# Supplement to "Irregular N2SLS and LASSO estimation of the matrix exponential spatial specification model"

Fei Jin[a] and Lung-fei Lee[b*]

[a]School of Economics, Fudan University, Shanghai 200433, China

[b]Department of Economics, The Ohio State University, Columbus, OH 43210, USA

June 30, 2018

## A Additional Monte Carlo results

### A.1 Additional Monte Carlo results for model (21) with a row-normalized $W_n$

Tables A.1–A.4 present Monte Carlo results for model (21) with a row-normalized $W_n$ in addition to those in the main text. The experimental designs are the same as in the main paper. Table A.1 reports the ratios of the RMSE when $n = 144$ to that when $n = 400$ for the N2SLS, N2SLS-r and AGLASSO estimators. Tables A.2–A.4 present biases, SEs and CPs when $n = 400$, while the main paper only presents those when $n = 144$.

### A.2 Monte Carlo results for a non-row-normalized $W_n$

We also conduct some Monte Carlo experiments for a MESS model with a spatial weights matrix $W_n$ that is not row-normalized:

$$e^{\alpha W_n} Y_n = X_{n1}\beta_{11} + l_n\beta_{12} + W_n l_n \beta_2 + W_n X_{n1}\beta_3 + Z_n\beta_4 + V_n, \tag{A.1}$$

where $W_n$ is an $n \times n$ matrix of group interactions. For $W_n$, individuals in a group interact with each other in the same group but do not interact with members in other groups. The $(i, j)$th element of $W_n$ is 1 if individual $i$ interacts with individual $j$; it is 0 otherwise. The group sizes are 9, 7, 5 and 3 in cycle. Since $W_n$ is not row-normalized, $W_n l_n$ is included in (A.1). We set $\delta_0 = (\beta_{11,0}, \beta_{12,0})'$ to $(1, 1)'$. The true parameter $\zeta_0 = (\beta_{20}, \beta_{30}, \beta_{40})'$ is $(0, 0, 0)'$, $(0, 1, 1)'$ or $(0, 0, 1)'$. We use the IV matrix $[l_n, W_n l_n, X_{n1}, W_n X_{n1}, W_n^2 X_{n1}, \bar{Z}_n, W_n \bar{Z}_n]$ in the estimation. For the investigation of powers of test statistics, the data are generated by MESS models with $\zeta_0$ values being $(0, 1, 0.5)'$, $(0, 1, 1)'$, $(0, 1, 1.5)'$, $(0, 1, 2)'$, $(0, 1, 2.5)'$, or $(0, 1, 3)'$. Other designs are the same as for model (21).

*Corresponding author. Tel.: +1 614 292 5508; fax: +1 614 292 4192. E-mail addresses: jin.fei@live.com (F. Jin), lee.1777@osu.edu (L.-F. Lee).

Table A.1: Ratios of the RMSE when $n = 144$ to that when $n = 400$ in model (21)

| | $\alpha$ | $\beta_1$ | $\beta_2$ | $\beta_3$ | $\beta_4$ |
|---|---|---|---|---|---|
| $\zeta_0 = 0$ | | | | | |
| queen, $R^2 = 0.2$, $\alpha_0 = -0.2$ | 1.308[1.750]1.940 | 1.728[1.715]1.723 | 1.237[1.756]1.959 | 1.300[—]— | 1.697[—]— |
| queen, $R^2 = 0.2$, $\alpha_0 = -1$ | 1.222[1.620]1.723 | 1.623[1.647]1.650 | 1.264[1.711]1.736 | 1.200[—]— | 1.680[—]— |
| rook, $R^2 = 0.2$, $\alpha_0 = -0.2$ | 1.258[1.594]1.841 | 1.664[1.669]1.676 | 1.278[1.611]1.933 | 1.289[—]— | 1.654[—]— |
| rook, $R^2 = 0.2$, $\alpha_0 = -1$ | 1.309[1.649]1.819 | 1.688[1.709]1.711 | 1.613[1.714]1.869 | 1.285[—]— | 1.695[—]— |
| queen, $R^2 = 0.8$, $\alpha_0 = -0.2$ | 1.355[1.700]1.933 | 1.703[1.648]1.648 | 1.332[1.751]2.219 | 1.327[—]— | 1.692[—]— |
| queen, $R^2 = 0.8$, $\alpha_0 = -1$ | 1.288[1.645]1.849 | 1.656[1.651]1.637 | 1.380[1.717]1.794 | 1.256[—]— | 1.730[—]— |
| rook, $R^2 = 0.8$, $\alpha_0 = -0.2$ | 1.302[1.580]1.816 | 1.674[1.684]1.688 | 1.368[1.609]1.854 | 1.308[—]— | 1.725[—]— |
| rook, $R^2 = 0.8$, $\alpha_0 = -1$ | 1.324[1.627]1.813 | 1.714[1.727]1.715 | 1.578[1.686]1.933 | 1.296[—]— | 1.708[—]— |
| $\zeta_0 = (1, 1)$ | | | | | |
| queen, $R^2 = 0.2$, $\alpha_0 = -0.2$ | 1.742[1.487]1.742 | 1.686[1.642]1.686 | 1.808[1.654]1.808 | 1.651[—]1.651 | 1.741[—]1.741 |
| queen, $R^2 = 0.2$, $\alpha_0 = -1$ | 1.723[1.484]1.722 | 1.684[1.636]1.684 | 1.758[1.611]1.758 | 1.729[—]1.729 | 1.781[—]1.781 |
| rook, $R^2 = 0.2$, $\alpha_0 = -0.2$ | 1.746[1.483]1.746 | 1.610[1.647]1.610 | 1.716[1.629]1.716 | 1.625[—]1.625 | 1.688[—]1.688 |
| rook, $R^2 = 0.2$, $\alpha_0 = -1$ | 1.710[1.468]1.710 | 1.692[1.563]1.692 | 1.697[1.618]1.697 | 1.730[—]1.730 | 1.725[—]1.725 |
| queen, $R^2 = 0.8$, $\alpha_0 = -0.2$ | 2.004[1.101]1.911 | 1.653[1.538]1.657 | 2.089[1.055]2.035 | 1.794[—]1.772 | 1.726[—]1.977 |
| queen, $R^2 = 0.8$, $\alpha_0 = -1$ | 1.908[1.111]1.977 | 1.685[1.575]1.686 | 2.230[1.042]1.831 | 1.693[—]1.695 | 1.807[—]2.165 |
| rook, $R^2 = 0.8$, $\alpha_0 = -0.2$ | 1.777[1.097]1.912 | 1.718[1.437]1.725 | 1.769[1.046]1.800 | 1.688[—]1.758 | 1.781[—]1.907 |
| rook, $R^2 = 0.8$, $\alpha_0 = -1$ | 1.834[1.045]1.855 | 1.694[1.418]1.695 | 1.762[1.026]1.781 | 1.810[—]1.815 | 1.712[—]1.817 |
| $\zeta_0 = (0, 1)$ | | | | | |
| queen, $R^2 = 0.2$, $\alpha_0 = -0.2$ | 1.771[1.716]1.771 | 1.688[1.666]1.688 | 1.743[1.824]1.743 | 1.648[—]1.648 | 1.711[—]1.711 |
| queen, $R^2 = 0.2$, $\alpha_0 = -1$ | 1.712[1.639]1.712 | 1.639[1.655]1.639 | 1.751[1.745]1.751 | 1.747[—]1.747 | 1.719[—]1.719 |
| rook, $R^2 = 0.2$, $\alpha_0 = -0.2$ | 1.776[1.652]1.776 | 1.678[1.665]1.678 | 1.740[1.663]1.739 | 1.759[—]1.759 | 1.688[—]1.688 |
| rook, $R^2 = 0.2$, $\alpha_0 = -1$ | 1.744[1.667]1.744 | 1.655[1.633]1.655 | 1.800[1.710]1.800 | 1.710[—]1.710 | 1.672[—]1.672 |
| queen, $R^2 = 0.8$, $\alpha_0 = -0.2$ | 1.534[1.692]1.488 | 1.710[1.662]1.717 | 1.449[1.966]1.650 | 1.481[—]1.439 | 1.674[—]1.965 |
| queen, $R^2 = 0.8$, $\alpha_0 = -1$ | 1.553[1.657]1.492 | 1.741[1.643]1.757 | 1.672[2.118]1.451 | 1.511[—]1.425 | 1.702[—]2.192 |
| rook, $R^2 = 0.8$, $\alpha_0 = -0.2$ | 1.641[1.633]1.571 | 1.680[1.578]1.689 | 1.978[1.771]1.861 | 1.592[—]1.537 | 1.685[—]2.005 |
| rook, $R^2 = 0.8$, $\alpha_0 = -1$ | 1.683[1.661]1.585 | 1.814[1.675]1.827 | 2.476[1.910]1.798 | 1.625[—]1.570 | 1.728[—]2.078 |

The numbers show the ratios of the RMSE when $n = 144$ to that when $n = 400$ in each case. The three numbers in each cell correspond to: N2SLS[N2SLS-r]AGLASSO. $\beta_{10} = 1$ and $\beta_{20} = 1$.

Table A.2: Biases, SEs and CPs when $\zeta_0 = 0$ and $n = 400$ in model (21)

| | $\alpha$ | $\beta_1$ | $\beta_2$ | $\beta_3$ | $\beta_4$ |
|---|---|---|---|---|---|
| queen, $R^2 = 0.2$, $\alpha_0 = -0.2$ | | | | | |
| N2SLS | -0.142[0.537]0.954 | 0.018[0.092]0.947 | 0.005[0.595]0.875 | -0.090[0.540]0.963 | -0.000[0.025]0.958 |
| N2SLS-r | -0.006[0.132]0.949 | -0.003[0.087]0.935 | 0.002[0.143]0.945 | 0.000[0.000]— | 0.000[0.000]— |
| AGLASSO | -0.007[0.136]0.948 | -0.002[0.087]0.935 | 0.002[0.145]0.946 | -0.001[0.026]— | 0.000[0.000]— |
| queen, $R^2 = 0.2$, $\alpha_0 = -1$ | | | | | |
| N2SLS | 0.009[0.608]0.915 | 0.021[0.093]0.947 | 0.231[0.908]0.905 | 0.073[0.626]0.910 | -0.000[0.024]0.958 |
| N2SLS-r | -0.013[0.135]0.944 | -0.003[0.086]0.928 | -0.004[0.143]0.937 | 0.000[0.000]— | 0.000[0.000]— |
| AGLASSO | -0.014[0.137]0.943 | -0.002[0.086]0.929 | -0.005[0.143]0.934 | -0.001[0.027]— | 0.000[0.003]— |
| rook, $R^2 = 0.2$, $\alpha_0 = -0.2$ | | | | | |
| N2SLS | 0.119[0.360]0.957 | 0.015[0.092]0.946 | 0.205[0.465]0.972 | 0.115[0.357]0.982 | -0.000[0.025]0.950 |
| N2SLS-r | -0.001[0.098]0.947 | -0.004[0.088]0.931 | 0.005[0.112]0.938 | 0.000[0.000]— | 0.000[0.000]— |
| AGLASSO | -0.001[0.098]0.947 | -0.004[0.088]0.931 | 0.005[0.112]0.939 | 0.000[0.000]— | 0.000[0.002]— |
| rook, $R^2 = 0.2$, $\alpha_0 = -1$ | | | | | |
| N2SLS | 0.164[0.382]0.929 | 0.019[0.090]0.948 | 0.270[0.535]0.973 | 0.154[0.364]0.977 | 0.000[0.024]0.957 |
| N2SLS-r | 0.003[0.097]0.947 | -0.001[0.085]0.939 | 0.008[0.110]0.953 | 0.000[0.000]— | 0.000[0.000]— |
| AGLASSO | 0.004[0.099]0.946 | -0.001[0.085]0.938 | 0.009[0.117]0.953 | 0.001[0.026]— | 0.000[0.002]— |
| queen, $R^2 = 0.8$, $\alpha_0 = -0.2$ | | | | | |
| N2SLS | -0.118[0.524]0.961 | 0.018[0.091]0.947 | 0.026[0.608]0.886 | -0.071[0.527]0.964 | 0.001[0.102]0.946 |
| N2SLS-r | -0.005[0.131]0.960 | -0.001[0.088]0.934 | 0.004[0.145]0.954 | 0.000[0.000]— | 0.000[0.000]— |
| AGLASSO | -0.006[0.135]0.958 | -0.001[0.088]0.934 | 0.004[0.146]0.953 | -0.001[0.025]— | 0.000[0.011]— |
| queen, $R^2 = 0.8$, $\alpha_0 = -1$ | | | | | |
| N2SLS | 0.014[0.590]0.923 | 0.018[0.094]0.939 | 0.220[0.864]0.913 | 0.068[0.613]0.920 | 0.004[0.098]0.955 |
| N2SLS-r | -0.004[0.137]0.951 | -0.004[0.088]0.928 | 0.004[0.149]0.947 | 0.000[0.000]— | 0.000[0.000]— |
| AGLASSO | -0.004[0.138]0.949 | -0.004[0.088]0.929 | 0.005[0.153]0.946 | 0.001[0.022]— | 0.000[0.012]— |
| rook, $R^2 = 0.8$, $\alpha_0 = -0.2$ | | | | | |
| N2SLS | 0.126[0.364]0.957 | 0.017[0.091]0.949 | 0.212[0.464]0.973 | 0.123[0.361]0.982 | 0.003[0.097]0.955 |
| N2SLS-r | -0.001[0.100]0.941 | -0.002[0.086]0.941 | 0.004[0.111]0.938 | 0.000[0.000]— | 0.000[0.000]— |
| AGLASSO | -0.000[0.101]0.941 | -0.002[0.086]0.941 | 0.004[0.113]0.938 | 0.001[0.018]— | 0.000[0.010]— |
| rook, $R^2 = 0.8$, $\alpha_0 = -1$ | | | | | |
| N2SLS | 0.166[0.377]0.937 | 0.018[0.089]0.946 | 0.270[0.533]0.981 | 0.161[0.359]0.971 | 0.002[0.097]0.952 |
| N2SLS-r | 0.001[0.096]0.951 | -0.001[0.085]0.941 | 0.004[0.108]0.947 | 0.000[0.000]— | 0.000[0.000]— |
| AGLASSO | 0.001[0.096]0.952 | -0.001[0.084]0.941 | 0.004[0.108]0.948 | 0.000[0.004]— | 0.000[0.006]— |

"N2SLS" denotes the unrestricted N2SLS estimator and "N2SLS-r" denotes the restricted N2SLS estimator with the restriction $\zeta = 0$ imposed. The three numbers in each cell are: bias[SE]CP. $\beta_{10} = 1$ and $\beta_{20} = 1$.

Table A.3: Biases, SEs and CPs when $\zeta_0 = (1, 1)'$ and $n = 400$ in model (21)

| | $\alpha$ | $\beta_1$ | $\beta_2$ | $\beta_3$ | $\beta_4$ |
|---|---|---|---|---|---|
| queen, $R^2 = 0.2$, $\alpha_0 = -0.2$ | | | | | |
| N2SLS | 0.000[0.064]0.944 | 0.001[0.087]0.942 | 0.004[0.082]0.944 | -0.004[0.164]0.941 | 0.000[0.023]0.952 |
| N2SLS-r | -0.109[0.176]1.000 | -0.015[0.219]0.884 | -0.088[0.245]0.965 | -1.000[0.000]— | -1.000[0.000]— |
| AGLASSO | 0.000[0.064]0.944 | 0.001[0.087]0.942 | 0.004[0.082]0.944 | -0.005[0.164]0.941 | 0.000[0.023]0.952 |
| queen, $R^2 = 0.2$, $\alpha_0 = -1$ | | | | | |
| N2SLS | 0.002[0.066]0.943 | 0.003[0.088]0.931 | 0.003[0.084]0.947 | 0.004[0.163]0.944 | -0.001[0.023]0.954 |
| N2SLS-r | -0.107[0.178]1.000 | -0.006[0.222]0.883 | -0.093[0.253]0.970 | -1.000[0.000]— | -1.000[0.000]— |
| AGLASSO | 0.002[0.066]0.943 | 0.003[0.088]0.931 | 0.003[0.084]0.948 | 0.004[0.163]0.944 | -0.001[0.023]0.954 |
| rook, $R^2 = 0.2$, $\alpha_0 = -0.2$ | | | | | |
| N2SLS | -0.000[0.041]0.958 | 0.002[0.088]0.939 | 0.000[0.066]0.950 | 0.000[0.108]0.944 | -0.000[0.022]0.952 |
| N2SLS-r | -0.083[0.130]1.000 | -0.016[0.234]0.876 | -0.078[0.234]0.966 | -1.000[0.000]— | -1.000[0.000]— |
| AGLASSO | -0.000[0.041]0.958 | 0.002[0.088]0.939 | 0.000[0.066]0.950 | 0.000[0.108]0.944 | -0.000[0.022]0.952 |
| rook, $R^2 = 0.2$, $\alpha_0 = -1$ | | | | | |
| N2SLS | 0.001[0.043]0.940 | 0.001[0.085]0.942 | 0.003[0.067]0.947 | -0.002[0.105]0.947 | 0.001[0.023]0.942 |
| N2SLS-r | -0.078[0.137]1.000 | -0.035[0.234]0.880 | -0.068[0.240]0.961 | -1.000[0.000]— | -1.000[0.000]— |
| AGLASSO | 0.001[0.043]0.940 | 0.001[0.085]0.942 | 0.003[0.067]0.947 | -0.002[0.105]0.947 | 0.001[0.023]0.942 |
| queen, $R^2 = 0.8$, $\alpha_0 = -0.2$ | | | | | |
| N2SLS | -0.015[0.209]0.952 | 0.000[0.091]0.945 | 0.007[0.221]0.955 | -0.018[0.307]0.957 | -0.005[0.096]0.944 |
| N2SLS-r | -0.584[0.202]0.287 | -0.055[0.146]0.862 | -0.429[0.131]0.249 | -1.000[0.000]— | -1.000[0.000]— |
| AGLASSO | -0.017[0.222]0.951 | 0.000[0.091]0.945 | 0.006[0.222]0.954 | -0.020[0.311]0.956 | -0.005[0.096]0.944 |
| queen, $R^2 = 0.8$, $\alpha_0 = -1$ | | | | | |
| N2SLS | -0.009[0.217]0.944 | 0.000[0.092]0.947 | 0.013[0.228]0.939 | -0.005[0.327]0.944 | -0.004[0.093]0.954 |
| N2SLS-r | -0.593[0.199]0.269 | -0.054[0.142]0.867 | -0.437[0.131]0.221 | -1.000[0.000]— | -1.000[0.000]— |
| AGLASSO | -0.009[0.217]0.944 | 0.000[0.092]0.947 | 0.013[0.228]0.939 | -0.005[0.327]0.944 | -0.004[0.093]0.954 |
| rook, $R^2 = 0.8$, $\alpha_0 = -0.2$ | | | | | |
| N2SLS | -0.005[0.138]0.950 | 0.001[0.092]0.941 | 0.004[0.141]0.953 | -0.007[0.170]0.953 | -0.004[0.089]0.956 |
| N2SLS-r | -0.503[0.188]0.275 | -0.105[0.150]0.798 | -0.384[0.130]0.282 | -1.000[0.000]— | -1.000[0.000]— |
| AGLASSO | -0.005[0.137]0.949 | 0.001[0.092]0.941 | 0.004[0.141]0.953 | -0.007[0.169]0.953 | -0.004[0.091]0.956 |
| rook, $R^2 = 0.8$, $\alpha_0 = -1$ | | | | | |
| N2SLS | -0.004[0.131]0.952 | -0.001[0.091]0.950 | 0.005[0.138]0.957 | -0.009[0.165]0.952 | -0.005[0.092]0.950 |
| N2SLS-r | -0.505[0.204]0.282 | -0.107[0.148]0.796 | -0.385[0.139]0.285 | -1.000[0.000]— | -1.000[0.000]— |
| AGLASSO | -0.004[0.130]0.952 | -0.001[0.091]0.950 | 0.005[0.137]0.957 | -0.009[0.164]0.952 | -0.005[0.094]0.950 |

"N2SLS" denotes the unrestricted N2SLS estimator and "N2SLS-r" denotes the restricted N2SLS estimator with the restriction $\zeta = 0$ imposed. The three numbers in each cell are: bias[SE]CP. $\beta_{10} = 1$ and $\beta_{20} = 1$.

Table A.4: Biases, SEs and CPs when $\zeta_0 = (0, 1)'$ and $n = 400$ in model (21)

| | $\alpha$ | $\beta_1$ | $\beta_2$ | $\beta_3$ | $\beta_4$ |
|---|---|---|---|---|---|
| queen, $R^2 = 0.2$, $\alpha_0 = -0.2$ | | | | | |
| N2SLS | 0.000[0.067]0.952 | 0.001[0.087]0.943 | 0.002[0.087]0.946 | -0.005[0.153]0.946 | -0.001[0.025]0.951 |
| N2SLS-r | 0.011[0.179]1.000 | 0.001[0.206]0.890 | 0.035[0.270]0.989 | 0.000[0.000]— | -1.000[0.000]— |
| AGLASSO | 0.000[0.067]0.952 | 0.001[0.087]0.943 | 0.002[0.087]0.946 | -0.005[0.153]0.946 | -0.001[0.025]0.951 |
| queen, $R^2 = 0.2$, $\alpha_0 = -1$ | | | | | |
| N2SLS | 0.000[0.070]0.948 | 0.000[0.089]0.932 | 0.003[0.087]0.955 | 0.006[0.150]0.953 | -0.001[0.025]0.945 |
| N2SLS-r | 0.007[0.192]1.000 | 0.003[0.215]0.866 | 0.030[0.274]0.989 | 0.000[0.000]— | -1.000[0.000]— |
| AGLASSO | 0.000[0.070]0.948 | 0.000[0.089]0.932 | 0.003[0.087]0.955 | 0.006[0.150]0.954 | -0.001[0.025]0.945 |
| rook, $R^2 = 0.2$, $\alpha_0 = -0.2$ | | | | | |
| N2SLS | -0.001[0.048]0.958 | -0.001[0.086]0.934 | 0.001[0.069]0.956 | 0.002[0.105]0.960 | -0.000[0.025]0.942 |
| N2SLS-r | 0.005[0.136]1.000 | -0.008[0.213]0.875 | 0.009[0.233]0.978 | 0.000[0.000]— | -1.000[0.000]— |
| AGLASSO | -0.001[0.048]0.958 | -0.001[0.086]0.934 | 0.001[0.069]0.956 | 0.002[0.105]0.960 | -0.000[0.025]0.942 |
| rook, $R^2 = 0.2$, $\alpha_0 = -1$ | | | | | |
| N2SLS | 0.001[0.049]0.943 | 0.003[0.086]0.944 | 0.002[0.070]0.952 | 0.001[0.107]0.952 | -0.000[0.026]0.942 |
| N2SLS-r | 0.009[0.133]1.000 | 0.013[0.215]0.880 | 0.020[0.229]0.979 | 0.000[0.000]— | -1.000[0.000]— |
| AGLASSO | 0.001[0.049]0.943 | 0.003[0.086]0.944 | 0.002[0.070]0.952 | 0.001[0.107]0.952 | -0.000[0.026]0.941 |
| queen, $R^2 = 0.8$, $\alpha_0 = -0.2$ | | | | | |
| N2SLS | 0.001[0.304]0.947 | 0.006[0.089]0.943 | 0.057[0.469]0.940 | 0.017[0.343]0.945 | -0.008[0.105]0.948 |
| N2SLS-r | -0.004[0.213]0.953 | -0.002[0.135]0.901 | 0.020[0.245]0.946 | 0.000[0.000]— | -1.000[0.000]— |
| AGLASSO | -0.000[0.301]0.948 | 0.006[0.089]0.943 | 0.052[0.420]0.940 | 0.016[0.341]0.944 | -0.009[0.108]0.947 |
| queen, $R^2 = 0.8$, $\alpha_0 = -1$ | | | | | |
| N2SLS | 0.013[0.307]0.946 | 0.006[0.088]0.941 | 0.068[0.421]0.938 | 0.027[0.341]0.952 | -0.010[0.104]0.947 |
| N2SLS-r | -0.001[0.214]0.952 | 0.002[0.135]0.891 | 0.022[0.245]0.944 | 0.000[0.000]— | -1.000[0.000]— |
| AGLASSO | 0.011[0.304]0.947 | 0.006[0.088]0.941 | 0.064[0.407]0.938 | 0.025[0.337]0.952 | -0.010[0.104]0.947 |
| rook, $R^2 = 0.8$, $\alpha_0 = -0.2$ | | | | | |
| N2SLS | -0.002[0.216]0.932 | 0.006[0.088]0.939 | 0.021[0.231]0.938 | 0.005[0.231]0.945 | -0.011[0.105]0.953 |
| N2SLS-r | -0.004[0.155]0.944 | -0.002[0.136]0.892 | 0.008[0.179]0.944 | 0.000[0.000]— | -1.000[0.000]— |
| AGLASSO | -0.002[0.216]0.932 | 0.006[0.088]0.939 | 0.021[0.231]0.937 | 0.005[0.231]0.943 | -0.011[0.109]0.952 |
| rook, $R^2 = 0.8$, $\alpha_0 = -1$ | | | | | |
| N2SLS | 0.005[0.216]0.934 | 0.004[0.087]0.944 | 0.030[0.242]0.941 | 0.002[0.231]0.946 | -0.008[0.104]0.951 |
| N2SLS-r | 0.006[0.159]0.944 | -0.002[0.135]0.888 | 0.018[0.187]0.933 | 0.000[0.000]— | -1.000[0.000]— |
| AGLASSO | 0.004[0.214]0.934 | 0.004[0.087]0.944 | 0.028[0.240]0.941 | 0.001[0.229]0.946 | -0.008[0.108]0.951 |

"N2SLS" denotes the unrestricted N2SLS estimator and "N2SLS-r" denotes the restricted N2SLS estimator with the restriction $\zeta = 0$ imposed. The three numbers in each cell are: bias[SE]CP. $\beta_{10} = 1$ and $\beta_{20} = 1$.

The Monte Carlo results are reported in Tables A.5–A.11. The patterns are similar to those for model (21) with a row-normalized spatial weights matrix.

Table A.5: Probabilities that the AGLASSO estimator selects the right model in model (A.1)

| | $n = 144$ | | | $n = 400$ | | |
|---|---|---|---|---|---|---|
| | $\zeta_0 = 0$ | $\zeta_0 = (0,1,1)'$ | $\zeta_0 = (0,0,1)'$ | $\zeta_0 = 0$ | $\zeta_0 = (0,1,1)'$ | $\zeta_0 = (0,0,1)'$ |
| $R^2 = 0.2$, $\alpha_0 = -0.2$ | 0.982 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 |
| $R^2 = 0.2$, $\alpha_0 = -1$ | 0.986 | 1.000 | 0.995 | 1.000 | 1.000 | 1.000 |
| $R^2 = 0.8$, $\alpha_0 = -0.2$ | 0.972 | 1.000 | 0.961 | 1.000 | 1.000 | 0.995 |
| $R^2 = 0.8$, $\alpha_0 = -1$ | 0.978 | 1.000 | 0.941 | 0.999 | 1.000 | 0.991 |

The numbers denote the proportions of Monte Carlo repetitions where the AGLASSO estimate $\hat{\zeta}_n = 0$ when $\zeta_0 = 0$, or $\hat{\zeta}_n \neq 0$ when $\zeta_0 \neq 0$. $\beta_{11,0} = 1$, and $\beta_{12,0} = 1$ and $\beta_{20} = 1$.

# B Low level conditions for Assumptions 3, 5 and 7

In Assumptions 3, 5 and 7, high level conditions are assumed for laws of large numbers on some terms related to the exogenous variables $X_{n1}$, the endogenous regressors $Z_n$ and the IVs $F_n$. In this section, we discuss low level conditions for them.

For spatial variables, it is appropriate to allow for spatial dependence. For generality, we assumed that observations of $X_{n1}$, $Z_n$ and $F_n$ are near-epoch dependent (NED), as developed for spatial processes in Jenish and Prucha (2012). NED processes are general and have several advantages (e.g., they include important classes of dependent processes such as linear processes with discrete innovations and nonlinear infinite moving average random fields under mild conditions). In addition to spatial dependence, it allows also heterogeneity. We first introduce some notations and then give the definition of near-epoch dependence. Let $M \subset \mathbb{R}^m, m \geq 1$, be a lattice of (possibly) unevenly placed locations in $\mathbb{R}^m$, and let $H = \{h_{ni}, i \in M_n, n \geq 1\}$ and $\epsilon = \{\epsilon_{ni}, i \in M_n, n \geq 1\}$ be triangular arrays of random fields defined on a probability space $(\Omega, \mathscr{F}, P)$ with $M_n \subset M$.[1] Equip $\mathbb{R}^m$ with the metric $\rho(i,j) = \max_{1 \leq l \leq m} |j_l - i_l|$, where $i_l$ is the $l$th component of $i$. Denote the cardinality of a finite subset $U \subset M$ as $|U|$. Let $h_{ni}$'s and $\epsilon_{ni}$'s take values in $\mathbb{R}^{k_h}$ and $\mathbb{R}^{k_\epsilon}$, respectively. Assume that $\mathbb{R}^{k_h}$ and $\mathbb{R}^{k_\epsilon}$ are normed spaces equipped with the Euclidean norm, denoted as $\|\cdot\|_E$. Denote the $L_p$-norm of any random vector $Y$ as $\|Y\|_p = [\mathrm{E}(\|Y\|_E)^p]^{1/p}$. Furthermore, $\mathscr{F}_{ni}(s) = \sigma(\epsilon_{nj}; j \in M_n, \rho(i,j) \leq s)$ denotes the $\sigma$-field generated by the random vectors $\epsilon_{nj}$'s located in the $s$-neighborhood of location $i$. As in Jenish and Prucha (2012), we make the following assumption to ensure growth of the sample size as the sample region $M_n$ expands, which is referred to as increasing domain asymptotics in the literature.

**Assumption B.1.** *The lattice $M \subset \mathbb{R}^m$, $m \geq 1$, is infinitely countable. The distance between any two elements in $M$ is at least $\rho_0 > 0$ from each other, i.e., $\rho(i,j) \geq \rho_0$ for any $i, j \in M$. We assume w.l.o.g. that $\rho_0 = 1$.*

[1] Each $i$ is located as a point $l(i)$ in $M_n$. $i \in M_n$ is a simplified notation for $l(i) \in M_n$.

Table A.6: Ratios of the SE when $n = 144$ to that when $n = 400$ in model (A.1)

| | $\alpha$ | $\beta_{11}$ | $\beta_{12}$ | $\beta_2$ | $\beta_3$ | $\beta_4$ |
|---|---|---|---|---|---|---|
| $\zeta_0 = 0$ | | | | | | |
| $R^2 = 0.2$, $\alpha_0 = -0.2$ | 1.118[1.724]2.178 | 1.579[1.597]1.606 | 1.028[1.599]1.672 | 0.950[—]— | 1.005[—]— | 1.782[—]— |
| $R^2 = 0.2$, $\alpha_0 = -1$ | 1.130[1.720]2.039 | 1.612[1.710]1.708 | 1.030[1.607]1.635 | 0.941[—]— | 1.074[—]— | 1.772[—]— |
| $R^2 = 0.8$, $\alpha_0 = -0.2$ | 1.195[1.733]2.246 | 1.643[1.704]1.706 | 1.039[1.588]1.617 | 0.975[—]— | 1.105[—]— | 1.771[—]— |
| $R^2 = 0.8$, $\alpha_0 = -1$ | 1.108[1.765]2.193 | 1.589[1.678]1.685 | 1.080[1.727]1.766 | 0.987[—]— | 1.057[—]— | 1.697[—]— |
| $\zeta_0 = (0,1,1)'$ | | | | | | |
| $R^2 = 0.2$, $\alpha_0 = -0.2$ | 2.092[1.726]2.092 | 1.670[1.599]1.670 | 1.692[1.645]1.692 | 1.796[—]1.796 | 1.928[—]1.928 | 1.794[—]1.794 |
| $R^2 = 0.2$, $\alpha_0 = -1$ | 2.036[1.604]2.036 | 1.797[1.608]1.797 | 1.568[1.625]1.568 | 1.625[—]1.625 | 1.984[—]1.984 | 1.812[—]1.812 |
| $R^2 = 0.8$, $\alpha_0 = -0.2$ | 1.910[1.798]1.961 | 1.667[1.590]1.671 | 1.796[1.582]1.798 | 1.988[—]1.988 | 1.897[—]1.906 | 1.771[—]1.773 |
| $R^2 = 0.8$, $\alpha_0 = -1$ | 1.806[1.674]1.991 | 1.691[1.628]1.694 | 1.748[1.615]1.746 | 1.816[—]1.816 | 1.712[—]1.739 | 1.766[—]1.766 |
| $\zeta_0 = (0,0,1)'$ | | | | | | |
| $R^2 = 0.2$, $\alpha_0 = -0.2$ | 1.975[1.707]1.975 | 1.646[1.627]1.646 | 1.719[1.782]1.719 | 1.987[—]1.987 | 1.950[—]1.950 | 1.692[—]1.692 |
| $R^2 = 0.2$, $\alpha_0 = -1$ | 2.081[1.663]2.081 | 1.661[1.667]1.661 | 1.735[1.703]1.735 | 2.034[—]2.034 | 1.973[—]1.973 | 1.748[—]1.748 |
| $R^2 = 0.8$, $\alpha_0 = -0.2$ | 1.611[1.756]1.556 | 1.685[1.562]1.697 | 1.396[1.587]1.470 | 1.215[—]1.243 | 1.545[—]1.477 | 1.820[—]2.138 |
| $R^2 = 0.8$, $\alpha_0 = -1$ | 1.685[1.712]1.592 | 1.762[1.688]1.780 | 1.523[1.602]1.671 | 1.470[—]1.370 | 1.599[—]1.405 | 1.805[—]2.249 |

The numbers show the ratios of the SE when $n = 144$ to that when $n = 400$ in each case. The three numbers in each cell correspond to: N2SLS[N2SLS-r]AGLASSO. The ratios for the N2SLS-r estimates of $\beta_2$, $\beta_3$ and $\beta_4$ are not reported, because those estimates are restricted to zero. $\beta_{11,0} = 1$ and $\beta_{12,0} = 1$. The ratios for the AGLASSO estimates of $\beta_2$, $\beta_3$ and $\beta_4$ when $\zeta_0 = 0$ are not reported either, because Table A.5 shows that those estimates are zero with very high probabilities.

Table A.7: Ratios of the RMSE when $n = 144$ to that when $n = 400$ in model (A.1)

| | $\alpha$ | $\beta_{11}$ | $\beta_{12}$ | $\beta_2$ | $\beta_3$ | $\beta_4$ |
|---|---|---|---|---|---|---|
| $\zeta_0 = 0$ | | | | | | |
| $R^2 = 0.2$, $\alpha_0 = -0.2$ | 1.359[1.784]2.233 | 1.574[1.597]1.606 | 1.070[1.629]1.699 | 0.948[—]— | 1.091[—]— | 1.783[—]— |
| $R^2 = 0.2$, $\alpha_0 = -1$ | 1.327[1.778]2.092 | 1.594[1.709]1.708 | 1.067[1.642]1.668 | 0.935[—]— | 1.127[—]— | 1.774[—]— |
| $R^2 = 0.8$, $\alpha_0 = -0.2$ | 1.335[1.806]2.345 | 1.602[1.705]1.705 | 1.086[1.623]1.656 | 0.974[—]— | 1.144[—]— | 1.771[—]— |
| $R^2 = 0.8$, $\alpha_0 = -1$ | 1.296[1.823]2.257 | 1.567[1.677]1.685 | 1.089[1.745]1.783 | 0.982[—]— | 1.113[—]— | 1.696[—]— |
| $\zeta_0 = (0,1,1)'$ | | | | | | |
| $R^2 = 0.2$, $\alpha_0 = -0.2$ | 2.092[1.099]2.092 | 1.671[1.429]1.671 | 1.691[1.448]1.691 | 1.795[—]1.795 | 1.928[—]1.928 | 1.795[—]1.795 |
| $R^2 = 0.2$, $\alpha_0 = -1$ | 2.036[1.079]2.036 | 1.798[1.447]1.798 | 1.568[1.413]1.568 | 1.625[—]1.625 | 1.986[—]1.986 | 1.813[—]1.813 |
| $R^2 = 0.8$, $\alpha_0 = -0.2$ | 1.911[1.033]1.962 | 1.667[1.064]1.671 | 1.798[1.020]1.800 | 1.990[—]1.990 | 1.897[—]1.906 | 1.771[—]1.773 |
| $R^2 = 0.8$, $\alpha_0 = -1$ | 1.814[1.030]2.002 | 1.692[1.080]1.695 | 1.749[1.025]1.747 | 1.816[—]1.816 | 1.713[—]1.741 | 1.764[—]1.764 |
| $\zeta_0 = (0,0,1)'$ | | | | | | |
| $R^2 = 0.2$, $\alpha_0 = -0.2$ | 1.974[1.709]1.974 | 1.645[1.627]1.645 | 1.719[1.783]1.719 | 1.987[—]1.987 | 1.951[—]1.951 | 1.693[—]1.693 |
| $R^2 = 0.2$, $\alpha_0 = -1$ | 2.080[1.664]2.080 | 1.662[1.667]1.662 | 1.735[1.707]1.735 | 2.036[—]2.036 | 1.973[—]1.973 | 1.749[—]1.749 |
| $R^2 = 0.8$, $\alpha_0 = -0.2$ | 1.797[1.817]1.750 | 1.705[1.561]1.713 | 1.407[1.610]1.479 | 1.214[—]1.248 | 1.583[—]1.531 | 1.821[—]2.137 |
| $R^2 = 0.8$, $\alpha_0 = -1$ | 1.837[1.747]1.769 | 1.776[1.689]1.789 | 1.537[1.614]1.679 | 1.467[—]1.378 | 1.627[—]1.467 | 1.805[—]2.251 |

The numbers show the ratios of the RMSE when $n = 144$ to that when $n = 400$ in each case. The three numbers in each cell correspond to: N2SLS[N2SLS-r]AGLASSO. $\beta_{11,0} = 1$ and $\beta_{12,0} = 1$.

Table A.8: Biases, SEs and CPs when $\zeta_0 = 0$ in model (A.1)

| | $\alpha$ | $\beta_{11}$ | $\beta_{12}$ | $\beta_2$ | $\beta_3$ | $\beta_4$ |
|---|---|---|---|---|---|---|
| $n = 144,\ R^2 = 0.2,\ \alpha_0 = -0.2$ | | | | | | |
| N2SLS | -0.118[0.120]0.946 | 0.064[0.158]0.951 | -0.220[0.381]0.937 | -0.010[0.193]0.922 | -0.062[0.134]0.939 | 0.001[0.045]0.943 |
| N2SLS-r | -0.010[0.027]0.000 | 0.001[0.139]0.000 | -0.042[0.165]0.930 | 0.000[0.000]— | 0.000[0.000]— | 0.000[0.000]— |
| AGLASSO | -0.012[0.034]0.930 | 0.002[0.140]0.906 | -0.043[0.174]0.911 | -0.001[0.016]— | -0.001[0.015]— | 0.000[0.012]— |
| $n = 144,\ R^2 = 0.2,\ \alpha_0 = -1$ | | | | | | |
| N2SLS | -0.110[0.124]0.950 | 0.064[0.166]0.945 | -0.233[0.397]0.938 | -0.000[0.203]0.910 | -0.053[0.150]0.930 | -0.002[0.046]0.938 |
| N2SLS-r | -0.010[0.026]0.000 | 0.001[0.147]0.000 | -0.045[0.162]0.936 | 0.000[0.000]— | 0.000[0.000]— | 0.000[0.000]— |
| AGLASSO | -0.011[0.032]0.936 | 0.002[0.147]0.893 | -0.046[0.166]0.918 | -0.001[0.011]— | -0.001[0.015]— | -0.000[0.009]— |
| $n = 144,\ R^2 = 0.8,\ \alpha_0 = -0.2$ | | | | | | |
| N2SLS | -0.111[0.124]0.947 | 0.060[0.165]0.939 | -0.226[0.388]0.930 | -0.002[0.196]0.905 | -0.053[0.145]0.936 | 0.002[0.186]0.930 |
| N2SLS-r | -0.010[0.027]0.000 | -0.004[0.145]0.000 | -0.045[0.167]0.937 | 0.000[0.000]— | 0.000[0.000]— | 0.000[0.000]— |
| AGLASSO | -0.013[0.036]0.937 | -0.001[0.145]0.895 | -0.047[0.170]0.916 | -0.001[0.015]— | -0.003[0.020]— | 0.006[0.054]— |
| $n = 144,\ R^2 = 0.8,\ \alpha_0 = -1$ | | | | | | |
| N2SLS | -0.110[0.123]0.955 | 0.059[0.164]0.949 | -0.212[0.393]0.940 | -0.004[0.201]0.910 | -0.054[0.146]0.939 | -0.000[0.177]0.953 |
| N2SLS-r | -0.009[0.027]0.000 | -0.002[0.145]0.000 | -0.035[0.171]0.937 | 0.000[0.000]— | 0.000[0.000]— | 0.000[0.000]— |
| AGLASSO | -0.011[0.033]0.937 | -0.000[0.146]0.898 | -0.037[0.176]0.910 | -0.001[0.022]— | -0.002[0.022]— | 0.005[0.041]— |
| $n = 400,\ R^2 = 0.2,\ \alpha_0 = -0.2$ | | | | | | |
| N2SLS | -0.062[0.107]0.918 | 0.041[0.100]0.951 | -0.178[0.371]0.907 | 0.018[0.203]0.875 | -0.023[0.134]0.900 | 0.000[0.025]0.955 |
| N2SLS-r | -0.004[0.016]0.000 | -0.001[0.087]0.000 | -0.016[0.103]0.942 | 0.000[0.000]— | 0.000[0.000]— | 0.000[0.000]— |
| AGLASSO | -0.004[0.016]0.942 | 0.000[0.087]0.934 | -0.018[0.104]0.931 | 0.000[0.000]— | 0.000[0.000]— | 0.000[0.000]— |
| $n = 400,\ R^2 = 0.2,\ \alpha_0 = -1$ | | | | | | |
| N2SLS | -0.060[0.109]0.918 | 0.043[0.103]0.949 | -0.194[0.386]0.907 | 0.024[0.215]0.872 | -0.021[0.140]0.890 | -0.000[0.026]0.945 |
| N2SLS-r | -0.004[0.015]0.000 | -0.001[0.086]0.000 | -0.017[0.101]0.944 | 0.000[0.000]— | 0.000[0.000]— | 0.000[0.000]— |
| AGLASSO | -0.004[0.016]0.944 | -0.000[0.086]0.933 | -0.018[0.101]0.941 | 0.000[0.000]— | 0.000[0.000]— | 0.000[0.000]— |
| $n = 400,\ R^2 = 0.8,\ \alpha_0 = -0.2$ | | | | | | |
| N2SLS | -0.069[0.104]0.929 | 0.044[0.100]0.958 | -0.178[0.373]0.919 | 0.010[0.201]0.893 | -0.031[0.131]0.910 | 0.001[0.105]0.942 |
| N2SLS-r | -0.003[0.016]0.000 | 0.001[0.085]0.000 | -0.017[0.105]0.937 | 0.000[0.000]— | 0.000[0.000]— | 0.000[0.000]— |
| AGLASSO | -0.003[0.016]0.937 | 0.002[0.085]0.935 | -0.017[0.105]0.931 | 0.000[0.000]— | 0.000[0.000]— | 0.000[0.000]— |
| $n = 400,\ R^2 = 0.8,\ \alpha_0 = -1$ | | | | | | |
| N2SLS | -0.063[0.111]0.912 | 0.041[0.103]0.949 | -0.189[0.364]0.902 | 0.021[0.204]0.869 | -0.023[0.138]0.888 | -0.004[0.104]0.954 |
| N2SLS-r | -0.003[0.015]0.000 | -0.003[0.087]0.000 | -0.014[0.099]0.942 | 0.000[0.000]— | 0.000[0.000]— | 0.000[0.000]— |
| AGLASSO | -0.003[0.015]0.942 | -0.002[0.086]0.930 | -0.015[0.100]0.949 | 0.000[0.002]— | -0.000[0.002]— | 0.000[0.008]— |

"N2SLS" denotes the unrestricted N2SLS estimator and "N2SLS-r" denotes the restricted N2SLS estimator with the restriction $\zeta = 0$ imposed. The three numbers in each cell are: bias[SE]CP. $\beta_{11,0} = 1$ and $\beta_{12,0} = 1$.

Table A.9: Biases, SEs and CPs when $\zeta_0 = (0, 1, 1)'$ in model (A.1)

| | $\alpha$ | $\beta_{11}$ | $\beta_{12}$ | $\beta_2$ | $\beta_3$ | $\beta_4$ |
|---|---|---|---|---|---|---|
| $n = 144,\ R^2 = 0.2,\ \alpha_0 = -0.2$ | | | | | | |
| N2SLS | 0.000[0.008]0.935 | 0.005[0.151]0.917 | -0.001[0.252]0.937 | 0.001[0.043]0.936 | 0.000[0.066]0.933 | -0.000[0.018]0.938 |
| N2SLS-r | -0.066[0.036]0.000 | -0.335[0.760]0.000 | -0.286[0.694]0.935 | 0.000[0.000]— | -1.000[0.000]— | -1.000[0.000]— |
| AGLASSO | 0.000[0.008]0.935 | 0.005[0.151]0.918 | -0.001[0.252]0.937 | 0.001[0.043]0.931 | 0.000[0.066]0.929 | -0.000[0.018]0.935 |
| $n = 144,\ R^2 = 0.2,\ \alpha_0 = -1$ | | | | | | |
| N2SLS | 0.000[0.009]0.934 | 0.006[0.156]0.924 | -0.001[0.242]0.951 | 0.000[0.041]0.951 | 0.003[0.069]0.940 | -0.001[0.018]0.947 |
| N2SLS-r | -0.065[0.035]0.000 | -0.321[0.767]0.000 | -0.298[0.680]0.934 | 0.000[0.000]— | -1.000[0.000]— | -1.000[0.000]— |
| AGLASSO | 0.000[0.009]0.934 | 0.006[0.156]0.924 | -0.001[0.242]0.951 | 0.000[0.041]0.949 | 0.003[0.069]0.938 | -0.001[0.018]0.945 |
| $n = 144,\ R^2 = 0.8,\ \alpha_0 = -0.2$ | | | | | | |
| N2SLS | -0.001[0.019]0.938 | 0.002[0.179]0.931 | -0.012[0.272]0.930 | 0.003[0.053]0.935 | 0.000[0.133]0.926 | -0.002[0.070]0.941 |
| N2SLS-r | -0.233[0.043]0.000 | -0.706[0.345]0.000 | -0.704[0.195]0.937 | 0.000[0.000]— | -1.000[0.000]— | -1.000[0.000]— |
| AGLASSO | -0.001[0.020]0.937 | 0.002[0.179]0.931 | -0.013[0.272]0.930 | 0.003[0.053]0.933 | -0.000[0.133]0.924 | -0.002[0.070]0.940 |
| $n = 144,\ R^2 = 0.8,\ \alpha_0 = -1$ | | | | | | |
| N2SLS | -0.002[0.019]0.939 | -0.005[0.183]0.916 | -0.011[0.263]0.922 | 0.000[0.048]0.923 | -0.005[0.126]0.937 | -0.000[0.068]0.944 |
| N2SLS-r | -0.234[0.040]0.000 | -0.714[0.346]0.000 | -0.710[0.204]0.939 | 0.000[0.000]— | -1.000[0.000]— | -1.000[0.000]— |
| AGLASSO | -0.002[0.021]0.939 | -0.006[0.183]0.916 | -0.010[0.263]0.922 | 0.000[0.048]0.920 | -0.007[0.128]0.933 | -0.000[0.068]0.940 |
| $n = 400,\ R^2 = 0.2,\ \alpha_0 = -0.2$ | | | | | | |
| N2SLS | -0.000[0.004]0.953 | -0.001[0.090]0.938 | 0.005[0.149]0.950 | -0.001[0.024]0.950 | -0.000[0.034]0.942 | -0.000[0.010]0.938 |
| N2SLS-r | -0.066[0.021]0.000 | -0.335[0.475]0.000 | -0.301[0.422]0.953 | 0.000[0.000]— | -1.000[0.000]— | -1.000[0.000]— |
| AGLASSO | -0.000[0.004]0.953 | -0.001[0.090]0.938 | 0.005[0.149]0.950 | -0.001[0.024]0.947 | -0.000[0.034]0.940 | -0.000[0.010]0.936 |
| $n = 400,\ R^2 = 0.2,\ \alpha_0 = -1$ | | | | | | |
| N2SLS | 0.000[0.004]0.940 | 0.001[0.087]0.944 | 0.002[0.155]0.939 | -0.000[0.025]0.937 | 0.000[0.035]0.943 | -0.000[0.010]0.949 |
| N2SLS-r | -0.065[0.022]0.000 | -0.321[0.477]0.000 | -0.318[0.418]0.940 | 0.000[0.000]— | -1.000[0.000]— | -1.000[0.000]— |
| AGLASSO | 0.000[0.004]0.940 | 0.001[0.087]0.944 | 0.002[0.155]0.939 | -0.000[0.025]0.935 | 0.000[0.035]0.941 | -0.000[0.010]0.947 |
| $n = 400,\ R^2 = 0.8,\ \alpha_0 = -0.2$ | | | | | | |
| N2SLS | -0.000[0.010]0.947 | -0.001[0.107]0.939 | -0.000[0.152]0.955 | 0.000[0.027]0.948 | -0.000[0.070]0.946 | -0.001[0.040]0.948 |
| N2SLS-r | -0.229[0.024]0.000 | -0.706[0.217]0.000 | -0.705[0.123]0.947 | 0.000[0.000]— | -1.000[0.000]— | -1.000[0.000]— |
| AGLASSO | -0.000[0.010]0.947 | -0.001[0.107]0.939 | -0.000[0.152]0.955 | 0.000[0.027]0.947 | -0.000[0.070]0.946 | -0.001[0.040]0.948 |
| $n = 400,\ R^2 = 0.8,\ \alpha_0 = -1$ | | | | | | |
| N2SLS | -0.000[0.011]0.938 | 0.002[0.108]0.927 | -0.003[0.151]0.949 | 0.001[0.027]0.944 | 0.002[0.073]0.941 | -0.002[0.039]0.948 |
| N2SLS-r | -0.229[0.024]0.000 | -0.704[0.213]0.000 | -0.710[0.126]0.938 | 0.000[0.000]— | -1.000[0.000]— | -1.000[0.000]— |
| AGLASSO | -0.000[0.011]0.938 | 0.002[0.108]0.927 | -0.003[0.151]0.949 | 0.001[0.027]0.942 | 0.002[0.073]0.939 | -0.002[0.039]0.946 |

"N2SLS" denotes the unrestricted N2SLS estimator and "N2SLS-r" denotes the restricted N2SLS estimator with the restriction $\zeta = 0$ imposed. The three numbers in each cell are: bias[SE]CP. $\beta_{11,0} = 1$ and $\beta_{12,0} = 1$.

Table A.10: Biases, SEs and CPs when $\zeta_0 = (0, 0, 1)'$ in model (A.1)

| | $\alpha$ | $\beta_{11}$ | $\beta_{12}$ | $\beta_2$ | $\beta_3$ | $\beta_4$ |
|---|---|---|---|---|---|---|
| $n = 144,\ R^2 = 0.2,\ \alpha_0 = -0.2$ | | | | | | |
| N2SLS | -0.000[0.022]0.944 | 0.003[0.142]0.924 | 0.002[0.264]0.943 | 0.000[0.054]0.946 | 0.001[0.048]0.943 | -0.001[0.043]0.941 |
| N2SLS-r | -0.003[0.047]0.000 | -0.006[0.344]0.000 | 0.018[0.434]0.944 | 0.000[0.000]— | 0.000[0.000]— | -1.000[0.000]— |
| AGLASSO | -0.000[0.022]0.944 | 0.003[0.142]0.924 | 0.002[0.264]0.943 | 0.000[0.054]0.944 | 0.001[0.048]0.941 | -0.001[0.043]0.939 |
| $n = 144,\ R^2 = 0.2,\ \alpha_0 = -1$ | | | | | | |
| N2SLS | -0.000[0.023]0.941 | 0.005[0.144]0.921 | -0.005[0.264]0.940 | 0.003[0.056]0.943 | -0.001[0.047]0.951 | -0.002[0.044]0.945 |
| N2SLS-r | -0.001[0.045]0.000 | 0.002[0.351]0.000 | 0.033[0.419]0.942 | 0.000[0.000]— | 0.000[0.000]— | -1.000[0.000]— |
| AGLASSO | -0.000[0.023]0.942 | 0.005[0.144]0.921 | -0.005[0.264]0.940 | 0.003[0.056]0.941 | -0.001[0.047]0.949 | -0.002[0.044]0.943 |
| $n = 144,\ R^2 = 0.8,\ \alpha_0 = -0.2$ | | | | | | |
| N2SLS | -0.042[0.079]0.958 | 0.025[0.154]0.937 | -0.070[0.286]0.976 | -0.008[0.109]0.936 | -0.023[0.103]0.935 | 0.010[0.185]0.939 |
| N2SLS-r | -0.018[0.046]0.000 | 0.006[0.211]0.000 | -0.066[0.277]0.960 | 0.000[0.000]— | 0.000[0.000]— | -1.000[0.000]— |
| AGLASSO | -0.041[0.073]0.960 | 0.021[0.152]0.936 | -0.063[0.267]0.971 | -0.011[0.094]0.902 | -0.024[0.088]0.901 | -0.015[0.257]0.906 |
| $n = 144,\ R^2 = 0.8,\ \alpha_0 = -1$ | | | | | | |
| N2SLS | -0.038[0.079]0.954 | 0.025[0.160]0.933 | -0.069[0.318]0.981 | -0.003[0.133]0.925 | -0.020[0.104]0.928 | 0.008[0.183]0.944 |
| N2SLS-r | -0.016[0.045]0.000 | 0.001[0.222]0.000 | -0.057[0.277]0.956 | 0.000[0.000]— | 0.000[0.000]— | -1.000[0.000]— |
| AGLASSO | -0.038[0.071]0.956 | 0.021[0.157]0.929 | -0.056[0.283]0.975 | -0.011[0.093]0.889 | -0.024[0.079]0.892 | -0.017[0.257]0.906 |
| $n = 400,\ R^2 = 0.2,\ \alpha_0 = -0.2$ | | | | | | |
| N2SLS | -0.000[0.011]0.942 | 0.002[0.086]0.943 | -0.000[0.154]0.944 | 0.000[0.027]0.946 | 0.000[0.025]0.944 | -0.001[0.025]0.948 |
| N2SLS-r | -0.001[0.027]0.000 | 0.004[0.212]0.000 | 0.006[0.244]0.942 | 0.000[0.000]— | 0.000[0.000]— | -1.000[0.000]— |
| AGLASSO | -0.000[0.011]0.942 | 0.002[0.086]0.943 | -0.000[0.154]0.944 | 0.000[0.027]0.944 | 0.000[0.025]0.942 | -0.001[0.025]0.946 |
| $n = 400,\ R^2 = 0.2,\ \alpha_0 = -1$ | | | | | | |
| N2SLS | -0.000[0.011]0.941 | 0.000[0.087]0.931 | -0.004[0.152]0.942 | 0.000[0.027]0.947 | 0.000[0.024]0.951 | 0.001[0.025]0.946 |
| N2SLS-r | -0.001[0.027]0.000 | 0.001[0.210]0.000 | 0.009[0.246]0.941 | 0.000[0.000]— | 0.000[0.000]— | -1.000[0.000]— |
| AGLASSO | -0.000[0.011]0.941 | 0.000[0.087]0.931 | -0.004[0.152]0.942 | 0.000[0.027]0.943 | 0.000[0.024]0.947 | 0.001[0.025]0.941 |
| $n = 400,\ R^2 = 0.8,\ \alpha_0 = -0.2$ | | | | | | |
| N2SLS | -0.008[0.049]0.953 | 0.004[0.091]0.936 | -0.043[0.205]0.967 | 0.008[0.090]0.935 | -0.001[0.067]0.943 | -0.005[0.102]0.953 |
| N2SLS-r | -0.007[0.026]0.000 | -0.006[0.135]0.000 | -0.029[0.174]0.956 | 0.000[0.000]— | 0.000[0.000]— | -1.000[0.000]— |
| AGLASSO | -0.009[0.047]0.956 | 0.003[0.090]0.937 | -0.038[0.182]0.967 | 0.005[0.075]0.930 | -0.002[0.060]0.939 | -0.008[0.120]0.947 |
| $n = 400,\ R^2 = 0.8,\ \alpha_0 = -1$ | | | | | | |
| N2SLS | -0.008[0.047]0.954 | 0.008[0.091]0.934 | -0.036[0.209]0.971 | 0.006[0.091]0.940 | -0.001[0.065]0.947 | -0.004[0.102]0.957 |
| N2SLS-r | -0.007[0.026]0.000 | -0.000[0.131]0.000 | -0.029[0.173]0.958 | 0.000[0.000]— | 0.000[0.000]— | -1.000[0.000]— |
| AGLASSO | -0.009[0.044]0.958 | 0.008[0.088]0.936 | -0.029[0.169]0.973 | 0.003[0.068]0.935 | -0.003[0.056]0.942 | -0.006[0.114]0.951 |

"N2SLS" denotes the unrestricted N2SLS estimator and "N2SLS-r" denotes the restricted N2SLS estimator with the restriction $\zeta = 0$ imposed. The three numbers in each cell are: bias[SE]CP. $\beta_{11,0} = 1$ and $\beta_{12,0} = 1$.

Table A.11: Size and power of the distance difference and gradient tests in model (A.1)

| | distance difference test | | | | | | | gradient test | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | size | power | | | | | | size | power | | | | | |
| | | (1) | (2) | (3) | (4) | (5) | (6) | | (1) | (2) | (3) | (4) | (5) | (6) |
| $W_n$, $R^2$, $\alpha_0$ | n=144 | | | | | | | | | | | | | |
| queen, 0.2, $-0.2$ | 0.080 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 | 0.040 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 |
| queen, 0.2, $-1$ | 0.070 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 | 0.040 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 | 0.999 |
| queen, 0.8, $-0.2$ | 0.075 | 0.986 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 | 0.036 | 0.975 | 0.999 | 1.000 | 1.000 | 1.000 | 1.000 |
| queen, 0.8, $-1$ | 0.084 | 0.981 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 | 0.049 | 0.974 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 |
| | n=400 | | | | | | | | | | | | | |
| queen, 0.2, $-0.2$ | 0.074 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 | 0.042 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 |
| queen, 0.2, $-1$ | 0.087 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 | 0.050 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 |
| queen, 0.8, $-0.2$ | 0.079 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 | 0.045 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 |
| queen, 0.8, $-1$ | 0.074 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 | 0.034 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 |

For the power, (1), (2), (3), (4), (5) and (6) in the table mean that in the DGP $\zeta_0 = (0, 1, 0.5)'$, $\zeta_0 = (0, 1, 1)'$, $\zeta_0 = (0, 1, 1.5)'$, $\zeta_0 = (0, 1, 2)'$, $\zeta_0 = (0, 1, 2.5)'$ and $\zeta_0 = (0, 1, 3)'$, respectively. $\beta_{11,0} = 1$ and $\beta_{12,0} = 1$.

**Definition B.1.** *Let $H = \{h_{ni}, i \in M_n, n \geq 1\}$ be a random field with $\|h_{ni}\|_p < \infty$ for some $p \geq 1$, let $\epsilon = \{\epsilon_{ni}, i \in M_n, n \geq 1\}$ be a random field, where $|M_n| \to \infty$ as $n \to \infty$, and let $c = \{c_{ni}, i \in M_n, n \geq 1\}$ be an array of finite positive constants. Then the random field $H$ is said to be $L_p(c)$-NED on the random field $\epsilon$ if $\|h_{ni} - \mathrm{E}(h_{ni}|\mathscr{F}_{ni}(s))\|_p \leq c_{ni} f(s)$ for some sequence $f(s) \geq 0$ with $\lim_{s\to\infty} f(s) = 0$. The $f(s)$, which are w.l.o.g. assumed to be non-increasing, are called NED coefficients, and $c_{ni}$'s are called NED scaling factors. The $H$ is said to be $L_p$-NED on $\epsilon$ of size $-\lambda$ if $f(s) = O(s^{-\mu})$ for some $\mu > \lambda > 0$. Furthermore, if $\sup_n \sup_{i\in M_n} c_{ni} < \infty$, then $H$ is said to be uniformly $L_p$-NED on $\epsilon$. If $f(s) = O(\varrho^s)$, where $0 < \varrho < 1$, then $H$ is called geometrically $L_p$-NED on $\epsilon$.*

Let the $i$th row of $X_{n1}$, $Z_n$ and $F_n$ be, respectively, $X_{n1,i}$, $Z_{ni}$ and $F_{ni}$. We assume that $X_{n1,i}$'s, $Z_{ni}$'s, $F_{ni}$'s and $\sigma_{ni}^2$'s, where $\sigma_{ni}^2 = \mathrm{E}(v_{ni}^2|F_n)$, are NED on some $\alpha$-mixing random field $\epsilon$. The $\alpha$-mixing coefficients employed are defined below.

**Definition B.2.** *Let $\mathscr{A}$ and $\mathscr{B}$ be two $\sigma$-algebras of $\mathscr{F}$, and let $\alpha(\mathscr{A}, \mathscr{B}) = \sup_{A\in\mathscr{A}, B\in\mathscr{B}} |P(AB) - P(A)P(B)|$. For $U \subset M_n$ and $V \subset M_n$, let $\sigma_n(U) = \sigma(\epsilon_{ni} : i \in U)$ and $\alpha_n(U, V) = \alpha(\sigma_n(U), \sigma_n(V))$. Then the $\alpha$-mixing coefficients for the random field $\epsilon$ are defined as $\bar{\alpha}(u, v, r) = \sup_n \sup_{U,V}(\alpha_n(U, V), |U| \leq u, |V| \leq v, \rho(U, V) \geq r)$, where $\rho(U, V) = \inf\{\rho(i, j) : i \in U, j \in V\}$ is the distance between $U$ and $V$.*

**Assumption B.2.** *$\{X_{n1,i}, i \in M_n\}$, $\{Z_{ni}, i \in M_n\}$, $\{F_{ni}, i \in M_n\}$ and $\{\sigma_{ni}^2, i \in M_n\}$ are uniformly $L_p$ bounded for some $p > 2$, i.e., $\sup_n \sup_{1\leq i\leq n} \mathrm{E}(\|A_{ni}\|_E)^p < \infty$ for $A_{ni} = X_{1n,i}$, $Z_{ni}$, $F_{ni}$ and $\sigma_{ni}^2$, and uniformly and geometrically $L_2$-NED on the $\alpha$-mixing random field $\epsilon = \{\epsilon_{ni}, i \in M_n\}$, for which the $\alpha$-mixing coefficients satisfy $\bar{\alpha}(u, v, r) \leq \varphi(u, v)\hat{\alpha}(r)$ for some function $\varphi(u, v)$ which is nondecreasing in each argument and some $\hat{\alpha}(r)$ such that*

$\sum_{r=1}^{\infty} r^{m-1}\hat{\alpha}(r) < \infty$.

Jenish and Prucha (2012) point out that SAR processes are NED under some weak conditions on the spatial weights matrix. This also applies to the MESS process.

For Assumption 3, note that the $(r,s)$th element of $\frac{1}{n}\Pi_n = \frac{1}{n}F_n'\Sigma_n F_n$ is $\frac{1}{n}\sum_{i=1}^n \sigma_{ni}^2 f_{n,ir}f_{n,is}$. By Lemma A.1 in Xu and Lee (2015), the random field formed by the product of two uniformly and geometrically $L_2$-NED random fields is still uniformly and geometrically $L_2$-NED. Thus, $\{\sigma_{ni}^2 f_{n,ir}f_{n,is}\}$ is uniformly and geometrically $L_2$-NED. Then the law of large number in Theorem 1 of Jenish and Prucha (2012) implies that $\frac{1}{n}\sum_{i=1}^n[\sigma_{ni}^2 f_{n,ir}f_{n,is} - \mathrm{E}(\sigma_{ni}^2 f_{n,ir}f_{n,is})] = o_p(1)$. Thus, $\frac{1}{n}\Pi_n - \frac{1}{n}\mathrm{E}(\Pi_n) = o_p(1)$. For Assumption 7, $\frac{1}{n}F_n'W_nD_n = [\frac{1}{n}F_n'W_nX_n^*, \frac{1}{n}F_n'W_n^2 l_n, \frac{1}{n}F_n'W_n^2X_{n1}, \frac{1}{n}F_n'W_nZ_n]$. The $(r,s)$th element of $\frac{1}{n}F_n'W_nZ_n$ is $\frac{1}{n}\sum_{i=1}^n f_{n,ir}(\sum_{j=1}^n w_{n,ij}z_{n,js})$, where $\sum_{j=1}^n w_{n,ij}z_{n,js}$ is the $(i,s)$th element of $W_nZ_n$. If $\{\sum_{j=1}^n w_{n,ij}z_{n,js}\}$ is uniformly and geometrically $L_2$-NED, then as argued above, $\frac{1}{n}\sum_{i=1}^n f_{n,ir}(\sum_{j=1}^n w_{n,ij}z_{n,js}) - \frac{1}{n}\mathrm{E}\sum_{i=1}^n f_{n,ir}(\sum_{j=1}^n w_{n,ij}z_{n,js}) = o_p(1)$. By Proposition 1 in Jenish and Prucha (2012), the uniform and geometric $L_2$-NED property of $\{\sum_{j=1}^n w_{n,ij}z_{n,js}\}$ is guaranteed by the follow condition.

**Assumption B.3.** $\sup_n \sup_{1\le i\le n}\sum_{j:\rho(i,j)>s}|w_{n,ij}| = O(\varrho^s)$ *for some* $0<\varrho<1$.

For the term $\frac{1}{n}F_n'W_n^2X_{n1}$ in $\frac{1}{n}F_n'W_nD_n$, as the random field formed by the rows of $W_nX_{n1}$ is uniformly and geometrically $L_2$-NED under Assumption B.3, so is the random field formed by the rows of $W_n^2X_{n1}$ under Assumption B.3. Thus, $\frac{1}{n}F_n'W_n^2X_{n1} - \frac{1}{n}\mathrm{E}(F_n'W_n^2X_{n1}) = o_p(1)$. It follows that $\frac{1}{n}F_n'W_n^2D_n - \frac{1}{n}\mathrm{E}(F_n'W_n^2D_n) = o_p(1)$.

Assumption 5 states a law of large numbers that $\frac{1}{n}F_n'e^{(\alpha-\alpha_0)W_n}D_n - \frac{1}{n}\mathrm{E}(F_n'e^{(\alpha-\alpha_0)W_n}D_n) = o_p(1)$ for any $\alpha\in[-\eta,\eta]$. For this to hold, we may show that the random field formed by the rows of $e^{(\alpha-\alpha_0)W_n}D_n$ is uniformly and geometrically $L_2$-NED under the following condition.

**Assumption B.4.** *Only individuals whose distances are less than or equal to some specific constant* $d_0$ *may affect each other. We assume that* $d_0>1$ *w.l.o.g.*

This assumption is also adopted in Xu and Lee (2015). It simplifies that argument for the NED property of $e^{(\alpha-\alpha_0)W_n}D_n$. Let $abs(A)$ be the matrix formed by the absolute values of corresponding elements of a matrix $A$.

**Lemma 1.** *Under Assumptions B.1, B.2 and B.4,* $\sum_{j=1}^n\{[e^{(\alpha-\alpha_0)W_n}]_{ij}d_{n,js}\}$ *for* $1\le s\le k_d$ *is uniformly and geometrically* $L_2$*-NED.*

*Proof.* Note that $[W_n^l]_{ij}=0$ if $\rho(i,j)>md_0$ for $m\ge l$. Then

$$\begin{aligned}\sum_{j:\rho(i,j)>md_0}[abs(e^{(\alpha-\alpha_0)W_n})]_{ij} &= \sum_{j:\rho(i,j)>md_0}\sum_{l=0}^{\infty}\frac{[abs((\alpha-\alpha_0)^lW_n^l)]_{ij}}{l!}\\ &= \sum_{j:\rho(i,j)>md_0}\sum_{l=m+1}^{\infty}\frac{[abs((\alpha-\alpha_0)^lW_n^l)]_{ij}}{l!}\\ &= \sum_{l=m+1}^{\infty}\sum_{j:\rho(i,j)>md_0}\frac{[abs((\alpha-\alpha_0)^lW_n^l)]_{ij}}{l!}\end{aligned}$$

$$
\begin{aligned}
&\leq \sum_{l=m+1}^{\infty}\sum_{j=1}^{n}\frac{[abs((\alpha-\alpha_0)^l W_n^l)]_{ij}}{l!}\\
&\leq \frac{1}{(m+1)!}\|(\alpha-\alpha_0)W_n\|_\infty^{m+1}e^{\|(\alpha-\alpha_0)W_n\|_\infty}\\
&\leq [3/(m+1)]^{m+1}\|(\alpha-\alpha_0)W_n\|_\infty^{m+1}e^{\|(\alpha-\alpha_0)W_n\|_\infty}\\
&\leq \varrho^{m+1}
\end{aligned}
$$

for any $0 < \varrho < 1$ and large enough $m$, as $3\|(\alpha-\alpha_0)W_n\|_\infty e^{\|(\alpha-\alpha_0)W_n\|_\infty/(m+1)}/(m+1)$ converges to zero as $m$ goes to infinity, where the third inequality follows by the fact that $m! > \sqrt{2\pi}m^{m+1/2}e^{-m}e^{1/(12m+1)} > (m/3)^m$ (Robbins, 1955). Thus, by Proposition 1 in Jenish and Prucha (2012), $\sum_{j=1}^n\{[e^{(\alpha-\alpha_0)W_n}]_{ij}d_{n,js}\}$ is uniformly and geometrically $L_2$-NED. □

We collect the results discussed above in the follow proposition.

**Proposition B.1.** *(i) Under Assumptions 2, B.1 and B.2,* $\frac{1}{n}\Pi_n - \frac{1}{n}\mathrm{E}(\Pi_n) = o_p(1)$ *in Assumption 3 holds;*

*(ii) Under Assumptions B.1, B.2 and B.3, Assumption 7 that* $\frac{1}{n}F_n'W_nD_n - \frac{1}{n}\mathrm{E}(F_n'W_nD_n) = o_p(1)$ *holds;*

*(iii) Under Assumptions B.1, B.2 and B.4, Assumption 5 that* $\frac{1}{n}F_n'e^{(\alpha-\alpha_0)W_n}D_n - \frac{1}{n}\mathrm{E}(F_n'e^{(\alpha-\alpha_0)W_n}D_n) = o_p(1)$ *for any* $\alpha \in [-\eta, \eta]$ *holds.*

# C Lemma and proofs

**Lemma 2.** *Let* $A_n = (a_{n,ij})$ *and* $B_n = (b_{n,ij})$ *be* $n \times n$ *nonstochastic matrices that are bounded in both row and column sum norms, and* $c_n = (c_{ni})$, $d_n = (d_{ni})$ *and* $\epsilon_n = (\epsilon_{ni})$ *be* $n\times 1$ *stochastic vectors such that* $\frac{1}{n}\mathrm{E}(c_n'c_n) = O(1)$, $\frac{1}{n}\mathrm{E}(d_n'd_n) = O(1)$, $\sup_n \sup_{1\leq i,j\leq n}\mathrm{E}\,|\epsilon_{ni}c_{nj}|^\tau < \infty$ *for some* $\tau > 2$*, and* $\epsilon_{ni}$*'s are independently distributed with mean zero conditional on* $c_n$*. Then*

*(i)* $\frac{1}{n}c_n'A_nd_n = O_p(1)$*, and* $\frac{1}{n}c_n'A_ne^{\dot\alpha_nB_n}d_n = \frac{1}{n}c_n'A_nd_n + o_p(1)$ *for* $\dot\alpha_n = o_p(1)$*;*

*(ii) if* $\operatorname{plim}_{n\to\infty}\frac{1}{n}c_n'A_n\Omega_nA_n'c_n$ *exists and is nonzero, where* $\Omega_n = \mathrm{diag}(\mathrm{E}(\epsilon_{n1}^2|c_n), \ldots, \mathrm{E}(\epsilon_{nn}^2|c_n))$*, then* $\frac{1}{\sqrt{n}}c_n'A_n\epsilon_n \xrightarrow{d} N(0, \operatorname{plim}_{n\to\infty}\frac{1}{n}c_n'A_n\Omega_nA_n'c_n)$*;*

*(iii)* $\frac{1}{n}c_n'A_ne^{\alpha B_n}d_n$ *is stochastically equicontinuous for* $|\alpha| \leq \eta < \infty$*,* $\frac{1}{n}\mathrm{E}(c_n'A_ne^{\alpha B_n}d_n)$ *is uniformly equicontinuous for* $|\alpha| \leq \eta$*, and* $\frac{1}{n}c_n'A_ne^{\alpha B_n}d_n - \frac{1}{n}\mathrm{E}(c_n'A_ne^{\alpha B_n}d_n) = o_p(1)$ *uniformly on* $[-\eta, \eta]$ *if* $\frac{1}{n}c_n'A_ne^{\alpha B_n}d_n - \frac{1}{n}\mathrm{E}(c_n'A_ne^{\alpha B_n}d_n) = o_p(1)$ *for each* $\alpha \in [-\eta, \eta]$*;*

*(iv)* $\frac{1}{n}c_n'A_ne^{\alpha B_n}\epsilon_n = o_p(1)$ *uniformly in* $\alpha \in [-\eta, \eta]$ *for a finite* $\eta > 0$*.*

*Proof.* $(i)$ By the Cauchy-Schwarz inequality, $\frac{1}{n}|c_n'A_nd_n| \leq \sqrt{\frac{1}{n}c_n'A_nA_n'c_n}\sqrt{\frac{1}{n}d_n'd_n}$. For $\frac{1}{n}c_n'A_nA_n'c_n$, by the spectral radius theorem, $\frac{1}{n}c_n'A_nA_n'c_n \leq \frac{1}{n}\|A_nA_n'\|_\infty c_n'c_n \leq \frac{1}{n}\|A_n\|_\infty\|A_n'\|_\infty c_n'c_n$. By the Markov inequality, $\frac{1}{n}c_n'c_n = O_p(1)$ and $\frac{1}{n}d_n'd_n = O_p(1)$. Thus, $\frac{1}{n}c_n'A_nd_n = O_p(1)$. Note that $\frac{1}{n}c_n'A_ne^{\dot\alpha_nB_n}d_n - \frac{1}{n}c_n'A_nd_n = \frac{1}{n}c_n'A_n(e^{\dot\alpha_nB_n} - I_n)d_n$, where $\|e^{\dot{\alpha}_n B_n} - I_n\|_\infty = o_p(1)$ and $\|e^{\dot{\alpha}_n B_n} - I_n\|_1 = o_p(1)$ by Lemma A.8 in Debarsy et al. (2015). By a similar argument as for $\frac{1}{n}c_n' A_n d_n = O_p(1)$, we have $\frac{1}{n}c_n' A_n(e^{\dot{\alpha}_n B_n} - I_n)d_n = o_p(1)$.

(*ii*) Note that $\frac{1}{\sqrt{n}}c_n' A_n \epsilon_n = \frac{1}{\sqrt{n}}\sum_{i=1}^n(\epsilon_{ni}\sum_{j=1}^n a_{n,ji}c_{nj})$. Consider the $\sigma$-fields $\mathcal{F}_{n0} = \{\emptyset, \Omega\}$, and $\mathcal{F}_{ni} = \sigma(c_{n1}, \ldots, c_{nn})$ for $1 \le i \le n$. Then $\{\epsilon_{ni}\sum_{j=1}^n a_{n,ji}c_{nj}, \mathcal{F}_{ni}, 1 \le i \le n, n \ge 1\}$ forms a martingale difference array. As a result,

$$\frac{1}{n}\sum_{i=1}^n \mathrm{E}\Big[\Big(\epsilon_{ni}\sum_{j=1}^n a_{n,ji}c_{nj}\Big)^2|\mathcal{F}_{n,i-1}\Big] = \frac{1}{n}\sum_{i=1}^n\Big(\sum_{j=1}^n a_{n,ji}c_{nj}\Big)^2 \mathrm{E}(\epsilon_{ni}^2|c_n) = \frac{1}{n}c_n' A_n \Omega_n A_n' c_n$$

is the variance of $\frac{1}{\sqrt{n}}c_n' A_n \epsilon_n$ conditional on $c_n$, which is assumed to have a nonzero probability limit. Since

$$\begin{aligned}
\frac{1}{n}\sum_{i=1}^n \mathrm{E}\Big|\epsilon_{ni}\sum_{j=1}^n a_{n,ji}c_{nj}\Big|^\tau &\le \frac{1}{n}\sum_{i=1}^n \mathrm{E}\Big(\sum_{j=1}^n |a_{n,ji}|^{1-1/\tau}|a_{n,ji}|^{1/\tau}|\epsilon_{ni}c_{nj}|\Big)^\tau \\
&\le \frac{1}{n}\sum_{i=1}^n\Big(\sum_{j=1}^n|a_{n,ji}|\Big)^{(\tau-1)}\sum_{j=1}^n|a_{n,ji}|\,\mathrm{E}\,|\epsilon_{ni}c_{nj}|^\tau \\
&= O(1),
\end{aligned}$$

where the second inequality follows by Hölder's inequality, $\frac{1}{\sqrt{n}}c_n' A_n \epsilon_n \xrightarrow{d} N(0, \mathrm{plim}_{n\to\infty}\frac{1}{n}c_n' A_n \Omega_n A_n' c_n)$ by Theorem A.1 in Kelejian and Prucha (2001).

(*iii*) For any $\alpha_1, \alpha_2 \in [-\eta, \eta]$, by the mean value theorem, $\frac{1}{n}c_n' A_n e^{\alpha_1 B_n}d_n - \frac{1}{n}c_n' A_n e^{\alpha_2 B_n}d_n = \frac{1}{n}c_n' A_n B_n e^{\tilde{\alpha} B_n}d_n(\alpha_1 - \alpha_2)$, where $\tilde{\alpha}$ lies between $\alpha_1$ and $\alpha_2$. As in the proof of (*i*), $\frac{1}{n}|c_n' A_n B_n e^{\tilde{\alpha}B_n}d_n| \le \sqrt{\frac{1}{n}c_n' A_n B_n e^{\tilde{\alpha}B_n}e^{\tilde{\alpha}B_n'}B_n' A_n' c_n}\sqrt{\frac{1}{n}d_n' d_n}$, where $\frac{1}{n}c_n' A_n B_n e^{\tilde{\alpha}B_n}e^{\tilde{\alpha}B_n'}B_n' A_n' c_n \le \frac{1}{n}\|A_n\|_\infty\|B_n\|_\infty e^{\eta\|B_n\|_\infty}e^{\eta\|B_n'\|_\infty}\|B_n'\|_\infty\|A_n'\|_\infty c_n' c_n = O_p(1)$, and $\frac{1}{n}d_n' d_n = O_p(1)$ by the Markov inequality. Thus, $\frac{1}{n}c_n' A_n e^{\alpha B_n}d_n$ is stochastically equicontinuous (Davidson, 1994, p. 339, Theorem 21.10). Similarly, $\frac{1}{n}\mathrm{E}(c_n' A_n e^{\alpha B_n}d_n)$ is uniformly equicontinuous. The uniform convergence of $\frac{1}{n}c_n' A_n e^{\alpha B_n}d_n - \frac{1}{n}\mathrm{E}(c_n' A_n e^{\alpha B_n}d_n)$ to zero on $[-\eta, \eta]$ follows from the point-wise convergence and stochastic equicontinuity (Davidson, 1994, p. 337, Theorem 21.9).

(*iv*) This follows directly by (*ii*) and (*iii*). □

## For Section 2: N2SLS estimator

*Proof of Proposition 2.1.* For a given $\alpha$, the N2SLS estimator for $\beta$ is $\check{\beta}_n(\alpha) = (D_n' H_n D_n)^{-1}D_n' H_n e^{\alpha W_n}Y_n$, where $H_n = F_n \Pi_n^{-1}F_n'$. Substituting $\check{\beta}_n(\alpha)$ into the N2SLS criterion function yields the function

$$Q_n(\alpha) = Y_n' e^{\alpha W_n'}[H_n - H_n D_n(D_n' H_n D_n)^{-1}D_n' H_n]e^{\alpha W_n}Y_n.$$

Let $\bar{Q}_n(\alpha) = \mathrm{E}(Y_n' e^{\alpha W_n'}F_n)A_n\,\mathrm{E}(F_n' e^{\alpha W_n}Y_n)$, where $A_n = \bar{\Pi}_n^{-1} - \bar{\Pi}_n^{-1}\,\mathrm{E}(F_n' D_n)[\mathrm{E}(D_n' F_n)\bar{\Pi}_n^{-1}\,\mathrm{E}(F_n' D_n)]^{-1}\,\mathrm{E}(D_n' F_n)\bar{\Pi}_n^{-1}$, with $\bar{\Pi}_n = \mathrm{E}(\Pi_n)$. Note that $\frac{1}{n}F_n' e^{\alpha W_n}Y_n - \frac{1}{n}\mathrm{E}(F_n' e^{\alpha W_n}Y_n) = \frac{1}{n}[F_n' e^{(\alpha-\alpha_0)W_n}D_n - \mathrm{E}(F_n' e^{(\alpha-\alpha_0)W_n}D_n)]\beta_0 + \frac{1}{n}F_n' e^{(\alpha-\alpha_0)W_n}V_n$. By Lemma 2(*iv*), $\frac{1}{n}F_n' e^{(\alpha-\alpha_0)W_n}V_n = o_p(1)$ uniformly in $\alpha \in [-\eta, \eta]$. In addition, $\frac{1}{n}[F_n' e^{(\alpha-\alpha_0)W_n}D_n - \mathrm{E}(F_n' e^{(\alpha-\alpha_0)W_n}D_n)] = o_p(1)$ uniformly in $\alpha \in [-\eta, \eta]$ by Lemma 2(*iii*) and Assumption 5, and $\frac{1}{n}\Pi_n - \frac{1}{n}\bar{\Pi}_n = o_p(1)$ under Assumption 3. Thus, $\frac{1}{n}F_n' e^{\alpha W_n}Y_n - \frac{1}{n}\mathrm{E}(F_n' e^{\alpha W_n}Y_n) = o_p(1)$ uniformly in $\alpha \in [-\eta, \eta]$, and $\frac{1}{n}[Q_n(\alpha) -$

$\bar{Q}_n(\alpha)] = o_p(1)$ uniformly in $\alpha \in [-\eta, \eta]$. Notice that $A_n = \bar{\Pi}_n^{-1/2} B_n \bar{\Pi}_n^{-1/2}$, where $B_n$ is the projection matrix

$$I_n - \bar{\Pi}_n^{-1/2} \mathrm{E}(F_n' D_n)[\mathrm{E}(D_n' F_n)\bar{\Pi}_n^{-1} \mathrm{E}(F_n' D_n)]^{-1} \mathrm{E}(D_n' F_n)\bar{\Pi}_n^{-1/2}.$$

Then by the partitioned matrix formula, Assumption 6 implies that $\frac{1}{n}\bar{Q}_n(\alpha)$ is uniquely zero at $\alpha_0$ for large enough $n$. In addition, $\frac{1}{n}\bar{Q}_n(\alpha)$ is uniformly equicontinuous as $\frac{1}{n}\mathrm{E}(F_n' e^{\alpha W_n} Y_n)$ is uniformly equicontinuous by Lemma 2$(iii)$. The identification condition and the uniform equicontinuity of $\frac{1}{n}\bar{Q}_n(\alpha)$ imply that the identification uniqueness condition for $\frac{1}{n}\bar{Q}_n(\alpha)$ holds. The consistency of $\check{\alpha}_n$ follows from the uniform convergence and identification uniqueness conditions (White, 1994, p. 28, Theorem 3.4). The consistency of $\check{\beta}_n$ can be seen by applying the mean value theorem to $\check{\beta}_n = \check{\beta}_n(\check{\alpha}_n) = (D_n' H_n D_n)^{-1} D_n' H_n e^{\check{\alpha}_n W_n} Y_n$ and stochastic boundedness of $\sup_{\alpha\in[-\eta,\eta]} \frac{1}{n} F_n' W_n e^{\alpha W_n} Y_n$. □

*Proof of Proposition 2.2.* The first order condition of the N2SLS estimation is $G_n'(\check{\theta}_n)\Pi_n^{-1} g_n(\check{\theta}_n) = 0$. By the mean value theorem, $0 = G_n'(\check{\theta}_n)\Pi_n^{-1} g_n(\check{\theta}_n) = G_n'(\check{\theta}_n)\Pi_n^{-1}[g_n(\theta_0) + G_n(\ddot{\theta}_n)(\check{\theta}_n - \theta_0)]$, where $\ddot{\theta}_n$ is between $\check{\theta}_n$ and $\theta_0$. Note that $\frac{1}{n}G_n(\theta) = \frac{1}{n}F_n'[W_n e^{(\alpha-\alpha_0)W_n}(D_n\beta_0 + V_n), -D_n]$. Since $\check{\alpha}_n = \alpha_0 + o_p(1)$, by Lemma 2$(i)$ and Assumption 7, $\frac{1}{n}F_n' W_n e^{(\check{\alpha}_n-\alpha_0)W_n} D_n = \frac{1}{n}F_n' W_n D_n + o_p(1) = \frac{1}{n}\mathrm{E}(F_n' W_n D_n) + o_p(1)$. Similarly, $\frac{1}{n}F_n' W_n e^{(\check{\alpha}_n-\alpha_0)W_n} V_n = \frac{1}{n}F_n' W_n V_n + o_p(1) = o_p(1)$. As $\frac{1}{n}\Pi_n - \frac{1}{n}\bar{\Pi}_n = o_p(1)$ under Assumption 3, and $\frac{1}{n}F_n' D_n - \frac{1}{n}\mathrm{E}(F_n' D_n) = o_p(1)$ under Assumption 6,

$$\sqrt{n}(\check{\theta}_n - \theta_0) = -\big[\frac{1}{n}G_n'(\check{\theta}_n)\Pi_n^{-1}G_n(\ddot{\theta}_n)\big]^{-1}\frac{1}{\sqrt{n}}G_n'(\check{\theta}_n)\Pi_n^{-1}g_n(\theta_0) = -\big(\frac{1}{n}\bar{G}_n'\bar{\Pi}_n^{-1}\bar{G}_n\big)^{-1}\frac{1}{\sqrt{n}}\bar{G}_n'\bar{\Pi}_n^{-1}g_n(\theta_0) + o_p(1).$$

By the central limit theorem in Lemma 2$(ii)$, $\frac{1}{\sqrt{n}}g_n(\theta_0) = \frac{1}{\sqrt{n}}F_n' V_n \xrightarrow{d} N(0, \lim_{n\to\infty}\frac{1}{n}\bar{\Pi}_n)$. Hence, $\sqrt{n}(\check{\theta}_n - \theta_0) \xrightarrow{d} N\big(0, \lim_{n\to\infty}\big(\frac{1}{n}\bar{G}_n'\bar{\Pi}_n^{-1}\bar{G}_n\big)^{-1}\big)$, under the assumption that $\lim_{n\to\infty}\bar{G}_n$ has full rank. Since $\bar{\Pi}_n = \mathrm{E}(F_n'\Sigma_n F_n)$ and $\bar{G}_n = [\mathrm{E}(F_n' W_n D_n)\beta_0, -\mathrm{E}(F_n' D_n)]$, $\bar{G}_n'\bar{\Pi}_n^{-1}\bar{G}_n \leq \mathrm{E}[(W_n D_n\beta_0, -D_n)'\Sigma_n^{-1}(W_n D_n\beta_0, -D_n)]$ by the generalized Cauchy-Schwarz inequality. The equality holds when $F_n$ is equal to $\Sigma_n^{-1}\mathrm{E}(W_n D_n\beta_0, -D_n|\mathbb{X}_n)$, or equivalently, the matrix formed by the independent columns of $\Sigma_n^{-1}\mathrm{E}(D_n, W_n D_n|\mathbb{X}_n)$ or more compactly the independent columns of $\Sigma_n^{-1}[X_n^*, W_n X_n, W_n^2 X_n, \mathrm{E}(Z_n, W_n Z_n|\mathbb{X}_n)]$, i.e., such a matrix provides the best IV's. □

*Proof of Proposition 2.3.* By taking a third order Taylor expansion of the first order condition $\frac{\partial Q_n^*(\check{\omega}_n)}{\partial\omega} = 0$, because $Q_n^*(\omega)$ is quadratic in $\psi$, by eliminating higher order derivative terms with zero values as in Appendix A, we have

$$\begin{aligned} 0 = \frac{\partial Q_n^*(\check{\omega}_n)}{\partial\phi} &= \frac{\partial Q_n^*(\omega_0)}{\partial\phi} + \frac{\partial^2 Q_n^*(\omega_0)}{\partial\phi^2}(\check{\phi}_n - \phi_0) + \frac{\partial^2 Q_n^*(\omega_0)}{\partial\phi\partial\psi'}(\check{\psi}_n - \psi_0) + \frac{1}{2}\frac{\partial^3 Q_n^*(\omega_0)}{\partial\phi^3}(\check{\phi}_n - \phi_0)^2 \\ &+ \frac{\partial^3 Q_n^*(\omega_0)}{\partial\phi^2\partial\psi'}(\check{\psi}_n - \psi_0)(\check{\phi}_n - \phi_0) + \frac{1}{6}\frac{\partial^4 Q_n^*(\omega_0)}{\partial\phi^4}(\check{\phi}_n - \phi_0)^3 + \frac{1}{2}\frac{\partial^4 Q_n^*(\omega_0)}{\partial\phi^3\partial\psi'}(\check{\psi}_n - \psi_0)(\check{\phi}_n - \phi_0)^2 \\ &+ \frac{1}{24}\frac{\partial^5 Q_n^*(\ddot{\omega}_n)}{\partial\phi^5}(\check{\phi}_n - \phi_0)^4 + \frac{1}{6}\frac{\partial^5 Q_n^*(\ddot{\omega}_n)}{\partial\phi^4\partial\psi'}(\check{\psi}_n - \psi_0)(\check{\phi}_n - \phi_0)^3, \end{aligned} \tag{C.1}$$

and

$$\begin{aligned} 0 = \frac{\partial Q_n^*(\check{\omega}_n)}{\partial\psi} &= \frac{\partial Q_n^*(\omega_0)}{\partial\psi} + \frac{\partial^2 Q_n^*(\omega_0)}{\partial\phi\partial\psi}(\check{\phi}_n - \phi_0) + \frac{\partial^2 Q_n^*(\omega_0)}{\partial\psi\partial\psi'}(\check{\psi}_n - \psi_0) + \frac{1}{2}\frac{\partial^3 Q_n^*(\omega_0)}{\partial\phi^2\partial\psi}(\check{\phi}_n - \phi_0)^2 \\ &+ \frac{1}{6}\frac{\partial^4 Q_n^*(\ddot{\omega}_n)}{\partial\phi^3\partial\psi}(\check{\phi}_n - \phi_0)^3, \end{aligned} \tag{C.2}$$

where $\ddot{\omega}_n$ lies between $\omega_0$ and $\check{\omega}_n$ elementwise. Since the N2SLS estimator $\check{\theta}_n$ is consistent, so is $\check{\omega}_n$. Let $\bar{o}_p(\cdot)$ denote terms with order smaller than those of some terms on the r.h.s. in the above equations. Using the consistency of $\ddot{\omega}_n$ and $\check{\omega}_n$, and relevant orders of derivatives (Appendix A gives expressions of derivatives and their orders), by keeping only possible leading order terms in (C.1), but dropping terms with surely relatively smaller orders into $\bar{o}_p(\cdot)$, we have

$$\begin{aligned} 0 = &\frac{\partial Q_n^*(\omega_0)}{\partial \phi} + \frac{\partial^2 Q_n^*(\omega_0)}{\partial \phi^2}(\check{\phi}_n - \phi_0) + \frac{\partial^2 Q_n^*(\omega_0)}{\partial \phi \partial \psi'}(\check{\psi}_n - \psi_0) + \frac{\partial^3 Q_n^*(\omega_0)}{\partial \phi^2 \partial \psi'}(\check{\psi}_n - \psi_0)(\check{\phi}_n - \phi_0) \\ &+ \frac{1}{6}\frac{\partial^4 Q_n^*(\omega_0)}{\partial \phi^4}(\check{\phi}_n - \phi_0)^3 + \bar{o}_p(\cdot), \end{aligned} \tag{C.1$'$}$$

which follows because the fourth term on the r.h.s. of (C.1) is dominated by the second term, the seventh term is dominated by the fifth term, and the last two terms are dominated by the sixth term. Furthermore,

$$0 = \frac{1}{\sqrt{n}}\frac{\partial Q_n^*(\omega_0)}{\partial \psi} + \frac{1}{n}\frac{\partial^2 Q_n^*(\omega_0)}{\partial \psi \partial \psi'}\sqrt{n}(\check{\psi}_n - \psi_0) + \frac{1}{2n}\frac{\partial^3 Q_n^*(\omega_0)}{\partial \phi^2 \partial \psi}\sqrt{n}(\check{\phi}_n - \phi_0)^2 + \bar{o}_p(\cdot), \tag{C.2$'$}$$

because the second term on the r.h.s. of (C.2) is dominated by the first term, and the fifth term is dominated by the fourth term. Because $(\frac{1}{n}\frac{\partial^2 Q_n^*(\omega_0)}{\partial \psi \partial \psi'})^{-1} = O_p(1)$,

$$\sqrt{n}(\check{\psi}_n - \psi_0) = -\Big(\frac{1}{n}\frac{\partial^2 Q_n^*(\omega_0)}{\partial \psi \partial \psi'}\Big)^{-1}\frac{1}{\sqrt{n}}\frac{\partial Q_n^*(\omega_0)}{\partial \psi} - \Big(\frac{1}{n}\frac{\partial^2 Q_n^*(\omega_0)}{\partial \psi \partial \psi'}\Big)^{-1}\frac{1}{2n}\frac{\partial^3 Q_n^*(\omega_0)}{\partial \phi^2 \partial \psi}\sqrt{n}(\check{\phi}_n - \phi_0)^2 + \bar{o}_p(\cdot). \tag{C.3}$$

Substituting (C.3) into (C.1$'$), and by multiplying $n^{-1/4}$ on the whole equation, we have, after rearrangement,

$$\begin{aligned} 0 = &n^{-1/4}\frac{\partial Q_n^*(\omega_0)}{\partial \phi} - n^{-1/4}\frac{1}{\sqrt{n}}\frac{\partial^2 Q_n^*(\omega_0)}{\partial \phi \partial \psi'}\Big(\frac{1}{n}\frac{\partial^2 Q_n^*(\omega_0)}{\partial \psi \partial \psi'}\Big)^{-1}\Big[\frac{1}{\sqrt{n}}\frac{\partial Q_n^*(\omega_0)}{\partial \psi} + \frac{1}{2n}\frac{\partial^3 Q_n^*(\omega_0)}{\partial \phi^2 \partial \psi}\sqrt{n}(\check{\phi}_n - \phi_0)^2\Big] \\ &+ n^{1/4}(\check{\phi}_n - \phi_0)\Big\{\frac{1}{\sqrt{n}}\frac{\partial^2 Q_n^*(\omega_0)}{\partial \phi^2} - \frac{1}{n}\frac{\partial^3 Q_n^*(\omega_0)}{\partial \phi^2 \partial \psi'}\Big(\frac{1}{n}\frac{\partial^2 Q_n^*(\omega_0)}{\partial \psi \partial \psi'}\Big)^{-1}\frac{1}{\sqrt{n}}\frac{\partial Q_n^*(\omega_0)}{\partial \psi} \\ &+ \Big[\frac{1}{6n}\frac{\partial^4 Q_n^*(\omega_0)}{\partial \phi^4} - \frac{1}{n}\frac{\partial^3 Q_n^*(\omega_0)}{\partial \phi^2 \partial \psi'}\Big(\frac{1}{n}\frac{\partial^2 Q_n^*(\omega_0)}{\partial \psi \partial \psi'}\Big)^{-1}\frac{1}{2n}\frac{\partial^3 Q_n^*(\omega_0)}{\partial \phi^2 \partial \psi}\Big]\sqrt{n}(\check{\phi}_n - \phi_0)^2\Big\} + \bar{o}_p(\cdot). \end{aligned} \tag{C.4}$$

Note that:

*i)* $n^{-1/4}\frac{\partial Q_n^*(\omega_0)}{\partial \phi} - n^{-1/4}\frac{1}{\sqrt{n}}\frac{\partial^2 Q_n^*(\omega_0)}{\partial \phi \partial \psi'}\big(\frac{1}{n}\frac{\partial^2 Q_n^*(\omega_0)}{\partial \psi \partial \psi'}\big)^{-1}\frac{1}{\sqrt{n}}\frac{\partial Q_n^*(\omega_0)}{\partial \psi} = 2n^{-1/4}V_n' W_n' \mathbb{M}_D V_n = O_p(n^{-1/4})$,

*ii)* $n^{-1/4}\frac{1}{\sqrt{n}}\frac{\partial^2 Q_n^*(\omega_0)}{\partial \phi \partial \psi'}\big(\frac{1}{n}\frac{\partial^2 Q_n^*(\omega_0)}{\partial \psi \partial \psi'}\big)^{-1}\frac{1}{2n}\frac{\partial^3 Q_n^*(\omega_0)}{\partial \phi^2 \partial \psi} = n^{-3/4}(W_n^2 X_n \delta_0)' \mathbb{P}_D W_n V_n + O_p(n^{-3/4}) = O_p(n^{-1/4})$,

*iii)* $\frac{1}{\sqrt{n}}\frac{\partial^2 Q_n^*(\omega_0)}{\partial \phi^2} - \frac{1}{n}\frac{\partial^3 Q_n^*(\omega_0)}{\partial \phi^2 \partial \psi'}(\frac{1}{n}\frac{\partial^2 Q_n^*(\omega_0)}{\partial \psi \partial \psi'})^{-1}\frac{1}{\sqrt{n}}\frac{\partial Q_n^*(\omega_0)}{\partial \psi} = R_n + O_p(n^{-1/2})$, where $R_n = \frac{2}{\sqrt{n}}(W_n^2 X_n \delta_0)' \mathbb{M}_D V_n = O_p(1)$, and

*iv)* $\frac{1}{6n}\frac{\partial^4 Q_n^*(\omega_0)}{\partial \phi^4} - \frac{1}{n}\frac{\partial^3 Q_n^*(\omega_0)}{\partial \phi^2 \partial \psi'}(\frac{1}{n}\frac{\partial^2 Q_n^*(\omega_0)}{\partial \psi \partial \psi'})^{-1}\frac{1}{2n}\frac{\partial^3 Q_n^*(\omega_0)}{\partial \phi^2 \partial \psi} = S_n + O_p(n^{-1/2})$, where $S_n = \frac{1}{n}(W_n^2 X_n \delta_0)' \mathbb{M}_D W_n^2 X_n \delta_0 = O(1)$.

Hence, (C.4) implies that

$$0 = O_p(n^{-1/4}) + O_p(n^{-1/4})[n^{1/4}(\check{\phi}_n - \phi_0)]^2 + n^{1/4}(\check{\phi}_n - \phi_0)R_n + [n^{1/4}(\check{\phi}_n - \phi_0)]^3 S_n + \bar{o}_p(\cdot). \tag{C.4$'$}$$

As $S_n > 0$ for large enough $n$, $n^{1/4}(\check{\phi}_n - \phi_0)$ cannot grow with a rate as an increasing function of $n$. This is because, otherwise, $S_n[n^{1/4}(\check{\phi}_n - \phi_0)]^3$ would be the dominating term of (C.4$'$) on the r.h.s., which grows to infinity. Hence, (C.4$'$) implies $\sqrt{n}(\check{\phi}_n - \phi_0)^2 = O_p(1)$. On the other hand, it follows from (C.3) that $\sqrt{n}(\check{\psi}_n - \psi_0) = O_p(1)$.

When $R_n > 0$, $R_n + S_n\sqrt{n}(\check{\phi}_n - \phi_0)^2 \geq R_n > 0$. Thus, conditional on $R_n > 0$, $R_n + S_n\sqrt{n}(\check{\phi}_n - \phi_0)^2$ cannot converge in distribution to a random variable with an atom of probability at 0 along any subsequence of $n$. Hence, the equality (C.4′) is possible only if $n^{1/4}(\check{\phi}_n - \phi_0) = o_p(1)$. Therefore, conditional on $R_n > 0$, $\sqrt{n}(\check{\phi}_n - \phi_0)^2 = o_p(1)$. Then, when $R_n > 0$, (C.3) becomes

$$\sqrt{n}(\check{\psi}_n - \psi_0) = -(\frac{1}{n}\frac{\partial^2 Q_n^*(\omega_0)}{\partial\psi\partial\psi'})^{-1}\frac{1}{\sqrt{n}}\frac{\partial Q_n^*(\omega_0)}{\partial\psi} + o_p(1) = L_n + o_p(1),$$

where $L_n = (\frac{1}{n}D_n'H_nD_n)^{-1}\frac{1}{\sqrt{n}}D_n'H_nV_n$. Note that $L_n \xrightarrow{d} L$ by Lemma 2($ii$), where $L$ is the normal random vector $N(0, \lim_{n\to\infty}[\frac{1}{n}\,\mathrm{E}(D_n'F_n)\bar{\Pi}_n^{-1}\,\mathrm{E}(F_n'D_n)]^{-1})$.

Next, when $R_n < 0$, we can prove that $\sqrt{n}(\check{\phi}_n - \phi_0)^2 = J_{1n} + o_p(1)$, where $J_{1n} = -S_n^{-1}R_n$, by showing that there exists no subsequence $n'$ of $n$ such that ${n'}^{1/4}(\check{\phi}_{n'} - \phi_0)$ converges in distribution to a random variable with an atom of probability at 0. That is, if ${n'}^{1/4}(\check{\phi}_{n'} - \phi_0) \xrightarrow{d} U$, then for any $\delta > 0$, there exists an $\epsilon > 0$ such that $\mathrm{P}(|U| > \epsilon) > 1 - \delta$. We show this by contradiction.[2] Suppose that there exists a subsequence $n'$ such that ${n'}^{1/4}(\check{\phi}_{n'} - \phi_0) \xrightarrow{d} U$ and $\mathrm{P}(U = 0) = \delta > 0$. By a fourth order Taylor expansion and with the orders of derivatives in Appendix A,

$$\begin{aligned}Q_n^*(\check{\omega}_n) - Q_n^*(\omega_0) = {} & \big[\frac{1}{\sqrt{n}}\frac{\partial Q_n^*(\omega_0)}{\partial\psi'} + \frac{1}{2n}\frac{\partial^3 Q_n^*(\omega_0)}{\partial\phi^2\partial\psi'}\sqrt{n}(\check{\phi}_n - \phi_0)^2\big]\sqrt{n}(\check{\psi}_n - \psi_0) + \frac{1}{2\sqrt{n}}\frac{\partial^2 Q_n^*(\omega_0)}{\partial\phi^2}\sqrt{n}(\check{\phi}_n - \phi_0)^2 \\ & + \sqrt{n}(\check{\psi}_n - \psi_0)'\frac{1}{2n}\frac{\partial^2 Q_n^*(\omega_0)}{\partial\psi\partial\psi'}\sqrt{n}(\check{\psi}_n - \psi_0) + \frac{1}{24n}\frac{\partial^4 Q_n^*(\omega_0)}{\partial\phi^4}n(\check{\phi}_n - \phi_0)^4 + O_p(n^{-1/4}).\end{aligned} \tag{C.5}$$

Note that the order $\bar{o}_p(\cdot)$ in (C.3) is $O_p(n^{-1/4})$. Substituting the expression for $\sqrt{n}(\check{\psi}_n - \psi_0)$ in (C.3) into (C.5) yields

$$\begin{aligned}& Q_n^*(\check{\omega}_n) - Q_n^*(\omega_0) \\ & = -\frac{1}{2\sqrt{n}}\frac{\partial Q_n^*(\omega_0)}{\partial\psi'}\big(\frac{1}{n}\frac{\partial^2 Q_n^*(\omega_0)}{\partial\psi\partial\psi'}\big)^{-1}\frac{1}{\sqrt{n}}\frac{\partial Q_n^*(\omega_0)}{\partial\psi} \\ & \quad + \sqrt{n}(\check{\phi}_n - \phi_0)^2\Big\{\Big[\frac{1}{2\sqrt{n}}\frac{\partial^2 Q_n^*(\omega_0)}{\partial\phi^2} - \frac{1}{2n}\frac{\partial^3 Q_n^*(\omega_0)}{\partial\phi^2\partial\psi'}\big(\frac{1}{n}\frac{\partial^2 Q_n^*(\omega_0)}{\partial\psi\partial\psi'}\big)^{-1}\frac{1}{\sqrt{n}}\frac{\partial Q_n^*(\omega_0)}{\partial\psi}\Big] \\ & \quad + \Big[\frac{1}{24n}\frac{\partial^4 Q_n^*(\omega_0)}{\partial\phi^4} - \frac{1}{8n}\frac{\partial^3 Q_n^*(\omega_0)}{\partial\phi^2\partial\psi'}\big(\frac{1}{n}\frac{\partial^2 Q_n^*(\omega_0)}{\partial\psi\partial\psi'}\big)^{-1}\frac{1}{n}\frac{\partial^3 Q_n^*(\omega_0)}{\partial\phi^2\partial\psi}\Big]\sqrt{n}(\check{\phi}_n - \phi_0)^2\Big\} + O_p(n^{-1/4}) \\ & = -V_n'\mathbb{P}_D V_n + \sqrt{n}(\check{\phi}_n - \phi_0)^2\big[\frac{1}{2}R_n + \frac{1}{4}S_n\sqrt{n}(\check{\phi}_n - \phi_0)^2\big] + O_p(n^{-1/4}).\end{aligned} \tag{C.6}$$

Since ${n'}^{1/4}(\check{\phi}_{n'} - \phi_0) \xrightarrow{d} U$ and $\mathrm{P}(U = 0) = \delta > 0$, there exists a negative constant $M$ such that for all $\epsilon > 0$,

$$Q_{n'}^*(\check{\omega}_{n'}) - Q_{n'}^*(\omega_0) > -V_{n'}'\mathbb{P}_D V_{n'} + \epsilon M$$

with probability converging to a number greater than $\delta/2$ along the subsequence. Note that (C.6) still holds if we replace $\check{\omega}_n$ by any $\bar{\omega}_n$ satisfying $\sqrt{n}(\bar{\phi}_n - \phi_0)^2 = O_p(1)$ and (C.3) with $\check{\omega}_n$ replaced by $\bar{\omega}_n$. In particular, if we let $\sqrt{n}(\bar{\phi}_n - \phi_0)^2 = -S_n^{-1}R_n$ and define $\bar{\psi}_n$ according to (C.3) after $\check{\phi}_n$ is replaced by $\bar{\phi}_n$ in that formula, then

$$Q_{n'}^*(\bar{\omega}_{n'}) - Q_{n'}^*(\omega_0) = -V_{n'}'\mathbb{P}_D V_{n'} - \frac{1}{4}S_{n'}^{-1}R_{n'}^2 + O_p(n^{-1/4}).$$

[2] Such an argument appears in Rotnitzky et al. (2000). We adopt the analysis for our model.

Hence, by taking $\epsilon$ small enough, $Q^*_{n'}(\bar{\omega}_{n'}) - Q^*_{n'}(\check{\omega}_{n'}) < -\epsilon M - \frac{1}{4}S^{-1}_{n'}R^2_{n'} < -\frac{1}{8}S^{-1}_{n'}R^2_{n'} < 0$ with probability converging along the subsequence to a strictly positive number. This is a contradiction since $\check{\omega}_n$ is the N2SLS estimator that minimizes $Q^*_n(\omega)$. Therefore, (C.4′) holds only if $\sqrt{n}(\check{\phi}_n - \phi_0)^2 = J_{1n} + o_p(1)$ when $R_n < 0$, where $J_{1n} = -S^{-1}_n R_n$. Then by (C.3), $\sqrt{n}(\check{\psi}_n - \psi_0) = J_{2n} + o_p(1)$ when $R_n < 0$, where

$$J_{2n} = L_n + \big(\frac{2}{n}D'_n H_n D_n\big)^{-1}\frac{1}{n}D'_n H_n W^2_n X_n \delta_0 J_{1n}.$$

Alternatively, when $R_n < 0$, by (C.1′) and (C.2′), we are essentially solving (8). Thus, the leading order term of $[\sqrt{n}(\check{\phi}_n - \phi_0)^2, \sqrt{n}(\check{\psi}_n - \psi_0)']'$ is $J_n = (J_{1n}, J'_{2n})'$ in (9). Under Assumptions 5 and 7, $J_n = \mathbb{J}_n + o_p(1)$. By Lemma 2$(ii)$, $\mathbb{J}_n \xrightarrow{d} J$, where $J = N(0, \lim_{n\to\infty}\Delta_n)$ with $\Delta_n$ being the covariance matrix of $\mathbb{J}_n$.

We next calculate the statistic that asymptotically determines the sign of $n^{1/4}(\check{\phi}_n - \phi_0)$ conditional on $R_n < 0$. By a fifth order Taylor expansion of $Q^*_n(\check{\omega}_n)$ at $\omega_0$, we have with the orders of derivatives in Appendix A that

$$Q^*_n(\check{\omega}_n) = \Xi_n + K_n + O_p(n^{-1/2}),$$

where

$$\begin{aligned}\Xi_n &= V'_n H_n V_n + \frac{\partial Q^*_n(\omega_0)}{\partial \psi'}(\check{\psi}_n - \psi_0) + \frac{1}{2}\frac{\partial^2 Q^*_n(\omega_0)}{\partial \phi^2}(\check{\phi}_n - \phi_0)^2 + \frac{1}{2}(\check{\psi}_n - \psi_0)'\frac{\partial^2 Q^*_n(\omega_0)}{\partial \psi \partial \psi'}(\check{\psi}_n - \psi_0)\\ &\quad + \frac{1}{2}\frac{\partial^3 Q^*_n(\omega_0)}{\partial \phi^2 \partial \psi'}(\check{\psi}_n - \psi_0)(\check{\phi}_n - \phi_0)^2 + \frac{1}{24}\frac{\partial^4 Q^*_n(\omega_0)}{\partial \phi^4}(\check{\phi}_n - \psi_0)^4\\ &= O_p(1),\end{aligned}$$

and

$$\begin{aligned}K_n &= \frac{\partial Q^*_n(\omega_0)}{\partial \phi}(\check{\phi}_n - \phi_0) + \frac{\partial^2 Q^*_n(\omega_0)}{\partial \phi \partial \psi'}(\check{\psi}_n - \psi_0)(\check{\phi}_n - \phi_0) + \frac{1}{6}\frac{\partial^3 Q^*_n(\omega_0)}{\partial \phi^3}(\check{\phi}_n - \phi_0)^3\\ &\quad + \frac{1}{6}\frac{\partial^4 Q^*_n(\omega_0)}{\partial \phi^3 \partial \psi'}(\check{\psi}_n - \psi_0)(\check{\phi}_n - \phi_0)^3 + \frac{1}{120}\frac{\partial^5 Q^*_n(\omega_0)}{\partial \phi^5}(\check{\phi}_n - \phi_0)^5\\ &= O_p(n^{-1/4}).\end{aligned} \tag{C.7}$$

Since the sign of $n^{1/4}(\check{\phi}_n - \phi_0)$ does not affect the value of $\Xi_n$, the sign of $n^{1/4}(\check{\phi}_n - \phi_0)$ must be chosen to minimize $K_n$. Note that $K_n = (\check{\phi}_n - \phi_0)\mathbb{K}_n + o_p(n^{-1/4})$, where

$$\begin{aligned}\mathbb{K}_n &= 2V'_n W'_n H_n V_n + \begin{pmatrix}\sqrt{n}(\check{\phi}_n - \phi_0)^2\\ \sqrt{n}(\check{\psi}_n - \psi_0)\end{pmatrix}'\Bigg[\begin{pmatrix}\frac{1}{3\sqrt{n}}(X_n\delta_0)'(W'^3_n H_n + 3W'^2_n H_n W_n)V_n\\ -\frac{2}{\sqrt{n}}D'_n H_n W_n V_n\end{pmatrix}\\ &\quad - \begin{pmatrix}\frac{1}{6n}(W^3_n X_n\delta_0)'H_n(-W^2_n X_n \delta_0, 2D_n)\\ 0\end{pmatrix}\begin{pmatrix}\sqrt{n}(\check{\phi}_n - \phi_0)^2\\ \sqrt{n}(\check{\psi}_n - \psi_0)\end{pmatrix}\Bigg]\\ &= 2V'_n W'_n H_n V_n + J'_n\Bigg[\begin{pmatrix}\frac{1}{3\sqrt{n}}(X_n\delta_0)'(W'^3_n H_n + 3W'^2_n H_n W_n)V_n\\ -\frac{2}{\sqrt{n}}D'_n H_n W_n V_n\end{pmatrix} - \begin{pmatrix}\frac{1}{6n}(W^3_n X_n\delta_0)'H_n(-W^2_n X_n\delta_0, 2D_n)\\ 0\end{pmatrix}J_n\Bigg] + o_p(1).\end{aligned}$$

Thus, $\mathrm{P}(n^{1/4}(\check{\phi}_n - \phi_0)\mathbb{K}_n < 0 | R_n < 0) \to 1$ as $n \to \infty$. Equivalently, $\mathrm{P}(I(n^{1/4}(\check{\phi}_n - \phi_0) < 0) = I(\mathbb{K}_n > 0) | R_n < 0) \to 1$ as $n \to \infty$. Using $U_{jn}$ and $\Upsilon_{jn}$ for $j = 1, 2$ defined above Proposition 2.3, we have $K^*_n$ in (10). Since $U_n$ is uncorrelated with $\mathbb{J}_n$, by Lemma 2$(ii)$, $(U'_n, J'_n)' \xrightarrow{d} (U', J')'$, where $U = N(0, \lim_{n\to\infty}\mathrm{E}(U_n U'_n))$ is independent of $J$. Hence, $K^*_n \xrightarrow{d} K^*$. As $J_{1n} = -S^{-1}_n R_n$, $J_{1n}$ has a sign opposite to that of $R_n$. Therefore, the asymptotic distribution of $\check{\omega}_n$ in the proposition follows. □

*Proof of Proposition 2.4.* As $\hat V_n = e^{\dot\alpha_n W_n} y_n - D_n\dot\beta_n = e^{(\dot\alpha_n-\alpha_0)W_n}(D_n\beta_0 + V_n) - D_n\dot\beta_n = V_n + (e^{(\dot\alpha_n-\alpha_0)W_n} - I_n)V_n + (e^{(\dot\alpha_n-\alpha_0)W_n} - I_n)D_n\beta_0 + D_n(\beta_0 - \dot\beta_n)$, $\hat v_{ni} = v_{ni} + a_{ni} + b_{ni} + c_{ni}$, where $a_{ni} = e_{ni}'(e^{(\dot\alpha_n-\alpha_0)W_n} - I_n)V_n$, $b_{ni} = e_{ni}'(e^{(\dot\alpha_n-\alpha_0)W_n} - I_n)D_n\beta_0$, and $c_{ni} = e_{ni}'D_n(\beta_0 - \dot\beta_n)$, with $e_{ni}$ being the $i$th column of the $n\times n$ identity matrix. Then the $(r,s)$th element of $\frac{1}{n}\hat\Pi_n - \frac{1}{n}\Pi_n = \frac{1}{n}F_n'\hat\Sigma_n F_n - \frac{1}{n}F_n'\Sigma_n F_n$ is

$$\frac{1}{n}\sum_{i=1}^n f_{n,ir}f_{n,is}(\hat v_{ni}^2 - \sigma_{ni}^2) = \frac{1}{n}\sum_{i=1}^n f_{n,ir}f_{n,is}[(v_{ni}^2 - \sigma_{ni}^2) + 2a_{ni}v_{ni} + 2b_{ni}v_{ni} + 2c_{ni}v_{ni} + a_{ni}^2 + b_{ni}^2 + c_{ni}^2 + 2a_{ni}b_{ni} + 2b_{ni}c_{ni} + 2c_{ni}a_{ni}],$$

where $\sigma_{ni}^2 = \mathrm{E}(v_{ni}^2|F_n)$. We shall show that every sample average over $i$ of a product of $f_{n,ir}f_{n,is}$ with each term in the square brackets of the above equation goes to zero in probability. Since $v_{ni}$'s are independent conditional on $F_n$, and $\mathrm{E}\,|f_{n,ir}f_{n,is}(v_{ni}^2 - \sigma_{ni}^2)| \le (\mathrm{E}\, f_{n,ir}^4)^{1/4}(\mathrm{E}\, f_{n,is}^4)^{1/4}[\mathrm{E}(v_{ni}^2 - \sigma_{ni}^2)^2]^{1/2} < \infty$ by the generalized Hölder's inequality and Assumption 9, $\frac{1}{n}\sum_{i=1}^n f_{n,ir}f_{n,is}(v_{ni}^2 - \sigma_{ni}^2) = o_p(1)$ by the martingale law of large numbers (Davidson, 1994, p. 299, Corollary 19.8). We next show that $\frac{1}{n}\sum_{i=1}^n f_{n,ir}f_{n,is}a_{ni}^2 = o_p(1)$. By a fourth order Taylor expansion at $\alpha_0$,

$$\frac{1}{n}\sum_{i=1}^n f_{n,ir}f_{n,is}a_{ni}^2 = \frac{1}{n}\sum_{i=1}^n f_{n,ir}f_{n,is}a_{ni}^{(1)}(\dot\alpha_n - \alpha_0)^2 + \frac{1}{n}\sum_{i=1}^n f_{n,ir}f_{n,is}a_{ni}^{(2)}(\dot\alpha_n - \alpha_0)^3 + \frac{1}{n}\sum_{i=1}^n f_{n,ir}f_{n,is}a_{ni}^{(3)}(\dot\alpha_n - \alpha_0)^4 + \frac{1}{n}\sum_{i=1}^n f_{n,ir}f_{n,is}a_{ni}^{(4)}(\dot\alpha_n - \alpha_0)^5,$$

where $a_{ni}^{(1)} = V_n'W_n'e_{ni}e_{ni}'W_nV_n$, $a_{ni}^{(2)} = V_n'W_n'^2e_{ni}e_{ni}'W_nV_n$, $a_{ni}^{(3)} = \frac{1}{3}V_n'W_n'^3e_{ni}e_{ni}'W_nV_n + \frac{1}{4}V_n'W_n'^2e_{ni}e_{ni}'W_n^2V_n$, and

$$a_{ni}^{(4)} = \frac{1}{60}V_n'W_n'^5e^{(\ddot\alpha_n-\alpha_0)W_n'}e_{ni}e_{ni}'(e^{(\ddot\alpha_n-\alpha_0)W_n} - I_n)V_n + \frac{1}{12}V_n'W_n'^4e^{(\ddot\alpha_n-\alpha_0)W_n'}e_{ni}e_{ni}'e^{(\ddot\alpha_n-\alpha_0)W_n}W_nV_n + \frac{1}{6}V_n'W_n'^3e^{(\ddot\alpha_n-\alpha_0)W_n'}e_{ni}e_{ni}'e^{(\ddot\alpha_n-\alpha_0)W_n}W_n^2V_n,$$

with $\ddot\alpha_n$ between $\dot\alpha_n$ and $\alpha_0$. Let $B_n = (b_{n,ij})$ and $C_n = (c_{n,ij})$ be $n\times n$ nonstochastic matrices which are bounded in both row and column sum norms. We show the general result that $\frac{1}{n}\sum_{i=1}^n f_{n,ir}f_{n,is}V_n'B_ne_{ni}e_{ni}'C_nV_n = O_p(1)$. Note that

$$\begin{aligned}\frac{1}{n}\mathrm{E}\Big|\sum_{i=1}^n f_{n,ir}f_{n,is}V_n'B_ne_{ni}e_{ni}'C_nV_n\Big| &= \frac{1}{n}\mathrm{E}\Big|\sum_{i=1}^n\sum_{j=1}^n\sum_{k=1}^n f_{n,ir}f_{n,is}b_{n,ji}v_{nj}c_{n,ik}v_{nk}\Big| \\ &\le \frac{1}{n}\sum_{i=1}^n\sum_{j=1}^n\sum_{k=1}^n |b_{n,ji}c_{n,ik}|\,\mathrm{E}\,|f_{n,ir}f_{n,is}v_{nj}v_{nk}| \\ &\le \frac{1}{n}\sum_{i=1}^n\sum_{j=1}^n\sum_{k=1}^n |b_{n,ji}c_{n,ik}|(\mathrm{E}\,|f_{n,ir}|^4)^{1/4}(\mathrm{E}\,|f_{n,is}|^4)^{1/4}(\mathrm{E}\,|v_{nj}|^4)^{1/4}(\mathrm{E}\,|v_{nk}|^4)^{1/4} \\ &\le \frac{c_1}{n}\sum_{i=1}^n\sum_{j=1}^n |b_{n,ji}|\sum_{k=1}^n |c_{n,ik}| \\ &\le c_2\end{aligned}$$

for some finite constants $c_1$ and $c_2$, where the second inequality follows by the generalized Hölder inequality. Thus $\frac{1}{n}\sum_{i=1}^n f_{n,ir}f_{n,is}V_n'B_ne_{ni}e_{ni}'C_nV_n = O_p(1)$ by the Markov inequality. Then, as $\dot\alpha_n - \alpha_0 = o_p(1)$, $\frac{1}{n}\sum_{i=1}^n f_{n,ir}f_{n,is}a_{ni}^{(j)}(\dot\alpha_n - \alpha_0)^{j+1} = o_p(1)$ for $j = 1,2,3$. For $\frac{1}{n}\sum_{i=1}^n f_{n,ir}f_{n,is}a_{ni}^{(4)}(\dot\alpha_n - \alpha_0)^5$, notice that for any $n\times n$ matrix $T_n$, $\frac{1}{n}|V_n'T_nV_n| \le$

$\sqrt{\frac{1}{n}V_n'T_nT_n'V_n}\sqrt{\frac{1}{n}V_n'V_n} \leq \sqrt{\frac{1}{n}\|T_n\|_\infty\|T_n'\|_\infty V_n'V_n}\sqrt{\frac{1}{n}V_n'V_n} = O_p(1)$, if $\|T_n\|_\infty < \infty$ and $\|T_n'\|_\infty < \infty$. Thus, $\sup_{1\leq i\leq n}\frac{1}{n}|a_{ni}^{(4)}| = O_p(1)$. Since $\frac{1}{n}|\sum_{i=1}^n f_{n,ir}f_{n,is}a_{ni}^{(4)}(\dot{\alpha}_n-\alpha_0)^5| \leq \frac{1}{n}|\dot{\alpha}_n-\alpha_0|^5\sum_{i=1}^n|f_{n,ir}f_{n,is}|\sup_{1\leq i\leq n}|a_{ni}^{(4)}|$, where $\frac{1}{n}\sum_{i=1}^n|f_{n,ir}f_{n,is}| = O_p(1)$ by the Markov inequality, and $\dot{\alpha}_n-\alpha_0 = O_p(n^{-1/4})$, $\frac{1}{n}\sum_{i=1}^n f_{n,ir}f_{n,is}a_{ni}^{(4)}(\dot{\alpha}_n-\alpha_0)^5 = o_p(1)$. Hence, $\frac{1}{n}\sum_{i=1}^n f_{n,ir}f_{n,is}a_{ni}^2 = o_p(1)$. Similarly, we can show that $\frac{1}{n}\sum_{i=1}^n f_{n,ir}f_{n,is}\iota_{ni} = o_p(1)$ for $\iota_{ni} = 2a_{ni}v_{ni}$, $2b_{ni}v_{ni}$, $2c_{ni}v_{ni}$, $b_{ni}^2$, $c_{ni}^2$, $2a_{ni}b_{ni}$, $2b_{ni}c_{ni}$, and $2c_{ni}a_{ni}$. Thus, $\frac{1}{n}\sum_{i=1}^n f_{n,ir}f_{n,is}(\hat{v}_{ni}^2-\sigma_{ni}^2)$ for any $r,s$. It follows that $\frac{1}{n}\hat{\Pi}_n = \frac{1}{n}\Pi_n + o_p(1)$.

Since $\hat{Q}_n(\theta)$ is quadratic in $F_n'(e^{\alpha W_n}Y_n - D_n\beta)$ and $\frac{1}{n}\hat{\Pi}_n - \frac{1}{n}\Pi_n = o_p(1)$, the arguments in the proof of Proposition 2.1 hold and the feasible N2SLS estimator $\hat{\theta}_n$ is consistent. Replacing $\Pi_n$ by $\hat{\Pi}_n$ in $Q_n(\theta)$ and $Q_n^*(\omega)$ does not affect the analyses for Propositions 2.2 and 2.3, because neither the orders of terms nor asymptotic distributions for these propositions will change. Thus, Proposition 2.2 still holds if $\check{\theta}_n$ is replaced by $\hat{\theta}_n$, and Proposition 2.3 still holds if $\check{\omega}_n$ is replaced by $\hat{\omega}_n$. □

## For Section 3: Testing for the irrelevance of the Durbin and endogenous regressors

*Proof of Proposition 3.1.* Let $P_{\Pi^{-1/2}F'D} = \Pi_n^{-1/2}F_n'D_n(D_n'H_nD_n)^{-1}D_n'F_n\Pi_n^{-1/2}$ be the orthogonal projector onto the column space of $\Pi_n^{-1/2}F_n'D_n$. Similarly, let

$$P_{\Pi^{-1/2}F'(-WX\delta_0,X)} = \Pi_n^{-1/2}F_n'(-W_nX_n\delta_0,X_n)[(-W_nX_n\delta_0,X_n)'H_n(-W_nX_n\delta_0,X_n)]^{-1}(-W_nX_n\delta_0,X_n)'F_n\Pi_n^{-1/2}$$

and $P_{\Pi^{-1/2}F'(-W^2X\delta_0,D)} = \Pi_n^{-1/2}F_n'(-W_n^2X_n\delta_0,D_n)[(-W_n^2X_n\delta_0,D_n)'H_n(-W_n^2X_n\delta_0,D_n)]^{-1}(-W_n^2X_n\delta_0,D_n)'F_n\Pi_n^{-1/2}$. Then $\mathbb{P}_D = F_n\Pi_n^{-1/2}P_{\Pi^{-1/2}F'D}\Pi_n^{-1/2}F_n'$, $\mathbb{P}_{(-WX\delta_0,X)} = F_n\Pi_n^{-1/2}P_{\Pi^{-1/2}F'(-WX\delta_0,X)}\Pi_n^{-1/2}F_n'$, and

$$\mathbb{P}_{(-W^2X\delta_0,D)} = F_n\Pi_n^{-1/2}P_{\Pi^{-1/2}F'(-W_n^2X_n\delta_0,D)}\Pi_n^{-1/2}F_n'.$$

Let the $k_f\times(k_{x^*}+k_x+k_z)$ matrix $[\Pi_n^{-1/2}F_n'(-W_nX_n\delta_0,X_n),A_n]$ be a basis matrix for the column space of $\Pi_n^{-1/2}F_n'D_n$, where $A_n$ is an $k_f\times(k_{x^*}+k_z-1)$ matrix perpendicular to $\Pi_n^{-1/2}F_n'(-W_nX_n\delta_0,X_n)$. Then, $P_{\Pi^{-1/2}F'D} = P_{\Pi^{-1/2}F'(-WX\delta_0,X)} + P_A$, where $P_A = A_n(A_n'A_n)^{-1}A_n'$. By (3.25) on p. 71 in Ruud (2000), $P_{\Pi^{-1/2}F'(-W^2X\delta_0,D)} = P_{\Pi^{-1/2}F'D} + P_B$, where $P_B$ is the orthogonal projector onto the column space of $B_n = M_{\Pi^{-1/2}F'D}\Pi_n^{-1/2}F_n'W_n^2X_n\delta_0$ with $M_{\Pi^{-1/2}F'D} = I_{k_f} - P_{\Pi^{-1/2}F'D}$, and $B_n$ is perpendicular to $A_n$. Thus, $P_{\Pi^{-1/2}F'(-W^2X\delta_0,D)} - P_{\Pi^{-1/2}F'(-WX\delta_0,X)} = P_B + P_A$. It follows that (15) becomes

$$\hat{Q}_n(\tilde{\theta}_n) - \hat{Q}_n(\hat{\theta}_n) = I(R_n<0)V_n'F_n\Pi_n^{-1/2}(P_B+P_A)\Pi_n^{-1/2}F_n'V_n + I(R_n>0)V_n'F_n\Pi_n^{-1/2}P_A\Pi_n^{-1/2}F_n'V_n + o_p(1).$$

By the central limit theorem in Lemma 2(v), $\Pi_n^{-1/2}F_n'V_n \xrightarrow{d} N(0,I_{k_f})$. Since $R_n = \frac{2}{\sqrt{n}}B_n'\Pi_n^{-1/2}F_n'V_n$, $\hat{Q}_n(\tilde{\theta}_n) - \hat{Q}_n(\hat{\theta}_n) \xrightarrow{d} T$, where $T = \sum_{i=1}^{k_{x^*}+k_z} r_i^2 I(r_1<0) + \sum_{i=2}^{k_{x^*}+k_z} r_i^2 I(r_1>0)$, with $r_1,\ldots,r_{k_{x^*}}$ being i.i.d. standard normal random variables. Because the probability density function of $r_1$ is symmetric, $T$ is a mixture of a $\chi^2(k_{x^*}+k_z-1)$ variable and a $\chi^2(k_{x^*}+k_z)$ variable with mixing probabilities equal to 1/2. □

*Proof of Proposition 3.2.* Note that $\mathbb{M}_{(-WX\delta_0,X)} = F_n\Pi_n^{-1/2}M_{\Pi^{-1/2}F'(-WX\delta_0,X)}\Pi_n^{-1/2}F_n'$, where $M_{\Pi^{-1/2}F'(-WX\delta_0,X)} = I_{k_f} - P_{\Pi^{-1/2}F'(-WX\delta_0,X)}$ with $P_{\Pi^{-1/2}F'(-WX\delta_0,X)}$ being defined in the proof of Proposition 3.1. As in the proof of Proposition 3.1, $\Pi_n^{-1/2}F_n \xrightarrow{d} N(0, I_{k_f})$. Thus, the asymptotic distribution of the gradient vector in (16) holds. Furthermore, by the partitioned matrix formula,

$$\begin{aligned}
&\mathrm{rk}\big(\mathrm{plim}_{n\to\infty}\frac{1}{n}(W_nX_n^{**}, Z_n)'F_n\Pi_n^{-1/2}M_{\Pi^{-1/2}F'(-WX\delta_0,X)}\Pi_n^{-1/2}F_n'(W_nX_n^{**}, Z_n)\big)\\
&= \mathrm{rk}\big(\mathrm{plim}_{n\to\infty}\frac{1}{n}(W_nX_n^{**}, Z_n, -W_nX_n\delta_0, X_n)'H_n(W_nX_n^{**}, Z_n, -W_nX_n\delta_0, X_n)\big)\\
&\quad - \mathrm{rk}\big(\mathrm{plim}_{n\to\infty}\frac{1}{n}(-W_nX_n\delta_0, X_n)'H_n(-W_nX_n\delta_0, X_n)\big)\\
&= \mathrm{rk}\big(\mathrm{plim}_{n\to\infty}\frac{1}{n}(W_nX_n^{*}, Z_n, X_n)'H_n(W_nX_n^{*}, Z_n, X_n)\big) - k_x - 1\\
&= k_{x^*} + k_z - 1.
\end{aligned}$$

Similarly, $\mathrm{rk}\big(\frac{1}{n}(W_nX_n^{**}, Z_n)'\tilde{\mathbb{M}}_{(-WX\tilde\delta,X)}(W_nX_n^{**}, Z_n)\big) = \mathrm{rk}\big(\frac{1}{n}(W_nX_n^{*}, Z_n, X_n)'\hat H_n(W_nX_n^{*}, Z_n, X_n)\big) - k_x - 1$. Thus, w.p.a.1. $\mathrm{rk}\big(\frac{1}{n}(W_nX_n^{**}, Z_n)'\tilde{\mathbb{M}}_{(-WX\tilde\delta,X)}(W_nX_n^{**}, Z_n)\big) = \mathrm{rk}\big(\mathrm{plim}_{n\to\infty}\frac{1}{n}(W_nX_n^{**}, Z_n)'\mathbb{M}_{(-WX\delta_0,X)}(W_nX_n^{**}, Z_n)\big)$. By Theorem 1 in Andrews (1987), the result in the proposition follows. □

*Proof of Proposition 3.3.* The proof is similar to that of Proposition 2.3, thus we pay attention to the differences brought by the drift in Assumption 10. Proposition 2.1 shows the consistency of the N2SLS estimator for model (1) when the true parameter vector $\theta_0$ is fixed regardless of the sample size $n$. Suppose that the true parameter vector $\theta_{0n}$ changes with $n$ but still satisfies $\theta_{0n} \to \theta_0$ as $n \to \infty$. Because

$$\lim_{n\to\infty}\frac{1}{n}[\mathrm{E}(F_n'e^{(\alpha-\alpha_0)W_n}D_n)\delta_{0n}, \mathrm{E}(F_n'D_n)] = \lim_{n\to\infty}\frac{1}{n}[\mathrm{E}(F_n'e^{(\alpha-\alpha_0)W_n}D_n)\delta_0, \mathrm{E}(F_n'D_n)],$$

$\frac{1}{n}\bar Q_n(\alpha)$ in the proof of Proposition 2.1 is uniquely zero at $\alpha_0$ for large enough $n$ under Assumption 6. Following almost the same argument as the proof of Proposition 2.1, Theorem 3.4 in White (1994, p. 28) applies under Assumption 10, so the N2SLS estimator $\hat\theta_n = \theta_{0n} + o_p(1)$. With the reparameterization $\omega$, we have $\hat\omega_n - \omega_{0n} = o_p(1)$. The derivatives of $\hat Q_n^*(\omega)$ at $\omega_{0n}$ still have the same orders as in the case without the drift. Then $\sqrt{n}(\hat\phi_n - \phi_{0n})^2 = O_p(1)$, $\sqrt{n}(\hat\psi_n - \psi_{0n}) = O_p(1)$, and (C.1′)–(C.4) with $\omega_0$ replaced by $\omega_{0n}$ hold. Note that $\frac{1}{\sqrt{n}}\frac{\partial^2\hat Q_n^*(\omega_{0n})}{\partial\phi^2} - \frac{1}{n}\frac{\partial^3\hat Q_n^*(\omega_{0n})}{\partial\phi^2\partial\psi'}\big(\frac{1}{n}\frac{\partial^2\hat Q_n^*(\omega_{0n})}{\partial\psi\partial\psi'}\big)^{-1}\frac{1}{\sqrt{n}}\frac{\partial\hat Q_n^*(\omega_{0n})}{\partial\psi} = R_n + O_p(n^{-1/2})$, where $R_n = \frac{2}{\sqrt{n}}(W_n^2X_n\delta_0)'\mathbb{M}_DV_n = O_p(1)$, and $\frac{1}{6n}\frac{\partial^4\hat Q_n^*(\omega_{0n})}{\partial\phi^4} - \frac{1}{n}\frac{\partial^3\hat Q_n^*(\omega_{0n})}{\partial\phi^2\partial\psi'}\big(\frac{1}{n}\frac{\partial^2\hat Q_n^*(\omega_{0n})}{\partial\psi\partial\psi'}\big)^{-1}\frac{1}{2n}\frac{\partial^3\hat Q_n^*(\omega_{0n})}{\partial\phi^2\partial\psi} = S_n + O_p(n^{-1/2})$, where

$$S_n = \frac{1}{n}(W_n^2X_n\delta_0)'\mathbb{M}_DW_n^2X_n\delta_0 = O(1)$$

and $S_n \geq 0$. Similar to the proof of Proposition 2.3, when $R_n > 0$, $\sqrt{n}(\hat\phi_n - \phi_{0n})^2 = o_p(1)$, and

$$\sqrt{n}(\hat\psi_n - \psi_{0n}) = -\big(\frac{1}{n}\frac{\partial^2\hat Q_n^*(\omega_{0n})}{\partial\psi\partial\psi'}\big)^{-1}\frac{1}{\sqrt{n}}\frac{\partial\hat Q_n^*(\omega_{0n})}{\partial\psi} + o_p(1) = L_n + o_p(1),$$

where $L_n = (\frac{1}{n}D_n'H_nD_n)^{-1}\frac{1}{\sqrt{n}}D_n'H_nV_n \xrightarrow{d} L$ with $L = N(0, (\lim_{n\to\infty}\frac{1}{n}\mathrm{E}(D_n'F_n)\bar\Pi_n^{-1}\mathrm{E}(F_n'D_n))^{-1})$. For $R_n < 0$, $R_n + S_n\sqrt{n}(\hat\phi_n - \phi_{0n})^2 = O_p(n^{-1/4})$ and thus $\sqrt{n}(\hat\phi_n - \phi_{0n})^2 = J_{1n} + o_p(1)$, where $J_{1n} = -S_n^{-1}R_n$. Then by (C.3), $\sqrt{n}(\hat\psi_n - \psi_{0n}) = J_{2n} + o_p(1)$ when $R_n < 0$, where

$$J_{2n} = L_n + \big(\frac{2}{n}D_n'H_nD_n\big)^{-1}\frac{1}{n}D_n'H_nW_n^2X_n\delta_0J_{1n}.$$

Alternatively, when $R_n < 0$, by (C.1′) and (C.2′), we are essentially solving (8) with $\omega_0$ replaced by $\omega_{0n}$. Then the leading order term of $[\sqrt{n}(\hat{\phi}_n - \phi_{0n})^2, \sqrt{n}(\hat{\psi}_n - \psi_{0n})']'$ is $J_n = \mathbb{J}_n + o_p(1)$, where $J_n = (J_{1n}, J'_{2n})' \xrightarrow{d} J$, with $J = (J_1, J'_2)' = N(0, \lim_{n\to\infty} \mathrm{E}(\mathbb{J}_n\mathbb{J}'_n))$. When $R_n < 0$, the sign of $n^{1/4}(\hat{\phi}_n - \phi_{0n})$ must be chosen to minimize $K_n$ in (C.7) (with $\omega_0$ replaced by $\omega_{0n}$). Note that $K_n = (\hat{\phi}_n - \phi_{0n})\mathbb{K}_n + o_p(n^{-1/4})$, where

$$
\begin{aligned}
\mathbb{K}_n &= 2\big(\frac{1}{\sqrt{n}}(W_nX_n^{**}, Z_n)\kappa + V_n\big)'W'_nH_nV_n \\
&\quad + \begin{pmatrix} \sqrt{n}(\hat{\phi}_n - \phi_{0n})^2 \\ \sqrt{n}(\hat{\psi}_n - \psi_{0n}) \end{pmatrix}' \Bigg[\begin{pmatrix} \frac{1}{3\sqrt{n}}(X_n\delta_0)'(W_n'^3H_n + 3W_n'^2H_nW_n)V_n + \frac{1}{n}(W_n^2X_n\delta_0)'H_nW_n(W_nX_n^{**}, Z_n)\kappa \\ -2D'_nH_nW_n\big(\frac{1}{n}(W_nX_n^{**}, Z_n)\kappa + \frac{1}{\sqrt{n}}V_n\big) \end{pmatrix} \\
&\quad - \begin{pmatrix} \frac{1}{6n}(W_n^3X_n\delta_0)'H_n(-W_n^2X_n\delta_0, 2D_n) \\ 0 \end{pmatrix}\begin{pmatrix} \sqrt{n}(\hat{\phi}_n - \phi_{0n})^2 \\ \sqrt{n}(\hat{\psi}_n - \psi_{0n}) \end{pmatrix}\Bigg] \\
&= 2\big(\frac{1}{\sqrt{n}}(W_nX_n^{**}, Z_n)\kappa + V_n\big)'W'_nH_nV_n \\
&\quad + J'_n\Bigg[\begin{pmatrix} \frac{1}{3\sqrt{n}}(X_n\delta_0)'(W_n'^3H_n + 3W_n'^2H_nW_n)V_n + \frac{1}{n}(W_n^2X_n\delta_0)'H_nW_n(W_nX_n^{**}, Z_n)\kappa \\ -2D'_nH_nW_n\big(\frac{1}{n}(W_nX_n^{**}, Z_n)\kappa + \frac{1}{\sqrt{n}}V_n\big) \end{pmatrix} \\
&\quad - \begin{pmatrix} \frac{1}{6n}(W_n^3X_n\delta_0)'H_n(-W_n^2X_n\delta_0, 2D_n) \\ 0 \end{pmatrix} J_n\Bigg] + o_p(1).
\end{aligned}
$$

Thus, $\mathrm{P}(n^{1/4}(\hat{\phi}_n - \phi_{0n})\mathbb{K}_n < 0 | R_n < 0) \to 1$ as $n \to \infty$. Equivalently, $\mathrm{P}(I(n^{1/4}(\hat{\phi}_n - \phi_{0n}) < 0) = I(\mathbb{K}_n > 0) | R_n < 0) \to 1$ as $n \to \infty$. Comparing the $\mathbb{K}_n$ above with that in the proof of Proposition 2.3, additional terms appear due to the drift in Assumption 10. Accounting for those additional terms, the asymptotic distribution of $\hat{\omega}_n$ in the proposition follows. □

*Proof of Proposition 3.4.* When $R_n < 0$, taking a fourth order Taylor expansion of $\hat{Q}_n^*(\hat{\omega}_n)$ at $\omega_0$, we have

$$
\begin{aligned}
\hat{Q}_n^*(\hat{\omega}_n) - \hat{Q}_n^*(\omega_0) &= \big(\frac{1}{2\sqrt{n}}\frac{\partial \hat{Q}_n^*(\omega_0)}{\partial \phi^2}, \frac{1}{\sqrt{n}}\frac{\partial \hat{Q}_n^*(\omega_0)}{\partial \psi'}\big)\begin{pmatrix} \sqrt{n}(\hat{\phi}_n - \phi_0)^2 \\ \sqrt{n}(\hat{\psi}_n - \psi_0) \end{pmatrix} \\
&\quad + \begin{pmatrix} \sqrt{n}(\hat{\phi}_n - \phi_0)^2 \\ \sqrt{n}(\hat{\psi}_n - \psi_0) \end{pmatrix}' \begin{pmatrix} \frac{1}{24n}\frac{\partial^4 \hat{Q}_n^*(\omega_0)}{\partial \phi^4} & \frac{1}{4n}\frac{\partial^3 \hat{Q}_n^*(\omega_0)}{\partial \phi^2 \partial \psi'} \\ \frac{1}{4n}\frac{\partial^3 \hat{Q}_n^*(\omega_0)}{\partial \phi^2 \partial \psi} & \frac{1}{2n}\frac{\partial^2 \hat{Q}_n^*(\omega_0)}{\partial \psi \partial \psi'} \end{pmatrix} \begin{pmatrix} \sqrt{n}(\hat{\phi}_n - \phi_0)^2 \\ \sqrt{n}(\hat{\psi}_n - \psi_0) \end{pmatrix} + o_p(1) \\
&= -\big(\frac{1}{\sqrt{n}}(W_nX_n^{**}, Z_n)\kappa + V_n\big)'\mathbb{P}_{(-W^2X\delta_0, D)}\big(\frac{1}{\sqrt{n}}(W_nX_n^{**}, Z_n)\kappa + V_n\big) + o_p(1);
\end{aligned}
$$

when $R_n > 0$, we have

$$
\begin{aligned}
\hat{Q}_n^*(\hat{\omega}_n) - \hat{Q}_n^*(\omega_0) &= \frac{\partial \hat{Q}_n^*(\omega_0)}{\partial \psi'}(\hat{\psi}_n - \psi_0) + \frac{1}{2}(\hat{\psi}_n - \psi_0)'\frac{\partial^2 \hat{Q}_n^*(\omega_0)}{\partial \psi \partial \psi'}(\hat{\psi}_n - \psi_0) + o_p(1) \\
&= -\big(\frac{1}{\sqrt{n}}(W_nX_n^{**}, Z_n)\kappa + V_n\big)'\mathbb{P}_D\big(\frac{1}{\sqrt{n}}(W_nX_n^{**}, Z_n)\kappa + V_n\big) + o_p(1).
\end{aligned}
$$

By the mean value theorem and (17), we have

$$
\begin{aligned}
\hat{Q}_n(\Psi_0, 0) - \hat{Q}_n(\tilde{\Psi}_n, 0) &= \frac{1}{2}\sqrt{n}(\Psi_0 - \tilde{\Psi}_n)'\frac{1}{n}\frac{\partial^2 \hat{Q}_n(\check{\Psi}_n, 0)}{\partial \Psi \partial \Psi'}\sqrt{n}(\Psi_0 - \tilde{\Psi}_n) \\
&= \big(\frac{1}{\sqrt{n}}(W_nX_n^{**}, Z_n)\kappa + V_n\big)'\mathbb{P}_{(-WX\delta_0, X)}\big(\frac{1}{\sqrt{n}}(W_nX_n^{**}, Z_n)\kappa + V_n\big) + o_p(1),
\end{aligned}
$$

where $\check{\Psi}_n$ lies between $\Psi_0$ and $\tilde{\Psi}_n$. Thus,

$$\begin{aligned}&\hat{Q}_n(\tilde{\theta}_n) - \hat{Q}_n(\hat{\theta}_n)\\&= I(R_n < 0)\big(\frac{1}{\sqrt{n}}(W_nX_n^{**}, Z_n)\kappa + V_n\big)'(\mathbb{P}_{(-W^2X\delta_0,D)} - \mathbb{P}_{(-WX\delta_0,X)})\big(\frac{1}{\sqrt{n}}(W_nX_n^{**}, Z_n)\kappa + V_n\big)\\&\quad + I(R_n > 0)\big(\frac{1}{\sqrt{n}}(W_nX_n^{**}, Z_n)\kappa + V_n\big)'(\mathbb{P}_D - \mathbb{P}_{(-WX\delta_0,X)})\big(\frac{1}{\sqrt{n}}(W_nX_n^{**}, Z_n)\kappa + V_n\big) + o_p(1).\end{aligned}$$

According to the proof of Proposition 3.1,

$$\begin{aligned}&\hat{Q}_n(\tilde{\theta}_n) - \hat{Q}_n(\hat{\theta}_n)\\&= I(R_n < 0)\big[\big(\frac{1}{\sqrt{n}}(W_nX_n^{**}, Z_n)\kappa + V_n\big)'F_n\Pi_n^{-1/2}(P_B + P_A)\Pi_n^{-1/2}F_n'\big(\frac{1}{\sqrt{n}}(W_nX_n^{**}, Z_n)\kappa + V_n\big)\big]\\&\quad + I(R_n > 0)\big(\frac{1}{\sqrt{n}}(W_nX_n^{**}, Z_n)\kappa + V_n\big)'F_n\Pi_n^{-1/2}P_A\Pi_n^{-1/2}F_n'\big(\frac{1}{\sqrt{n}}(W_nX_n^{**}, Z_n)\kappa + V_n\big) + o_p(1)\\&= I(R_n < 0)\big[V_n'F_n\Pi_n^{-1/2}P_B\Pi_n^{-1/2}F_n'V_n + \big(\frac{1}{\sqrt{n}}(W_nX_n^{**}, Z_n)\kappa + V_n\big)'F_n\Pi_n^{-1/2}P_A\Pi_n^{-1/2}F_n'\big(\frac{1}{\sqrt{n}}(W_nX_n^{**}, Z_n)\kappa + V_n\big)\big]\\&\quad + I(R_n > 0)\big(\frac{1}{\sqrt{n}}(W_nX_n^{**}, Z_n)\kappa + V_n\big)'F_n\Pi_n^{-1/2}P_A\Pi_n^{-1/2}F_n'\big(\frac{1}{\sqrt{n}}(W_nX_n^{**}, Z_n)\kappa + V_n\big) + o_p(1),\end{aligned}$$

where the second equality holds because $B_n'\Pi_n^{-1/2}F_n'(W_nX_n^{**}, Z_n) = (W_n^2X_n\delta_0)'F_n\Pi_n^{-1/2}M_{\Pi^{-1/2}F'D}\Pi_n^{-1/2}F_n'(W_nX_n^{**}, Z_n) = 0$. The expression for $\hat{Q}_n(\tilde{\theta}_n) - \hat{Q}_n(\hat{\theta}_n)$ is the same as that in the proof of Proposition 3.1 except for the additional terms due to the drift $\frac{1}{\sqrt{n}}\kappa$. Note that $(W_nX_n^{**}, Z_n)'F_n\Pi_n^{-1/2}P_A\Pi_n^{-1/2}F_n'(W_nX_n^{**}, Z_n) = (W_nX_n^{**}, Z_n)'F_n\Pi_n^{-1/2}(P_{\Pi^{-1/2}F'D} - P_{\Pi^{-1/2}F'(-WX\delta_0,X)})\Pi_n^{-1/2}F_n'(W_nX_n^{**}, Z_n) = (W_nX_n^{**}, Z_n)'F_n\Pi_n^{-1/2}(I_{k_f} - P_{\Pi^{-1/2}F'(-WX\delta_0,X)})\Pi_n^{-1/2}F_n'(W_nX_n^{**}, Z_n) = (W_nX_n^{**}, Z_n)'\mathbb{M}_{(-WX\delta_0,X)}(W_nX_n^{**}, Z_n)$, where the second equality uses the fact that $P_{\Pi^{-1/2}F'D}\Pi_n^{-1/2}F_n'(W_nX_n^{**}, Z_n) = \Pi_n^{-1/2}F_n'(W_nX_n^{**}, Z_n)$. Hence, the result in the proposition follows. ☐

*Proof of Proposition 3.5.* This proposition is proved in the main text in front of the proposition. ☐

*Proof of Proposition 3.6.* A first order Taylor expansion of $\frac{\partial \hat{Q}_n(\tilde{\Psi}_n,0)}{\partial \Psi} = 0$ at $(\alpha_0 - n^{1/4}, \delta_0')'$ yields

$$0 = \frac{\partial \hat{Q}_n(\tilde{\Psi}_n, 0)}{\partial \Psi} = \frac{\partial \hat{Q}_n(\alpha_0 - n^{-1/4}, \delta_0, 0)}{\partial \Psi} + \frac{\partial^2 \hat{Q}_n(\alpha_0 - n^{-1/4}, \delta_0, 0)}{\partial \Psi \partial \Psi'}\begin{pmatrix}\tilde{\alpha}_n - \alpha_0 + n^{1/4}\\ \tilde{\delta}_n - \delta_0\end{pmatrix} + O_p(1).$$

By using the derivatives in (4)–(5) and (22)–(24), and the definition of a matrix exponential,

$$0 = 2(W_nX_n\delta_0, -X_n)'H_n\big(V_n - \frac{1}{2}n^{-1/2}W_n^2X_n\delta_0\big) + 2(-W_nX_n\delta_0, X_n)'H_n(-W_nX_n\delta_0, X_n)\begin{pmatrix}\tilde{\alpha}_n - \alpha_0 + n^{-1/4}\\ \tilde{\delta}_n - \delta_0\end{pmatrix} + O_p(1).$$

Thus,

$$\sqrt{n}\begin{pmatrix}\tilde{\alpha}_n - \alpha_0 + n^{-1/4}\\ \tilde{\delta}_n - \delta_0\end{pmatrix} = \big[\frac{1}{n}(-W_nX_n\delta_0, X_n)'H_n(-W_nX_n\delta_0, X_n)\big]^{-1}(-W_nX_n\delta_0, X_n)'H_n\big(\frac{1}{\sqrt{n}}V_n - \frac{1}{2n}W_n^2X_n\delta_0\big) + o_p(1). \tag{C.8}$$

A first order Taylor expansion of $\frac{\partial \hat{Q}_n(\tilde{\theta}_n)}{\partial \zeta}$ at $(\alpha_0 - n^{-1/4}, \delta_0', 0)'$ yields

$$\begin{aligned}\frac{1}{\sqrt{n}}\frac{\partial \hat{Q}_n(\tilde{\theta}_n)}{\partial \zeta} &= \frac{1}{\sqrt{n}}\frac{\partial \hat{Q}_n(\alpha_0 - n^{-1/4}, \delta_0, 0)}{\partial \zeta} + \frac{1}{n}\frac{\partial^2 \hat{Q}_n(\alpha_0 - n^{-1/4}, \delta_0, 0)}{\partial \zeta \partial \Psi'}\sqrt{n}\begin{pmatrix}\tilde{\alpha}_n - \alpha_0 + n^{-1/4}\\ \tilde{\delta}_n - \delta_0\end{pmatrix} + o_p(1)\\&= -\frac{2}{\sqrt{n}}(W_nX_n^{**}, Z_n)'H_n(e^{(\alpha_0 - n^{-1/4})W_n}Y_n - X_n\delta_0)\\&\quad + \frac{2}{n}(W_nX_n^{**}, Z_n)'H_n[-W_ne^{(\alpha_0 - n^{-1/4})W_n}Y_n, X_n]\sqrt{n}\begin{pmatrix}\tilde{\alpha}_n - \alpha_0 + n^{-1/4}\\ \tilde{\delta}_n - \delta_0\end{pmatrix} + o_p(1).\end{aligned}$$

Substituting (C.8) into the above equation yields

$$\frac{1}{\sqrt{n}}\frac{\partial \hat{Q}_n(\tilde{\theta}_n)}{\partial \zeta} = -2(W_nX_n^{**}, Z_n)'\mathbb{M}_{(-WX\delta_0,X)}\big(\frac{1}{\sqrt{n}}V_n - \frac{1}{2n}W_n^2X_n\delta_0\big) + o_p(1)$$
$$\xrightarrow{d} N\Big(\text{plim}_{n\to\infty}\frac{1}{n}(W_nX_n^{**}, Z_n)'\mathbb{M}_{(-WX\delta_0,X)}W_n^2X_n\delta_0, \text{plim}_{n\to\infty}\frac{4}{n}(W_nX_n^{**}, Z_n)'\mathbb{M}_{(-WX\delta_0,X)}(W_nX_n^{**}, Z_n)\Big).$$

Hence, the result in the proposition follows. □

## For Section 4: AGLASSO estimator

*Proof of Proposition 4.1.* Let $\bar{g}_n(\theta) = \mathrm{E}[g_n(\theta)]$. By the definition of $\hat{\theta}_n$,

$$\frac{1}{n}\hat{Q}_n(\hat{\theta}_n) + \lambda_n\|\tilde{\zeta}_n\|^{-\mu}\|\hat{\zeta}_n\| \le \frac{1}{n}\hat{Q}_n(\theta_0) + \lambda_n\|\tilde{\zeta}_n\|^{-\mu}\|\zeta_0\|. \tag{C.9}$$

Note that $\frac{1}{n}\hat{Q}_n(\theta_0) = \frac{1}{n}V_n'F_n(\frac{1}{n}\hat{\Pi}_n)^{-1}\frac{1}{n}F_n'V_n = o_p(1)$. If $\zeta_0 \neq 0$, as $\tilde{\zeta}_n = \zeta_0 + o_p(1)$, $\lambda_n\|\tilde{\zeta}_n\|^{-\mu}\|\zeta_0\| = O_p(\lambda_n) = o_p(1)$ under Assumption 13; if $\zeta_0 = 0$, $\lambda_n\|\tilde{\zeta}_n\|^{-\mu}\|\zeta_0\| = 0$. As $\lambda_n > 0$ and $\hat{Q}_n(\hat{\theta}_n) \ge 0$, $\frac{1}{n}\hat{Q}_n(\hat{\theta}_n) = o_p(1)$ by (C.9). By Proposition 2.4 and Assumption 3, $\frac{1}{n}\hat{\Pi}_n - \frac{1}{n}\mathrm{E}(\Pi_n) = [\frac{1}{n}\hat{\Pi}_n - \frac{1}{n}\Pi_n] + [\frac{1}{n}\Pi_n - \frac{1}{n}\mathrm{E}(\Pi_n)] = o_p(1)$, where $\lim_{n\to\infty}\frac{1}{n}\mathrm{E}(\Pi_n)$ is nonsingular. Thus $\frac{1}{n}\hat{Q}_n(\hat{\theta}_n) \ge C\|\frac{1}{n}g_n(\hat{\theta}_n)\|^2$ w.p.a.1., where $C$ is a finite positive constant. Then $\|\frac{1}{n}g_n(\hat{\theta}_n)\| = o_p(1)$. Since $\|\frac{1}{n}g_n(\hat{\theta}_n)\| \ge \|\frac{1}{n}\bar{g}_n(\hat{\theta}_n)\| - \|\frac{1}{n}g_n(\hat{\theta}_n) - \frac{1}{n}\bar{g}_n(\hat{\theta}_n)\|$ and $\sup_{\theta\in\Theta}\|\frac{1}{n}g_n(\theta) - \frac{1}{n}\bar{g}_n(\theta)\| = \sup_{\theta\in\Theta}\|\frac{1}{n}F_n'e^{(\alpha-\alpha_0)W_n}D_n\beta_0 - \frac{1}{n}\mathrm{E}(F_n'e^{(\alpha-\alpha_0)W_n}D_n)\beta_0 + \frac{1}{n}F_n'e^{(\alpha-\alpha_0)W_n}V_n - \frac{1}{n}[F_n'D_n - \mathrm{E}(F_n'D_n)]\beta\| = o_p(1)$ by Assumption 5 and Lemma 2, $\|\frac{1}{n}\bar{g}_n(\hat{\theta}_n)\| = o_p(1)$. As $\|\frac{1}{n}\bar{g}_n(\hat{\theta}_n)\| = \|\frac{1}{n}[\mathrm{E}(F_n'e^{(\hat{\alpha}_n-\alpha_0)W_n}D_n)\beta_0, \mathrm{E}(F_n'D_n)]\binom{1}{-\hat{\beta}_n}\|$, Assumption 6 implies that $\hat{\alpha}_n = \alpha_0 + o_p(1)$. Then $\|\frac{1}{n}\bar{g}_n(\hat{\theta}_n)\| = \|\frac{1}{n}\mathrm{E}(F_n'D_n)(\beta_0 - \hat{\beta}_n) + o_p(1)\|$. Assumption 6 further implies that $\hat{\beta}_n = \beta_0 + o_p(1)$. Thus the result in the proposition holds. □

*Proof of Proposition 4.2.* With $\zeta_0 = 0$, by the definition of $\hat{\theta}_n$,

$$\frac{1}{n}\hat{Q}_n(\hat{\theta}_n) + \lambda_n\|\tilde{\zeta}_n\|^{-\mu}\|\hat{\zeta}_n\| \le \frac{1}{n}\hat{Q}_n(\theta_0) + \lambda_n\|\tilde{\zeta}_n\|^{-\mu}\|\zeta_0\| = \frac{1}{n}\hat{Q}_n(\theta_0). \tag{C.10}$$

Note that $\lambda_n > 0$, $\frac{1}{n}\hat{Q}_n(\hat{\theta}_n) \ge C\|\frac{1}{n}g_n(\hat{\theta}_n)\|^2$ w.p.a.1. for some finite positive constant $C$, and

$$\frac{1}{n}\hat{Q}_n(\theta_0) = \frac{1}{n}V_n'F_n\big(\frac{1}{n}\hat{\Pi}\big)^{-1}\frac{1}{n}F_n'V_n = O_p(n^{-1}).$$

Then $\frac{1}{n}g_n(\hat{\theta}_n) = O_p(n^{-1/2})$. If $\hat{\zeta}_n \neq 0$, then the AGLASSO criterion function is differentiable at $\hat{\theta}_n = (\hat{\alpha}_n, \hat{\delta}_n', \hat{\zeta}_n')'$ and we have the following first order condition with respect to $\zeta$:

$$-\frac{2}{n}(W_nX_n^{**}, Z_n)'F_n\big(\frac{1}{n}\hat{\Pi}_n\big)^{-1}\frac{1}{n}g_n(\hat{\theta}_n) + \lambda_n\|\tilde{\zeta}_n\|^{-\mu}\hat{\zeta}_n\|\hat{\zeta}_n\|^{-1} = 0. \tag{C.11}$$

Note that $\frac{1}{n}(W_nX_n^{**}, Z_n)'F_n = O_p(1)$. As $\hat{\zeta}_n \neq 0$, there must be some component $\hat{\zeta}_{nj}$ of $\hat{\zeta}_n = (\hat{\zeta}_{n1}, \ldots, \hat{\zeta}_{np})'$, where $p$ is the length of $\zeta$, such that $|\hat{\zeta}_{nj}| = \max\{|\hat{\zeta}_{nk}| : 1 \le k \le p\}$. Then $|\hat{\zeta}_{nj}|/\|\hat{\zeta}_n\| \ge 1/\sqrt{p} > 0$. Under Assumption 14, (C.11) cannot hold w.p.a.1., which is a contradiction to the first order condition. Hence the result in the proposition follows. □

*Proof of Proposition 4.3.* The first order derivative of the AGLASSO criterion function with respect to $\Psi$ evaluated at $\hat{\theta}_n$ is

$$\frac{2}{n}(W_ne^{\hat{\alpha}_nW_n}y_n, -X_n)'F_n(\frac{1}{n}\hat{\Pi}_n)^{-1}\frac{1}{n}F_n'(e^{\hat{\alpha}_nW_n}y_n - D_n\hat{\beta}_n) = 0.$$

With $\zeta_0 = 0$, Proposition 4.2 shows that $\hat{\zeta}_n = 0$ w.p.a.1. Hence, the following equation holds w.p.a.1:

$$\frac{2}{n}(W_n e^{\hat{\alpha}_n W_n} y_n, -X_n)' F_n(\frac{1}{n}\hat{\Pi}_n)^{-1}\frac{1}{n}F_n'(e^{\hat{\alpha}_n W_n} y_n - X_n\hat{\delta}_n) = 0.$$

This first order condition is exactly the same as the one derived from the corresponding N2SLS criterion function by imposing the constraint $\zeta = 0$. Thus the oracle property becomes apparent. By the mean value theorem,

$$0 = \frac{2}{n}(W_n e^{\hat{\alpha}_n W_n} y_n, -X_n)' F_n\big(\frac{1}{n}\hat{\Pi}_n\big)^{-1}\frac{1}{n}F_n'[e^{\alpha_0 W_n} y_n - X_n\delta_0 + (W_n e^{\ddot{\alpha}_n W_n} y_n, -X_n)(\hat{\Psi}_n - \Psi_0)],$$

where $\ddot{\Psi}_n$ lies between $\hat{\Psi}_n$ and $\Psi_0$. Thus,

$$\begin{aligned}
&\sqrt{n}(\hat{\Psi}_n - \Psi_0)\\
&= -\big[\frac{1}{n}(W_n e^{\hat{\alpha}_n W_n} y_n, -X_n)' F_n(\frac{1}{n}\hat{\Pi}_n)^{-1}\frac{1}{n}F_n'(W_n e^{\ddot{\alpha}_n W_n} y_n, -X_n)\big]^{-1}\frac{1}{n}(W_n e^{\hat{\alpha}_n W_n} y_n, -X_n)' F_n\big(\frac{1}{n}\hat{\Pi}_n\big)^{-1}\frac{1}{\sqrt{n}}F_n' V_n\\
&= \big[\frac{1}{n}(-W_n X_n\delta_0, X_n)' F_n\big(\frac{1}{n}\Pi_n\big)^{-1}\frac{1}{n}F_n'(-W_n X_n\delta_0, X_n)\big]^{-1}\frac{1}{n}(-W_n X_n\delta_0, X_n)' F_n(\frac{1}{n}\Pi_n)^{-1}\frac{1}{\sqrt{n}}F_n' V_n + o_p(1)\\
&\xrightarrow{d} N\big(0, \lim_{n\to\infty}\frac{1}{n}\big\{\mathrm{E}[(-W_n X_n\delta_0, X_n)' F_n]\bar{\Pi}_n^{-1}\,\mathrm{E}[F_n'(-W_n X_n\delta_0, X_n)]\big\}^{-1}\big),
\end{aligned}$$

where the asymptotic distribution follows by Lemma 2($ii$). □

*Proof of Proposition 4.4.* Note that

$$\begin{aligned}
&\frac{1}{n}g_n'(\hat{\theta}_n)\big(\frac{1}{n}\hat{\Pi}_n\big)^{-1}\frac{1}{n}g_n(\hat{\theta}_n) - \frac{1}{n}g_n'(\theta_0)\big(\frac{1}{n}\hat{\Pi}_n\big)^{-1}\frac{1}{n}g_n(\theta_0)\\
&= \frac{1}{n}[g_n(\hat{\theta}_n) - g_n(\theta_0)]'\big(\frac{1}{n}\hat{\Pi}_n\big)^{-1}\frac{1}{n}[g_n(\hat{\theta}_n) - g_n(\theta_0)] + \frac{2}{n}[g_n(\hat{\theta}_n) - g_n(\theta_0)]'\big(\frac{1}{n}\hat{\Pi}_n\big)^{-1}\frac{1}{n}g_n(\theta_0) \qquad \text{(C.12)}\\
&\geq C_1\big\|\frac{1}{n}g_n(\hat{\theta}_n) - \frac{1}{n}g_n(\theta_0)\big\|^2 - C_2\big\|\frac{1}{n}g_n(\hat{\theta}_n) - \frac{1}{n}g_n(\theta_0)\big\|\big\|\frac{1}{n}g_n(\theta_0)\big\|,
\end{aligned}$$

w.p.a.1., where $C_1$ and $C_2$ are finite positive constants, and the inequality follows by Assumption 2 and the Cauchy-Schwarz inequality. By the mean value theorem,

$$\frac{1}{n}g_n(\hat{\theta}_n) - \frac{1}{n}g_n(\theta_0) = \frac{1}{n}\mathrm{E}\big(\frac{\partial g_n(\theta_0)}{\partial\theta'}\big)(\hat{\theta}_n - \theta_0) + \frac{1}{n}\big[\frac{\partial g_n(\ddot{\theta}_n)}{\partial\theta'} - \mathrm{E}\big(\frac{\partial g_n(\theta_0)}{\partial\theta'}\big)\big](\hat{\theta}_n - \theta_0), \qquad \text{(C.13)}$$

where $\ddot{\theta}_n$ lies between $\hat{\theta}_n$ and $\theta_0$. In addition, $\frac{1}{n}\frac{\partial g_n(\theta)}{\partial\theta'} - \frac{1}{n}\frac{\partial g_n(\theta_0)}{\partial\theta'} = [\frac{1}{n}F_n' W_n^2 e^{\alpha_1 W_n} Y_n(\alpha - \alpha_0), 0_{k_f\times k_d}]$ for some $\alpha_1$ between $\alpha$ and $\alpha_0$, where $\frac{1}{n}F_n' W_n^2 e^{\alpha W_n} Y_n = O_p(1)$ uniformly in a neighborhood of $\alpha_0$ by Lemma 2. Furthermore, $\frac{1}{n}\frac{\partial g_n(\theta_0)}{\partial\theta'} - \frac{1}{n}\mathrm{E}\big(\frac{\partial g_n(\theta_0)}{\partial\theta'}\big) = [\frac{1}{n}F_n' W_n D_n\beta_0 - \frac{1}{n}\mathrm{E}(F_n' W_n D_n)\beta_0 + \frac{1}{n}F_n' W_n V_n, -\frac{1}{n}F_n' D_n + \frac{1}{n}\mathrm{E}(F_n' D_n)] = o_p(1)$ by Assumption 5 and Lemma 2. Thus, $\frac{1}{n}[\frac{\partial g_n(\ddot{\theta}_n)}{\partial\theta'} - \mathrm{E}(\frac{\partial g_n(\theta_0)}{\partial\theta'})] = o_p(1)$. Since $\mathrm{E}(\frac{1}{n}\frac{\partial g_n(\theta_0)}{\partial\theta'}) = O(1)$ has full rank when $\zeta_0 \neq 0$ for large enough $n$, (C.13) implies that

$$C_3(1 - a_n)\|\hat{\theta}_n - \theta_0\| \leq \frac{1}{n}\|g_n(\hat{\theta}_n) - g_n(\theta_0)\| \leq C_3(1 + a_n)\|\hat{\theta}_n - \theta_0\|, \qquad \text{(C.14)}$$

where $C_3$ is a finite positive constant, $a_n \geq 0$ and $a_n = o_p(1)$. By (C.9),

$$\frac{1}{n}g_n(\hat{\theta}_n)'\hat{\Pi}_n^{-1}g_n(\hat{\theta}_n) - \frac{1}{n}g_n(\theta_0)'\hat{\Pi}_n^{-1}g_n(\theta_0) \leq \lambda_n\|\tilde{\zeta}_n\|^{-\mu}(\|\zeta_0\| - \|\hat{\zeta}_n\|) \leq \lambda_n\|\tilde{\zeta}_n\|^{-\mu}\|\hat{\zeta}_n - \zeta_0\| \leq C_4\lambda_n\|\hat{\theta}_n - \theta_0\| \qquad \text{(C.15)}$$

w.p.a.1., where $C_4$ is a finite positive constant, and the last inequality follows since $\zeta_0 \neq 0$ and $\tilde{\zeta}_n = \zeta_0 + o_p(1)$. Combining (C.12), (C.14) and (C.15) yields

$$C_1 C_3^2 (1-a_n)^2 \|\hat{\theta}_n - \theta_0\|^2 - C_2 C_3 (1+a_n) \|\hat{\theta}_n - \theta_0\| \cdot \frac{1}{n}\|g_n(\theta_0)\| \leq C_4 \lambda_n \|\hat{\theta}_n - \theta_0\|.$$

The above inequality can be written as

$$\|\hat{\theta}_n - \theta_0\| \cdot \big[C_1 C_3^2 (1-a_n)^2 \|\hat{\theta}_n - \theta_0\| - C_2 C_3 (1+a_n) \big\|\frac{1}{n} g_n(\theta_0)\big\| - C_4 \lambda_n\big] \leq 0.$$

As $\frac{1}{n} g_n(\theta_0) = O_p(n^{-1/2})$, $\|\hat{\theta}_n - \theta_0\| = O_p(n^{-1/2} + \lambda_n)$. □

*Proof of Proposition 4.5.* If $\zeta_0 \neq 0$, the first order condition of the AGLASSO criterion function with respect to $\theta$ is:

$$\frac{2}{n}(W_n e^{\hat{\alpha}_n W_n} y_n, -D_n)' F_n \big(\frac{1}{n}\hat{\Pi}_n\big)^{-1} \frac{1}{n} g_n(\hat{\theta}_n) + \lambda_n \|\tilde{\zeta}_n\|^{-\mu} \|\hat{\zeta}_n\|^{-1} \begin{pmatrix} 0 \\ \hat{\zeta}_n \end{pmatrix} = 0. \qquad \text{(C.16)}$$

By (C.14), Proposition 4.4 and Assumption 15, $\frac{1}{n} g_n(\hat{\theta}_n) = O_p(n^{-1/2})$. Then the first term on the l.h.s. of (C.16) has the order $O_p(n^{-1/2})$. As $\zeta_0 \neq 0$, $\tilde{\zeta}_n = \zeta_0 + o_p(1) = O_p(1)$ and $\hat{\zeta}_n = \zeta_0 + o_p(1) = O_p(1)$. By Assumption 15, the second term on the l.h.s. of (C.16) has the order $o_p(n^{-1/2})$. Hence, by the mean value theorem,

$$\frac{2}{n}(W_n e^{\hat{\alpha}_n W_n} y_n, -D_n)' F_n \big(\frac{1}{n}\hat{\Pi}_n\big)^{-1} \big[\frac{1}{\sqrt{n}} g_n(\theta_0) + \frac{1}{n} F_n' (W_n e^{\ddot{\alpha}_n W_n} y_n, -D_n) \sqrt{n}(\hat{\theta}_n - \theta_0)\big] + o_p(1) = 0,$$

where $\ddot{\alpha}_n$ lies between $\alpha_0$ and $\hat{\alpha}_n$. If follows that

$$\begin{aligned}
\sqrt{n}(\hat{\theta}_n - \theta_0) &= -\big[\frac{1}{n}(W_n e^{\hat{\alpha}_n W_n} y_n, -D_n)' H_n (W_n e^{\ddot{\alpha}_n W_n} y_n, -D_n)\big]^{-1} \frac{1}{\sqrt{n}} (W_n e^{\hat{\alpha}_n W_n} y_n, -D_n)' H_n V_n + o_p(1) \\
&= -\big[\frac{1}{n}(W_n D_n \beta_0, -D_n)' H_n (W_n D_n \beta_0, -D_n)\big]^{-1} \frac{1}{\sqrt{n}} (W_n D_n \beta_0, -D_n)' H_n V_n + o_p(1) \\
&\xrightarrow{d} N\big(0, \lim_{n\to\infty} \big\{\frac{1}{n} \mathrm{E}[(-W_n D_n \beta_0, D_n)' F_n] \bar{\Pi}_n^{-1} \mathrm{E}[F_n'(-W_n D_n \beta_0, D_n)]\big\}^{-1}\big).
\end{aligned}$$

□

*Proof of Proposition 4.6.* We consider the following two cases separately: (1) $\zeta_0 \neq 0$, but $\hat{\zeta}_\lambda = 0$; (2) $\zeta_0 = 0$, but $\hat{\zeta}_\lambda \neq 0$.

*Case 1: $\zeta_0 \neq 0$, but $\hat{\zeta}_\lambda = 0$.* Let $\ddot{\theta}_n = (\ddot{\Psi}_n', 0)'$ be the restricted N2SLS estimator with the restriction $\zeta = 0$ imposed, where $\ddot{\Psi}_n = \arg\min_\Psi [(e^{\alpha W_n} y_n - X_n \delta)' \hat{H}_n (e^{\alpha W_n} y_n - X_n \delta)]$. As $\zeta_0 \neq 0$, $\bar{\theta} \equiv \mathrm{plim}_{n\to\infty} \ddot{\theta}_n \neq \theta_0$. Note that $\frac{1}{n} g_n(\ddot{\theta}_n) = \frac{1}{n}\bar{g}_n(\ddot{\theta}_n) + o_p(1) = \frac{1}{n}\bar{g}_n(\bar{\theta}) + o_p(1)$, where the first equality follows since $\sup_{\theta\in\Theta} \frac{1}{n}\|g_n(\theta) - \bar{g}_n(\theta)\| = o_p(1)$ as shown in the proof of Proposition 2.1, and the second equality follows by the mean value theorem. By Assumption 6, $\lim_{n\to\infty} \frac{1}{n}\bar{g}_n(\bar{\theta}) \neq 0$. Then $\frac{1}{n}\hat{Q}_n(\ddot{\theta}_n) \xrightarrow{p} C > 0$ for a constant $C$. By the definition of $\ddot{\theta}_n$ and the setting of Case 1, $h_n(\lambda) = \frac{1}{n}\hat{Q}_n(\hat{\theta}_\lambda) - \Gamma_n \geq \frac{1}{n}\hat{Q}_n(\ddot{\theta}_n) - \Gamma_n$. Thus, under Assumption 16, $h_n(\lambda) > C/2 > 0$ w.p.a.1. Furthermore, by Proposition 4.1, $\hat{\theta}_{\bar{\lambda}_n} = \theta_0 + o_p(1)$. Hence, $\frac{1}{n} g_n(\hat{\theta}_{\bar{\lambda}_n}) = \frac{1}{n}\bar{g}_n(\theta_0) + o_p(1) = o_p(1)$ and $\frac{1}{n}\hat{Q}_n(\hat{\theta}_{\bar{\lambda}_n}) = o_p(1)$. It follows that $h_n(\bar{\lambda}_n) = o_p(1)$. Therefore, $\mathrm{P}(\inf_{\lambda\in\{\lambda\in\Lambda:\zeta_0\neq 0,\ \text{but}\ \hat{\zeta}_\lambda = 0\}} h_n(\lambda) > h_n(\bar{\lambda}_n)) \to 1$ as $n \to \infty$.

*Case 2: $\zeta_0 = 0$, but $\hat{\zeta}_\lambda \neq 0$.* Under this setting, $h_n(\lambda) = \frac{1}{n}\hat{Q}_n(\hat{\theta}_\lambda)$. By Proposition 4.2, $\mathrm{P}(\hat{\zeta}_{\bar{\lambda}_n} = 0) \to 1$ as $n \to \infty$. Then w.p.a.1.,

$$\begin{aligned}
n^{1/2}[h_n(\lambda) - h_n(\bar{\lambda}_n)] &= n^{-1/2}\hat{Q}_n(\hat{\theta}_\lambda) - n^{-1/2}\hat{Q}_n(\hat{\theta}_{\bar{\lambda}_n}) + n^{1/2}\Gamma_n \\
&\geq n^{-1/2}\hat{Q}_n(\tilde{\theta}_n) - n^{-1/2}\hat{Q}_n(\hat{\theta}_{\bar{\lambda}_n}) + n^{1/2}\Gamma_n,
\end{aligned} \qquad \text{(C.17)}$$

where $\tilde{\theta}_n$ is the feasible N2SLS estimator (without penalty). By Proposition 2.4, $\tilde{\theta}_n = \theta_0 + O_p(n^{-1/4})$ when $\zeta_0 = 0$. Then by the mean value theorem, $n^{-3/4} g_n(\tilde{\theta}_n) = n^{-3/4} g_n(\theta_0) + \frac{1}{n}\frac{\partial g_n(\dot{\theta}_n)}{\partial \theta'} n^{1/4}(\tilde{\theta}_n - \theta_0) = n^{-3/4} F_n' V_n + \frac{1}{n}\frac{\partial g_n(\dot{\theta}_n)}{\partial \theta'} n^{1/4}(\tilde{\theta}_n - \theta_0)$, where $\dot{\theta}_n$ lies between $\theta_0$ and $\tilde{\theta}_n$. As in the proof of Proposition 4.5, $\frac{1}{n}\frac{\partial g_n(\dot{\theta}_n)}{\partial \theta'} = \frac{1}{n}\mathrm{E}(\frac{\partial g_n(\theta_0)}{\partial \theta'}) + o_p(1) = O_p(1)$. Thus, $n^{-3/4} g_n(\tilde{\theta}_n) = O_p(1)$ and $n^{-1/2}\hat{Q}_n(\tilde{\theta}_n) = O_p(1)$. Since $\mathrm{P}(\hat{\zeta}_{\bar{\lambda}_n} = 0) \to 1$ as $n \to \infty$, $n^{-1/2} g_n(\hat{\theta}_{\bar{\lambda}_n}) = n^{-1/2} F_n'(e^{\hat{\alpha}_{\bar{\lambda}_n} W_n} y_n - X_n \hat{\delta}_{\bar{\lambda}_n})$ w.p.a.1. By Proposition 4.3, $\hat{\Psi}_{\bar{\lambda}_n} = \Psi_0 + O_p(n^{-1/2})$. Then by the mean value theorem, $n^{-1/2} g_n(\hat{\theta}_{\bar{\lambda}_n}) = n^{-1/2} g_n(\theta_0) + \frac{1}{n}\frac{\partial g_n(\ddot{\theta}_n)}{\partial \Psi'} n^{1/2}(\hat{\Psi}_{\bar{\lambda}_n} - \Psi_0) = n^{-1/2} F_n' V_n + \frac{1}{n}\frac{\partial g_n(\ddot{\theta}_n)}{\partial \Psi'} n^{1/2}(\hat{\Psi}_{\bar{\lambda}_n} - \Psi_0) = O_p(1)$, where $\ddot{\theta}_n$ lies between $\theta_0$ and $\hat{\theta}_{\bar{\lambda}_n}$. Thus, $n^{-1/2}\hat{Q}_n(\hat{\theta}_{\bar{\lambda}_n}) = o_p(1)$. Since $n^{1/2}\Gamma_n \to \infty$ as $n \to \infty$ under Assumption 16, (C.17) implies that $\mathrm{P}(\inf_{\lambda \in \{\lambda \in \Lambda : \zeta_0 = 0, \text{ but } \hat{\zeta}_\lambda \neq 0\}} h_n(\lambda) > h_n(\bar{\lambda}_n)) \to 1$ as $n \to \infty$.

Combining the results in the above two cases, we have the result in the proposition. □

# References


Andrews, D.W.K., 1987. Asymptotic results for generalized Wald tests. Econometric Theory 3, 348–358.

Davidson, J., 1994. Stochastic limit theory: An introduction for econometricians. Oxford University Press, USA.

Debarsy, N., Jin, F., Lee, L.F., 2015. Large sample properties of the matrix exponential spatial specification with an application to FDI. Journal of Econometrics 188, 1–21.

Jenish, N., Prucha, I.R., 2012. On spatial processes and asymptotic inference under near-epoch dependence. Journal of Econometrics 170, 178–190.

Kelejian, H.H., Prucha, I.R., 2001. On the asymptotic distribution of the Moran *I* test statistic with applications. Journal of Econometrics 104, 219–257.

Robbins, H., 1955. A remark on Stirling's formula. The American Mathematical Monthly 62, 26–29.

Rotnitzky, A., Cox, D.R., Bottai, M., Robins, J., 2000. Likelihood-based inference with singular information matrix. Bernoulli 6, 243–284.

Ruud, P.A., 2000. An Introduction to Classical Econometric Theory. Oxford University Press.

White, H., 1994. Estimation, Inference and Specification Analysis. New York: Cambridge University Press.

Xu, X., Lee, L.F., 2015. A spatial autoregressive model with a nonlinear transformation of the dependent variable. Journal of Econometrics 186, 1–18.