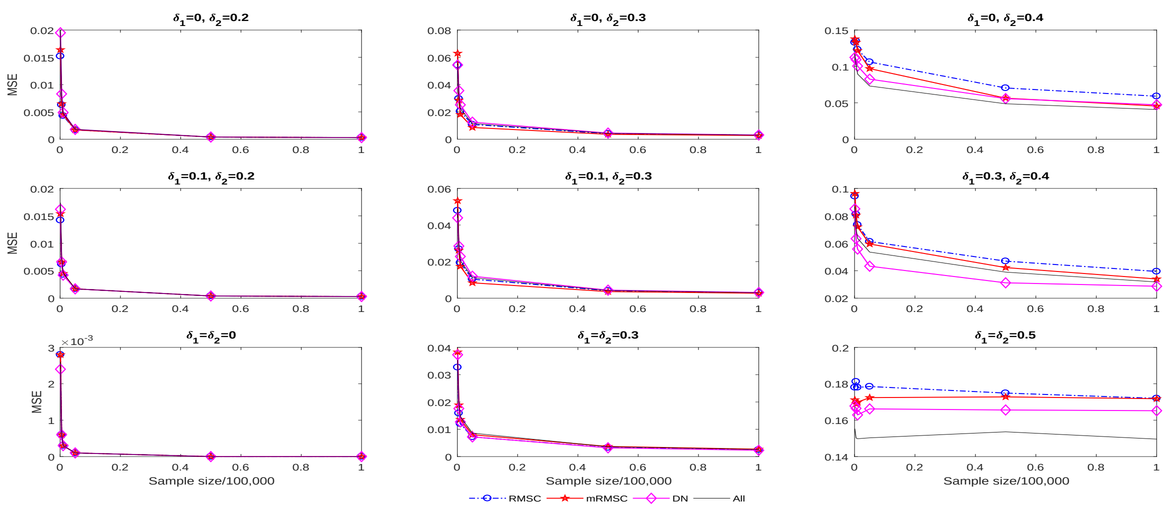


(b) Model with $p = 2$: bias of $\hat{\theta}_2$

Figure S1.3: MSE of GMM post selection with mRMSC, RMSC and DN, and GMM with full set of IVs – Sample size T = 100; 200; 500; 1,000; 5,000; 10,000; 20,000; 50,000; 100,000; number of replications: 5,000.

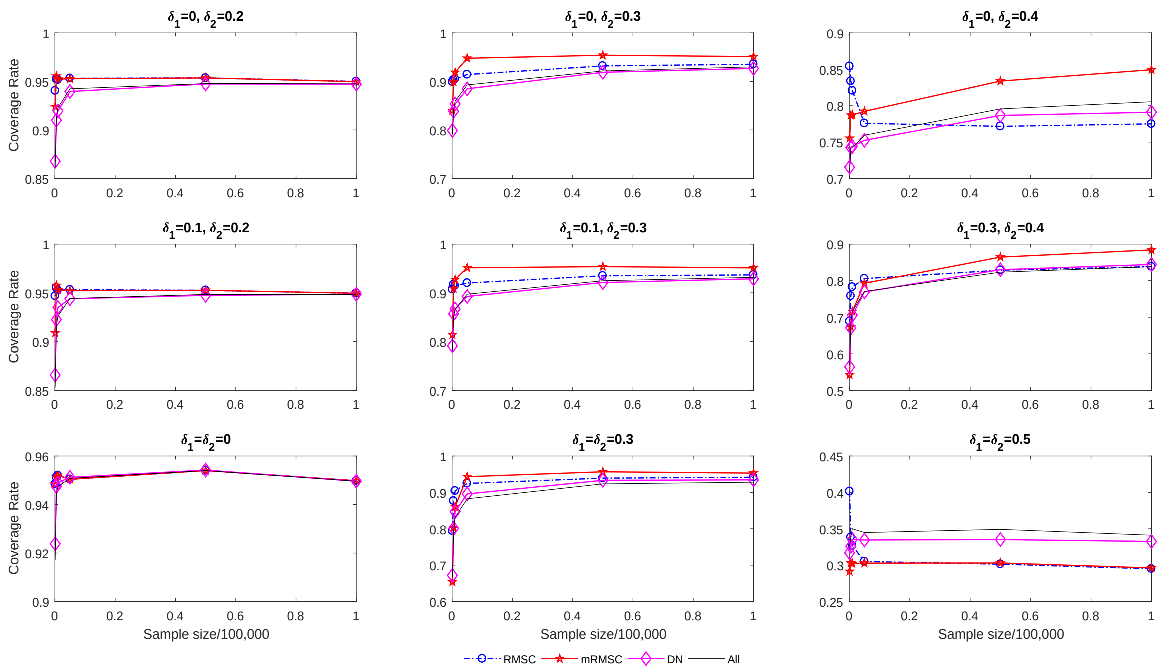


(a) Model with $p = 2, \rho_z = 0.5$: MSE of $\hat{\theta}_1$



(b) Model with $p = 2, \rho_z = 0.5$: MSE of $\hat{\theta}_2$

Figure S1.4: Coverage rate of confidence sets at nominal level 95%: GMM post selection with mRMSC, RMSC and DN, and GMM with full set of IVs – Sample size T = 100; 200; 500; 1,000; 5,000; 10,000; 20,000; 50,000; 100,000; number of replications: 5,000.



(a) Model with $p = 2, \rho_z = 0.5$

Table S1.1: Empirical Selection Probabilities with correlated IVs: Two endogenous regressors ($p = 2$)

$\rho_z = 0.5$			$T = 100$								$T = 1,000$							
	δ_1	δ_2	z1+z2	z1+I	z2+I	z1+z2+I	z1+z2+2I	all I	zj+more I	All	z1+z2	z1+I	z2+I	z1+z2+I	z1+z2+2I	all I	zj+more I	All
	$\delta_1 < \delta_2$																	
RMSC	0	0.2	0.51	0.00	0.00	0.37	0.10	0.00	0.02	0.00	0.82	0.00	0.00	0.16	0.02	0.00	0.00	0.00
	0.1	0.2	0.45	0.00	0.00	0.39	0.12	0.00	0.03	0.00	0.85	0.00	0.00	0.14	0.01	0.00	0.00	0.00
	0	0.3	0.19	0.01	0.00	0.39	0.26	0.00	0.15	0.00	0.16	0.00	0.00	0.34	0.31	0.00	0.19	0.00
	0.1	0.3	0.16	0.00	0.00	0.36	0.29	0.00	0.18	0.00	0.12	0.00	0.00	0.31	0.33	0.00	0.24	0.00
	0	0.4	0.06	0.02	0.00	0.26	0.29	0.00	0.37	0.00	0.02	0.00	0.00	0.11	0.28	0.00	0.59	0.00
	0.3	0.4	0.01	0.00	0.00	0.05	0.17	0.05	0.72	0.00	0.00	0.00	0.00	0.02	0.08	0.01	0.90	0.00
mRMSC	0	0.2	0.56	0.00	0.00	0.10	0.05	0.00	0.28	0.01	0.99	0.00	0.00	0.00	0.00	0.00	0.00	0.00
	0.1	0.2	0.45	0.00	0.00	0.06	0.03	0.00	0.42	0.04	1.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
	0	0.3	0.20	0.01	0.00	0.08	0.06	0.00	0.63	0.03	0.55	0.00	0.00	0.10	0.07	0.00	0.28	0.00
	0.1	0.3	0.14	0.00	0.00	0.04	0.03	0.00	0.72	0.07	0.51	0.00	0.00	0.07	0.04	0.00	0.35	0.02
	0	0.4	0.07	0.01	0.00	0.04	0.04	0.00	0.80	0.04	0.12	0.01	0.00	0.08	0.07	0.00	0.71	0.01
	0.3	0.4	0.00	0.00	0.00	0.00	0.00	0.01	0.82	0.16	0.02	0.00	0.00	0.01	0.01	0.01	0.82	0.14
DN	0	0.2	0.00	0.00	0.00	0.00	0.00	0.00	0.70	0.30	0.00	0.00	0.00	0.00	0.00	0.00	0.80	0.20
	0.1	0.2	0.00	0.00	0.00	0.00	0.00	0.00	0.75	0.25	0.00	0.00	0.00	0.00	0.00	0.00	0.76	0.24
	0	0.3	0.00	0.00	0.00	0.00	0.00	0.00	0.65	0.35	0.00	0.00	0.00	0.00	0.00	0.00	0.68	0.32
	0.1	0.3	0.00	0.00	0.00	0.00	0.00	0.00	0.72	0.28	0.00	0.00	0.00	0.00	0.00	0.00	0.74	0.26
	0	0.4	0.00	0.00	0.00	0.00	0.00	0.01	0.63	0.35	0.00	0.00	0.00	0.00	0.00	0.00	0.60	0.40
	0.3	0.4	0.00	0.00	0.00	0.00	0.00	0.01	0.82	0.17	0.00	0.00	0.00	0.00	0.00	0.01	0.79	0.20
	$\delta_1 = \delta_2$																	
RMSC	0	0	1.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	1.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
	0.3	0.3	0.02	0.00	0.00	0.13	0.26	0.03	0.57	0.00	0.02	0.00	0.00	0.10	0.24	0.00	0.65	0.00
	0.5	0.5	0.00	0.00	0.00	0.01	0.05	0.16	0.78	0.00	0.00	0.00	0.00	0.00	0.00	0.08	0.92	0.00
mRMSC	0	0	1.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	1.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
	0.3	0.3	0.01	0.00	0.00	0.00	0.00	0.01	0.81	0.16	0.16	0.00	0.00	0.03	0.02	0.00	0.70	0.10
	0.5	0.5	0.00	0.00	0.00	0.00	0.00	0.03	0.79	0.18	0.00	0.00	0.00	0.00	0.00	0.04	0.80	0.16
DN	0	0	0.00	0.00	0.00	0.00	0.00	0.00	0.79	0.21	0.00	0.00	0.00	0.00	0.00	0.00	0.79	0.21
	0.3	0.3	0.00	0.00	0.00	0.00	0.00	0.01	0.83	0.17	0.00	0.00	0.00	0.00	0.00	0.00	0.82	0.18
	0.5	0.5	0.00	0.00	0.00	0.00	0.00	0.03	0.80	0.17	0.00	0.00	0.00	0.00	0.00	0.03	0.81	0.15

$\rho_z = 0.5$			$T = 5,000$								$T = 50,000$							
	δ_1	δ_2	z1+z2	z1+I	z2+I	z1+z2+I	z1+z2+2I	all I	zj+more I	All	z1+z2	z1+I	z2+I	z1+z2+I	z1+z2+2I	all I	zj+more I	All
	$\delta_1 < \delta_2$																	
RMSC	0	0.2	0.94	0.00	0.00	0.06	0.00	0.00	0.00	0.00	0.99	0.00	0.00	0.01	0.00	0.00	0.00	0.00
	0.1	0.2	0.95	0.00	0.00	0.05	0.00	0.00	0.00	0.00	1.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
	0	0.3	0.14	0.00	0.00	0.31	0.30	0.00	0.25	0.00	0.10	0.00	0.00	0.27	0.30	0.00	0.34	0.00
	0.1	0.3	0.13	0.00	0.00	0.30	0.31	0.00	0.25	0.00	0.11	0.00	0.00	0.28	0.29	0.00	0.32	0.00
	0	0.4	0.01	0.00	0.00	0.04	0.15	0.00	0.80	0.00	0.00	0.00	0.00	0.01	0.04	0.00	0.95	0.00
	0.3	0.4	0.00	0.00	0.00	0.00	0.03	0.00	0.96	0.00	0.00	0.00	0.00	0.00	0.01	0.00	0.99	0.00
mRMSC	0	0.2	1.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	1.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
	0.1	0.2	1.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	1.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
	0	0.3	0.84	0.00	0.00	0.06	0.03	0.00	0.08	0.00	1.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
	0.1	0.3	0.86	0.00	0.00	0.05	0.02	0.00	0.07	0.00	1.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
	0	0.4	0.16	0.01	0.00	0.09	0.08	0.00	0.65	0.01	0.25	0.00	0.00	0.11	0.09	0.00	0.54	0.00
	0.3	0.4	0.07	0.00	0.00	0.02	0.02	0.00	0.79	0.10	0.24	0.00	0.00	0.08	0.05	0.00	0.59	0.04
DN	0	0.2	0.00	0.00	0.00	0.00	0.00	0.00	0.83	0.17	0.00	0.00	0.00	0.00	0.00	0.00	0.84	0.16
	0.1	0.2	0.00	0.00	0.00	0.00	0.00	0.00	0.81	0.19	0.00	0.00	0.00	0.00	0.00	0.00	0.83	0.17
	0	0.3	0.00	0.00	0.00	0.00	0.00	0.00	0.78	0.22	0.00	0.00	0.00	0.00	0.00	0.00	0.83	0.17
	0.1	0.3	0.00	0.00	0.00	0.00	0.00	0.00	0.76	0.24	0.00	0.00	0.00	0.00	0.00	0.00	0.80	0.20
	0	0.4	0.00	0.00	0.00	0.00	0.00	0.00	0.65	0.35	0.00	0.00	0.00	0.00	0.00	0.00	0.71	0.29
	0.3	0.4	0.00	0.00	0.00	0.00	0.00	0.00	0.77	0.23	0.00	0.00	0.00	0.00	0.00	0.00	0.72	0.28
	$\delta_1 = \delta_2$																	
RMSC	0	0	1.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	1.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
	0.3	0.3	0.02	0.00	0.00	0.09	0.21	0.00	0.69	0.00	0.01	0.00	0.00	0.07	0.18	0.00	0.74	0.00
	0.5	0.5	0.00	0.00	0.00	0.00	0.00	0.06	0.93	0.01	0.00	0.00	0.00	0.00	0.00	0.04	0.88	0.08
mRMSC	0	0	1.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	1.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
	0.3	0.3	0.64	0.00	0.00	0.04	0.02	0.00	0.28	0.03	1.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
	0.5	0.5	0.00	0.00	0.00	0.00	0.00	0.04	0.78	0.18	0.00	0.00	0.00	0.00	0.00	0.04	0.79	0.17
DN	0	0	0.00	0.00	0.00	0.00	0.00	0.00	0.80	0.20	0.00	0.00	0.00	0.00	0.00	0.00	0.79	0.21
	0.3	0.3	0.00	0.00	0.00	0.00	0.00	0.00	0.81	0.19	0.00	0.00	0.00	0.00	0.00	0.00	0.80	0.20
	0.5	0.5	0.00	0.00	0.00	0.00	0.00	0.04	0.80	0.16	0.00	0.00	0.00	0.00	0.00	0.03	0.81	0.15

Note: 'z1 + z2' denotes models with the 2 instruments z1 and z2; 'zj+I' ($j = 1, 2$) denotes models with zj + 1 irrelevant (i.e. completely unrelated) instrument; 'z1 + z2 + I' denotes models with the 2 instruments z1 and z2 + 1 irrelevant instrument; 'z1 + z2 + 2I' denotes models with the 2 instruments z1 and z2 + 2 irrelevant instruments; 'all I' denotes models with irrelevant instruments only; 'zj + more I' denotes models with zj ($j = 1, 2$) + more than 1 irrelevant instrument; 'All' denotes model with all instruments. The highlighted column correspond to the best subset of instruments. This subset depends on the combination of strengths (δ_1, δ_2) and the number p of estimated parameters.

Table S1.2: Bias, MSE, and Coverage Rate of post-selection GMM estimator: Two endogenous regressors ($p = 2$) and correlated instruments ($\rho_z = 0.5$)

$\rho_z = 0.5$			Bias $\hat{\theta}_1$				Bias $\hat{\theta}_2$					MSE $\hat{\theta}_1$				MSE $\hat{\theta}_2$				Joint Cov Rate			
	δ_1	δ_2	RMSC	mRMSC	DN	All	RMSC	mRMSC	DN	All		RMSC	mRMSC	DN	All	RMSC	mRMSC	DN	All	RMSC	mRMSC	DN	All
	$\delta_1 = \delta_2$																						
T=100	0	0	0.00	0.00	0.02	0.02	0.00	0.00	0.02	0.02		0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.95	0.95	0.92	0.92
	0.3	0.3	0.17	0.19	0.19	0.17	0.17	0.19	0.19	0.17		0.03	0.04	0.04	0.04	0.03	0.04	0.04	0.04	0.79	0.65	0.67	0.67
	0.5	0.5	0.42	0.41	0.41	0.39	0.42	0.41	0.41	0.39		0.18	0.17	0.17	0.16	0.18	0.17	0.17	0.16	0.40	0.29	0.32	0.33
T=1,000	0	0	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00		0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.95	0.95	0.95	0.95
	0.3	0.3	0.07	0.08	0.09	0.09	0.07	0.08	0.09	0.08		0.01	0.01	0.01	0.02	0.01	0.01	0.01	0.01	0.90	0.86	0.85	0.83
	0.5	0.5	0.42	0.41	0.41	0.39	0.42	0.41	0.40	0.39		0.18	0.17	0.17	0.15	0.18	0.17	0.16	0.15	0.33	0.30	0.34	0.35
T=5,000	0	0	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00		0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.95	0.95	0.95	0.95
	0.3	0.3	0.04	0.02	0.05	0.05	0.04	0.02	0.05	0.05		0.01	0.01	0.01	0.01	0.01	0.01	0.01	0.01	0.92	0.94	0.90	0.88
	0.5	0.5	0.42	0.41	0.41	0.39	0.42	0.42	0.41	0.39		0.18	0.17	0.17	0.15	0.18	0.17	0.17	0.15	0.31	0.30	0.33	0.34
T=100,000	0	0	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00		0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.95	0.95	0.95	0.95
	0.3	0.3	0.01	0.00	0.02	0.02	0.01	0.00	0.02	0.02		0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.94	0.95	0.93	0.93
	0.5	0.5	0.42	0.42	0.41	0.39	0.41	0.41	0.41	0.39		0.18	0.17	0.17	0.16	0.17	0.17	0.17	0.15	0.30	0.30	0.33	0.34
	$\delta_1 < \delta_2$																						
T=100	0	0.2	-0.01	-0.01	0.00	0.00	0.07	0.09	0.12	0.12		0.00	0.00	0.00	0.00	0.02	0.02	0.02	0.02	0.94	0.92	0.87	0.86
	0.1	0.2	-0.01	0.01	0.03	0.03	0.06	0.09	0.11	0.10		0.01	0.01	0.00	0.01	0.01	0.02	0.02	0.02	0.95	0.91	0.87	0.85
	0	0.3	-0.02	-0.01	0.00	0.00	0.23	0.25	0.23	0.22		0.00	0.00	0.00	0.00	0.05	0.06	0.05	0.06	0.90	0.84	0.80	0.79
	0.1	0.3	-0.02	0.01	0.04	0.02	0.21	0.23	0.20	0.21		0.00	0.00	0.01	0.00	0.05	0.05	0.04	0.05	0.91	0.81	0.79	0.78
	0	0.4	-0.02	-0.01	0.01	0.00	0.36	0.37	0.33	0.34		0.00	0.00	0.00	0.00	0.13	0.14	0.11	0.12	0.85	0.76	0.72	0.71
	0.3	0.4	0.18	0.20	0.21	0.17	0.31	0.31	0.29	0.29		0.03	0.04	0.05	0.04	0.09	0.10	0.09	0.09	0.69	0.54	0.56	0.57
T=1,000	0	0.2	0.00	0.00	0.00	0.00	0.01	0.00	0.05	0.04		0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.95	0.95	0.92	0.92
	0.1	0.2	0.00	0.00	0.00	0.00	0.01	0.00	0.04	0.03		0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.95	0.95	0.94	0.93
	0	0.3	-0.01	-0.01	-0.01	-0.01	0.12	0.11	0.15	0.13		0.00	0.00	0.00	0.00	0.02	0.02	0.03	0.02	0.90	0.92	0.85	0.86
	0.1	0.3	-0.01	-0.01	-0.01	-0.01	0.12	0.10	0.14	0.12		0.00	0.00	0.00	0.00	0.02	0.02	0.02	0.02	0.91	0.93	0.87	0.87
	0	0.4	-0.01	-0.01	-0.01	-0.01	0.35	0.35	0.31	0.29		0.00	0.00	0.00	0.00	0.12	0.12	0.10	0.09	0.82	0.79	0.74	0.74
	0.3	0.4	0.06	0.08	0.11	0.07	0.27	0.27	0.23	0.24		0.01	0.01	0.02	0.01	0.07	0.07	0.06	0.06	0.78	0.72	0.71	0.72
T=5,000	0	0.2	0.00	0.00	0.00	0.00	0.00	0.00	0.02	0.02		0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.95	0.95	0.94	0.94
	0.1	0.2	0.00	0.00	0.00	0.00	0.00	0.00	0.02	0.01		0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.95	0.95	0.94	0.94
	0	0.3	0.00	0.00	0.00	0.00	0.07	0.03	0.09	0.08		0.00	0.00	0.00	0.00	0.01	0.01	0.01	0.01	0.91	0.95	0.88	0.89
	0.1	0.3	-0.01	0.00	-0.01	0.00	0.07	0.03	0.09	0.08		0.00	0.00	0.00	0.00	0.01	0.01	0.01	0.01	0.92	0.95	0.89	0.90
	0	0.4	0.00	0.00	0.00	0.00	0.33	0.31	0.29	0.26		0.00	0.00	0.00	0.00	0.11	0.10	0.08	0.07	0.78	0.79	0.75	0.76
	0.3	0.4	0.02	0.02	0.06	0.03	0.24	0.24	0.20	0.22		0.01	0.01	0.01	0.01	0.06	0.06	0.04	0.05	0.81	0.79	0.77	0.77
T=100,000	0	0.2	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00		0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.95	0.95	0.95	0.95
	0.1	0.2	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00		0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.95	0.95	0.95	0.95
	0	0.3	0.00	0.00	0.00	0.00	0.03	0.00	0.03	0.03		0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.94	0.95	0.93	0.93
	0.1	0.3	0.00	0.00	0.00	0.00	0.03	0.00	0.03	0.03		0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.94	0.95	0.93	0.93
	0	0.4	0.00	0.00	0.00	0.00	0.24	0.21	0.21	0.19		0.00	0.00	0.00	0.00	0.06	0.05	0.05	0.04	0.78	0.85	0.79	0.81
	0.3	0.4	-0.01	-0.01	0.00	0.00	0.19	0.18	0.16	0.16	0	0.00	0.00	0.00	0.00	0.04	0.03	0.03	0.03	0.84	0.88	0.84	0.84

Table S1.3: Bias, MSE, and Coverage Rate of post-selection GMM estimator: Two endogenous regressors ($p = 2$) and uncorrelated instruments ($\rho_z = 0$)

$\rho_z = 0$			Bias $\hat{\theta}_1$				Bias $\hat{\theta}_2$				MSE $\hat{\theta}_1$				MSE $\hat{\theta}_2$				Joint Cov Prob			
	δ_1	δ_2	RMSC	mRMSC	DN	All	RMSC	mRMSC	DN	All	RMSC	mRMSC	DN	All	RMSC	mRMSC	DN	All	RMSC	mRMSC	DN	All
	$\delta_1 = \delta_2$																					
T=100	0	0	0.00	0.00	0.02	0.02	0.00	0.00	0.02	0.02	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.95	0.95	0.91	0.91
	0.3	0.3	0.23	0.25	0.25	0.22	0.22	0.24	0.24	0.22	0.05	0.06	0.06	0.05	0.05	0.06	0.06	0.05	0.70	0.55	0.56	0.57
	0.5	0.5	0.45	0.44	0.44	0.43	0.45	0.44	0.44	0.42	0.20	0.19	0.19	0.18	0.20	0.20	0.19	0.18	0.37	0.25	0.28	0.29
T=1,000	0	0	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.95	0.95	0.95	0.95
	0.3	0.3	0.11	0.11	0.12	0.12	0.10	0.11	0.12	0.11	0.01	0.02	0.02	0.02	0.02	0.02	0.02	0.02	0.84	0.80	0.77	0.77
	0.5	0.5	0.45	0.44	0.43	0.42	0.45	0.43	0.43	0.42	0.20	0.19	0.19	0.18	0.20	0.19	0.19	0.17	0.29	0.26	0.29	0.31
T=5,000	0	0	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.95	0.95	0.95	0.95
	0.3	0.3	0.06	0.04	0.07	0.07	0.06	0.04	0.07	0.07	0.01	0.01	0.01	0.01	0.01	0.01	0.01	0.01	0.88	0.92	0.84	0.84
	0.5	0.5	0.45	0.44	0.44	0.42	0.45	0.44	0.43	0.42	0.20	0.20	0.19	0.18	0.20	0.20	0.19	0.18	0.27	0.27	0.30	0.31
T=100,000	0	0	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.95	0.95	0.95	0.95
	0.3	0.3	0.02	0.00	0.02	0.02	0.02	0.00	0.02	0.02	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.93	0.95	0.91	0.91
	0.5	0.5	0.44	0.44	0.43	0.42	0.44	0.44	0.44	0.42	0.19	0.19	0.19	0.18	0.20	0.20	0.19	0.18	0.27	0.27	0.30	0.31
	$\delta_1 < \delta_2$																					
T=100	0	0.2	0.01	0.01	0.02	0.02	0.06	0.08	0.13	0.12	0.00	0.00	0.00	0.00	0.01	0.01	0.02	0.02	0.93	0.92	0.84	0.84
	0.1	0.2	0.02	0.04	0.06	0.06	0.07	0.09	0.13	0.12	0.00	0.01	0.01	0.01	0.01	0.01	0.02	0.02	0.93	0.88	0.81	0.81
	0	0.3	0.01	0.02	0.03	0.03	0.21	0.24	0.23	0.22	0.00	0.00	0.00	0.00	0.05	0.06	0.06	0.05	0.88	0.81	0.76	0.76
	0.1	0.3	0.02	0.05	0.06	0.06	0.21	0.24	0.23	0.22	0.00	0.01	0.01	0.01	0.05	0.06	0.06	0.05	0.87	0.76	0.73	0.73
	0	0.4	0.01	0.02	0.03	0.03	0.36	0.38	0.34	0.33	0.00	0.00	0.00	0.00	0.13	0.14	0.12	0.11	0.82	0.71	0.67	0.67
	0.3	0.4	0.23	0.25	0.26	0.23	0.37	0.37	0.35	0.34	0.05	0.06	0.07	0.05	0.14	0.13	0.12	0.11	0.60	0.45	0.47	0.48
T=1,000	0	0.2	0.00	0.00	0.00	0.00	0.01	0.00	0.04	0.03	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.95	0.95	0.92	0.92
	0.1	0.2	0.00	0.00	0.01	0.01	0.01	0.00	0.04	0.03	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.95	0.95	0.92	0.92
	0	0.3	0.00	0.00	0.00	0.00	0.09	0.07	0.13	0.11	0.00	0.00	0.00	0.00	0.01	0.01	0.02	0.02	0.91	0.93	0.86	0.86
	0.1	0.3	0.00	0.00	0.01	0.01	0.09	0.08	0.13	0.11	0.00	0.00	0.00	0.00	0.01	0.01	0.02	0.02	0.91	0.93	0.85	0.86
	0	0.4	0.00	0.00	0.00	0.00	0.32	0.32	0.29	0.27	0.00	0.00	0.00	0.00	0.10	0.10	0.09	0.08	0.81	0.79	0.74	0.75
	0.3	0.4	0.11	0.13	0.14	0.12	0.32	0.31	0.29	0.27	0.02	0.02	0.02	0.02	0.10	0.09	0.08	0.08	0.68	0.62	0.62	0.63
T=5,000	0	0.2	0.00	0.00	0.00	0.00	0.00	0.00	0.01	0.01	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.95	0.95	0.94	0.94
	0.1	0.2	0.00	0.00	0.00	0.00	0.00	0.00	0.02	0.01	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.95	0.95	0.94	0.94
	0	0.3	0.00	0.00	0.00	0.00	0.05	0.01	0.07	0.07	0.00	0.00	0.00	0.00	0.01	0.01	0.01	0.01	0.92	0.95	0.89	0.90
	0.1	0.3	0.00	0.00	0.00	0.00	0.05	0.01	0.07	0.07	0.00	0.00	0.00	0.00	0.01	0.01	0.01	0.01	0.93	0.95	0.89	0.90
	0	0.4	0.00	0.00	0.00	0.00	0.28	0.27	0.25	0.23	0.00	0.00	0.00	0.00	0.08	0.07	0.06	0.06	0.79	0.81	0.77	0.77
	0.3	0.4	0.06	0.07	0.08	0.07	0.27	0.27	0.25	0.23	0.01	0.01	0.01	0.01	0.08	0.07	0.06	0.06	0.72	0.72	0.70	0.71
T=100,000	0	0.2	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.95	0.95	0.95	0.95
	0.1	0.2	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.95	0.95	0.95	0.95
	0	0.3	0.00	0.00	0.00	0.00	0.02	0.00	0.02	0.02	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.94	0.95	0.93	0.94
	0.1	0.3	0.00	0.00	0.00	0.00	0.02	0.00	0.02	0.02	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.94	0.95	0.93	0.94
	0	0.4	0.00	0.00	0.00	0.00	0.20	0.16	0.18	0.16	0.00	0.00	0.00	0.00	0.04	0.03	0.03	0.03	0.80	0.88	0.82	0.83
	0.3	0.4	0.02	0.01	0.02	0.02	0.19	0.16	0.18	0.16	0.00	0.00	0.00	0.00	0.04	0.03	0.03	0.03	0.80	0.86	0.81	0.81

Appendix S2: Optimal choice of the tuning parameter α

In this section, we conduct a cross-validation exercise to determine the best choice of the tuning α (see Section 4.3, Equation (24) in the main paper). The goal of this exercise is to select α that minimizes the mean squared forecast error (MSFE) of the best model selected by our mRMSC criterion. To do this, we consider the framework of Sections 3 and 5 in the main paper with two endogenous regressors ($p = 2$), and we split the data into a training set and a validation set. This allows us to estimate the tuning parameter α by: (i) identifying the best model with our mRMSC selection criterion on the training set, and (ii) finding α that minimizes the MSFE on the validation set.

Standard leave one out cross-validation approaches are usually intuitive but are often computationally intensive. We consider instead the K-fold cross-validation technique commonly employed in applied work. To be more precise, we use the following K-fold cross-validation algorithm from James et al. (2013).[1]

1. Divide the set of sample size T into K subsets (K folds) with equal size $T_1, T_2, \cdots, T_K$ (we use the notation T_k to denote both the sub-sample and its size).

2. For $k = 1, 2, \cdots, K$, consider a training set (X_i, Y_i, Z_i), $i \in T_{-k} := \{T_1, \ldots, T_{k-1}, T_{k+1}, \ldots, T_K\}$, and a validation set (X_i, Y_i, Z_i) for $i \in T_k$, where T_{-k} represents the $K-1$ folds used as a training set with the remaining fold used as a validation set. In this way each fold will be used only once as the validation set throughout the procedure. For each $\alpha \in \{\alpha_1, \alpha_2, \cdots, \alpha_m\}$, use the mRMSC to identify the best model on the training set. Then compute the GMM estimate and its associated predicted value $\hat{Y}_\alpha^k$ on the validation set using this best model, and then record the total mean squared forecast error
$$MSE_k(\alpha) = \sum_{i \in T_k} \left(Y_i - \hat{Y}_\alpha^k \right)^2.$$

3. Compute the average total mean squared forecast error over all folds:
$$MSE(\alpha) = \frac{1}{K} \sum_{k=1}^{K} MSE_k(\alpha).$$
The optimal $\hat{\alpha}$ is the one that minimizes $MSE(\alpha)$, i.e.
$$\hat{\alpha} = \underset{\alpha \in \{\alpha_1, \alpha_2, \cdots, \alpha_m\}}{\arg\min} MSE(\alpha).$$

[1] See lecture notes at "http://stat.cmu.edu/~ryantibs/datamining/lectures."

In the simulations, we use a grid of 20 points for α, i.e., $\alpha = 0.05 \cdot j : j = 1, \ldots, 20$ but the results do not change if we extend the grid or use a finer grid. Table S2.1 shows, for various identification strengths ($\delta_1 < \delta_2$ and $\delta_1 = \delta_2$), the simulated MSFEs across values of α and sample sizes of $T = 100, 500, 5000$. The results are obtained with 5,000 replications and $K = 5$. With the exception of the case where α is close to zero ($\alpha = 0.05$), the results clearly show that the simulated MSFEs are similar across sample sizes and values of $\alpha \in [0.1, 1.0]$, irrespective of the identification strength (δ_1, δ_2). This suggests that any choice of α above 0.1 yields qualitatively the same performance of mRMSC, which supports our choice of $\alpha = 0.1$ in the paper. .

Table S2.1: Choice of the tuning parameter α by cross validation, T = 100; 500; 5,000

			α																			
	δ_1	δ_2	0.05	0.10	0.15	0.20	0.25	0.30	0.35	0.40	0.45	0.50	0.55	0.60	0.65	0.70	0.75	0.80	0.85	0.90	0.95	1.00
	$\delta_1 < \delta_2$																					
T=100	0	0.2	1.01	0.69	0.69	0.69	0.69	0.69	0.69	0.69	0.69	0.69	0.69	0.69	0.69	0.69	0.69	0.69	0.69	0.69	0.69	0.69
	0.1	0.2	0.93	0.64	0.64	0.64	0.64	0.64	0.64	0.64	0.64	0.64	0.64	0.64	0.64	0.64	0.64	0.64	0.65	0.65	0.65	0.65
	0	0.3	0.89	0.66	0.66	0.66	0.66	0.66	0.66	0.66	0.66	0.66	0.66	0.66	0.66	0.66	0.66	0.66	0.66	0.66	0.66	0.66
	0.1	0.3	0.79	0.61	0.61	0.61	0.61	0.61	0.61	0.61	0.61	0.61	0.61	0.61	0.61	0.61	0.61	0.61	0.61	0.61	0.61	0.61
	0	0.4	0.81	0.65	0.65	0.65	0.65	0.65	0.65	0.65	0.65	0.65	0.65	0.65	0.65	0.65	0.65	0.65	0.65	0.65	0.65	0.65
	0.3	0.4	0.57	0.51	0.51	0.51	0.51	0.51	0.51	0.51	0.51	0.51	0.51	0.51	0.52	0.52	0.52	0.52	0.52	0.52	0.52	0.52
T=500	0	0.2	1.10	1.10	0.85	0.85	0.85	0.84	0.84	0.84	0.84	0.84	0.84	0.84	0.84	0.84	0.84	0.84	0.84	0.84	0.84	0.84
	0.1	0.2	1.12	0.77	0.76	0.76	0.76	0.76	0.76	0.76	0.76	0.76	0.76	0.76	0.76	0.76	0.76	0.76	0.76	0.76	0.76	0.76
	0	0.3	1.91	0.75	0.75	0.75	0.75	0.75	0.75	0.75	0.75	0.75	0.75	0.75	0.75	0.75	0.75	0.75	0.75	0.75	0.75	0.75
	0.1	0.3	1.86	0.72	0.72	0.72	0.72	0.72	0.72	0.72	0.72	0.72	0.72	0.72	0.72	0.72	0.72	0.72	0.72	0.72	0.72	0.72
	0	0.4	2.07	0.74	0.74	0.74	0.74	0.74	0.74	0.74	0.74	0.74	0.74	0.74	0.74	0.74	0.74	0.74	0.74	0.74	0.74	0.74
	0.3	0.4	0.62	0.57	0.57	0.57	0.57	0.57	0.57	0.57	0.57	0.57	0.57	0.57	0.57	0.57	0.57	0.57	0.57	0.57	0.57	0.57
T=5,000	0	0.2	1.02	1.02	1.00	1.02	0.99	0.99	0.98	1.02	0.97	0.97	0.96	1.02	0.95	0.95	0.95	1.02	0.95	0.95	0.95	0.95
	0.1	0.2	1.02	1.02	1.00	1.02	0.99	0.98	0.98	1.02	0.96	0.96	0.95	1.02	0.95	0.95	0.94	1.02	0.94	0.94	0.94	0.94
	0	0.3	1.18	1.17	0.84	1.15	0.84	0.83	0.83	1.09	0.83	0.83	0.83	1.02	0.83	0.83	0.83	0.97	0.83	0.83	0.83	0.83
	0.1	0.3	1.19	1.18	0.84	1.15	0.83	0.83	0.83	1.08	0.83	0.83	0.83	1.01	0.83	0.83	0.83	0.96	0.83	0.83	0.83	0.83
	0	0.4	2.30	2.00	0.78	1.57	0.78	0.78	0.78	1.11	0.78	0.78	0.78	0.97	0.78	0.78	0.78	0.88	0.78	0.78	0.78	0.78
	0.3	0.4	1.26	1.10	0.62	0.89	0.62	0.62	0.62	0.73	0.62	0.62	0.62	0.71	0.62	0.62	0.62	0.70	0.62	0.62	0.62	0.62
	$\delta_1 = \delta_2$																					
T=100	0	0	0.99	0.82	0.81	0.81	0.80	0.80	0.80	0.79	0.79	0.79	0.79	0.79	0.79	0.79	0.79	0.79	0.79	0.79	0.79	0.79
	0.3	0.3	0.60	0.53	0.53	0.53	0.53	0.53	0.53	0.53	0.53	0.53	0.53	0.53	0.53	0.53	0.53	0.53	0.53	0.53	0.53	0.53
	0.5	0.5	0.53	0.48	0.48	0.48	0.48	0.48	0.49	0.49	0.49	0.49	0.49	0.49	0.49	0.49	0.49	0.49	0.49	0.49	0.49	0.49
T=500	0	0	0.99	0.99	0.99	0.99	0.99	0.99	0.99	0.99	0.99	0.99	0.98	0.98	0.98	0.97	0.97	0.96	0.96	0.96	0.96	0.95
	0.3	0.3	1.14	1.01	0.59	0.59	0.59	0.59	0.59	0.59	0.59	0.59	0.59	0.59	0.59	0.59	0.59	0.59	0.59	0.59	0.59	0.59
	0.5	0.5	0.64	0.64	0.54	0.54	0.54	0.54	0.54	0.54	0.54	0.54	0.54	0.54	0.54	0.54	0.54	0.54	0.55	0.55	0.55	0.55
T=5,000	0	0	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
	0.3	0.3	1.30	1.19	0.67	1.04	0.67	0.67	0.67	0.85	0.68	0.68	0.68	0.78	0.68	0.68	0.68	0.76	0.68	0.68	0.68	0.68
	0.5	0.5	0.69	0.67	0.55	0.66	0.56	0.56	0.56	0.63	0.56	0.56	0.56	0.63	0.56	0.56	0.56	0.63	0.56	0.56	0.56	0.56

References

James, G., D. Witten, T. Hastie, and R. Tibshirani (2013). *An introduction to statistical learning: with applications in R*. New York : Springer.