# Online Supplement to “Identification and Estimation of Categorical Random Coefficient Models”

Zhan Gao* M. Hashem Pesaran†

March 20, 2023

## S.1 Introduction

This online supplement is composed of four sections. Section S.2 provides additional proofs and technical details omitted from the main text. Section S.3 provides additional simulation results. Section S.4 gives additional empirical results. Details of the computational algorithm used are described in Section S.5.

## S.2 Proofs

We include omitted proofs and technical details in this section.

**Proof of Theorem 3.** From (3.1), we have

$$\frac{1}{n}\sum_{i=1}^{n}\mathbf{w}_i y_i = \frac{1}{n}\sum_{i=1}^{n}\mathbf{w}_i\mathbf{w}_i'\boldsymbol{\phi} + \frac{1}{n}\sum_{i=1}^{n}\mathbf{w}_i\xi_i,$$

where $\boldsymbol{\phi} = \mathrm{E}\left(\boldsymbol{\phi}_i\right) = \left(\mathrm{E}\left(\beta_i\right), \boldsymbol{\gamma}'\right)'$, and $\xi_i = u_i + x_i v_i$, which can be written equivalently as

$$\mathbf{q}_{n,wy} = \mathbf{Q}_{n,ww}\boldsymbol{\phi} + \frac{1}{n}\sum_{i=1}^{n}\mathbf{w}_i\xi_i.$$

Taking expectations of both sides and rearrange terms, we have

$$\boldsymbol{\phi} = \mathrm{E}\left(\mathbf{Q}_{n,ww}\right)^{-1}\mathrm{E}\left(\mathbf{q}_{n,wy}\right).$$

*Ph.D. student, Department of Economics, University of Southern California, 3620 South Vermont Avenue, Los Angeles, CA 90089, USA. Email: zhangao@usc.edu.

†1. Department of Economics, University of Southern California, 3620 South Vermont Avenue, Los Angeles, CA 90089, USA. Email: pesaran@usc.edu. 2. Trinity College, Cambridge, United Kingdom

Consider

$$
\begin{aligned}
\hat{\boldsymbol{\phi}} - \boldsymbol{\phi} &= \mathbf{Q}_{n,ww}^{-1}\mathbf{q}_{n,wy} - \mathrm{E}\left(\mathbf{Q}_{n,ww}\right)^{-1}\mathrm{E}\left(\mathbf{q}_{n,wy}\right) \\
&= \left[\mathbf{Q}_{n,ww}^{-1} - E\left(\mathbf{Q}_{n,ww}\right)^{-1} + E\left(\mathbf{Q}_{n,ww}\right)^{-1}\right]\left[\mathbf{q}_{n,wy} - \mathrm{E}\left(\mathbf{q}_{n,wy}\right) + \mathrm{E}\left(\mathbf{q}_{n,wy}\right)\right] - \mathrm{E}\left(\mathbf{Q}_{n,ww}\right)^{-1}\mathrm{E}\left(\mathbf{q}_{n,wy}\right) \\
&= \left[\mathbf{Q}_{n,ww}^{-1} - E\left(\mathbf{Q}_{n,ww}\right)^{-1}\right]\left[\mathbf{q}_{n,wy} - \mathrm{E}\left(\mathbf{q}_{n,wy}\right)\right] + \left[\mathbf{Q}_{n,ww}^{-1} - \mathrm{E}\left(\mathbf{Q}_{n,ww}\right)^{-1}\right]\mathrm{E}\left(\mathbf{q}_{n,wy}\right) \\
&\qquad + \mathrm{E}\left(\mathbf{Q}_{n,ww}\right)^{-1}\left[\mathbf{q}_{n,wy} - \mathrm{E}\left(\mathbf{q}_{n,wy}\right)\right].
\end{aligned}
$$

Then,

$$
\begin{aligned}
\left\|\hat{\boldsymbol{\phi}} - \boldsymbol{\phi}\right\| \leq& \left\|\mathbf{Q}_{n,ww}^{-1} - \mathrm{E}\left(\mathbf{Q}_{n,ww}\right)^{-1}\right\|\left\|\mathbf{q}_{n,wy} - \mathrm{E}\left(\mathbf{q}_{n,wy}\right)\right\| + \left\|\mathbf{Q}_{n,ww}^{-1} - \mathrm{E}\left(\mathbf{Q}_{n,ww}\right)^{-1}\right\|\left\|\mathrm{E}\left(\mathbf{q}_{n,wy}\right)\right\| \\
&+ \left\|\mathrm{E}\left(\mathbf{Q}_{n,ww}\right)^{-1}\right\|\left\|\mathbf{q}_{n,wy} - \mathrm{E}\left(\mathbf{q}_{n,wy}\right)\right\|.
\end{aligned}
$$

By Assumption 1(c), we have $\left\|\mathbf{Q}_{n,ww}^{-1} - \mathrm{E}\left(\mathbf{Q}_{n,ww}\right)^{-1}\right\| = O_p\left(n^{-1/2}\right)$, $\left\|\mathbf{q}_{n,wy} - \mathrm{E}\left(\mathbf{q}_{n,wy}\right)\right\| = O_p\left(n^{-1/2}\right)$, and by Assumption 1(b), $\left\|\mathrm{E}\left(\mathbf{q}_{n,wy}\right)\right\|$ and $\left\|\mathrm{E}\left(\mathbf{Q}_{n,ww}\right)^{-1}\right\|$ are bounded. Thus,

$$
\left\|\hat{\boldsymbol{\phi}} - \boldsymbol{\phi}\right\| = O_p\left(n^{-1/2}\right). \tag{S.2.1}
$$

To establish the asymptotic distribution of $\hat{\boldsymbol{\phi}}$, we first note that

$$
\sqrt{n}\left(\hat{\boldsymbol{\phi}} - \boldsymbol{\phi}\right) = \mathbf{Q}_{n,ww}^{-1}\left(n^{-1/2}\sum_{i=1}^{n}\mathbf{w}_i\xi_i\right).
$$

By Assumption 3, we have

$$
\mathrm{var}\left(n^{-1/2}\sum_{i=1}^{n}\mathbf{w}_i\xi_i\right) = \frac{1}{n}\sum_{i=1}^{n}\mathrm{var}\left(\mathbf{w}_i\xi_i\right) = \frac{1}{n}\sum_{i=1}^{n}\mathrm{E}\left(\mathbf{w}_i\mathbf{w}_i'\xi_i^2\right) \to \mathbf{V}_{w\xi} \succ 0.
$$

Note that $\xi_i = u_i + x_i v_i$, and $\mathbf{w}_i$ is distributed independently of $u_i$ and $v_i$. Then

$$
\mathbf{w}_i\xi_i = \mathbf{w}_i\left(u_i + x_i v_i\right) = \mathbf{w}_i u_i + \left(\mathbf{w}_i x_i\right)v_i,
$$

and by Minkowski's inequality, for $r = 2 + \delta$ with $0 < \delta < 1$,

$$
\left[E\left\|\mathbf{w}_i\xi_i\right\|^r\right]^{1/r} \leq \left[E\left\|\mathbf{w}_i u_i\right\|^r\right]^{1/r} + \left[E\left\|\left(\mathbf{w}_i x_i\right)v_i\right\|^r\right]^{1/r}.
$$

Due to the independence of $u_i$ and $v_i$ from $\mathbf{w}_i$, we have

$$
\mathrm{E}\left(\left\|\mathbf{w}_i u_i\right\|^r\right) \leq E\left\|\mathbf{w}_i\right\|^r E\left\|u_i\right\|^r, \text{ and } E\left\|\left(\mathbf{w}_i x_i'\right)v_i\right\|^r \leq E\left\|\mathbf{w}_i x_i\right\|^r E\left\|v_i\right\|^r.
$$

Also, $E\left\|\mathbf{w}_i x_i\right\|^r \leq E\left\|\left(x_i^2, x_i\mathbf{z}_i'\right)'\right\|^r \leq E\left\|\mathbf{w}_i\mathbf{w}_i'\right\|^r \leq E\left\|\mathbf{w}_i\right\|^{2r}$, where $2 < r < 3$, and hence

$2r < 6$. By Assumptions 1(a.ii) and 1(b.ii), we have $\sup_i \mathrm{E}\left(\|\mathbf{w}_i\|^6\right) < C$, $\sup_i \mathrm{E}\left(\|u_i\|^3\right) < C$, and $\mathrm{E}\left(\|v_i\|^3\right) \leq \max_{1\leq k\leq K} |b_k - \mathrm{E}(\beta_i)|^3 < C$. Then, we verified that $\sup_i \mathrm{E}(\|\mathbf{w}_i u_i\|^r) < C$, and $E\left\|(\mathbf{w}_i x_i')\, v_i\right\|^r < C$, and hence the Lyapunov condition that $\sup_i \mathrm{E}(\|\mathbf{w}_i \xi_i\|^r) < C$, where $r = 2 + \delta \in (2, 3)$. By the central limit theorem for independent but not necessarily identically distributed random vectors (see Pesaran (2015, Theorem 18) or Hansen (2022, Theorem 6.5)), we have

$$\frac{1}{\sqrt{n}} \sum_{i=1}^{n} \mathbf{w}_i \xi_i \to_d N(\mathbf{0}, \mathbf{V}_{w\xi}),$$

as $n \to \infty$, and by Assumption 1 and continuous mapping theorem,

$$\sqrt{n}(\hat{\boldsymbol{\phi}} - \boldsymbol{\phi}) \to_d N\left(\mathbf{0}, \mathbf{Q}_{ww}^{-1} \mathbf{V}_{w\xi} \mathbf{Q}_{ww}^{-1}\right).$$

We then turn to the consistent estimation of the variance matrix. By Assumption 3, we have

$$\begin{aligned}
\left\|\hat{\mathbf{V}}_{w\xi} - \mathbf{V}_{w\xi}\right\| &= \left\| \frac{1}{n}\sum_{i=1}^{n} \mathbf{w}_i \mathbf{w}_i' \hat{\xi}_i^2 - \frac{1}{n}\sum_{i=1}^{n} \mathrm{E}\left(\mathbf{w}_i \mathbf{w}_i \xi_i^2\right) + \frac{1}{n}\sum_{i=1}^{n} \mathrm{E}\left(\mathbf{w}_i \mathbf{w}_i \xi_i^2\right) - \mathbf{V}_{w\xi} \right\| \\
&\leq \left\| \frac{1}{n}\sum_{i=1}^{n} \mathbf{w}_i \mathbf{w}_i' \xi_i^2 - \frac{1}{n}\sum_{i=1}^{n} \mathrm{E}\left(\mathbf{w}_i \mathbf{w}_i \xi_i^2\right) \right\| + \left\| \frac{1}{n}\sum_{i=1}^{n} \mathrm{E}\left(\mathbf{w}_i \mathbf{w}_i \xi_i^2\right) - \mathbf{V}_{w\xi} \right\| \\
&\quad + \left\| \frac{1}{n}\sum_{i=1}^{n} \mathbf{w}_i \mathbf{w}_i' \left(\hat{\xi}_i^2 - \xi_i^2\right) \right\| \\
&\leq \frac{1}{n}\sum_{i=1}^{n} \|\mathbf{w}_i\|^2 \left|\hat{\xi}_i^2 - \xi_i^2\right| + O_p(n^{-1/2}).
\end{aligned} \tag{S.2.2}$$

Note that $\hat{\xi}_i = \xi_i - \left(\hat{\boldsymbol{\phi}} - \boldsymbol{\phi}\right)' \mathbf{w}_i$, then

$$\begin{aligned}
\left|\hat{\xi}_i^2 - \xi_i^2\right| &\leq 2\left|\xi_i \mathbf{w}_i' \left(\hat{\boldsymbol{\phi}} - \boldsymbol{\phi}\right)\right| + \left(\hat{\boldsymbol{\phi}} - \boldsymbol{\phi}\right)' \left(\mathbf{w}_i \mathbf{w}_i'\right) \left(\hat{\boldsymbol{\phi}} - \boldsymbol{\phi}\right) \\
&\leq 2|\xi_i| \|\mathbf{w}_i\| \left\|\hat{\boldsymbol{\phi}} - \boldsymbol{\phi}\right\| + \|\mathbf{w}_i\|^2 \left\|\hat{\boldsymbol{\phi}} - \boldsymbol{\phi}\right\|^2.
\end{aligned} \tag{S.2.3}$$

Combine (S.2.2) and (S.2.3), we have

$$\left\|\hat{\mathbf{V}}_{w\xi} - \mathbf{V}_{w\xi}\right\| \leq 2\left(\frac{1}{n}\sum_{i=1}^{n} \|\mathbf{w}_i\|^3 |\xi_i|\right) \left\|\hat{\boldsymbol{\phi}} - \boldsymbol{\phi}\right\| + \left(\frac{1}{n}\sum_{i=1}^{n} \|\mathbf{w}_i\|^4\right) \left\|\hat{\boldsymbol{\phi}} - \boldsymbol{\phi}\right\|^2 + O_p\left(n^{-1/2}\right). \tag{S.2.4}$$

We showed that $\left\|\hat{\boldsymbol{\phi}} - \boldsymbol{\phi}\right\| = O_p\left(n^{-1/2}\right)$. By Hölder's inequality,

$$\frac{1}{n}\sum_{i=1}^{n} \|\mathbf{w}_i\|^3 |\xi_i| \leq \left(\frac{1}{n}\sum_{i=1}^{n} \|\mathbf{w}_i\|^4\right)^{3/4} \left(\frac{1}{n}\sum_{i=1}^{n} \xi_i^4\right)^{1/4}. \tag{S.2.5}$$

By Assumption 1(b.iii), $n^{-1}\sum_{i=1}^{n}\|\mathbf{w}_i\|^4 = O_p(1)$. By Minkowski inequality,

$$\begin{aligned}\left(\frac{1}{n}\sum_{i=1}^{n}\xi_i^4\right)^{1/4} &= \left(\frac{1}{n}\sum_{i=1}^{n}(u_i + x_i v_i)^4\right)^{1/4} \leq \left(\frac{1}{n}\sum_{i=1}^{n}u_i^4\right)^{1/4} + \left(\frac{1}{n}\sum_{i=1}^{n}x_i^4 v_i^4\right)^{1/4} \\ &\leq \left(\frac{1}{n}\sum_{i=1}^{n}u_i^4\right)^{1/4} + \max_k\left\{\left|b_k - \mathrm{E}\left(\beta_i\right)\right|\right\}\left(\frac{1}{n}\sum_{i=1}^{n}x_i^4\right)^{1/4} \\ &= O_p(1),\end{aligned}$$

where the last inequality is from Assumptions 1(a.iii) and (b.iii) that $n^{-1}\sum_{i=1}^{n}u_i^4 = O_p\left(1\right)$, and $n^{-1}\sum_{i=1}^{n}x_i^4 \leq n^{-1}\sum_{i=1}^{n}\|\mathbf{w}_i\|^4 = O_p(1)$. Then we verified in (S.2.5) that $n^{-1}\sum_{i=1}^{n}\|\mathbf{w}_i\|^3\left|\xi_i\right| = O_p(1)$. Then using the above results in (S.2.4), and noting from (S.2.1), we have $\left\|\hat{\mathbf{V}}_{w\xi} - \mathbf{V}_{w\xi}\right\| = O_p\left(n^{-1/2}\right)$, as required. ■

## S.3 Monte Carlo Simulation

### S.3.1 Results with $S = 5$ and $S = 6$

Tables S.1 and S.2 present the summary results corresponding to $S = 5$ and $S = 6$, for the data generating processes described in Section 5.1. These results show that adding more moments does not necessarily improve the estimation accuracy but could be counter-productive.

### S.3.2 GMM Estimation of Moments of $\beta_i$

With the data generating processes described in Section 5.1, we report the bias, RMSE and size of the GMM estimator for moments of $\beta_i$ in Table S.3. The GMM estimator for moments of $\beta_i$ achieve better small sample performance as compared to those for the distributional parameters $\pi, \beta_L$ and $\beta_H$.

### S.3.3 Three Estimators of $\mathrm{E}\left(\beta_i\right)$

Table S.4 compares the finite sample performance of three estimators of $\mathrm{E}\left(\beta_i\right)$ with the data generating processes described in Section 5.1.

- The OLS estimator $\hat{\boldsymbol{\phi}}$ studied in Section 3.1.
- The GMM estimator of $\mathrm{E}\left(\beta_i\right)$ with moment conditions given by (3.7).
- $\widehat{\mathrm{E}\left(\beta_i\right)} = \hat{\pi}\hat{\beta}_L + \left(1 - \hat{\pi}\right)\hat{\beta}_H$, where $\hat{\pi}, \hat{\beta}_L, \hat{\beta}_H$ are the GMM estimators of $\pi, \beta_L$, and $\beta_H$.

According to Table S.4, three estimators perform comparably well in terms of bias and RMSE, whereas the OLS estimator, along with the standard error from Theorem 3, controls size well when $n$ is small.

Table S.1: Bias, RMSE and size of the GMM estimator for distributional parameters of $\beta$ with $S = 5$

| | DGP | Baseline | | | Categorical $x$ | | | Categorical $u$ | | |
|---|---|---|---|---|---|---|---|---|---|---|
| | Sample size $n$ | Bias | RMSE | Size | Bias | RMSE | Size | Bias | RMSE | Size |
| | | *high variance*: $\text{var}(\beta_i) = 0.25$ | | | | | | | | |
| $\pi = 0.5$ | 100 | 0.0308 | 0.1869 | 0.1021 | 0.0259 | 0.1986 | 0.1276 | 0.0106 | 0.1944 | 0.1050 |
| | 1,000 | 0.0048 | 0.1235 | 0.1950 | 0.0054 | 0.1334 | 0.2112 | -0.0364 | 0.1638 | 0.2239 |
| | 2,000 | -0.0006 | 0.0875 | 0.1641 | -0.0009 | 0.0962 | 0.1887 | -0.0238 | 0.1172 | 0.2059 |
| | 5,000 | -0.0005 | 0.0484 | 0.1339 | -0.0001 | 0.0591 | 0.1602 | -0.0125 | 0.0740 | 0.1667 |
| | 10,000 | -0.0002 | 0.0334 | 0.1152 | -0.0005 | 0.0373 | 0.1246 | -0.0080 | 0.0519 | 0.1386 |
| | 100,000 | -0.0002 | 0.0096 | 0.0636 | 0.0001 | 0.0116 | 0.0738 | -0.0008 | 0.0174 | 0.0766 |
| $\beta_L = 1$ | 100 | 0.2234 | 0.4541 | 0.3205 | 0.1992 | 0.4777 | 0.2843 | 0.1780 | 0.5090 | 0.2519 |
| | 1,000 | 0.0503 | 0.1609 | 0.3060 | 0.0475 | 0.1812 | 0.2963 | 0.0100 | 0.2024 | 0.2141 |
| | 2,000 | 0.0265 | 0.1148 | 0.2501 | 0.0257 | 0.1262 | 0.2501 | 0.0088 | 0.1337 | 0.1905 |
| | 5,000 | 0.0108 | 0.0606 | 0.1926 | 0.0130 | 0.0702 | 0.2042 | 0.0031 | 0.0803 | 0.1641 |
| | 10,000 | 0.0054 | 0.0409 | 0.1408 | 0.0061 | 0.0456 | 0.1510 | 0.0008 | 0.0527 | 0.1338 |
| | 100,000 | 0.0004 | 0.0114 | 0.0716 | 0.0006 | 0.0134 | 0.0790 | 0.0002 | 0.0184 | 0.0834 |
| $\beta_H = 2$ | 100 | -0.1956 | 0.5486 | 0.2448 | -0.1941 | 0.5638 | 0.2386 | -0.2029 | 0.5801 | 0.2269 |
| | 1,000 | -0.0418 | 0.2080 | 0.3299 | -0.0414 | 0.2300 | 0.3384 | -0.0752 | 0.2583 | 0.3620 |
| | 2,000 | -0.0264 | 0.1379 | 0.2799 | -0.0286 | 0.1554 | 0.2860 | -0.0529 | 0.1789 | 0.3048 |
| | 5,000 | -0.0113 | 0.0696 | 0.2008 | -0.0116 | 0.0883 | 0.2170 | -0.0254 | 0.1038 | 0.2411 |
| | 10,000 | -0.0053 | 0.0432 | 0.1502 | -0.0064 | 0.0520 | 0.1642 | -0.0156 | 0.0690 | 0.2002 |
| | 100,000 | -0.0007 | 0.0113 | 0.0662 | -0.0004 | 0.0135 | 0.0764 | -0.0016 | 0.0209 | 0.0818 |
| | | *low variance*: $\text{var}(\beta_i) = 0.15$ | | | | | | | | |
| $\pi = 0.3$ | 100 | 0.2214 | 0.2820 | 0.1063 | 0.2291 | 0.2942 | 0.1328 | 0.2212 | 0.2876 | 0.1221 |
| | 1,000 | 0.0477 | 0.1746 | 0.2235 | 0.0605 | 0.1928 | 0.2430 | 0.0348 | 0.2039 | 0.2900 |
| | 2,000 | 0.0217 | 0.1198 | 0.2020 | 0.0262 | 0.1331 | 0.2246 | -0.0080 | 0.1608 | 0.2822 |
| | 5,000 | 0.0112 | 0.0709 | 0.1732 | 0.0154 | 0.0828 | 0.1956 | -0.0115 | 0.1072 | 0.2289 |
| | 10,000 | 0.0063 | 0.0465 | 0.1588 | 0.0106 | 0.0576 | 0.1649 | -0.0075 | 0.0761 | 0.1890 |
| | 100,000 | 0.0001 | 0.0130 | 0.0810 | 0.0014 | 0.0158 | 0.0882 | 0.0040 | 0.0280 | 0.0978 |
| $\beta_L = 0.5$ | 100 | 0.4245 | 0.5722 | 0.2938 | 0.4048 | 0.5818 | 0.2612 | 0.3827 | 0.6052 | 0.2278 |
| | 1,000 | 0.1300 | 0.2692 | 0.3058 | 0.1300 | 0.2890 | 0.3057 | 0.0882 | 0.3673 | 0.1970 |
| | 2,000 | 0.0763 | 0.1746 | 0.3147 | 0.0735 | 0.1903 | 0.2820 | 0.0149 | 0.2523 | 0.1964 |
| | 5,000 | 0.0378 | 0.1018 | 0.2690 | 0.0410 | 0.1155 | 0.2695 | 0.0034 | 0.1417 | 0.1905 |
| | 10,000 | 0.0202 | 0.0674 | 0.2344 | 0.0257 | 0.0822 | 0.2404 | 0.0013 | 0.0961 | 0.1690 |
| | 100,000 | 0.0013 | 0.0184 | 0.0952 | 0.0026 | 0.0221 | 0.1042 | 0.0060 | 0.0347 | 0.1112 |
| $\beta_H = 1.345$ | 100 | -0.0646 | 0.3773 | 0.1781 | -0.0616 | 0.4058 | 0.1668 | -0.0564 | 0.4357 | 0.1688 |
| | 1,000 | -0.0180 | 0.1523 | 0.2496 | -0.0119 | 0.1804 | 0.2615 | -0.0476 | 0.2022 | 0.2721 |
| | 2,000 | -0.0104 | 0.1021 | 0.2375 | -0.0101 | 0.1147 | 0.2414 | -0.0381 | 0.1448 | 0.2830 |
| | 5,000 | -0.0027 | 0.0549 | 0.1680 | -0.0016 | 0.0680 | 0.1936 | -0.0193 | 0.0927 | 0.2369 |
| | 10,000 | -0.0001 | 0.0368 | 0.1458 | 0.0007 | 0.0438 | 0.1458 | -0.0115 | 0.0634 | 0.1976 |
| | 100,000 | -0.0002 | 0.0102 | 0.0726 | 0.0005 | 0.0120 | 0.0688 | 0.0021 | 0.0214 | 0.0902 |

*Notes:* The data generating process is (5.1). *high variance* and *low variance* parametrization are described in (5.2). "Baseline", "Categorical $x$" and "Categorical $u$" refer to DGP 1 to 3 as in Section 5.1. Generically, bias, RMSE and size are calculated by $R^{-1}\sum_{r=1}^{R}\left(\hat{\theta}^{(r)} - \theta_0\right)$, $\sqrt{R^{-1}\sum_{r=1}^{R}\left(\hat{\theta}^{(r)} - \theta_0\right)^2}$, and $R^{-1}\sum_{r=1}^{R}\mathbf{1}\left[\left|\hat{\theta}^{(r)} - \theta_0\right| / \hat{\sigma}_{\hat{\theta}}^{(r)} > \text{cv}_{0.05}\right]$, respectively, for true parameter $\theta_0$, its estimate $\hat{\theta}^{(r)}$, the estimated standard error of $\hat{\theta}^{(r)}$, $\hat{\sigma}_{\hat{\theta}}^{(r)}$, and the critical value $\text{cv}_{0.05} = \Phi^{-1}(0.975)$ across $R = 5,000$ replications, where $\Phi(\cdot)$ is the cumulative distribution function of standard normal distribution.

Table S.2: Bias, RMSE and size of the GMM estimator for distributional parameters of $\beta$ with $S = 6$

| | DGP | Baseline | | | Categorical $x$ | | | Categorical $u$ | | |
|---|---|---|---|---|---|---|---|---|---|---|
| | Sample size $n$ | Bias | RMSE | Size | Bias | RMSE | Size | Bias | RMSE | Size |
| | | *high variance*: $\text{var}(\beta_i) = 0.25$ | | | | | | | | |
| $\pi = 0.5$ | 100 | 0.0337 | 0.1472 | 0.0456 | 0.0293 | 0.1645 | 0.0695 | 0.0227 | 0.1498 | 0.0469 |
| | 1,000 | 0.0021 | 0.1405 | 0.2545 | 0.0015 | 0.1469 | 0.2543 | -0.0265 | 0.1635 | 0.2551 |
| | 2,000 | 0.0008 | 0.1071 | 0.2614 | 0.0006 | 0.1185 | 0.2789 | -0.0201 | 0.1281 | 0.2732 |
| | 5,000 | -0.0020 | 0.0661 | 0.2261 | -0.0016 | 0.0765 | 0.2518 | -0.0142 | 0.0836 | 0.2510 |
| | 10,000 | -0.0005 | 0.0444 | 0.1844 | -0.0011 | 0.0505 | 0.2155 | -0.0093 | 0.0587 | 0.2323 |
| | 100,000 | 0.0000 | 0.0097 | 0.0732 | 0.0000 | 0.0118 | 0.0912 | -0.0020 | 0.0178 | 0.1162 |
| $\beta_L = 1$ | 100 | 0.2226 | 0.4373 | 0.3341 | 0.2151 | 0.4658 | 0.3237 | 0.1879 | 0.4841 | 0.2896 |
| | 1,000 | 0.0721 | 0.2081 | 0.4485 | 0.0780 | 0.2197 | 0.4318 | 0.0531 | 0.2283 | 0.3576 |
| | 2,000 | 0.0443 | 0.1464 | 0.4056 | 0.0455 | 0.1609 | 0.4157 | 0.0342 | 0.1536 | 0.3271 |
| | 5,000 | 0.0175 | 0.0806 | 0.3035 | 0.0203 | 0.0923 | 0.3341 | 0.0150 | 0.0933 | 0.2770 |
| | 10,000 | 0.0092 | 0.0510 | 0.2350 | 0.0098 | 0.0594 | 0.2723 | 0.0081 | 0.0629 | 0.2403 |
| | 100,000 | 0.0010 | 0.0114 | 0.0850 | 0.0013 | 0.0136 | 0.0982 | 0.0002 | 0.0186 | 0.1116 |
| $\beta_H = 2$ | 100 | -0.2495 | 0.5629 | 0.2563 | -0.2580 | 0.5681 | 0.2608 | -0.2589 | 0.5782 | 0.2248 |
| | 1,000 | -0.0618 | 0.2530 | 0.4938 | -0.0686 | 0.2733 | 0.4867 | -0.0962 | 0.2814 | 0.4874 |
| | 2,000 | -0.0334 | 0.1729 | 0.4454 | -0.0365 | 0.1951 | 0.4461 | -0.0625 | 0.2017 | 0.4643 |
| | 5,000 | -0.0189 | 0.1010 | 0.3457 | -0.0203 | 0.1178 | 0.3638 | -0.0383 | 0.1223 | 0.3946 |
| | 10,000 | -0.0080 | 0.0634 | 0.2670 | -0.0109 | 0.0732 | 0.3011 | -0.0246 | 0.0830 | 0.3347 |
| | 100,000 | -0.0013 | 0.0114 | 0.0842 | -0.0012 | 0.0141 | 0.1070 | -0.0043 | 0.0220 | 0.1396 |
| | | *low variance*: $\text{var}(\beta_i) = 0.15$ | | | | | | | | |
| $\pi = 0.3$ | 100 | 0.2374 | 0.2757 | 0.0591 | 0.2352 | 0.2816 | 0.0829 | 0.2330 | 0.2771 | 0.0801 |
| | 1,000 | 0.1071 | 0.2107 | 0.2608 | 0.1114 | 0.2244 | 0.2775 | 0.0764 | 0.2158 | 0.2772 |
| | 2,000 | 0.0702 | 0.1661 | 0.2994 | 0.0786 | 0.1815 | 0.3258 | 0.0242 | 0.1806 | 0.3291 |
| | 5,000 | 0.0452 | 0.1101 | 0.3217 | 0.0519 | 0.1260 | 0.3466 | 0.0092 | 0.1263 | 0.3329 |
| | 10,000 | 0.0300 | 0.0816 | 0.3060 | 0.0390 | 0.0933 | 0.3389 | 0.0108 | 0.0954 | 0.3161 |
| | 100,000 | 0.0018 | 0.0164 | 0.1128 | 0.0041 | 0.0234 | 0.1482 | 0.0055 | 0.0298 | 0.1688 |
| $\beta_L = 0.5$ | 100 | 0.4146 | 0.5479 | 0.3137 | 0.4191 | 0.5636 | 0.2965 | 0.3844 | 0.5678 | 0.2532 |
| | 1,000 | 0.2445 | 0.3459 | 0.4601 | 0.2436 | 0.3579 | 0.4561 | 0.2080 | 0.3872 | 0.3187 |
| | 2,000 | 0.1663 | 0.2539 | 0.4809 | 0.1684 | 0.2620 | 0.4797 | 0.1108 | 0.2830 | 0.3203 |
| | 5,000 | 0.0977 | 0.1648 | 0.4800 | 0.1051 | 0.1788 | 0.4938 | 0.0590 | 0.1731 | 0.3606 |
| | 10,000 | 0.0613 | 0.1182 | 0.4230 | 0.0730 | 0.1315 | 0.4717 | 0.0417 | 0.1251 | 0.3667 |
| | 100,000 | 0.0050 | 0.0242 | 0.1420 | 0.0086 | 0.0333 | 0.1808 | 0.0101 | 0.0386 | 0.1906 |
| $\beta_H = 1.345$ | 100 | -0.0817 | 0.3703 | 0.1601 | -0.0883 | 0.3842 | 0.1687 | -0.0806 | 0.4136 | 0.1614 |
| | 1,000 | -0.0086 | 0.1726 | 0.3174 | -0.0144 | 0.1907 | 0.3295 | -0.0560 | 0.2029 | 0.3239 |
| | 2,000 | 0.0022 | 0.1194 | 0.3267 | 0.0029 | 0.1368 | 0.3401 | -0.0395 | 0.1582 | 0.3736 |
| | 5,000 | 0.0093 | 0.0722 | 0.2899 | 0.0099 | 0.0876 | 0.3254 | -0.0189 | 0.0998 | 0.3570 |
| | 10,000 | 0.0092 | 0.0535 | 0.2642 | 0.0117 | 0.0601 | 0.2889 | -0.0076 | 0.0733 | 0.3141 |
| | 100,000 | -0.0002 | 0.0116 | 0.0972 | 0.0012 | 0.0157 | 0.1326 | 0.0019 | 0.0220 | 0.1454 |

*Notes:* The data generating process is (5.1). *high variance* and *low variance* parametrization are described in (5.2). "Baseline", "Categorical $x$" and "Categorical $u$" refer to DGP 1 to 3 as in Section 5.1. Generically, bias, RMSE and size are calculated by $R^{-1}\sum_{r=1}^{R}\left(\hat{\theta}^{(r)} - \theta_0\right)$, $\sqrt{R^{-1}\sum_{r=1}^{R}\left(\hat{\theta}^{(r)} - \theta_0\right)^2}$, and $R^{-1}\sum_{r=1}^{R}\mathbf{1}\left[\left|\hat{\theta}^{(r)} - \theta_0\right| / \hat{\sigma}_{\hat{\theta}}^{(r)} > \text{cv}_{0.05}\right]$, respectively, for true parameter $\theta_0$, its estimate $\hat{\theta}^{(r)}$, the estimated standard error of $\hat{\theta}^{(r)}$, $\hat{\sigma}_{\hat{\theta}}^{(r)}$, and the critical value $\text{cv}_{0.05} = \Phi^{-1}(0.975)$ across $R = 5,000$ replications, where $\Phi(\cdot)$ is the cumulative distribution function of standard normal distribution.

Table S.3: Bias, RMSE and size of the GMM estimator for moments of $\beta$

| DGP | | Baseline | | | Categorical $x$ | | | Categorical $u$ | | |
|---|---|---|---|---|---|---|---|---|---|---|
| | Sample size $n$ | Bias | RMSE | Size | Bias | RMSE | Size | Bias | RMSE | Size |
| *high variance*: $\text{var}(\beta_i) = 0.25$ | | | | | | | | | | |
| $\text{E}(\beta_i) = 1.5$ | 100 | -0.0080 | 0.2262 | 0.1922 | -0.0117 | 0.2297 | 0.1940 | -0.0030 | 0.2418 | 0.1800 |
| | 1,000 | -0.0029 | 0.0663 | 0.0936 | -0.0015 | 0.0673 | 0.0848 | -0.0037 | 0.0725 | 0.0804 |
| | 2,000 | -0.0012 | 0.0431 | 0.0688 | -0.0015 | 0.0463 | 0.0700 | -0.0021 | 0.0494 | 0.0656 |
| | 5,000 | -0.0003 | 0.0263 | 0.0566 | -0.0009 | 0.0276 | 0.0588 | -0.0013 | 0.0303 | 0.0622 |
| | 10,000 | 0.0004 | 0.0183 | 0.0530 | -0.0001 | 0.0186 | 0.0498 | -0.0003 | 0.0206 | 0.0492 |
| | 100,000 | 0.0000 | 0.0056 | 0.0434 | 0.0000 | 0.0058 | 0.0472 | 0.0000 | 0.0066 | 0.0514 |
| $\text{E}(\beta_i^2) = 2.5$ | 100 | -0.0627 | 0.9082 | 0.3464 | -0.0826 | 0.8821 | 0.3166 | -0.0629 | 0.9459 | 0.3122 |
| | 1,000 | -0.0300 | 0.2909 | 0.1518 | -0.0275 | 0.2837 | 0.1382 | -0.0362 | 0.3112 | 0.1512 |
| | 2,000 | -0.0160 | 0.1751 | 0.0976 | -0.0188 | 0.1868 | 0.1074 | -0.0255 | 0.1900 | 0.1048 |
| | 5,000 | -0.0067 | 0.0916 | 0.0658 | -0.0090 | 0.0993 | 0.0710 | -0.0124 | 0.1091 | 0.0754 |
| | 10,000 | -0.0015 | 0.0580 | 0.0506 | -0.0036 | 0.0609 | 0.0530 | -0.0061 | 0.0704 | 0.0566 |
| | 100,000 | -0.0005 | 0.0179 | 0.0462 | -0.0005 | 0.0185 | 0.0498 | -0.0011 | 0.0219 | 0.0542 |
| $\text{E}(\beta_i^3) = 4.5$ | 100 | -0.2511 | 2.3755 | 0.3698 | -0.2990 | 2.3416 | 0.3424 | -0.2940 | 2.6179 | 0.3522 |
| | 1,000 | -0.1155 | 0.7641 | 0.1734 | -0.1092 | 0.7613 | 0.1606 | -0.1478 | 0.8856 | 0.1904 |
| | 2,000 | -0.0667 | 0.4683 | 0.1166 | -0.0745 | 0.5058 | 0.1234 | -0.1066 | 0.5485 | 0.1378 |
| | 5,000 | -0.0290 | 0.2475 | 0.0800 | -0.0365 | 0.2696 | 0.0788 | -0.0507 | 0.3178 | 0.0942 |
| | 10,000 | -0.0099 | 0.1559 | 0.0516 | -0.0163 | 0.1699 | 0.0602 | -0.0282 | 0.2088 | 0.0660 |
| | 100,000 | -0.0020 | 0.0488 | 0.0462 | -0.0023 | 0.0515 | 0.0526 | -0.0052 | 0.0653 | 0.0520 |
| *low variance*: $\text{var}(\beta_i) = 0.15$ | | | | | | | | | | |
| $\text{E}(\beta_i) = 1.0915$ | 100 | 0.0165 | 0.1943 | 0.1618 | 0.0089 | 0.1983 | 0.1514 | 0.0169 | 0.2112 | 0.1416 |
| | 1,000 | 0.0045 | 0.0577 | 0.0800 | 0.0042 | 0.0584 | 0.0702 | 0.0033 | 0.0655 | 0.0734 |
| | 2,000 | 0.0019 | 0.0384 | 0.0594 | 0.0016 | 0.0410 | 0.0698 | 0.0010 | 0.0452 | 0.0632 |
| | 5,000 | 0.0008 | 0.0243 | 0.0562 | 0.0003 | 0.0250 | 0.0540 | -0.0003 | 0.0283 | 0.0574 |
| | 10,000 | 0.0007 | 0.0171 | 0.0502 | 0.0001 | 0.0175 | 0.0476 | 0.0000 | 0.0194 | 0.0442 |
| | 100,000 | 0.0000 | 0.0052 | 0.0430 | 0.0000 | 0.0054 | 0.0476 | 0.0000 | 0.0062 | 0.0472 |
| $\text{E}(\beta_i^2) = 1.3413$ | 100 | -0.0121 | 0.5119 | 0.2440 | -0.0280 | 0.5095 | 0.2330 | -0.0236 | 0.5724 | 0.2340 |
| | 1,000 | -0.0061 | 0.1528 | 0.1232 | -0.0084 | 0.1566 | 0.1126 | -0.0163 | 0.1776 | 0.1246 |
| | 2,000 | -0.0072 | 0.0973 | 0.0836 | -0.0080 | 0.1053 | 0.0922 | -0.0143 | 0.1154 | 0.0964 |
| | 5,000 | -0.0037 | 0.0565 | 0.0658 | -0.0044 | 0.0603 | 0.0698 | -0.0088 | 0.0699 | 0.0720 |
| | 10,000 | -0.0018 | 0.0381 | 0.0582 | -0.0027 | 0.0401 | 0.0590 | -0.0054 | 0.0476 | 0.0618 |
| | 100,000 | -0.0004 | 0.0119 | 0.0496 | -0.0005 | 0.0125 | 0.0538 | -0.0009 | 0.0152 | 0.0506 |
| $\text{E}(\beta_i^3) = 1.7407$ | 100 | -0.0759 | 0.9761 | 0.2806 | -0.0995 | 1.0052 | 0.2672 | -0.1277 | 1.2814 | 0.2718 |
| | 1,000 | -0.0364 | 0.2925 | 0.1486 | -0.0396 | 0.3112 | 0.1456 | -0.0687 | 0.3973 | 0.1720 |
| | 2,000 | -0.0297 | 0.1927 | 0.1040 | -0.0310 | 0.2126 | 0.1178 | -0.0526 | 0.2650 | 0.1324 |
| | 5,000 | -0.0148 | 0.1141 | 0.0798 | -0.0168 | 0.1252 | 0.0860 | -0.0301 | 0.1619 | 0.0964 |
| | 10,000 | -0.0078 | 0.0771 | 0.0654 | -0.0097 | 0.0846 | 0.0722 | -0.0188 | 0.1126 | 0.0828 |
| | 100,000 | -0.0013 | 0.0242 | 0.0478 | -0.0016 | 0.0262 | 0.0554 | -0.0031 | 0.0360 | 0.0566 |

*Notes:* The data generating process is (5.1). $S = 4$ is used. *high variance* and *low variance* parametrization are described in (5.2). "Baseline", "Categorical $x$" and "Categorical $u$" refer to DGP 1 to 3 as in Section 5.1. Generically, bias, RMSE and size are calculated by $R^{-1}\sum_{r=1}^{R}\left(\hat{\theta}^{(r)} - \theta_0\right)$, $\sqrt{R^{-1}\sum_{r=1}^{R}\left(\hat{\theta}^{(r)} - \theta_0\right)^2}$, and $R^{-1}\sum_{r=1}^{R}\mathbf{1}\left[\left|\hat{\theta}^{(r)} - \theta_0\right| / \hat{\sigma}_{\hat{\theta}}^{(r)} > \text{cv}_{0.05}\right]$, respectively, for true parameter $\theta_0$, its estimate $\hat{\theta}^{(r)}$, the estimated standard error of $\hat{\theta}^{(r)}$, $\hat{\sigma}_{\hat{\theta}}^{(r)}$, and the critical value $\text{cv}_{0.05} = \Phi^{-1}(0.975)$ across $R = 5,000$ replications, where $\Phi(\cdot)$ is the cumulative distribution function of standard normal distribution.

Table S.4: Bias, RMSE and size of three estimators for $\mathrm{E}(\beta_i)$

| | DGP | Baseline | | | Categorical $x$ | | | Categorical $u$ | | |
|---|---|---|---|---|---|---|---|---|---|---|
| | Sample size $n$ | Bias | RMSE | Size | Bias | RMSE | Size | Bias | RMSE | Size |
| | | | | | *high variance*: $\mathrm{E}(\beta_i)=1.5$, $\mathrm{var}(\beta_i)=0.25$ | | | | | |
| OLS | 100 | -0.0024 | 0.2035 | 0.0966 | -0.0037 | 0.2035 | 0.0858 | -0.0042 | 0.2268 | 0.0920 |
| | 1,000 | -0.0017 | 0.0669 | 0.0568 | -0.0002 | 0.0657 | 0.0540 | -0.0019 | 0.0738 | 0.0540 |
| | 2,000 | -0.0008 | 0.0463 | 0.0512 | -0.0015 | 0.0475 | 0.0534 | -0.0010 | 0.0523 | 0.0522 |
| | 5,000 | -0.0004 | 0.0301 | 0.0540 | -0.0008 | 0.0300 | 0.0546 | -0.0007 | 0.0335 | 0.0560 |
| | 10,000 | 0.0002 | 0.0214 | 0.0508 | 0.0000 | 0.0212 | 0.0510 | 0.0000 | 0.0229 | 0.0456 |
| | 100,000 | -0.0001 | 0.0066 | 0.0472 | 0.0000 | 0.0066 | 0.0460 | 0.0000 | 0.0075 | 0.0506 |
| GMM | 100 | -0.0080 | 0.2262 | 0.1922 | -0.0117 | 0.2297 | 0.1940 | -0.0030 | 0.2418 | 0.1800 |
| | 1,000 | -0.0029 | 0.0663 | 0.0936 | -0.0015 | 0.0673 | 0.0848 | -0.0037 | 0.0725 | 0.0804 |
| | 2,000 | -0.0012 | 0.0431 | 0.0688 | -0.0015 | 0.0463 | 0.0700 | -0.0021 | 0.0494 | 0.0656 |
| | 5,000 | -0.0003 | 0.0263 | 0.0566 | -0.0009 | 0.0276 | 0.0588 | -0.0013 | 0.0303 | 0.0622 |
| | 10,000 | 0.0004 | 0.0183 | 0.0530 | -0.0001 | 0.0186 | 0.0498 | -0.0003 | 0.0206 | 0.0492 |
| | 100,000 | 0.0000 | 0.0056 | 0.0434 | 0.0000 | 0.0058 | 0.0472 | 0.0000 | 0.0066 | 0.0514 |
| $\hat{\pi}\hat{\beta}_L+(1-\hat{\pi})\hat{\beta}_H$ | 100 | -0.0087 | 0.2922 | 0.1961 | -0.1232 | 0.2347 | 0.1809 | -0.0037 | 0.2947 | 0.1894 |
| | 1,000 | -0.0012 | 0.0648 | 0.0709 | -0.0237 | 0.0783 | 0.0665 | -0.0023 | 0.0713 | 0.0652 |
| | 2,000 | -0.0004 | 0.0410 | 0.0556 | -0.0140 | 0.0537 | 0.0597 | -0.0015 | 0.0479 | 0.0558 |
| | 5,000 | 0.0000 | 0.0259 | 0.0536 | -0.0063 | 0.0296 | 0.0546 | -0.0011 | 0.0299 | 0.0590 |
| | 10,000 | 0.0004 | 0.0183 | 0.0526 | -0.0035 | 0.0205 | 0.0496 | -0.0003 | 0.0205 | 0.0488 |
| | 100,000 | 0.0000 | 0.0056 | 0.0436 | -0.0006 | 0.0062 | 0.0472 | 0.0000 | 0.0066 | 0.0514 |
| | | | | | *low variance*: $\mathrm{E}(\beta_i)=1.0915$, $\mathrm{var}(\beta_i)=0.15$ | | | | | |
| OLS | 100 | -0.0006 | 0.1829 | 0.0810 | -0.0023 | 0.1855 | 0.0766 | -0.0025 | 0.2094 | 0.0828 |
| | 1,000 | -0.0005 | 0.0597 | 0.0610 | 0.0005 | 0.0590 | 0.0478 | -0.0006 | 0.0670 | 0.0542 |
| | 2,000 | -0.0002 | 0.0408 | 0.0516 | -0.0007 | 0.0427 | 0.0606 | -0.0004 | 0.0475 | 0.0544 |
| | 5,000 | -0.0002 | 0.0264 | 0.0530 | -0.0006 | 0.0266 | 0.0480 | -0.0005 | 0.0302 | 0.0538 |
| | 10,000 | 0.0000 | 0.0189 | 0.0546 | -0.0002 | 0.0188 | 0.0486 | -0.0002 | 0.0208 | 0.0482 |
| | 100,000 | -0.0001 | 0.0059 | 0.0474 | 0.0000 | 0.0059 | 0.0494 | 0.0000 | 0.0068 | 0.0508 |
| GMM | 100 | -0.0121 | 0.5119 | 0.2440 | -0.0280 | 0.5095 | 0.2330 | -0.0236 | 0.5724 | 0.2340 |
| | 1,000 | -0.0061 | 0.1528 | 0.1232 | -0.0084 | 0.1566 | 0.1126 | -0.0163 | 0.1776 | 0.1246 |
| | 2,000 | -0.0072 | 0.0973 | 0.0836 | -0.0080 | 0.1053 | 0.0922 | -0.0143 | 0.1154 | 0.0964 |
| | 5,000 | -0.0037 | 0.0565 | 0.0658 | -0.0044 | 0.0603 | 0.0698 | -0.0088 | 0.0699 | 0.0720 |
| | 10,000 | -0.0018 | 0.0381 | 0.0582 | -0.0027 | 0.0401 | 0.0590 | -0.0054 | 0.0476 | 0.0618 |
| | 100,000 | -0.0004 | 0.0119 | 0.0496 | -0.0005 | 0.0125 | 0.0538 | -0.0009 | 0.0152 | 0.0506 |
| $\hat{\pi}\hat{\beta}_L+(1-\hat{\pi})\hat{\beta}_H$ | 100 | 0.0166 | 0.2392 | 0.1496 | 0.0063 | 0.2342 | 0.1412 | 0.0182 | 0.2432 | 0.1586 |
| | 1,000 | 0.0078 | 0.0621 | 0.0827 | 0.0068 | 0.0615 | 0.0677 | 0.0064 | 0.0674 | 0.0693 |
| | 2,000 | 0.0024 | 0.0388 | 0.0559 | 0.0021 | 0.0414 | 0.0672 | 0.0019 | 0.0454 | 0.0627 |
| | 5,000 | 0.0009 | 0.0241 | 0.0554 | 0.0003 | 0.0247 | 0.0524 | 0.0001 | 0.0282 | 0.0548 |
| | 10,000 | 0.0007 | 0.0170 | 0.0502 | 0.0002 | 0.0174 | 0.0478 | 0.0003 | 0.0193 | 0.0438 |
| | 100,000 | 0.0000 | 0.0052 | 0.0430 | 0.0000 | 0.0054 | 0.0480 | 0.0004 | 0.0063 | 0.0494 |

*Notes:* The data generating process is (5.1). *high variance* and *low variance* parametrization are described in (5.2). "Baseline", "Categorical $x$" and "Categorical $u$" refer to DGP 1 to 3 as in Section 5.1. Generically, bias, RMSE and size are calculated by $R^{-1}\sum_{r=1}^{R}\left(\hat{\theta}^{(r)}-\theta_0\right)$, $\sqrt{R^{-1}\sum_{r=1}^{R}\left(\hat{\theta}^{(r)}-\theta_0\right)^2}$, and $R^{-1}\sum_{r=1}^{R}\mathbf{1}\left[\left|\hat{\theta}^{(r)}-\theta_0\right|/\hat{\sigma}_{\hat{\theta}}^{(r)}>\mathrm{cv}_{0.05}\right]$, respectively, for true parameter $\theta_0$, its estimate $\hat{\theta}^{(r)}$, the estimated standard error of $\hat{\theta}^{(r)}$, $\hat{\sigma}_{\hat{\theta}}^{(r)}$, and the critical value $\mathrm{cv}_{0.05}=\Phi^{-1}(0.975)$ across $R=5,000$ replications, where $\Phi(\cdot)$ is the cumulative distribution function of standard normal distribution.

### S.3.4 Experiments with higher $\text{var}(\beta_i)$

Following the data generating processes in Section 5.1, we increase the variance of $\beta_i$ by considering the following two parametrizations:

$$(\pi, \beta_L, \beta_H, \text{E}(\beta_i), \text{var}(\beta_i)) = \begin{cases} (0.3, 0.5, 6, 4.35, 6.3525), \\ (0.3, 0.5, 10, 7.15, 18.9525). \end{cases} \tag{S.3.1}$$

Table S.5 presents the results, which show that using larger values of $\text{var}(\beta_i)$ improves the small sample performance of the GMM estimators.

### S.3.5 Experiments with three categories ($K = 3$)

#### S.3.5.1 Data generating processes

We generate $y_i$ as

$$y_i = \alpha + x_i\beta_i + z_{i1}\gamma_1 + z_{i2}\gamma_2 + u_i, \text{ for } i = 1, 2, ..., n, \tag{S.3.2}$$

with $\beta_i$ distributed as in (2.2) with $K = 3$,

$$\beta_i = \begin{cases} \beta_L, & \text{w.p. } \pi_L \\ \beta_M, & \text{w.p. } \pi_M \\ \beta_L, & \text{w.p. } 1 - \pi_L - \pi_M, \end{cases}$$

where w.p. denotes "with probability". The parameters take values $(\pi_L, \pi_M, \beta_L, \beta_M, \beta_H) = (0.3, 0.3, 1, 2, 3)$. Corresponding, the moments of $\beta_i$ are $\left(\text{E}(\beta_i), \text{E}(\beta_i^2), \text{E}(\beta_i^3), \text{E}(\beta_i^4), \text{E}(\beta_i^5)\right) = (2.1, 5.1, 13.5, 37.5, 107.1)$. The remaining parameters are set as $\alpha = 0.25$, and $\boldsymbol{\gamma} = (1, 1)'$.

We first generate $\tilde{x}_i \sim \text{IID}\chi^2(2)$, and then set $x_i = (\tilde{x}_i - 2)/2$ so that $x_i$ has 0 mean and unit variance. The additional regressors, $z_{ij}$, for $j = 1, 2$ with homogeneous slopes are generated as

$$z_{i1} = x_i + v_{i1} \text{ and } z_{i2} = z_{i1} + v_{i2},$$

with $v_{ij} \sim \text{IID } N(0, 1)$, for $j = 1, 2$. The error term, $u_i$, is generated as $u_i = \sigma_i\varepsilon_i$, where $\sigma_i^2$ are generated as $0.5(1 + \text{IID}\chi^2(1))$, and $\varepsilon_i \sim \text{IID}N(0, 1)$.

#### S.3.5.2 Results

Table S.6 reports the bias, RMSE and size of the GMM estimator for distributional parameters and moments of $\beta_i$. The results are based on 5,000 replications and $S = 6$. The results show that even larger sample sizes are needed for the GMM estimators (both the moments of $\beta_i$ and its distributional parameters) to achieve reasonable finite sample performance, since higher order of moments are involved.

In additional to the results of jointly estimating distributional parameters and moments of $\beta_i$

Table S.5: Bias, RMSE and size of the GMM estimator for distributional parameters of $\beta$

| | DGP | Baseline | | | Categorical $x$ | | | Categorical $u$ | | |
|---|---|---|---|---|---|---|---|---|---|---|
| | Sample size $n$ | Bias | RMSE | Size | Bias | RMSE | Size | Bias | RMSE | Size |
| | $\text{var}(\beta_i) = 6.35$ | | | | | | | | | |
| $\pi = 0.3$ | 100 | 0.0755 | 0.3014 | 0.1885 | 0.0628 | 0.2829 | 0.1601 | 0.0760 | 0.2967 | 0.1795 |
| | 1,000 | -0.0113 | 0.1058 | 0.1485 | -0.0002 | 0.0882 | 0.1406 | -0.0092 | 0.1043 | 0.1509 |
| | 2,000 | -0.0103 | 0.0646 | 0.1025 | -0.0016 | 0.0495 | 0.1072 | -0.0077 | 0.0598 | 0.1104 |
| | 5,000 | -0.0026 | 0.0276 | 0.0718 | -0.0009 | 0.0197 | 0.0726 | -0.0021 | 0.0245 | 0.0742 |
| | 10,000 | -0.0008 | 0.0095 | 0.0576 | -0.0005 | 0.0093 | 0.0608 | -0.0010 | 0.0099 | 0.0588 |
| | 100,000 | -0.0002 | 0.0027 | 0.0490 | -0.0001 | 0.0026 | 0.0518 | -0.0002 | 0.0028 | 0.0504 |
| $\beta_L = 0.5$ | 100 | 2.7277 | 3.5109 | 0.2385 | 2.3640 | 3.2861 | 0.2207 | 2.6810 | 3.4783 | 0.2292 |
| | 1,000 | 0.2951 | 1.1688 | 0.2743 | 0.1539 | 0.9017 | 0.2521 | 0.2473 | 1.1016 | 0.2725 |
| | 2,000 | 0.0933 | 0.6394 | 0.1916 | 0.0460 | 0.5158 | 0.1988 | 0.0698 | 0.5904 | 0.1951 |
| | 5,000 | 0.0159 | 0.2570 | 0.1236 | -0.0005 | 0.1786 | 0.1306 | 0.0066 | 0.2080 | 0.1225 |
| | 10,000 | 0.0009 | 0.0607 | 0.0884 | -0.0005 | 0.0504 | 0.0998 | -0.0014 | 0.0585 | 0.0830 |
| | 100,000 | 0.0000 | 0.0130 | 0.0572 | 0.0005 | 0.0135 | 0.0630 | -0.0003 | 0.0148 | 0.0622 |
| $\beta_H = 6$ | 100 | 0.1286 | 1.1700 | 0.0978 | 0.0482 | 1.1467 | 0.1057 | 0.1395 | 1.3662 | 0.0970 |
| | 1,000 | 0.0031 | 0.2840 | 0.1320 | 0.0062 | 0.2695 | 0.1200 | 0.0043 | 0.3197 | 0.1382 |
| | 2,000 | -0.0108 | 0.1392 | 0.0982 | 0.0007 | 0.1552 | 0.1094 | -0.0108 | 0.1519 | 0.1088 |
| | 5,000 | -0.0041 | 0.0621 | 0.0746 | -0.0024 | 0.0608 | 0.0736 | -0.0054 | 0.0652 | 0.0794 |
| | 10,000 | -0.0018 | 0.0340 | 0.0550 | -0.0012 | 0.0347 | 0.0678 | -0.0034 | 0.0386 | 0.0642 |
| | 100,000 | -0.0003 | 0.0109 | 0.0530 | 0.0001 | 0.0107 | 0.0518 | -0.0006 | 0.0125 | 0.0588 |
| | $\text{var}(\beta_i) = 18.95$ | | | | | | | | | |
| $\pi = 0.3$ | 100 | 0.0575 | 0.2896 | 0.1761 | 0.0530 | 0.2762 | 0.1524 | 0.0554 | 0.2889 | 0.1646 |
| | 1,000 | -0.0136 | 0.1070 | 0.1217 | -0.0025 | 0.0892 | 0.1306 | -0.0110 | 0.1024 | 0.1369 |
| | 2,000 | -0.0101 | 0.0650 | 0.0850 | -0.0032 | 0.0488 | 0.0969 | -0.0077 | 0.0610 | 0.0957 |
| | 5,000 | -0.0027 | 0.0291 | 0.0668 | -0.0010 | 0.0217 | 0.0625 | -0.0023 | 0.0247 | 0.0713 |
| | 10,000 | -0.0009 | 0.0122 | 0.0549 | -0.0005 | 0.0097 | 0.0600 | -0.0009 | 0.0100 | 0.0570 |
| | 100,000 | -0.0002 | 0.0025 | 0.0480 | -0.0001 | 0.0024 | 0.0514 | -0.0002 | 0.0025 | 0.0484 |
| $\beta_L = 0.5$ | 100 | 4.5691 | 5.9597 | 0.2001 | 4.0139 | 5.6053 | 0.1750 | 4.4575 | 5.8827 | 0.1991 |
| | 1,000 | 0.5104 | 1.8908 | 0.2327 | 0.2907 | 1.5133 | 0.2146 | 0.4062 | 1.7517 | 0.2522 |
| | 2,000 | 0.1678 | 1.0260 | 0.1683 | 0.0929 | 0.8581 | 0.1714 | 0.1178 | 0.9144 | 0.1736 |
| | 5,000 | 0.0292 | 0.3901 | 0.1069 | 0.0073 | 0.3040 | 0.1095 | 0.0186 | 0.3400 | 0.1036 |
| | 10,000 | 0.0058 | 0.1638 | 0.0719 | 0.0014 | 0.0899 | 0.0834 | 0.0000 | 0.0919 | 0.0740 |
| | 100,000 | 0.0000 | 0.0171 | 0.0572 | 0.0006 | 0.0171 | 0.0614 | -0.0004 | 0.0185 | 0.0576 |
| $\beta_H = 10$ | 100 | 0.0520 | 1.5471 | 0.0926 | -0.0530 | 1.4858 | 0.0944 | 0.0460 | 1.6879 | 0.0888 |
| | 1,000 | -0.0078 | 0.4047 | 0.1185 | -0.0108 | 0.4158 | 0.1020 | -0.0100 | 0.4178 | 0.1195 |
| | 2,000 | -0.0093 | 0.2058 | 0.0936 | -0.0005 | 0.2067 | 0.0975 | -0.0129 | 0.2546 | 0.0933 |
| | 5,000 | -0.0037 | 0.0944 | 0.0727 | -0.0034 | 0.0922 | 0.0709 | -0.0052 | 0.0871 | 0.0709 |
| | 10,000 | -0.0023 | 0.0512 | 0.0555 | -0.0010 | 0.0504 | 0.0684 | -0.0034 | 0.0529 | 0.0580 |
| | 100,000 | -0.0005 | 0.0160 | 0.0522 | 0.0002 | 0.0154 | 0.0526 | -0.0007 | 0.0171 | 0.0560 |

*Notes:* The data generating process is (5.1). Parametrization are described in (S.3.1). $S = 4$ is used. "Baseline", "Categorical $x$" and "Categorical $u$" refer to DGP 1 to 3 as in Section 5.1. Generically, bias, RMSE and size are calculated by $R^{-1}\sum_{r=1}^{R}\left(\hat{\theta}^{(r)} - \theta_0\right)$, $\sqrt{R^{-1}\sum_{r=1}^{R}\left(\hat{\theta}^{(r)} - \theta_0\right)^2}$, and $R^{-1}\sum_{r=1}^{R}\mathbf{1}\left[\left|\hat{\theta}^{(r)} - \theta_0\right| / \hat{\sigma}_{\hat{\theta}}^{(r)} > \text{cv}_{0.05}\right]$, respectively, for true parameter $\theta_0$, its estimate $\hat{\theta}^{(r)}$, the estimated standard error of $\hat{\theta}^{(r)}$, $\hat{\sigma}_{\hat{\theta}}^{(r)}$, and the critical value $\text{cv}_{0.05} = \Phi^{-1}(0.975)$ across $R = 5{,}000$ replications, where $\Phi(\cdot)$ is the cumulative distribution function of standard normal distribution.

by GMM, Table S.7 reports the results of GMM estimation of moments of $\beta_i$ up to order 3 using the moment conditions as in the $K = 2$ case where $S = 4$ in the left panel, and the results of OLS estimation of $\boldsymbol{\phi}$ in the right panel. These results show that we are still able to obtain accurate estimation of lower order moments of $\beta_i$ when the fourth and fifth moments of $\beta_i$ are not used, confirming the lower information content of the higher order moments for estimation of the lower order moments of $\beta_i$.

### S.3.6 Experiments with idiosyncratic heterogeneity

In addition to the existing results, the following Monte Carlo experiment is designed to examine the finite sample performance of the estimator under different degrees of idiosyncratic heterogeneity. Following DGP 1 in Section 5.1, we generate $\tilde{x}_i \sim \text{IID}\chi^2(2)$, and then set $x_i = (\tilde{x}_i - 2)/2$. The additional regressors, $z_{ij}$, for $j = 1, 2$ with homogeneous slopes are generated as

$$z_{i1} = x_i + v_{i1} \text{ and } z_{i2} = z_{i1} + v_{i2},$$

with $v_{ij} \sim \text{IID } N(0, 1)$, for $j = 1, 2$. The error term, $u_i$, is generated as

$$u_i = \begin{cases} \sigma_i \varepsilon_i + e_i & \text{if } i = 1, 2, \cdots, \lfloor n^\alpha \rfloor \\ \sigma_i \varepsilon_i & \text{if } i = \lfloor n^\alpha \rfloor + 1, \cdots, n \end{cases}$$

where $\sigma_i^2$ are generated as $0.5(1+\text{IID}\chi^2(1))$, $\varepsilon_i \sim \text{IID}N(0, 1)$, and $e_i$ is the idiosyncratic heterogeneity that is generated from the standard normal distribution and then set to be fixed across Monte Carlo replications. Then in this case we have

$$\left| n^{-1} \sum_{i=1}^{n} \mathrm{E}\left(u_i^2\right) - 1 \right| = \left| n^{-1} \sum_{i=1}^{\lfloor n^\alpha \rfloor} e_i^2 \right| \leq n^{-1} \sum_{i=1}^{\lfloor n^\alpha \rfloor} \left| e_i^2 \right| \leq \left( \max_{1 \leq i \leq \lfloor n^\alpha \rfloor} \left| e_i^2 \right| \right) n^{\alpha - 1}.$$

Similar arguments can be made for $r = 3$.

Following the same parametrization as in Section 5, we consider the degree of heterogeneity $\alpha = 0.25$, 0.4, and 0.5. The estimation results are reported in Table S.8. The results are similar to that of the Baseline DGP as reported in Table 3, which suggests that the GMM estimator is robust to limited degrees of idiosyncratic heterogeneity.

## S.4 Additional empirical results

In this section, we provide additional results for the empirical application. In addition to the quadratic in experience in Section 6, we further consider the following quartic in experience specification,

$$\log \text{wage}_i = \alpha + \beta_i \text{edu}_i + \rho_1 \text{exper}_i + \rho_2 \text{exper}_i^2 + \rho_3 \text{exper}_i^3 + \rho_4 \text{exper}_i^4 + \tilde{\mathbf{z}}_i' \tilde{\boldsymbol{\gamma}} + u_i, \qquad \text{(S.4.1)}$$

Table S.6: Bias, RMSE and size of the GMM estimator for distributional parameters and moments of $\beta$ with $K = 3$

| | | Distribution of $\beta_i$ | | | | Moments of $\beta_i$ | | |
|---|---|---|---|---|---|---|---|---|
| Sample size $n$ | | Bias | RMSE | Size | | Bias | RMSE | Size |
| 100 | $\pi_L = 0.3$ | -0.0405 | 0.1910 | 0.1319 | $\mathrm{E}(\beta_i) = 2.1$ | 0.1484 | 0.7471 | 0.6451 |
| 1,000 | | -0.0417 | 0.1633 | 0.1915 | | -0.0711 | 0.5415 | 0.6128 |
| 2,000 | | -0.0383 | 0.1474 | 0.2354 | | -0.1112 | 0.4408 | 0.5264 |
| 5,000 | | -0.0299 | 0.1186 | 0.3098 | | -0.0904 | 0.3712 | 0.4034 |
| 10,000 | | -0.0209 | 0.0949 | 0.3371 | | -0.0523 | 0.2740 | 0.2910 |
| 100,000 | | -0.0074 | 0.0314 | 0.2295 | | -0.0026 | 0.0400 | 0.0678 |
| 200,000 | | -0.0050 | 0.0208 | 0.1917 | | -0.0004 | 0.0202 | 0.0568 |
| 100 | $\beta_M = 0.3$ | 0.2166 | 0.2995 | 0.0492 | $\mathrm{E}(\beta_i^2) = 5.1$ | 0.2841 | 2.8452 | 0.7223 |
| 1,000 | | 0.1404 | 0.2378 | 0.1364 | | -0.6374 | 1.9507 | 0.6456 |
| 2,000 | | 0.1035 | 0.2117 | 0.1901 | | -0.7163 | 1.7408 | 0.5472 |
| 5,000 | | 0.0615 | 0.1645 | 0.2381 | | -0.5478 | 1.4628 | 0.4472 |
| 10,000 | | 0.0364 | 0.1292 | 0.2477 | | -0.3391 | 1.1394 | 0.3432 |
| 100,000 | | 0.0013 | 0.0322 | 0.1305 | | -0.0209 | 0.2300 | 0.0932 |
| 200,000 | | 0.0006 | 0.0185 | 0.1033 | | -0.0046 | 0.1128 | 0.0620 |
| 100 | $\beta_L = 1$ | 0.6881 | 1.1994 | 0.1110 | $\mathrm{E}(\beta_i^3) = 13.5$ | 0.4897 | 10.0757 | 0.7189 |
| 1,000 | | 0.2588 | 0.7438 | 0.1994 | | -2.7735 | 7.0573 | 0.6718 |
| 2,000 | | 0.1096 | 0.5372 | 0.2607 | | -2.9100 | 6.3988 | 0.5894 |
| 5,000 | | 0.0205 | 0.4184 | 0.3426 | | -2.1889 | 5.4307 | 0.5078 |
| 10,000 | | 0.0070 | 0.2733 | 0.3360 | | -1.3454 | 4.3382 | 0.4042 |
| 100,000 | | -0.0064 | 0.0556 | 0.2213 | | -0.0942 | 1.0263 | 0.1132 |
| 200,000 | | -0.0047 | 0.0320 | 0.1775 | | -0.0236 | 0.5035 | 0.0738 |
| 100 | $\beta_M = 2$ | 0.1249 | 0.7256 | 0.0642 | $\mathrm{E}(\beta_i^4) = 37.5$ | 0.9092 | 35.1538 | 0.7235 |
| 1,000 | | -0.1190 | 0.6298 | 0.1531 | | -10.1071 | 24.1521 | 0.6944 |
| 2,000 | | -0.1935 | 0.5762 | 0.2303 | | -10.7108 | 21.5751 | 0.6268 |
| 5,000 | | -0.1662 | 0.4777 | 0.3670 | | -8.2675 | 18.7735 | 0.5464 |
| 10,000 | | -0.1261 | 0.3703 | 0.4414 | | -5.5310 | 15.4382 | 0.4406 |
| 100,000 | | -0.0326 | 0.1175 | 0.2681 | | -0.4433 | 3.5927 | 0.1240 |
| 200,000 | | -0.0193 | 0.0682 | 0.2203 | | -0.1114 | 1.6644 | 0.0810 |
| 100 | $\beta_H = 3$ | 0.8514 | 3.1645 | 0.1064 | $\mathrm{E}(\beta_i^5) = 107.1$ | 2.4059 | 121.1286 | 0.6989 |
| 1,000 | | 1.6632 | 4.5208 | 0.3124 | | -34.0298 | 77.5508 | 0.7012 |
| 2,000 | | 1.7929 | 4.6701 | 0.4000 | | -35.4018 | 69.5876 | 0.6424 |
| 5,000 | | 1.3425 | 4.0152 | 0.4539 | | -27.3828 | 60.4373 | 0.5638 |
| 10,000 | | 0.9637 | 3.3831 | 0.4333 | | -18.1022 | 50.3990 | 0.4590 |
| 100,000 | | 0.0474 | 0.8321 | 0.2046 | | -1.5330 | 11.7796 | 0.1314 |
| 200,000 | | 0.0033 | 0.3237 | 0.1573 | | -0.4226 | 5.9529 | 0.0812 |

*Notes:* The data generating process is (S.3.2). Generically, bias, RMSE and size are calculated by $R^{-1}\sum_{r=1}^{R}\left(\hat{\theta}^{(r)}-\theta_0\right)$, $\sqrt{R^{-1}\sum_{r=1}^{R}\left(\hat{\theta}^{(r)}-\theta_0\right)^2}$, and $R^{-1}\sum_{r=1}^{R}\mathbf{1}\left[\left|\hat{\theta}^{(r)}-\theta_0\right|/\hat{\sigma}_{\hat{\theta}}^{(r)} > \mathrm{cv}_{0.05}\right]$, respectively, for true parameter $\theta_0$, its estimate $\hat{\theta}^{(r)}$, the estimated standard error of $\hat{\theta}^{(r)}$, $\hat{\sigma}_{\hat{\theta}}^{(r)}$, and the critical value $\mathrm{cv}_{0.05} = \Phi^{-1}(0.975)$ across $R = 5,000$ replications, where $\Phi(\cdot)$ is the cumulative distribution function of standard normal distribution.

Table S.7: Bias, RMSE and size of estimation of $\phi$ and moments of $\beta_i$ (using $S = 4$) with $K = 3$

| | | Moments of $\beta_i$ ($S = 4$) | | | | OLS Estimate $hat\phi i$ | | |
|---|---|---|---|---|---|---|---|---|
| $n$ | | Bias | RMSE | Size | | Bias | RMSE | Size |
| 100 | $\text{E}(\beta_i) = 2.1$ | 0.0025 | 0.2867 | 0.2088 | $\text{E}(\beta_i) = 2.1$ | -0.0031 | 0.2768 | 0.1042 |
| 1,000 | | -0.0006 | 0.0821 | 0.1008 | | -0.0008 | 0.0939 | 0.0588 |
| 2,000 | | 0.0004 | 0.0537 | 0.0734 | | 0.0000 | 0.0653 | 0.0550 |
| 5,000 | | 0.0004 | 0.0323 | 0.0610 | | -0.0008 | 0.0422 | 0.0506 |
| 10,000 | | 0.0007 | 0.0224 | 0.0572 | | -0.0001 | 0.0299 | 0.0510 |
| 100,000 | | 0.0000 | 0.0069 | 0.0454 | | -0.0001 | 0.0093 | 0.0462 |
| 200,000 | | 0.0000 | 0.0050 | 0.0550 | | 0.0000 | 0.0067 | 0.0498 |
| 100 | $\text{E}(\beta_i^2) = 5.1$ | -0.1195 | 1.8290 | 0.3948 | $\gamma_1 = 1$ | -0.0020 | 0.1817 | 0.0604 |
| 1,000 | | -0.0455 | 0.5965 | 0.1602 | | 0.0000 | 0.0581 | 0.0474 |
| 2,000 | | -0.0196 | 0.3454 | 0.0902 | | 0.0001 | 0.0409 | 0.0474 |
| 5,000 | | -0.0073 | 0.1630 | 0.0608 | | -0.0001 | 0.0259 | 0.0494 |
| 10,000 | | -0.0004 | 0.1028 | 0.0544 | | -0.0004 | 0.0183 | 0.0518 |
| 100,000 | | 0.0001 | 0.0311 | 0.0488 | | -0.0001 | 0.0058 | 0.0490 |
| 200,000 | | -0.0002 | 0.0217 | 0.0492 | | -0.0001 | 0.0041 | 0.0490 |
| 100 | $\text{E}(\beta_i^3) = 13.5$ | -0.7404 | 6.7772 | 0.4396 | $\gamma_2 = 1$ | 0.0011 | 0.1296 | 0.0672 |
| 1,000 | | -0.3116 | 2.2732 | 0.1964 | | 0.0000 | 0.0414 | 0.0570 |
| 2,000 | | -0.1433 | 1.3285 | 0.1110 | | 0.0000 | 0.0291 | 0.0478 |
| 5,000 | | -0.0524 | 0.6468 | 0.0702 | | -0.0001 | 0.0183 | 0.0506 |
| 10,000 | | -0.0117 | 0.4052 | 0.0568 | | 0.0002 | 0.0130 | 0.0526 |
| 100,000 | | 0.0001 | 0.1236 | 0.0528 | | 0.0001 | 0.0041 | 0.0494 |
| 200,000 | | -0.0009 | 0.0850 | 0.0462 | | 0.0000 | 0.0029 | 0.0542 |

*Notes:* The data generating process is (S.3.2). Generically, bias, RMSE and size are calculated by $R^{-1}\sum_{r=1}^{R}\left(\hat{\theta}^{(r)} - \theta_0\right)$, $\sqrt{R^{-1}\sum_{r=1}^{R}\left(\hat{\theta}^{(r)} - \theta_0\right)^2}$, and $R^{-1}\sum_{r=1}^{R}\mathbf{1}\left[\left|\hat{\theta}^{(r)} - \theta_0\right| / \hat{\sigma}_{\hat{\theta}}^{(r)} > \text{cv}_{0.05}\right]$, respectively, for true parameter $\theta_0$, its estimate $\hat{\theta}^{(r)}$, the estimated standard error of $\hat{\theta}^{(r)}$, $\hat{\sigma}_{\hat{\theta}}^{(r)}$, and the critical value $\text{cv}_{0.05} = \Phi^{-1}(0.975)$ across $R = 5,000$ replications, where $\Phi(\cdot)$ is the cumulative distribution function of standard normal distribution.

Table S.8: Bias, RMSE and size of the GMM estimator for distributional parameters of $\beta$

| | $\alpha$ | 0.25 | | | 0.40 | | | 0.50 | | |
|---|---|---|---|---|---|---|---|---|---|---|
| | Sample size $n$ | Bias | RMSE | Size | Bias | RMSE | Size | Bias | RMSE | Size |
| | | *high variance*: $\text{var}(\beta_i) = 0.25$ | | | | | | | | |
| $\pi = 0.5$ | 100 | 0.0292 | 0.2201 | 0.1957 | 0.0293 | 0.2177 | 0.1859 | 0.0297 | 0.2160 | 0.1609 |
| | 1,000 | 0.0020 | 0.1273 | 0.1943 | 0.0039 | 0.1293 | 0.2047 | 0.0037 | 0.1356 | 0.2150 |
| | 2,000 | 0.0014 | 0.0879 | 0.1585 | 0.0003 | 0.0812 | 0.1421 | 0.0020 | 0.0851 | 0.1455 |
| | 5,000 | 0.0002 | 0.0440 | 0.0980 | 0.0010 | 0.0457 | 0.0982 | -0.0003 | 0.0445 | 0.0946 |
| | 10,000 | -0.0007 | 0.0301 | 0.0764 | 0.0003 | 0.0304 | 0.0824 | -0.0001 | 0.0311 | 0.0910 |
| | 100,000 | 0.0000 | 0.0098 | 0.0610 | 0.0000 | 0.0097 | 0.0536 | -0.0002 | 0.0096 | 0.0556 |
| $\beta_L = 1$ | 100 | 0.2027 | 0.5686 | 0.1807 | 0.1993 | 0.5706 | 0.1738 | 0.2007 | 0.5662 | 0.1712 |
| | 1,000 | 0.0104 | 0.1711 | 0.2115 | 0.0136 | 0.1750 | 0.2156 | 0.0079 | 0.1827 | 0.2132 |
| | 2,000 | 0.0094 | 0.1121 | 0.1741 | 0.0069 | 0.1025 | 0.1529 | 0.0087 | 0.1109 | 0.1593 |
| | 5,000 | 0.0040 | 0.0543 | 0.1090 | 0.0052 | 0.0557 | 0.1136 | 0.0050 | 0.0546 | 0.1112 |
| | 10,000 | 0.0023 | 0.0365 | 0.0856 | 0.0024 | 0.0365 | 0.0882 | 0.0025 | 0.0367 | 0.0922 |
| | 100,000 | 0.0004 | 0.0116 | 0.0602 | 0.0005 | 0.0115 | 0.0604 | 0.0004 | 0.0115 | 0.0584 |
| $\beta_H = 2$ | 100 | -0.1947 | 0.5616 | 0.1307 | -0.1983 | 0.5545 | 0.1421 | -0.2094 | 0.5510 | 0.1358 |
| | 1,000 | -0.0096 | 0.1720 | 0.1682 | -0.0078 | 0.1729 | 0.1710 | -0.0066 | 0.1802 | 0.1751 |
| | 2,000 | -0.0060 | 0.1142 | 0.1445 | -0.0068 | 0.1066 | 0.1523 | -0.0070 | 0.1060 | 0.1405 |
| | 5,000 | -0.0047 | 0.0530 | 0.1130 | -0.0037 | 0.0545 | 0.1110 | -0.0054 | 0.0559 | 0.1088 |
| | 10,000 | -0.0031 | 0.0360 | 0.0922 | -0.0023 | 0.0370 | 0.0826 | -0.0024 | 0.0372 | 0.0896 |
| | 100,000 | -0.0004 | 0.0116 | 0.0592 | -0.0003 | 0.0115 | 0.0546 | -0.0005 | 0.0114 | 0.0600 |
| | | *low variance*: $\text{var}(\beta_i) = 0.15$ | | | | | | | | |
| $\pi = 0.3$ | 100 | 0.2132 | 0.2951 | 0.1851 | 0.2133 | 0.2912 | 0.1797 | 0.2132 | 0.2945 | 0.1716 |
| | 1,000 | 0.0133 | 0.1591 | 0.1894 | 0.0125 | 0.1613 | 0.1872 | 0.0163 | 0.1637 | 0.1840 |
| | 2,000 | -0.0051 | 0.1103 | 0.1619 | -0.0055 | 0.1048 | 0.1553 | -0.0027 | 0.1083 | 0.1559 |
| | 5,000 | -0.0046 | 0.0599 | 0.1198 | -0.0029 | 0.0607 | 0.1070 | -0.0046 | 0.0620 | 0.1208 |
| | 10,000 | -0.0038 | 0.0418 | 0.0900 | -0.0023 | 0.0418 | 0.0932 | -0.0022 | 0.0423 | 0.0930 |
| | 100,000 | -0.0003 | 0.0132 | 0.0622 | -0.0003 | 0.0130 | 0.0576 | -0.0004 | 0.0127 | 0.0532 |
| $\beta_L = 0.5$ | 100 | 0.3935 | 0.6293 | 0.1959 | 0.3900 | 0.6353 | 0.1853 | 0.3917 | 0.6236 | 0.1811 |
| | 1,000 | 0.0310 | 0.2598 | 0.1590 | 0.0357 | 0.2634 | 0.1589 | 0.0298 | 0.2653 | 0.1609 |
| | 2,000 | 0.0025 | 0.1590 | 0.1539 | 0.0004 | 0.1478 | 0.1274 | 0.0025 | 0.1565 | 0.1459 |
| | 5,000 | -0.0008 | 0.0849 | 0.1100 | 0.0018 | 0.0849 | 0.1122 | 0.0003 | 0.0854 | 0.1078 |
| | 10,000 | -0.0001 | 0.0586 | 0.0922 | 0.0004 | 0.0586 | 0.0958 | 0.0012 | 0.0576 | 0.0918 |
| | 100,000 | 0.0005 | 0.0183 | 0.0596 | 0.0002 | 0.0181 | 0.0582 | 0.0003 | 0.0177 | 0.0558 |
| $\beta_H = 1.345$ | 100 | -0.0463 | 0.4194 | 0.1128 | -0.0509 | 0.4224 | 0.1147 | -0.0489 | 0.4386 | 0.1239 |
| | 1,000 | -0.0097 | 0.1428 | 0.1498 | -0.0106 | 0.1427 | 0.1523 | -0.0094 | 0.1486 | 0.1467 |
| | 2,000 | -0.0107 | 0.0920 | 0.1443 | -0.0106 | 0.0917 | 0.1439 | -0.0093 | 0.0915 | 0.1389 |
| | 5,000 | -0.0065 | 0.0492 | 0.1166 | -0.0056 | 0.0500 | 0.1092 | -0.0063 | 0.0532 | 0.1134 |
| | 10,000 | -0.0045 | 0.0345 | 0.0910 | -0.0037 | 0.0344 | 0.0902 | -0.0035 | 0.0344 | 0.0900 |
| | 100,000 | -0.0006 | 0.0108 | 0.0602 | -0.0004 | 0.0107 | 0.0572 | -0.0005 | 0.0105 | 0.0560 |

*Notes: high variance* and *low variance* parametrization are described in (5.2). "Baseline", "Categorical $x$" and "Categorical $u$" refer to DGP 1 to 3 as in Section 5.1. Generically, bias, RMSE and size are calculated by $R^{-1}\sum_{r=1}^{R}\left(\hat{\theta}^{(r)} - \theta_0\right)$, $\sqrt{R^{-1}\sum_{r=1}^{R}\left(\hat{\theta}^{(r)} - \theta_0\right)^2}$, and $R^{-1}\sum_{r=1}^{R}\mathbf{1}\left[\left|\hat{\theta}^{(r)} - \theta_0\right|/\hat{\sigma}_{\hat{\theta}}^{(r)} > \text{cv}_{0.05}\right]$, respectively, for true parameter $\theta_0$, its estimate $\hat{\theta}^{(r)}$, the estimated standard error of $\hat{\theta}^{(r)}$, $\hat{\sigma}_{\hat{\theta}}^{(r)}$, and the critical value $\text{cv}_{0.05} = \Phi^{-1}(0.975)$ across $R = 5,000$ replications, where $\Phi(\cdot)$ is the cumulative distribution function of standard normal distribution.

where

$$\beta_i = \begin{cases} b_L & \text{w.p. } \pi, \\ b_H & \text{w.p. } 1-\pi. \end{cases}$$

Table S.9 and S.10 report the estimates of the distributional parameters of $\beta_i$ and the estimates of $\boldsymbol{\gamma}$ with the specification (S.4.1).

The estimates of parameter of interests with specification (S.4.1) are almost the same as that with quadratic in experience specification (6.3), reported in Table 5. The qualitative analysis and conclusion discussed in Section 6 remain robust to adding third and fourth order powers of $\text{exper}_i$ in the regressions.

## S.5 Computational algorithm

In this section, we describe the computational procedure used for estimation of $\boldsymbol{\gamma}$, moments of $\beta_i$, and distributional parameters of $\beta_i$.

1. Denote $\mathbf{w}_i = (x_i, \mathbf{z}_i')'$. Compute the OLS estimator

$$\left(\widehat{\mathrm{E}\left(\beta_i\right)}^{(0)}, \widehat{\boldsymbol{\gamma}}'\right)' = \left(\frac{1}{n}\sum_{i=1}^{n} \mathbf{w}_i \mathbf{w}_i'\right)^{-1} \left(\frac{1}{n}\sum_{i=1}^{n} \mathbf{w}_i' y_i\right),$$

   and $\hat{\tilde{y}}_i = y_i - \mathbf{z}_i'\widehat{\gamma}$.

2. For $r = 2, 3, \cdots, 2K-1$, compute the sample version of the moment conditions (2.8) and (2.9) in the main paper by replacing $\rho_{r,s}$ by $n^{-1}\sum_{i=1}^{n} \hat{\tilde{y}}_i^r x_i^s$, and solving for $\widehat{\mathrm{E}\left(\beta_i^r\right)}^{(0)}$ and $\widehat{\sigma_r}^{(0)}$, recursively.

3. Use the initial estimates $\left\{\widehat{\mathrm{E}\left(\beta_i^r\right)}^{(0)}\right\}_{r=1}^{2K-1}$ and $\left\{\widehat{\sigma_r}^{(0)}\right\}_{r=2}^{2K-1}$ to construct the weighting matrix $\hat{\mathbf{A}}_n$ in (3.10) and compute the GMM estimators $\left\{\widehat{\mathrm{E}\left(\beta_i^r\right)}^{(1)}\right\}_{r=1}^{2K-1}$ and $\left\{\widehat{\sigma_r}^{(1)}\right\}_{r=2}^{2K-1}$ to compute the moments of $\beta_i$ and $\sigma_r$. Iterate the GMM estimation one more time with $\left\{\widehat{\mathrm{E}\left(\beta_i^r\right)}^{(1)}\right\}_{r=1}^{2K-1}$ and $\left\{\widehat{\sigma_r}^{(1)}\right\}_{r=2}^{2K-1}$ as initial estimates to obtain $\left\{\widehat{\mathrm{E}\left(\beta_i^r\right)}\right\}_{r=1}^{2K-1}$ and $\{\widehat{\sigma_r}\}_{r=2}^{2K-1}$.

4. Solve

$$\min_{\pi_k, b_k} \left\{ \sum_{j=1}^{r} \left( \sum_{k=1}^{K} \pi_k b_k^r - \widehat{\mathrm{E}\left(\beta_i^r\right)} \right)^2 \right\}$$

   to get the initial estimates, $\widehat{\boldsymbol{\theta}}^{(0)} = \left(\widehat{\boldsymbol{\pi}}^{(0)\prime}, \widehat{\mathbf{b}}^{(0)\prime}\right)'$.

5. Using $\widehat{\boldsymbol{\theta}}^{(0)} = \left(\widehat{\boldsymbol{\pi}}^{(0)\prime}, \widehat{\mathbf{b}}^{(0)\prime}\right)'$ construct the weighting matrix $\hat{\mathbf{A}}_n$ and compute the GMM estimator as $\widehat{\boldsymbol{\theta}}^{(1)} = \left(\widehat{\boldsymbol{\pi}}^{(1)\prime}, \widehat{\mathbf{b}}^{(1)\prime}\right)'$ for $\boldsymbol{\theta}$. Iterate the GMM estimation one more time with

Table S.9: Estimates of the distribution of the return to education with specification (S.4.1) across two periods, 1973 - 75 and 2001 - 03, by years of education and gender

| | High School or Less | | Postsecondary Edu. | | All | |
|---|---|---|---|---|---|---|
| | 1973 - 75 | 2001 - 03 | 1973 - 75 | 2001 - 03 | 1973 - 75 | 2001 - 03 |
| | | | | | Both Male and Female | |
| $\pi$ | 0.4841 | 0.5081 | 0.4281 | 0.3576 | 0.4689 | 0.3559 |
| | *(5274.3)* | *(0.0267)* | *(0.0495)* | *(0.0089)* | *(0.0534)* | *(0.0046)* |
| $\beta_L$ | 0.0617 | 0.0392 | 0.0627 | 0.0859 | 0.0567 | 0.0658 |
| | *(5.9252)* | *(0.0013)* | *(0.0035)* | *(0.0009)* | *(0.0022)* | *(0.0004)* |
| $\beta_H$ | 0.0628 | 0.0928 | 0.1108 | 0.1397 | 0.0938 | 0.1270 |
| | *(5.5919)* | *(0.0019)* | *(0.0031)* | *(0.0007)* | *(0.0023)* | *(0.0004)* |
| $\beta_H/\beta_L$ | 1.0177 | 2.3645 | 1.7675 | 1.6267 | 1.6533 | 1.9299 |
| | *(7.1413)* | *(0.0400)* | *(0.0629)* | *(0.0111)* | *(0.0305)* | *(0.0076)* |
| $\mathrm{E}(\beta_i)$ | 0.0623 | 0.0656 | 0.0902 | 0.1205 | 0.0764 | 0.1053 |
| $\mathrm{var}(\beta_i)$ | 0.0005 | 0.0268 | 0.0238 | 0.0258 | 0.0185 | 0.0293 |
| $n$ | 77,899 | 216,136 | 33,733 | 295,683 | 111,632 | 511,819 |
| | | | | | | Male |
| $\pi$ | 0.4835 | 0.4968 | 0.4478 | 0.3007 | 0.4856 | 0.3550 |
| | n/a | *(0.0394)* | *(0.0676)* | *(0.0095)* | *(0.0936)* | *(0.0052)* |
| $\beta_L$ | 0.0648 | 0.0419 | 0.0520 | 0.0733 | 0.0553 | 0.0581 |
| | n/a | *(0.0019)* | *(0.0047)* | *(0.0012)* | *(0.0033)* | *(0.0005)* |
| $\beta_H$ | 0.0651 | 0.0927 | 0.0988 | 0.1321 | 0.0875 | 0.1220 |
| | n/a | *(0.0026)* | *(0.0041)* | *(0.0008)* | *(0.0034)* | *(0.0005)* |
| $\beta_H/\beta_L$ | 1.0048 | 2.2143 | 1.9002 | 1.8015 | 1.5816 | 2.1003 |
| | n/a | *(0.0495)* | *(0.1124)* | *(0.0210)* | *(0.0456)* | *(0.0124)* |
| $\mathrm{E}(\beta_i)$ | 0.0649 | 0.0675 | 0.0778 | 0.1144 | 0.0719 | 0.0993 |
| $\mathrm{var}(\beta_i)$ | 0.0002 | 0.0254 | 0.0233 | 0.0269 | 0.0161 | 0.0306 |
| $n$ | 44,299 | 116,129 | 20,851 | 144,138 | 65,150 | 260,267 |
| | | | | | | Female |
| $\pi$ | 0.5000 | 0.5210 | 0.4512 | 0.3849 | 0.4733 | 0.3773 |
| | *(0.5611)* | *(0.0281)* | *(0.0739)* | *(0.0167)* | *(0.0870)* | *(0.0083)* |
| $\beta_L$ | 0.0453 | 0.0352 | 0.0804 | 0.0956 | 0.0644 | 0.0762 |
| | *(0.0143)* | *(0.0016)* | *(0.0050)* | *(0.0013)* | *(0.0034)* | *(0.0006)* |
| $\beta_H$ | 0.0724 | 0.0969 | 0.1307 | 0.1449 | 0.1032 | 0.1338 |
| | *(0.0169)* | *(0.0025)* | *(0.0052)* | *(0.0011)* | *(0.0040)* | *(0.0007)* |
| $\beta_H/\beta_L$ | 1.5994 | 2.7540 | 1.6252 | 1.5154 | 1.6012 | 1.7564 |
| | *(0.1537)* | *(0.0666)* | *(0.0551)* | *(0.0125)* | *(0.0323)* | *(0.0084)* |
| $\mathrm{E}(\beta_i)$ | 0.0588 | 0.0648 | 0.1080 | 0.1260 | 0.0848 | 0.1121 |
| $\mathrm{var}(\beta_i)$ | 0.0136 | 0.0308 | 0.0250 | 0.0240 | 0.0193 | 0.0279 |
| $n$ | 33,600 | 100,007 | 12,882 | 151,545 | 46,482 | 251,552 |

*Notes:* This table reports the estimates of the distribution of $\beta_i$ with the quartic in experience specification (S.4.1), using $S = 4$ order moments of $\mathrm{edu}_i$. "Postsecondary Edu." stands for the sub-sample with years of education higher than 12 and "High School or Less" stands for those with years of education less than or equal to 12. $\mathrm{s.d.}(\beta_i)$ corresponds to the square root of estimated $\mathrm{var}(\beta_i)$. $n$ is the sample size. "n/a" is inserted when the estimates show homogeneity of $\beta_i$ and $\pi$ is not identified and cannot be estimated.

Table S.10: Estimates of $\boldsymbol{\gamma}$ associated with control variables $\mathbf{z}_i$ with specification (S.4.1) across two periods, 1973 - 75 and 2001 - 03, by years of education and gender, which complements Table S.9

| | High School or Less | | Postsecondary Edu. | | All | |
|---|---|---|---|---|---|---|
| | 1973 - 75 | 2001 - 03 | 1973 - 75 | 2001 - 03 | 1973 - 75 | 2001 - 03 |
| | | | *Both male and female* | | | |
| `exper.` | 0.0769 | 0.0526 | 0.0817 | 0.0763 | 0.0757 | 0.0603 |
| | *(0.0015)* | *(0.0009)* | *(0.0029)* | *(0.0012)* | *(0.0013)* | *(0.0007)* |
| `exper.`$^2$ | -0.0040 | -0.0020 | -0.0045 | -0.0039 | -0.0038 | -0.0024 |
| | *(0.0001)* | *(0.0001)* | *(0.0003)* | *(0.0001)* | *(0.0001)* | *(0.0001)* |
| `exper.`$^3$ $(\times 10^5)$ | 9.2470 | 3.4329 | 11.2100 | 8.9370 | 8.3625 | 3.6521 |
| | *(0.4146)* | *(0.2882)* | *(1.2538)* | *(0.4460)* | *(0.3677)* | *(0.2412)* |
| `exper.`$^4$ $(\times 10^5)$ | -0.0768 | -0.0236 | -0.1074 | -0.0777 | -0.0654 | -0.0169 |
| | *(0.0043)* | *(0.0031)* | *(0.0158)* | *(0.0054)* | *(0.0039)* | *(0.0027)* |
| `marriage` | 0.0819 | 0.0700 | 0.0728 | 0.0674 | 0.0799 | 0.0718 |
| | *(0.0037)* | *(0.0020)* | *(0.0060)* | *(0.0020)* | *(0.0031)* | *(0.0014)* |
| `nonwhite` | -0.1052 | -0.0808 | -0.0486 | -0.0613 | -0.0855 | -0.0719 |
| | *(0.0046)* | *(0.0024)* | *(0.0088)* | *(0.0025)* | *(0.0041)* | *(0.0018)* |
| `gender` | 0.4146 | 0.2272 | 0.2933 | 0.2008 | 0.3854 | 0.2150 |
| | *(0.0029)* | *(0.0017)* | *(0.0049)* | *(0.0018)* | *(0.0025)* | *(0.0013)* |
| $n$ | 77,899 | 216,136 | 33,733 | 295,683 | 111,632 | 511,819 |
| | | | *Male* | | | |
| `exper.` | 0.0823 | 0.0620 | 0.0859 | 0.0780 | 0.0825 | 0.0664 |
| | *(0.0020)* | *(0.0012)* | *(0.0040)* | *(0.0018)* | *(0.0017)* | *(0.0010)* |
| `exper.`$^2$ $(\times 10^2)$ | -0.0039 | -0.0024 | -0.0041 | -0.0036 | -0.0037 | -0.0025 |
| | *(0.0002)* | *(0.0001)* | *(0.0004)* | *(0.0002)* | *(0.0001)* | *(0.0001)* |
| `exper.`$^3$ $(\times 10^5)$ | 8.2014 | 4.3686 | 9.2747 | 7.3170 | 7.4306 | 3.6749 |
| | *(0.5321)* | *(0.3864)* | *(1.7422)* | *(0.6709)* | *(0.4700)* | *(0.3241)* |
| `exper.`$^4$ $(\times 10^5)$ | -0.0650 | -0.0314 | -0.0880 | -0.0582 | -0.0552 | -0.0161 |
| | *(0.0054)* | *(0.0042)* | *(0.0223)* | *(0.0081)* | *(0.0049)* | *(0.0036)* |
| `marriage` | 0.1493 | 0.1052 | 0.1310 | 0.1234 | 0.1421 | 0.1192 |
| | *(0.0056)* | *(0.0029)* | *(0.0088)* | *(0.0031)* | *(0.0048)* | *(0.0021)* |
| `nonwhite` | -0.1362 | -0.1191 | -0.1214 | -0.1040 | -0.1309 | -0.1136 |
| | *(0.0064)* | *(0.0035)* | *(0.0126)* | *(0.0039)* | *(0.0057)* | *(0.0027)* |
| $n$ | 44,299 | 116,129 | 20,851 | 144,138 | 65,150 | 260,267 |
| | | | *Female* | | | |
| `exper.` | 0.0713 | 0.0455 | 0.0911 | 0.0782 | 0.0729 | 0.0568 |
| | *(0.0022)* | *(0.0013)* | *(0.0040)* | *(0.0016)* | *(0.0019)* | *(0.0011)* |
| `exper.`$^2$ $(\times 10^2)$ | -0.0044 | -0.0018 | -0.0067 | -0.0045 | -0.0045 | -0.0025 |
| | *(0.0002)* | *(0.0001)* | *(0.0004)* | *(0.0002)* | *(0.0002)* | *(0.0001)* |
| `exper.`$^3$ $(\times 10^5)$ | 11.0325 | 3.4767 | 19.6859 | 11.2858 | 11.3406 | 4.4944 |
| | *(0.6649)* | *(0.4360)* | *(1.7412)* | *(0.5915)* | *(0.6095)* | *(0.3682)* |
| `exper.`$^4$ $(\times 10^5)$ | -0.0974 | -0.0264 | -0.1979 | -0.1046 | -0.0969 | -0.0272 |
| | *(0.0071)* | *(0.0048)* | *(0.0216)* | *(0.0071)* | *(0.0066)* | *(0.0042)* |
| `marriage` | -0.0078 | 0.0278 | -0.0175 | 0.0168 | -0.0082 | 0.0234 |
| | *(0.0048)* | *(0.0028)* | *(0.0080)* | *(0.0026)* | *(0.0041)* | *(0.0020)* |
| `nonwhite` | -0.0714 | -0.0479 | 0.0276 | -0.0291 | -0.0356 | -0.0375 |
| | *(0.0065)* | *(0.0033)* | *(0.0117)* | *(0.0033)* | *(0.0057)* | *(0.0024)* |
| $n$ | 33,600 | 100,007 | 12,882 | 151,545 | 46,482 | 251,552 |

*Notes:* This table reports the estimates of $\boldsymbol{\gamma}$ in (S.4.1). "Postsecondary Edu." stands for the sub-sample with years of education higher than 12 and "High School or Less" stands for those with years of education less than or equal to 12. The standard error of estimates of coefficients associated with control variables are estimated based on Theorem 3 and reported in parentheses. $n$ is the sample size.

$\widehat{\boldsymbol{\theta}}^{(1)} = \left(\widehat{\boldsymbol{\pi}}^{(1)\prime}, \widehat{\mathbf{b}}^{(1)\prime}\right)'$ as initial estimates to obtain $\widehat{\boldsymbol{\theta}} = \left(\widehat{\boldsymbol{\pi}}', \widehat{\mathbf{b}}'\right)'$. In the setup of the optimization problem for the optimization solver, imposing the constraint $b_1 < b_2 < \cdots < b_K$ is important to improve the numerical performance, particularly when $n$ is not sufficiently large (less than $5,000$).

# References


Hansen, E. B. (2022). *Econometrics.* Princeton University Press, Princeton.

Pesaran, M. H. (2015). *Time Series and Panel Data Econometrics.* Oxford University Press, Oxford.