Supplement to "Group Local Projection"

Not For Publication

Jiaming Huang

Universitat Pompeu Fabra, Barcelona GSE

September 29, 2023

Contents

S1 Bias correction

Theorem 2 shows that the infeasible GLP estimator admits

$$\sqrt{N_j T}(\hat{\beta}_{j,h} - \beta^0_{j,h}) \xrightarrow{d} \Sigma^{-1}_{j,h} N(\sqrt{\kappa_j}\mathcal{B}, \Psi_{j,h}) \tag{1}$$

This section derives the analytical formula to estimate the bias. We define the following estimates:

$$\begin{aligned}
\hat{\mathcal{O}}_{zc,i,h} =& \overline{d}'_{zc,i}\hat{\Omega}_{i,h}\overline{d}_{zc,i} \\
\overline{M}_{zc,i,h} =& \hat{\Omega}_{i,h} - \hat{\Omega}_{i,h}\overline{d}_{zc,i}\left(\overline{d}'_{zc,i}\hat{\Omega}_{i,h}\overline{d}_{zc,i}\right)^{-1}\overline{d}'_{zc,i}\hat{\Omega}_{i,h} \\
e_{i,t+h} =& y_{i,t+h} - x_{i,t}\hat{\beta}_{\hat{g}_i,h} - c_{i,t}\hat{\phi}_{i,h}
\end{aligned} \tag{2}$$

Results from Section S4.3 shows that the bias contains

$$\begin{aligned}
\mathcal{B}_1 =& \frac{1}{N_j T^2}\sum_{i\in\mathcal{S}^0_j}\sum_t\sum_s z_{i,t}x'_{i,t}M_{zc,i,h}z_{i,s}\epsilon_{i,s+h} \\
\mathcal{B}_2 =& \frac{1}{N_j}\sum_{i\in\mathcal{S}^0_j} d'_{zx,i}\left(\hat{\Omega}_{i,h} - \Omega_{i,h}\right)\Omega^{-1}_{i,h}M_{zc,i,h}\overline{d}_{z\epsilon,i,h} \\
&- \frac{1}{N_j}\sum_{i\in\mathcal{S}^0_j} d'_{zx,i}\Omega_{i,h}d_{zc,i}\mathcal{O}^{-1}_{zc,i,h}d_{zc,i}\left(\hat{\Omega}_{i,h} - \Omega_{i,h}\right)\overline{d}_{z\epsilon,i,h} \\
\mathcal{B}_3 =& \frac{1}{N_j}\sum_{i\in\mathcal{S}^0_j} d'_{zx,i}\Omega_{i,h}(\overline{d}_{zc,i} - d_{zc,i})\mathcal{O}^{-1}_{zc,i,h}d_{zc,i}\Omega_{i,h}\overline{d}_{z\epsilon,i,h} \\
\mathcal{B}_4 =& \frac{1}{N_j}\sum_{i\in\mathcal{S}^0_j} d'_{zx,i}\Omega_{i,h}d_{zc,i}\mathcal{O}^{-1}_{zc,i,h}\left(\bar{d}_{zc,i} - d_{zc,i}\right)'\Omega_{i,h}\overline{d}_{z\epsilon,i,h}
\end{aligned} \tag{3}$$

The above bias derivation generalizes the setup in Su et al. (2016), where they ignore the bias caused by the estimation of individual-level parameters. As a result, they only contain the first term $\mathcal{B}_1$. However, general formula for correcting the bias arising from the weighting matrix $\mathcal{B}_2$ is not available. Instead, here I assume individual-level 2SLS weighting matrix, i.e., $\hat{\Omega}_{i,h} = \frac{1}{T}z_{i,t}z'_{i,t}$.

We propose to estimate the bias by

$$
\begin{aligned}
\hat{\mathcal{B}}_1 =& \frac{1}{N_j T^2} \sum_{i\in\hat{\mathcal{S}}_j} \sum_t \sum_s k(|t-s|) z_{i,t} x'_{i,t} \overline{M}_{zc,i,h} z_{i,s} \epsilon_{i,s+h} \\
\hat{\mathcal{B}}_2 =& \frac{1}{N_j} \sum_{i\in\hat{\mathcal{S}}_j} \sum_t \sum_s k(|t-s|) \overline{d}'_{zx,i} z_{i,t} z'_{i,t} \hat{\Omega}^{-1}_{i,h} \overline{M}_{zc,i,h} z_{i,s} \epsilon_{i,s+h} \\
&- \frac{1}{N_j} \sum_{i\in\hat{\mathcal{S}}_j} \sum_t \sum_s k(|t-s|) \overline{d}'_{zx,i} \hat{\Omega}_{i,h} \overline{d}_{zc,i} \mathcal{O}^{-1}_{zc,i,h} \overline{d}_{zc,i} z_{i,t} z'_{i,t} z_{i,s} \epsilon_{i,s+h} \\
\hat{\mathcal{B}}_3 =& \frac{1}{N_j} \sum_{i\in\hat{\mathcal{S}}_j} \sum_t \sum_s k(|t-s|) \overline{d}'_{zx,i} \hat{\Omega}_{i,h} z_{i,t} c'_{i,t} \hat{\mathcal{O}}^{-1}_{zc,i,h} \overline{d}_{zc,i} \hat{\Omega}_{i,h} z_{i,s} \epsilon_{i,s+h} \\
\hat{\mathcal{B}}_4 =& \frac{1}{N_j} \sum_{i\in\hat{\mathcal{S}}_j} \sum_t \sum_s k(|t-s|) \overline{d}'_{zx,i} \hat{\Omega}_{i,h} \overline{d}_{zc,i} \mathcal{O}^{-1}_{zc,i,h} z_{i,t} c'_{i,t} \hat{\Omega}_{i,h} z_{i,s} \epsilon_{i,s+h}
\end{aligned}
\tag{4}
$$

where $k(|t-s|)$ is some kernel function, e.g., the Bartlett kernel $ker(|t-s|) = 1 - \frac{|t-s|}{H+2}$.

Having obtained estimates of $\mathcal{B}_1, \mathcal{B}_2, \mathcal{B}_3, \mathcal{B}_4$, we can correct the bias by

$$
\tilde{\beta}^{bc}_{j,h} = \tilde{\beta}_{j,h} - \hat{\Sigma}^{-1}_{j,h} \left(\hat{\mathcal{B}}_1 + \hat{\mathcal{B}}_2 + \hat{\mathcal{B}}_3 + \hat{\mathcal{B}}_4 \right) \tag{5}
$$

where $\hat{\Sigma}_{j,h} = \frac{1}{N_j} \sum_{i\in\hat{\mathcal{S}}_j} \overline{d}'_{zx,i} \overline{M}_{zc,i,h} \overline{d}_{zx,i}$.

Another popular way to correct the bias is to use half-panel jackknife (Dhaene and Jochmans, 2015). In particular, we can obtain the bias-corrected estimator as

$$
\tilde{\beta}^{bc1}_{j,h} = 2\tilde{\beta}_{j,h} - (\tilde{\beta}^1_{j,h} + \tilde{\beta}^2_{j,h})/2 \tag{6}
$$

where $\tilde{\beta}_{j,h}$ is the raw estimates, and $\tilde{\beta}^1_{j,h}, \tilde{\beta}^2_{j,h}$ are estimates based on the first half ($1 \le t \le \lfloor T/2 \rfloor$) and the second half ($1 + \lfloor T/2 \rfloor \le t \le T$) respectively.

Such bias correction, however, may not work well in the current setup due to the presence of misclassification. For example, when the group assignment is imprecisely estimated, even when the bias-corrected estimator recovers the "pseudo-true" parameter, it may still not be well centered around the true group parameters.

S1.1 Simulation studies

This section examines the performance of bias correction methods by replicating the simulation study. We consider two inference method. The first one estimate the asymptotic variance of the GLP estimator according to Theorem 2, without taking into account the potential incidental parameter bias. The second approach applies the half-panel jackknife correction (Dhaene and Jochmans, 2015).

Before presenting the result, one may wonder why should researchers even use the uncorrected confidence interval, knowing that in the current setup a $\sqrt{\frac{T}{N}}$ bias is present. However, estimating the bias, be it the analytical form or the more flexible jackknife alternative, can be extremely difficult in when the data has latent group structure. For example, the bias formula provided in (4) requries that we are working with the true group partition.

The results are reported in Table S2 and Table S3. Two patterns are noteworthy. For the infeasible GLP estimator, the bias correction approach produces coverage probabilities close to the nominal level. If we do not correct the bias, the coverage probabilities deteriorates especially at longer horizons; see for comparison Table S1.

In stark contrast, the bias correction method fails to improve the coverage of the GLP confidence intervals. Instead, it sometimes leads to even much lower coverage. This indicates that the incidental parameter bias is poorly estimated, confirming our conjecture that misclassification.

Finally, as expected, the bias correction is useful when the GLP recovers the latent group structure with high accuracy. This points to a case when correcting incidental parameter bias is particularly useful: the time series dimension is large enough for consistent group estimation, while the cross-sectional dimension is even much larger N.

S1.2 Bias-robust inference

When the time series dimension is limited, Theorem 2 may not provide an accurate approximation for the finite sample behavior of the GLP. This section provides a bias correction method that explicitly takes into account the *worst-case bias* in the presence of incidental parameters and possible misclassification.

Recall that by (9) in the main text (page 10), for any given group assignment vector $\hat{\gamma}$ we have

$$\hat{\beta}_{j,h} = \left(\sum_{i \in \hat{S}_j} \overline{d}'_{zx,i} \overline{M}_{zc,i,h} \overline{d}_{zx,i} \right)^{-1} \left(\sum_{i \in \hat{S}_j} \overline{d}'_{zx,i} \overline{M}_{zc,i,h} \overline{d}_{zy,i,h} \right) , \tag{7}$$

For notational convenience, let us denote $A_{j,h} = \frac{1}{N_j} \sum_{i=1}^{N} \mathbf{1} \{ \hat{g}_i = j \} \bar{d}'_{zx,i} M_{zc,i,h} \bar{d}_{zx,i}$. Then we can rewrite the above expression as

$$\begin{aligned}
\hat{\beta}_{j,h} - \beta^0_{j,h} = &\; A^{-1}_{j,h} \left(\tfrac{1}{N_j} \textstyle\sum_{i=1}^{N} \mathbf{1} \{ \hat{g}_i = j \} \bar{d}'_{zx,i} M_{zc,i,h} \bar{d}_{zy,i,h} \right) - \beta^0_{j,h} \\
= &\; A^{-1}_{j,h} \left(\tfrac{1}{N_j} \textstyle\sum_{i=1}^{N} \mathbf{1} \{ \hat{g}_i = j \} \bar{d}'_{zx,i} M_{zc,i,h} \bar{d}_{zx,i} \left[\beta^0_{g^0_i} - \beta^0_{j,h} \right] \right) \\
&+ A^{-1}_{j,h} \left(\tfrac{1}{N_j} \textstyle\sum_{i=1}^{N} \mathbf{1} \{ \hat{g}_i = j \} \bar{d}'_{zx,i} M_{zc,i,h} \bar{d}_{z\epsilon,i,h} \right) \\
= &\; A^{-1}_{j,h} \textstyle\sum_{l=1, l \neq j}^{G} \left(\tfrac{1}{N_j} \textstyle\sum_{i=1}^{N} \mathbf{1} \{ \hat{g}_i = j, g^0_i = l \} \bar{d}'_{zx,i} M_{zc,i,h} \bar{d}_{zx,i} \right) \left[\beta^0_{l,h} - \beta^0_{j,h} \right] \\
&+ A^{-1}_{j,h} \left(\tfrac{1}{N_j} \textstyle\sum_{i=1}^{N} \mathbf{1} \{ \hat{g}_i = j \} \bar{d}'_{zx,i} M_{zc,i,h} \bar{d}_{z\epsilon,i,h} \right)
\end{aligned} \tag{8}$$

As we can see from (8), the bias of the GLP estimator can be decomposed into two parts. For starters, the GLP estimator is subject to misclassification bias. In fact, (8) shows that when $\mathbf{1} \{ \hat{g}_i = j, g^0_i = l \} \neq 0$ (as is the case with small or fixed T) the GLP estimator converges to pseudo-true parameters that are weighted averages of the true ones [1]. As a result, the bias depends not only on the misclassification rates but also the discrepancy between true group parameters $\beta^0_{l,h} - \beta^0_{j,h}$. Without knowledge of the true group parameters, neither the sign nor the magnitude of the bias can be pinned down. This feature makes explicit bias correction for misclassification difficult.

Second, even with perfect knowledge of the group structure, e.g., γ^0 is known, the GLP estimator is subject to the classical incidental parameter bias (Fernández-Val and Lee, 2013), which is analyzed in the previous section.

To address the above concerns, I propose a bias-robust inference method that explicitly takes into account the worst-case bias. The key idea is to replace the true group differences

[1] A similar expression is derived by Bonhomme and Manresa (2015)

by our estimates. Specifically, by add and subtract the GLP estimates, we have

$$\beta^0_{l,h} - \beta^0_{j,h} = \beta^0_{l,h} - \hat{\beta}_{l,h} + \hat{\beta}_{l,h} - \hat{\beta}_{j,h} + \hat{\beta}_{j,h} - \beta^0_{j,h} \ . \tag{9}$$

Denote by $C_{j,l,h} = \frac{1}{N_j}\sum_{i=1}^{N} \mathbf{1}\{\hat{g}_i = j, g^0_i = l\} \bar{d}'_{zx,i} M_{zc,i,h} \bar{d}_{zx,i}$, and the incidental parameter bias by B_0, we can rewrite (8) as

$$\hat{\beta}_{j,h} - \beta^0_{j,h} = B_0 + A^{-1}_{j,h} \sum_{l=1, l\neq j}^{G} C_{j,l,h} \left[\beta^0_{l,h} - \hat{\beta}_{l,h} + \hat{\beta}_{l,h} - \hat{\beta}_{j,h} + \hat{\beta}_{j,h} - \beta^0_{j,h}\right] \ . \tag{10}$$

We can now rearrange the terms to obtain

$$\begin{aligned}
&\begin{bmatrix} I - \sum_{l\neq 1} A^{-1}_{j,h} C_{1,l,h} & A^{-1}_{j,h} C_{1,2,h} & \dots & A^{-1}_{j,h} C_{1,G,h} \\ A^{-1}_{j,h} C_{2,1,h} & I - \sum_{l\neq 2} A^{-1}_{j,h} C_{2,l,h} & \dots & A^{-1}_{j,h} C_{2,G,h} \\ \vdots & \vdots & \ddots & \vdots \\ A^{-1}_{j,h} C_{G,1,h} & A^{-1}_{j,h} C_{G,2,h} & \dots & I - \sum_{l\neq G} A^{-1}_{j,h} C_{G,l,h} \end{bmatrix} \begin{bmatrix} \hat{\beta}_{1,h} - \beta^0_{1,h} \\ \hat{\beta}_{2,h} - \beta^0_{2,h} \\ \dots \\ \hat{\beta}_{G,h} - \beta^0_{G,h} \end{bmatrix} \\
= &\begin{bmatrix} A^{-1}_{j,h} \sum_{l\neq 1}^{G} C_{1,l,h} \left[\hat{\beta}_{l,h} - \hat{\beta}_{1,h}\right] \\ A^{-1}_{j,h} \sum_{l\neq 2}^{G} C_{2,l,h} \left[\hat{\beta}_{l,h} - \hat{\beta}_{2,h}\right] \\ \dots \\ A^{-1}_{j,h} \sum_{l\neq G}^{G} C_{G,l,h} \left[\hat{\beta}_{l,h} - \hat{\beta}_{G,h}\right] \end{bmatrix} + \begin{bmatrix} B_0 \\ B_0 \\ \dots \\ B_0 \end{bmatrix}
\end{aligned} \tag{11}$$

Several comments are in order. First, the above system describes the relation between the parameter estimation errors $\hat{\beta}_{j,h} - \beta^0_{j,h}$, the misclassification errors (reflected in terms $C_{j,l,h}$), and the incidental parameter bias.

Second, $A_{j,h}, C_{j,l,h}, \hat{\beta}_{j,h}$ and B_0 in the above expression are all readily available or estimable. The remaining unknown is the misclassification errors $\mathbf{1}\{\hat{g}_i = j, g^0_i = l\}$.

When the time series dimension is moderately large, we may estimate the misclassification errors by the group assignment probabilities using bootstrap. Specifically, suppose we can generate bootstrapped data $\{(y^{(b)}_{i,t}, z^{(b)}_{i,t}, w^{(b)}_{i,t})\}$. We can apply the GLP estimation to the bootstrapped data, which leads to a $\boldsymbol{\gamma}^{(b)}$. We can then estimate the misclassification errors by

$$P\left(\hat{g}_i = j, g_i^0 = l\right) = \frac{1}{B}\sum_{b=1}^{B} \mathbf{1}\left\{\hat{g}_i^{(b)} = j, g_i^0 = l\right\} \ . \tag{12}$$

S2 Additional simulation evidence

S2.1 Alternative sample sizes

This section provides additional simulation evidence on different sample sizes. Specifically, I repeat the exercise in Section 6.2 with a short T panel ($T = 50$), a large T panel ($T = 300$), and a panel with sample size comparable to the empirical study ($T = 200, N = 400$).

Table S4 reports the results with relatively short T. Three observations are noteworthy. First, the performance of the GLP depends crucially on the true IRs: in Design 1 where the IRs are inseparable in short horizons, the GLP is unable to correctly classify entities; in contrast, in Design 2 where the IRs are separable in short horizons, the GLP works decently well. This observation illustrates clearly the difficulty in grouping IRs, and emphasizes the importance of using horizon-specific weighting matrix.

Second, the performance of the GLP remains stable as we fix T and increase N, which is in line with the findings in Section 6.2. In fact, although the consistency of the GLP depends on T going to infinity, the GLP still *gains* from the increase in the cross-sectional dimension.

Third, we see that the GLP outperforms the individual LP-IV by a wide margin under short T. Specifically, the RMSE of the GLP is, in the worse case, only 56.5% of the individual LP-IV counterpart. The gain comes from a large reduction in the variance, as Column BR suggests.

Next we consider the cases with both N and T large. Comparing Table S5 and Table S4, we see that the performance of the GLP improves substantially as the time series dimension gets large. Moreover, consistent with small sample case, the GLP performance remains stable as we increase the cross-section dimension. In terms of the coverage probabilities of the confidence interval, Table S8 shows that the coverage rates are satisfactorily, which is reassuring for the results in our empirical application.

Given the above patterns, it is unsurprising to see that the GLP performs extremely well when the time series dimension is large. Table S6 reports the results with $T = 300$ and N ranging from 500 to 1500. Even though we have $N \gg T$, the GLP still outperforms the panel LP-IV and the individual LP-IV in all cases. For example, in Design 1 with $T = 300$, $N = 1500$, $G^0 = 3$, the RMSE of the GLP is only 0.238, which is around half of the individual LP-IV counterpart (0.423) and one-third of the panel LP-IV one (0.809). Moreover, the classification accuracy remains well above 95%, corroborating that the time series dimension can grow slower than the cross-sectional dimension.

Importantly, Table S7 and Table S9 show that the GLP coverage probabilities remain satisfactory when N is large. In particular, the GLP coverage rates are comparable with the infeasible counterparts when N is around three times larger than T. However, researchers should be more careful when interpreting the confidence bands when the cross-sections are exceedingly large.

On the whole, the simulation evidence shows that the GLP performs satisfactory in finite samples, even when N is larger than T.

S2.2 Alternative sample sizes: unknown group number

This section reports additional simulation evidence when the number of groups is unknown and selected by the information criterion. The simulation setup is identical to Section 6.1.

As is shown in Table S10, the selected number of groups converge to the true as the sample size increases. Moreover, it generally minimizes the RMSE even when the group number is misspecified. When T is particularly small, however, it tends to under-select and may not minimize the RMSE, e.g. $T = 50, G^0 = 3$ in Design 2. Overall, the patterns are in line with the findings in Section 6.3.

S2.3 Alternative weighting matrix

One of the main merits of the GLP estimator is its flexibility in choosing weighting matrix, which is important in grouping IR estimates. This section studies various choices of the weighting matrix, including:

1. Unit and horizon specific weighting matrix (hereafter, UH). In particular, we set $\hat{\Omega}_{i,h} =$

$\hat{V}_{i,h}^{-1}$ with $\hat{V}_{i,h}^{-1}$ defined in (6).

2. Horizon specific weighting matrix (H). We set $\hat{\Omega}_{i,h} = \hat{\Omega}_h = (\frac{1}{N}\sum_i \hat{V}_{i,h})^{-1}$.

3. Unit specific weighting (U). We set $\hat{\Omega}_{i,h} = \hat{\Omega}_i = (\frac{1}{H}\sum_{h=0}^{H} \hat{V}_{i,h})^{-1}$.

4. 2SLS weighting matrix. We set $\hat{\Omega}_{i,h} = \hat{\Omega}_i^{\text{TSLS}} = \frac{1}{T}\sum_{t=1}^{T} z_{i,t} z_{i,t}'$.

5. IV weighting matrix. We set $\hat{\Omega}_{i,h} = \hat{\Omega}^{\text{IV}} = I_L$

As is clear, the alternative weighting matrices exploit the information in the data to varying degrees. Intuitively, the UH weight is most efficient as it not only downweights uninformative horizons but also downweights uninformative units. However, the efficiency comes at a cost of biasedness. To see this, consider the case without control variables. The GLP with known group partition is:

$$\widetilde{\beta}_{j,h} = \beta_{j,h}^0 + \left[\sum_{i\in\mathcal{S}_g^0} \bar{d}_{zx,i}' \hat{\Omega}_{i,h} \bar{d}_{zx,i}\right]^{-1} \left[\sum_{i\in\mathcal{S}_g^0} \bar{d}_{zx,i}' \hat{\Omega}_{i,h} \bar{d}_{z\epsilon,i,h}\right] . \tag{13}$$

Since $\hat{\Omega}_{i,h}$ is correlated with $\bar{d}_{z\epsilon,i,h}$, the resulting estimates do not converge to the mean effects but some weighted average of the individual effects. This problem is reminiscent of the biasedness of the weighted least squares estimator.

The horizon-specific weights (H) instead only weigh against uninformative horizons.[2] Although it leads to efficiency loss especially in the presence of heteroskedasticity, it is free of the aforementioned bias.

Given that, unit-specific weighting (U) tends to be both biased and inefficient. As is well known in the VAR and local projections literature (Kilian and Lütkepohl, 2017), impluse responses are less precisely estimated for the longer horizons. Therefore, grouping based on unit-specific weights is likely to perform worse.

[2] One may argue for an alternative horizon-specific weight $\hat{\Omega}_h = \hat{V}_{\alpha,h}^{-1}$ where $\hat{V}_{\alpha,h}$ is the estimated asymptotic variance matrix of α in the panel LP-IV model, with standard errors clustered in the unit-level (Cameron and Miller, 2015). However, since the pooled panel LP-IV is by construction inconsistent, this alternative choice may perform worse.

We now discuss the results of different weighting schemes, which are summarized in Table S11-Table S18. First, Table S11 shows that horizon-specific weights lead to the highest classification accuracy. Notice that the gain can be quite large. For example, in Design 2 with $N = 300, T = 100, G^0 = 3$, the H weight accuracy achieves 99.6%, while 2SLS results in only 89.8% accuracy. The unit-and-horizon specific weights are generally the second-best option. The remaining weighting choices generally lead to similar classification accuracy. To sum up, the horizon-specific weights are preferred for better group assignment.

Second, Table S12 shows the GLP outperforms the panel LP-IV and the individual LP-IV regardless of the choices of weighting matrices. Moreover, consistent with the previous results, the horizon-specific weighting yields lowest RMSE, followed closely by the UH weights. Besides, the U weights generally lead to slightly more precise IR estimates than the 2SLS and IV. Overall, the horizon-specific weights is most effective in reducing RMSE.

As for the length of confidence intervals, the UH weights generally have the narrowest bands, as Table S13 shows. The horizon-specific bands are instead the largest, although still only one-fifth of the IND counterparts. Given that some weighting schemes are more biased (e.g. 2SLS and IV leading to misclassification errors), we should expect large differences in the coverage probabilities.

This is confirmed by Table S14-Table S18. For ease of interpretation, I report the coverage probabilities of the infeasible GLP as the benchmark, and differences between the IGLP and the GLP with different weighting schemes. For example, the column for ($N = 100, T = 100, G^0 = 2, h = 0$, Design 1) shows that the IGLP coverage probabilities are 94.9% while the UH coverage rates are on average 14% lower than the IGLP.

On average, the tables show that coverage probabilities of the horizon-specific weights are closest to the infeasible counterparts. Alternative weights can lead to substantial under coverage especially when T is moderate. For example, consider the case with ($N = 300, T = 100, G^0 = 2$ Design 1). The coverage rates for H weights are on average 8.9% lower than the infeasible ones, whereas 2SLS weights are on average 16.1% lower. On the whole, the horizon-specific weighting scheme performs the best in terms of the coverage rates.

In conclusion, the evidence shows that the H weights are preferred. Unless otherwise noted, this will be the default weighting scheme used in this paper.

S2.4 Alternative inference methods

This section compares the large T inference methods (Theorem 2) and small T adjustment (Proposition **??**). We label the resulting outcomes as LT and ST.

Table S19 reports the band ratios of the corresponding confidence bands. As expected, the ST confidence intervals are always larger than the LT ones, although the differences are limited. Moreover, the two confidence bands converge as T increases, which is in line with Proposition **??**. Table S20 and Table S21 further confirm the previous results. As we can see, the coverage rates of the ST inference are on average slightly higher than the LT ones. Overall, the result suggests that ST is a slightly more conservative inference method. Given that, the large T inference is used by default in this paper.

S2.5 Alternative specification: first-differenced

Given our baseline DGP

$$\begin{array}{rcl} y_{i,t} & = & \mu_i + \rho_{g_i} y_{i,t-1} + \delta_{g_i} x_{i,t} + \epsilon_{i,t} \\ x_{i,t} & = & \mu_i + \pi \tilde{z}_{i,t} + u_{i,t} \end{array}, \tag{14}$$

model (19) (in the paper) introduces dynamic panel bias when projecting out unit fixed effects. To address this concern, I estimate the first-differenced model following Anderson and Hsiao (1982). Specifically, denote $\Delta y_{i,t} = y_{i,t} - y_{i,t-1}$, then

$$\Delta y_{i,t} = \rho_{g_i} \Delta y_{i,t-1} + \delta_{g_i} \Delta x_{i,t} + \Delta \epsilon_{i,t}. \tag{15}$$

Notice that $\Delta y_{i,t-1}$ is endogenous as it is correlated with $\epsilon_{i,t-1}$ (and thus $\Delta \epsilon_{i,t}$). However, we can instrument it with $y_{i,t-2}$. As for $\Delta x_{i,t} = \pi \Delta \tilde{z}_{i,t} + \Delta u_{i,t}$, we can instrument it with

$\tilde{z}_{i,t-1}$. To sum up, we estimate the impulse responses through

$$\Delta y_{i,t+h} = \beta_{g_i,h} \Delta x_{i,t} + \phi_{i,h} \Delta y_{i,t-1} + e_{i,t+h} \ , \tag{16}$$

with instrument $(\tilde{z}_{i,t-1}, y_{i,t-2})$.

The results are summarized in Table S22-Table S24. Three comments are in order. First, comparing Table S22 and Table 2, we see that all the conclusions drawn in Section 6.2 hold under this alternative specification: 1) the GLP serves as a good data-driven middle ground between the panel LP-IV and the individual LP-IV; 2) the GLP performs better as T increases.

Second, the RMSE of the baseline fixed effects estimation can be even smaller than the first-differenced alternative, because of the reduction in variance. As the Column BR suggests, the first-differenced estimator is, in the current setup, slightly less efficient than the fixed effects estimator and thus has larger confidence bands.

Third, Table S23 shows that the GLP coverage rates are overall more conservative under the first-differenced specification. Similarly, the coverage rates of the infeasible GLP also improves in the first-differenced setup especially for longer horizons (see Table S1 and Table S24). This is mainly because of the dynamic panel bias in the baseline fixed-effects setup.

Overall, the simulation evidence confirms the reliable performance of the GLP algorithm. Moreover, the first-differenced specification reduces the dynamic panel bias, while being slightly less efficient than the baseline fixed effects case.

S2.6 Alternative objective function

As is mentioned in the paper, the baseline GLP estimator defined in (4) is by construction different from the conventional panel GMM estimator. For illustrative purposes, let us compare the two objective functions given the true group partition assuming that there are no controls (and fixed effects) and we use identical weighting matrix $\Omega_{i,h} = I_L$. The conventional GMM criterion is

$$Q^{pool}(\beta) = \sum_{h=0}^{H} \left[\frac{1}{NT} \sum_{i=1}^{N} \sum_{t=1}^{T} z_{i,t}(y_{i,t+h} - x_{i,t}\beta_{g_i,h}) \right]' \left[\frac{1}{NT} \sum_{i=1}^{N} \sum_{t=1}^{T} z_{i,t}(y_{i,t+h} - x_{i,t}\beta_{g_i,h}) \right]$$

which gives

$$\check{\beta}_{j,h} = \beta^0_{j,h} + \left[\left(\sum_{i \in \mathcal{S}^0_j} X_i' Z_i \right) \left(\sum_{i \in \mathcal{S}^0_j} Z_i' X_i \right) \right]^{-1} \left[\left(\sum_{i \in \mathcal{S}^0_j} X_i' Z_i \right) \left(\sum_{i \in \mathcal{S}^0_j} Z_i' \mathcal{E}_{i,h} \right) \right] , \tag{17}$$

where $\beta^0_{j,h}$ is the true parameter, and $X_i = (x_{i,1}, \ldots, x_{i,T})'$ and $Z_i = (z_{i,1}, \ldots, z_{i,T})'$ are defined as in the paper. Under conventional dependence and moment assumptions, we have

$$\frac{1}{N_j T} \sum_{i \in \mathcal{S}^0_j} \sum_{t=1}^{T} z_{i,t} x_{i,t} \xrightarrow{p} \Sigma_j , \quad \frac{1}{\sqrt{N_j T}} \sum_{i \in \mathcal{S}^0_j} \sum_{t=1}^{T} z_{i,t} \epsilon_{i,t+h} \xrightarrow{d} N(0, V_{j,h}) ,$$

and thus

$$\sqrt{N_j T}(\check{\beta}_{j,h} - \beta^0_{j,h}) \sim N(0, (\Sigma_j' \Sigma_j)^{-1} \Sigma_j' V_{j,h} \Sigma_j (\Sigma_j' \Sigma_j)^{-1}) . \tag{18}$$

Instead, my objective function gives:

$$\hat{\beta}_{j,h} = \beta^0_{j,h} + \left[\sum_{i \in \mathcal{S}^0_j} (X_i' Z_i)(Z_i' X_i) \right]^{-1} \left[\sum_{i \in \mathcal{S}^0_j} (X_i' Z_i)(Z_i' \mathcal{E}_{i,h}) \right] \tag{19}$$

leading to

$$\sqrt{N_j T}(\hat{\beta}_{j,h} - \beta^0_{j,h}) = \left[\frac{1}{N_j} \sum_{i \in \mathcal{S}^0_j} \frac{\sum_{s=1}^{T} x_{i,t} z_{i,t}'}{T} \frac{\sum_{s=1}^{T} z_{i,t} x_{i,t}'}{T} \right]^{-1} \left[\frac{1}{\sqrt{N_j}} \sum_{i \in \mathcal{S}^0_j} \frac{\sum_{t=1}^{T} x_{i,t} z_{i,t}'}{T} \frac{\sum_{t=1}^{T} z_{i,t} \epsilon_{i,t+h}}{\sqrt{T}} \right] .$$

The first term converges to some positive definite matrix $\tilde{\Sigma}$ by standard assumptions, e.g. Assumptions 1 and 3. As for the second term, we can decompose it by

$$\frac{1}{\sqrt{N_j}} \sum_{i \in \mathcal{S}^0_j} \frac{\sum_{t=1}^{T} x_{i,t} z_{i,t}'}{T} \frac{\sum_{t=1}^{T} z_{i,t} \epsilon_{i,t+h}}{\sqrt{T}}$$

$$
\begin{aligned}
=&\frac{1}{T\sqrt{N_jT}}\sum_{i\in\mathcal{S}_j^0}\sum_{s=1}^{T}\sum_{t=1}^{T}\mathbb{E}(x_{i,s}z_{i,s}')z_{it}\epsilon_{i,t+h}\\
&+\frac{1}{T\sqrt{N_jT}}\sum_{i\in\mathcal{S}_j^0}\sum_{s=1}^{T}\sum_{t=1}^{T}\left[x_{i,s}z_{i,s}'-\mathbb{E}(x_{i,s}z_{i,s}')\right]z_{it}\epsilon_{i,t+h}\\
=&\frac{1}{\sqrt{N_jT}}\sum_{i\in\mathcal{S}_j^0}\sum_{t=1}^{T}\mathbb{E}(x_{i,s}z_{i,s}')z_{it}\epsilon_{i,t+h}+\frac{1}{T\sqrt{N_jT}}\sum_{i\in\mathcal{S}_j^0}\sum_{s=1}^{T}\sum_{t=1}^{T}\mathbb{E}[x_{i,s}z_{i,s}'z_{i,t}\epsilon_{i,t+h}]\\
&+\frac{1}{T\sqrt{N_jT}}\sum_{i\in\mathcal{S}_j^0}\sum_{s=1}^{T}\sum_{t=1}^{T}\left[\left(x_{i,s}z_{i,s}'-\mathbb{E}(x_{i,s}z_{i,s}')\right)z_{it}\epsilon_{i,t+h}-\mathbb{E}[x_{i,s}z_{i,s}'z_{i,t}\epsilon_{i,t+h}]\right]\ . \qquad (20)
\end{aligned}
$$

Now the first element is asymptotically normal, whereas the second term contains the bias from IV estimation, and the third term is asymptotically negligible under Assumption 1.B and Lemma 9.

In particular, assume that we have the first stage:

$$
x_{i,t}=\Pi z_{i,t}+u_{i,t} \qquad (21)
$$

where Π is a $K\times L$ matrix governing the instrument strength. Then we can rewrite the second term as

$$
\begin{aligned}
&\frac{1}{T\sqrt{N_jT}}\sum_{i\in\mathcal{S}_j^0}\sum_{s=1}^{T}\sum_{t=1}^{T}\mathbb{E}[x_{i,s}z_{i,s}'z_{i,t}\epsilon_{i,t+h}]\\
=&\frac{1}{T\sqrt{N_jT}}\sum_{i\in\mathcal{S}_j^0}\sum_{s=1}^{T}\sum_{t=1}^{T}\mathbb{E}[(\Pi z_{i,s}+u_{i,s})z_{i,s}'z_{i,t}\epsilon_{i,t+h}]\\
=&\frac{1}{T\sqrt{N_jT}}\sum_{i\in\mathcal{S}_j^0}\sum_{s=1}^{T}\sum_{t=1}^{T}\mathbb{E}[\Pi z_{i,s}z_{i,s}'z_{i,t}\epsilon_{i,t+h}]+\frac{1}{T\sqrt{N_jT}}\sum_{i\in\mathcal{S}_j^0}\sum_{s=1}^{T}\sum_{t=1}^{T}\mathbb{E}[u_{i,s}z_{i,s}'z_{i,t}\epsilon_{i,t+h}]\\
=&\frac{1}{T\sqrt{N_jT}}\sum_{i\in\mathcal{S}_j^0}\mathbb{E}[U_i'Z_iZ_i'\mathcal{E}_{i,h}]=\sqrt{\frac{N_j}{T}}\frac{1}{T}\mathbb{E}[U_i'Z_iZ_i'\mathcal{E}_{i,h}]\ , \qquad (22)
\end{aligned}
$$

where $U_i=(u_{i,1},\ldots,u_{i,T})'$. Two comments are in order. First, it is clear that the bias arises from the correlation between $u_{i,t}$ and $\epsilon_{i,t+h}$, which is well documented in the IV literature

(Nagar, 1959). Second, the magnitude of the bias is determined by the convergence rate of T relative to N. By Assumption 3.B, within each group we have $N/T \to 0$ and thus the bias is negligible, which leads to

$$\frac{1}{\sqrt{N_j}} \sum_{i \in \mathcal{S}_j^0} \frac{\sum_{t=1}^T x_{i,t} z_{i,t}'}{T} \frac{\sum_{t=1}^T z_{i,t} \epsilon_{i,t+h}}{\sqrt{T}} \xrightarrow{d} N(0, \tilde{V}) \tag{23}$$

and

$$\sqrt{N_j T}(\hat{\beta}_{j,h} - \beta_{j,h}^0) \xrightarrow{d} N(0, \tilde{\Sigma}^{-1} \tilde{V} \tilde{\Sigma}^{-1}) \ . \tag{24}$$

The difference between our estimator and the conventional one is twofold. First, our estimator is generally less efficient because it does not pool over both i and t. Second, our estimator can be biased when T is small relative to N_j. In our setup, we assume that T is large by Assumption 3.B so that the bias is negligible. Moreover, the bias is small enough in macroeconomic applications with large T. In this sense, the difference between the proposed GMM estimator and the conventional one is small.

To quantify the difference, I study the finite sample performance of the infeasible GLP under these two objective functions. In particular, I simulate the data from model (18) and estimate (19) using the true group partitions.

Table S25 reports the results. Three observations stand out. First, the fully-pooled criterion generally leads to more precise estimates with lower RMSE. Second, the baseline GLP criterion yields slightly narrower confidence bands. Third, however, the differences between the criteria remain sufficiently small through all sample sizes, with the biggest RMSE differential being only 0.02. Moreover, Table S27 and Table S26 report the coverage rates. Again, the coverage rates are slightly higher for the fully-pooled estimator, although the average difference in coverage probabilities is merely 1.3%.

Overall, finite sample analysis shows that the fully-pooled estimator in general performs better, although the marginal gains are quantitatively small.

S2.7 Horizon-by-horizon grouping

Various methods have been proposed to group coefficients (Su et al., 2016; Wang et al., 2018; Lewis et al., 2019). However, these approaches fail to recognize the fact that (a) IRs may overlap at some horizons, e.g. starting from close to zero in $h = 0$, and (b) IRs are very noisy for longer horizons. Therefore, it may be difficult to cluster impulse responses horizon-by-horizon (hereafter HBH) both in population (due to the violation of Assumption 2.B) and in samples (due to large estimation variance). In this section, I repeat the exercise in Section 6.2 but estimate it under HBH.

Table S30 and Table S31 present the results for classification accuracy, confidence bands ratios and RMSE. We discuss the three metrics in turn. First, grouping IRs horizon-by-horizon results in extremely imprecise group estimates. Second, the band ratios of the HBH remain largely the same as our baseline GLP. Combined with the low classification accuracy, we would expect that the confidence intervals of the HBH seriously undercover the true IRs. Finally, the RMSE of the HBH are inflated due to the misclassification biases. Overall, the baseline GLP performs much better than the HBH. Even in the worse case with ($N = 100, T = 100, G^0 = 3$, Design 1), our baseline GLP increases the accuracy by 18.3% and reduces the RMSE by 25.7%.

Given the above results, it is not surprising that the confidence intervals of the HBH results in much lower coverage rates. Table S32 reports the differences in coverage rates between HBH and our baseline. On average the coverage probabilities of the HBH are 56.2% lower than the baseline.

Interestingly, the results for Design 2 and horizon $h = 0$ clearly shows the condition under which the HBH can perform well. Recall from Table 1 that in this case the IRs are both informative and separable. Therefore, the cost of ignoring information from nearby horizons is relatively small. However, when moving on to horizon $h = 1$, the HBH fails to correctly identify groups as IRs are noiser. In stark contrast, the baseline GLP effectively weights across horizons.

In conclusion, the evidence shows that grouping IRs horizon-by-horizon would lead to imprecise group estimates and IR estimates. More importantly, the confidence intervals would be exceedingly short. As a consequence, existing methods in grouping coefficients are unlikely to correctly group impulse responses.

S3 Empirical application

S3.1 Data description

- *Housing prices.* The benchmark model is based on monthly MSA level house price index from Freddie Mac, which includes $N = 382$ MSAs from January 1975 to March 2019. The price index is a seasonally adjusted repeated-sales index, normalized to be 100 in December 2000. The level index is transformed into monthly growth rates by taking log differences, and is expressed in percentage terms.

- *Macroeconomic variables.* The macro variables used in benchmark model are monthly series of Fed Funds Rates (FFR), industrial production (IP), consumption expenditure (PCEPI), and real estate loans at all commercial banks (REALLN), all of them from FRED database, and 30-year fixed-rate for mortgage products (FRM30), collected from Freddie Mac. The weekly FRM30 series is averaged to monthly data. IP, PCEPI and REALLN are transformed into monthly growth rates in the same way as house prices.

- *External instruments.* The benchmark instrument for the monetary shocks over the period 1991:1 to 2009:12 is the informationally robust instrument in **?**.

- *MSA economic profiles.* MSA-level economic data, including per capita personal income (dollars), population (thousands of person), and total employment (thousands of jobs) are 2017 figures obtained from the MAINC30 from the US Bureau of Economic Analysis (BEA). The real GDP per capita in 2017 is obtained from MAGDP10. The and house price elasticity is provided by Saiz (2010). Household debt-to-income ratio is from Ahn et al. (2018). Precise ratio is not accessible. Instead, we obtain the high and low range of the ratio.

S3.2 Determining the group number of housing responses

To start with, Figure S1 shows that the information criterion is minimized with two groups ($\hat{G} = 2$). However, as is discussed in Section 6.3, the information criterion may under-select. Therefore, I consider also the distinctiveness of the estimated IRs to help select the group number.

Comparing Figure 3 and Figure S2, specifying $\hat{G} = 3$ separates Group 2 (mild negative responses) into groups with muted responses and negative responses, while keeping the positive responses largely unchanged. If we further increase the group number to four, however, Figure S3 shows that the IRs become less distinguishable, e.g. between Group 2 and Group 3.

Therefore, we choose three groups as our baseline case. Importantly, the group patterns discussed in Section 7.2 hold under two or even four groups.

S3.3 Determinants of the group structure of housing responses

To formally examine the relation between economic factors and group membership, I estimate a multinomial logit model on the estimated group assignment.

Table S33 presents the results with three groups ($\hat{G} = 3$). First, the signs of the estimated coefficients are fairly stable across all models, and they are consistent with the analysis in Section 7.2. Specifically, the first two columns suggest that the probability of belonging to Group 3 relative to Group 1 is significantly lower as real GDP or employment increase. In words, Group 3 are generally poorer and have less employment.

Second, columns (3)-(5) show that an MSA is more likely to be classified into Group 1 when the debt-to-income ratio is higher and when housing markets are less elastic. Moreover, the relation holds even when income levels are controlled for.

Finally, Table S34 shows that when we increase the group number to four ($\hat{G} = 4$) the group partition exhibits a similar pattern as described above.

On the whole, the evidence supports the qualitative analysis in Section 7.2. That is, rich, populated and indebted regions (Group 1) are likely to have house price appreciation after a tightening monetary shock. Moreover, reducing household debt-to-income ratio significantly increases the probability of having mild price responses (Group 2), whereas reducing the income level strongly increases the probability of large price depreciation.

S3.4 Alternative specifications

In this section I discuss the results excluding the lagged dependent variables as controls. For ease of notation, I call the benchmark model as FE and this alternative specification is labeled as FE'.

To start with, Figure S6-Figure S8 show that the estimated IRs are similar to the benchmark, albeit in smaller magnitudes. Moreover, Table S35 reports the economic profiles under FE'. As we can see, the group pattern under this alternative specification captures the same economic forces. Specifically, Group 1 with positive IRs is: 1) richer and more economically developed; 2) more populated; 3) more regulated and inelastic in the housing markets; and 4) more indebted (compared to Group 2).

As for the number of groups, Figure S5 shows that the information criterion is again minimized at $\hat{G} = 2$. As is discussed in Section S3.2, the group patterns hold under difference choices of the number of groups: MSAs with positive IRs are generally more economically developed, housing markets are more regulated, and debt levels are higher.

Importantly, the group partition still cannot be recovered by simple external criteria. For example, Figure S11 replicates Figure 4 under FE'. As we can see, the poorest MSAs in Group 1 are unambiguously distinct from the poor MSAs in Group 3. Besides, the richest 10% MSAs, though now resembles the responses of Group 2, have much moderate responses compared with Group 1. Hence, income level fails to recover the documented group pattern.

As a whole, the analysis shows that the main results in Section 7.2 hold under this alternative specification.

S3.5 Horizon-by-horizon grouping

We now examine the stability of the estimated group structure by estimating latent groups *horizon-by-horizon* (hereafter HBH). Four observations stand out.

First, as we can see from Figure S12, the HBH generates Group IRs that are qualitatively similar to our benchmark case (hereafter BASE). Specifically, it separates 1) Group 1 with positive IRs, 2) Group 2 with muted responses, and 3) Group 3 with negative IRs. However, now the IR estimates are highly jagged, as they are contaminated by horizon-specific noises (Barnichon and Brownlees, 2019). In fact, when comparing the estimates, Group 3 (HBH) has IRs around twice as large as Group 3 (BASE).

Second, compared to BASE, HBH yields much narrower confidence bands. Notice however, these bands generally have very low coverage rates as is illustrated in Section S3.5. Consider for example Charleston city of West Virginia. Figure S13 reports the impulse responses given by the individual LP-IV (IND), HBH and BASE respectively. Although the IND IRs tend to be positive before turning into negative for longer horizons, they are in most cases insignificant. The HBH keeps the sign and magnitude of the individual IRs largely unchanged, but substantially reduces the confidence bands —by grouping nearby MSAs. Moreover, there are several "breaks" in the HBH IRs that are hard to justify in economic theories, and is likely to result from misclassification.

In contrast, the BASE serves as a middle ground between the IND and HBH: First, it preserves the shape of the IRs. Second, it effectively takes into account the large confidence bands of the IND. As a consequence, the the IRs are not only more moderate but also free of "breaks".

Third, Figure S14-S18 report the group assignments for horizons $h = 1, 6, 12, 18, 24$. As we can see, the group structure is highly unstable across horizons; see for example the clusters in Florida. In fact, 80.4% of the MSAs change their group membership across horizons, among which 27 MSAs switch between Group 1 and Group 3, leading to breaks in IR estimates. The reason is that (a) IRs are hardly distinguishable around $h = 0$, and (b) IRs in the longer horizons are noisy, as illustrated in Figure 2, and grouping based on those horizons alone can be imprecise.

Finally, Table S36 shows that group patterns of our baseline results generally hold under this alternative specification. For example, Group 1 with positive IRs are generally more developed. However, the documented patterns are also unstable across horizons. For example, at horizon $h = 1$ there is a U-shaped relation between population and IRs, which converts into a monotonically decreasing relation for longer horizons. Similar patterns can be found for housing supply elasticity and debt-to-income ratios. The unstable relations between economic characteristics and the impulse responses are likely to result from the reasons discussed above.

Overall, the results demonstrate that estimating group IRs horizon-by-horizon can replicate some patterns of our baseline method, but the estimates are subject to noises and thus unstable. Moreover, modeling group structure horizon-by-horizon yields unusually small estimated standard errors. In light of the simulation evidence in Section S3.5, researchers should be cautious in using and interpreting horizon-by-horizon grouping results.

S4 Proof of Theorems and Propositions

It is useful to present here some (in)equalities:

$$|a'b| \leq \|a\| * \|b\|, \qquad \text{for } a, b \in \mathbb{R}^n , \tag{25}$$

$$\|Ab\| \leq \|A\| * \|b\|, \qquad \text{for } A \in \mathbb{R}^{m\times n},\ b \in \mathbb{R}^n , \tag{26}$$

$$|a'Ab| \leq \|a\| * \|A\| * \|b\|, \qquad \text{for } a \in \mathbb{R}^m,\ A \in \mathbb{R}^{m\times n},\ b \in \mathbb{R}^n , \tag{27}$$

$$\|AB\| \leq \|A\| * \|B\|, \qquad \text{for } A \in \mathbb{R}^{m\times n},\ B \in \mathbb{R}^{n\times r} , \tag{28}$$

where $\|a\|^2 = \text{tr}(aa') = \sum_{i=1}^n a_i^2$ for $a = (a_1, \ldots, a_n)'$, and $\|A\|^2 = \text{tr}(AA') = \sum_{i=1}^n \sum_{j=1}^n a_{i,j}^2$, for $A = [a_{i,j}]_{i,j=1,\ldots,n}$. Moreover, for any conformable matrix A and vectors a, b, we have

$$a'Aa - b'Ab = (a-b)'A(a-b) + 2b'A(a-b) \tag{29}$$

All the inequalities can be easily derived using the Cauchy-Schwarz inequality.

Notation. For notation convenience, summations are taken over all possible values unless otherwise stated, e.g. $\sum_t = \sum_{t=1}^T$. Moreover, consistent with the main text, we denote the following moments:

$$\begin{aligned}
\bar{d}_{zx,i} &= \tfrac{1}{T}\textstyle\sum_{t=1}^T z_{i,t}x'_{i,t}, & d_{zx,i} &= \mathbb{E}[z_{i,t}x'_{i,t}] \\
\bar{d}_{zc,i} &= \tfrac{1}{T}\textstyle\sum_{t=1}^T z_{i,t}c'_{i,t}, & d_{zc,i} &= \mathbb{E}[z_{i,t}c'_{i,t}] \\
\bar{d}_{zy,i,h} &= \tfrac{1}{T}\textstyle\sum_{t=1}^T z_{i,t}y_{i,t+h}, & & \\
\bar{d}_{z\epsilon,i,h} &= \tfrac{1}{T}\textstyle\sum_{t=1}^T z_{i,t}\epsilon'_{i,t+h}, & d_{z\epsilon,i,h} &= \mathbb{E}[z_{i,t}\epsilon_{i,t+h}] \\
\overline{M}_{zc,i,h} &= \hat{\Omega}_{i,h} - \hat{\Omega}_{i,h}\bar{d}_{zc,i}(\bar{d}'_{zc,i}\hat{\Omega}_{i,h}\bar{d}_{zc,i})^{-1}\bar{d}'_{zc,i}\hat{\Omega}_{i,h} & & \\
M_{zc,i,h} &= \Omega_{i,h} - \Omega_{i,h}d_{zc,i}(d'_{zc,i}\Omega_{i,h}d_{zc,i})^{-1}d'_{zc,i}\Omega_{i,h} & &
\end{aligned} \tag{30}$$

Then we can rewrite the individual objective function (equation (5) in the main text) as

$$\hat{Q}_{iTh}(\beta_{g_i,h}, \phi_{i,h}) = \hat{m}'_{i,h}\hat{\Omega}_{i,h}\hat{m}_{i,h}, \quad \hat{m}_{i,h} = \bar{d}_{zy,i,h} - \bar{d}_{zx,i}\beta_{g_i,h} - \bar{d}_{zc,i}\phi_{i,h} . \tag{31}$$

The GLP estimator given some generic grouping γ is then

$$\hat{\beta}_{j,h}(\gamma) = \left(\sum_{i\in S_j} \overline{d}'_{zx,i}\overline{M}_{zc,i,h}\overline{d}_{zx,i}\right)^{-1}\left(\sum_{i\in S_j} \overline{d}'_{zx,i}\overline{M}_{zc,i,h}\overline{d}_{zy,i,h}\right) \tag{32}$$

and for all $i = 1, \dots, N$, the individual level parameter given β_{g_i}

$$\hat{\phi}_{i,h}(\beta_{g_i}) = \left(\overline{d}'_{zc,i}\hat{\Omega}_{i,h}\overline{d}_{zc,i}\right)^{-1}\overline{d}'_{zc,i}\hat{\Omega}_{i,h}\left(\overline{d}_{zy,i,h} - \overline{d}_{zx,i}\beta_{g_i,h}\right) \ . \tag{33}$$

Finally, we define an auxiliary objective function as below.

Definition 1. The auxiliary objective function is defined as

$$Q_{NT}(\boldsymbol{\beta}, \boldsymbol{\phi}, \gamma) = \frac{1}{N}\sum_i \sum_h Q_{iTh}(\beta_{g_i,h}, \phi_{i,h}) \tag{34}$$

with $Q_{iTh}(\beta_{g_i,h}, \phi_{i,h}) = m'_{i,h}\Omega_{i,h}m_{i,h}$ and

$$\begin{aligned} m_{i,h} = & \ \mathbb{E}[\hat{m}_{i,h}] = \mathbb{E}\left[z_{it}x'_{it}\left(\beta^0_{g_i^0,h} - \beta_{g_i,h}\right) + z_{it}c'_{it}\left(\phi^0_{i,h} - \phi\right) + z_{it}\epsilon_{it+h}\right] \\ = & \ d_{zx,i}\left(\beta^0_{g_i^0,h} - \beta_{g_i,h}\right) + d_{zc,i}\left(\phi^0_{i,h} - \phi_{i,h}\right) \end{aligned} \tag{35}$$

Definition 2. For two collections of parameters $\tilde{\boldsymbol{\beta}} = (\tilde{\beta}_1, \dots, \tilde{\beta}_{G_1})$ and $\boldsymbol{\beta} = (\beta_1, \dots, \beta_{G_2})$ with possibly different G_1 and G_2, define the Hausdorff distance between them as

$$d_H(\boldsymbol{\beta}, \tilde{\boldsymbol{\beta}}) = \max\left\{\max_{g\in\{1,\dots,G_2\}}\min_{\tilde{g}\in\{1,\dots,G_1\}}\|\tilde{\beta}_{\tilde{g}} - \beta_g\|, \max_{\tilde{g}\in\{1,\dots,G_1\}}\min_{g\in\{1,\dots,G_2\}}\|\tilde{\beta}_{\tilde{g}} - \beta_g\|\right\} \ . \tag{36}$$

Definition 3. We work with the following the neighborhood of $\beta^0_{g,h}$:

$$\mathcal{N}_\eta = \left\{\boldsymbol{\beta} \in \Theta_G \colon d_H\left(\hat{\boldsymbol{\beta}}, \boldsymbol{\beta}^0\right) \leq \eta\right\} \ . \tag{37}$$

S4.1 List of all lemmas

I list all the relevant lemmas below.

Lemma 1. Under Assumptions 1.B-1.E, for any $\nu > 0$, we have for any $\delta > 0$

$$\begin{aligned} P\left(\sup_i \left\|\frac{1}{T}\sum_{t=1}^{T} z_{i,t}w'_{i,t} - \mathbb{E}[z_{i,t}w'_{i,t}]\right\| > \nu\right) &= o(NT^{-\delta}) \\ P\left(\sup_i \left\|\frac{1}{T}\sum_{t=1}^{T} z_{i,t}\epsilon_{i,t+h}\right\| > \nu\right) &= o(NT^{-\delta}) \end{aligned} . \tag{38}$$

Lemma 2. Under Assumption 1, we have

$$\begin{aligned} &\sup_i \sup_{\beta\in\Theta,\ \phi\in\Phi} \left|\hat{Q}_{iTh}(\beta,\phi) - Q_{iTh}(\beta,\phi)\right| = o_p(NT^{-\delta}), \quad \text{for all } h \in 0,\dots,H \\ &\sup_{\boldsymbol{\beta}\in\Theta_G,\boldsymbol{\phi}\in\Phi_N,\gamma\in\mathcal{G}} \left|\hat{Q}_{NT}(\boldsymbol{\beta},\boldsymbol{\phi},\gamma) - Q_{NT}(\boldsymbol{\beta},\boldsymbol{\phi},\gamma)\right| = o_p(NT^{-\delta}) \end{aligned} \tag{39}$$

Lemma 3. Under Assumptions 1 and 2, and assume that the number of groups G^0 is given, then the Hausdorff distance between the estimated group IRs and the true converges to zero as both N and T go to infinity, i.e.

$$d_H(\hat{\boldsymbol{\beta}},\boldsymbol{\beta}^0) \xrightarrow{p} 0 \ . \tag{40}$$

Lemma 4. Under Assumption 1, and for β_{g_i} such that $\max_i \|\beta_{g_i} - \beta_{g_i^0}\| = o_p(1)$, we have

$$\max_i \|\hat{\phi}_{i,h}(\beta_{g_i}) - \phi^0_{i,h}\| = o_p(1) \ . \tag{41}$$

Lemma 5. Under Assumptions 1-2, with correctly specified number of groups G^0. Then for $\eta > 0$ small enough, we have

$$\sup_{\boldsymbol{\beta}\in\mathcal{N}_\eta} \sup_i \mathbf{1}\{\hat{g}_i(\boldsymbol{\beta}) \neq g_i^0\} = o_p(NT^{-\delta}) \ . \tag{42}$$

Lemma 6. Under Assumptions 1, we have

$$\begin{aligned} \sup_i \bar{d}'_{zx,i}\hat{\Omega}_{i,h}\bar{d}'_{zx,i} &= \sup_i d_{zx,i}\Omega_{i,h}d_{zx,i} + o_p(1) \\ \sup_i \bar{d}'_{zc,i}\hat{\Omega}_{i,h}\bar{d}'_{zc,i} &= \sup_i d_{zx,i}\Omega_{i,h}d_{zc,i} + o_p(1) \\ \sup_i \bar{d}'_{zx,i}\hat{\Omega}_{i,h}\bar{d}'_{zc,i} &= \sup_i d_{zx,i}\Omega_{i,h}d_{zc,i} + o_p(1) \end{aligned} . \tag{43}$$

Lemma 7. [Equation (1.7) in Merlevède et al. (2011)] Let $\{X_t, t \geq 1\}$ be a sequence of strongly mixing real-valued and centered random variables. Assume that for some positive constants, c_1, c_2, c_3, c_4

(i) the mixing coefficient satisfies $\alpha(\tau) \leq \exp(-c_1 \tau^{c_2})$;

(ii) tail probabilities $\sup_t \mathbb{P}(|X_t| > \nu) \leq \exp(1 - (\nu/c_3)^{c_4})$.

Then let $c = \frac{c_2 c_4}{c_2 + c_4}$, there exists a positive constant C such that for any $\lambda > 0$ and $r \geq 1$,

$$\mathbb{P}\left(\sup_{1\leq s\leq T} \Big|\sum_{t=1}^{s} X_t\Big| \geq 4\lambda\right) \leq 4\left(1 + \frac{\lambda^2}{rTM}\right)^{-r/2} + 4CT\lambda^{-1}\exp\left[-c_1\frac{\lambda^c}{c_3^c r^c}\right] \tag{44}$$

where $M = \sup_{t>0}\left(\mathbb{E}(X_t^2) + 2\sum_{s>t}|\mathbb{E}(X_t X_s)|\right)$.

Lemma 8. [Corollary A.2 in Hall and Heyde (1980)] Suppose X and Y are two random variables which are $\mathcal{F}$- and $\mathcal{H}$-measurable. If $E\|X\|^p < \infty$ and $E\|Y\|^q < \infty$ where $p, q > 1$ and $p^{-1} + q^{-1} < 1$. We have

$$Cov(X, Y) \leq 8\|X\|_p \|Y\|_q \left(\alpha(\mathcal{G}, \mathcal{H})\right)^{1-p^{-1}-q^{-1}} \tag{45}$$

Lemma 9. [Lemma A.2 in Gao (2007)] Let $f(\cdot,\cdot)$ be a symmetric Borel function defined on $\mathbb{R}^L \times \mathbb{R}^L$. Let the process ξ_t be an L-dimensional strictly stationary and α-mixing process. Assume that for any fixed $r \in \mathbb{R}^L$, $\mathbb{E}[f(\xi_1, r)] = 0$ and $\mathbb{E}|f(\xi_t, \xi_s)|^{2(1+\delta)} < \infty$. Then

$$\mathbb{E}\left[\sum_{t=1}^{T}\sum_{s=1}^{T} f(\xi_t, \xi_s)\right]^2 \leq O(T^2)\ . \tag{46}$$

Remark S1. Lemma 1 is analogous to Lemma B.5 in Bonhomme and Manresa (2015). Since the uniform convergence result holds for any $\delta > 0$, they work under the assumption that there exists some positive number $a > 0$ such that $N/T^a \to 0$ and set $\delta = a$. In the large N, T asymptotics we consider, i.e., $\frac{N}{T} \to \kappa \in [0, \infty)$, uniform convergence still holds when we choose $\delta > 1$. Although the choice of δ is arbitrary, larger values of δ necessarily requires a larger sample size T to ensure finite sample performance.

Remark S2. Lemma 2 follows closely Lemma 3 in Fernández-Val and Lee (2013), which is crucial in establishing the consistency of β in Lemma 3 and of $\phi_{i,h}$ in Lemma 4. The key step is (124), which shows that the uniform convergence rate is determined by $\sup_i \|\overline{d}_{zw,i} - d_{zw,i}\|$, $\sup_i \|\overline{d}_{z\epsilon,i,h}\|$ and $\|\hat{\Omega}_{i,h} - \Omega_{i,h}\|$. Alternative assumptions can be made. For example, Assumption B1(iv) of Su et al. (2016) assumes that

$$P\left(\sup_i \|\hat{\Omega}_{i,h} - \Omega_{i,h}\| \geq \eta\right) = o(N^{-1}) \tag{47}$$

which holds, if we use 2SLS weighting matrix $\hat{\Omega}_{i,h} = \frac{1}{T} z_{i,t} z'_{i,t}$, and apply Lemma 1 with $\delta > 2$.

Remark S3. Notice that Lemma 3 is derived under the assumption that the true number of groups is known. To prove Proposition 1, we would like to study the behavior of the GLP estimator with $G > G^0$ (the behavior when underfitting $G < G^0$ is assumed in the high-level assumption 4.A). To this end, in Lemma **??** we re-establish Lemmas 3 with $G > G^0$.

Compared with the correct specification case, the difficulty arises from the fact that the Hausdorff distance in Lemma 3 cannot be easily established. To see this, consider the extreme case when we set $G = N$, then the model reduces to unit-level time series regression, which slows down the convergence rate. Moreover, the mapping argument in Lemma 3 is not applicable. Instead, we prove Lemma **??** which states that the estimation error of β_{g_i} is uniformly bounded.

Remark S4. Lemmas 8-9 are a set of results for strongly mixing processes, and the proof is omitted.

S4.2 Theorem 1

Proof. The theorem contains two parts. The first part shows that the misclassification probability converges to zero. The second part shows that the GLP estimator is asymptotically equivalent to the infeasible estimator under true grouping.

Part I: consistent group estimation. The last piece of Theorem 1 is the convergence of

$\hat{g}$ to the true group assignment. Notice that

$$\begin{aligned}
&\mathbb{P}\left(\sup_{i\in\{1,\dots,N\}} \mid \hat{g}(\hat{\beta}) - g_i^0 \mid > 0\right) \\
=&\mathbb{P}\left(\sup_{i\in\{1,\dots,N\}} \mid \hat{g}(\beta) - g_i^0 \mid > 0\right)\left(\mathbb{P}\left(\boldsymbol{\beta}\in\mathcal{N}_\eta\right) + \mathbb{P}\left(\boldsymbol{\beta}\notin\mathcal{N}_\eta\right)\right) \\
\leq&\mathbb{P}\left(\sup_{\boldsymbol{\beta}\in\mathcal{N}_\eta}\sup_{i}\mathbf{1}\left\{\hat{g}_i(\boldsymbol{\beta}) \neq g_i^0\right\} > 0\right) + \mathbb{P}\left(\boldsymbol{\beta}\notin\mathcal{N}_\eta\right) \\
\leq& o_p(NT^{-\delta}) + o_p(1)
\end{aligned} \tag{48}$$

where the last inequality follows from Lemma 5, which bounds the probability of misclassification given that $\beta \in \mathcal{N}_\eta$, and Lemma 3, which bounds the probability $\mathbb{P}\left(\boldsymbol{\beta}\notin\mathcal{N}_\eta\right)$.

Part II: asymptotic equivalence. We want to show that $\hat{\boldsymbol{\beta}} - \tilde{\boldsymbol{\beta}} \xrightarrow{p} 0$ where $\tilde{\boldsymbol{\beta}}$ is the infeasible estimator defined by

$$(\tilde{\boldsymbol{\beta}}, \tilde{\boldsymbol{\phi}}) = \underset{(\boldsymbol{\beta},\boldsymbol{\phi})\in(\Theta,\Phi)}{\arg\min}\ \hat{Q}_{NT}(\boldsymbol{\beta}, \boldsymbol{\phi}, \gamma^0)\ . \tag{49}$$

Notice that the individual parameters depend implicitly on $\tilde{\boldsymbol{\beta}}$, and I suppress the dependence for notational simplicity. The infeasible estimator satisfies the first order conditions:

$$\begin{aligned}
\sum_{i\in\mathcal{S}_g^0} \bar{d}'_{zx,i}\hat{\Omega}_{i,h}\left(\bar{d}_{zy,i,h} - \bar{d}_{zx,i}\tilde{\beta}_{g_i^0,h} - \bar{d}_{zc,i}\tilde{\phi}_{i,h}\right) &= 0 \\
\bar{d}'_{zc,i}\hat{\Omega}_{i,h}\left(\bar{d}_{zy,i,h} - \bar{d}_{zx,i}\tilde{\beta}_{g_i^0,h} - \bar{d}_{zc,i}\tilde{\phi}_{i,h}\right) &= 0
\end{aligned}\ . \tag{50}$$

For later use, let us denote the residuals of the auxiliary estimator by $\tilde{e}_{i,t+h} = y_{i,t+h} - x'_{i,t}\tilde{\beta}_{g_i^0,h} - c'_{i,t}\tilde{\phi}_{i,h}$ and $\bar{d}_{z\tilde{e},i,h} = \frac{1}{T}\sum_{t=1}^{T} z_{i,t}\tilde{e}_{i,t+h}$. The FOCs can then be rewritten compactly as

$$\sum_{i\in\mathcal{S}_g^0} \bar{d}'_{zx,i}\hat{\Omega}_{i,h}\bar{d}_{z\tilde{e},i,h} = 0, \quad \bar{d}'_{zc,i}\hat{\Omega}_{i,h}\bar{d}_{z\tilde{e},i,h} = 0\ . \tag{51}$$

Moreover, let us denote $\beta^d_{g_i^0,h} = \tilde{\beta}_{g_i^0,h} - \hat{\beta}_{g_i^0,h}$ and $\phi^d_{i,h} = \tilde{\phi}_{i,h} - \hat{\phi}_{i,h}$. Then by simple manipulations, we have

$$
\begin{aligned}
&\hat{Q}_{NT}(\hat{\boldsymbol{\beta}}, \hat{\boldsymbol{\phi}}, \gamma^0) - \hat{Q}_{NT}(\tilde{\boldsymbol{\beta}}, \tilde{\boldsymbol{\phi}}, \gamma^0) \\
=&\frac{1}{N}\sum_h\sum_i \left[\bar{d}'_{zy,i,h} - \bar{d}_{zx,i}\hat{\beta}_{g_i^0,h} - \bar{d}_{zc,i}\hat{\phi}_{i,h}\right]' \hat{\Omega}_{i,h} \left[\bar{d}'_{zy,i,h} - \bar{d}_{zx,i}\hat{\beta}_{g_i^0,h} - \bar{d}_{zc,i}\hat{\phi}_{i,h}\right] \\
&\quad - \frac{1}{N}\sum_h\sum_i \left[\bar{d}'_{zy,i,h} - \bar{d}_{zx,i}\tilde{\beta}_{g_i^0,h} - \bar{d}_{zc,i}\tilde{\phi}_{i,h}\right]' \hat{\Omega}_{i,h} \left[\bar{d}'_{zy,i,h} - \bar{d}_{zx,i}\tilde{\beta}_{g_i^0,h} - \bar{d}_{zc,i}\tilde{\phi}_{i,h}\right] \\
=&\frac{1}{N}\sum_h\sum_i \left[\bar{d}_{zx,i}\beta^d_{g_i^0,h} + \bar{d}_{zc,i}\phi^d_{i,h} + \bar{d}_{z\tilde{e},i,h}\right]' \hat{\Omega}_{i,h} \left[\bar{d}_{zx,i}\beta^d_{g_i^0,h} + \bar{d}_{zc,i}\phi^d_{i,h} + \bar{d}_{z\tilde{e},i,h}\right] \\
&\quad - \frac{1}{N}\sum_h\sum_i \bar{d}'_{z\tilde{e},i,h}\hat{\Omega}_{i,h}\bar{d}_{z\tilde{e},i,h} \\
=&\frac{1}{N}\sum_h\sum_i \left[\bar{d}_{zx,i}\beta^d_{g_i^0,h} + \bar{d}_{zc,i}\phi^d_{i,h}\right]' \hat{\Omega}_{i,h} \left[\bar{d}_{zx,i}\beta^d_{g_i^0,h} + \bar{d}_{zc,i}\phi^d_{i,h}\right] \\
&\quad + \underbrace{\frac{2}{N}\sum_h\sum_i \left[\bar{d}_{zx,i}\beta^d_{g_i^0,h} + \bar{d}_{zc,i}\phi^d_{i,h}\right]' \hat{\Omega}_{i,h}\bar{d}_{z\tilde{e},i,h}}_{=0}
\end{aligned} \tag{52}
$$

To show that the second term is numerically zero, notice that we can decompose it into

$$
\begin{aligned}
& \tfrac{2}{N}\textstyle\sum_h\sum_i \left[\bar{d}_{zx,i}\beta^d_{g_i^0,h} + \bar{d}_{zc,i}\phi^d_{i,h}\right]' \hat{\Omega}_{i,h}\bar{d}_{z\tilde{e},i,h} \\
= \quad & \tfrac{2}{N}\textstyle\sum_h\sum_g \beta^{d\prime}_{g,h}\left(\sum_{i\in\mathcal{S}_g^0} \bar{d}'_{zx,i}\hat{\Omega}_{i,h}\bar{d}_{z\tilde{e},i,h}\right) + \tfrac{2}{N}\sum_h\sum_i \phi^{d\prime}_{i,h}\left(\bar{d}'_{zc,i}\hat{\Omega}_{i,h}\bar{d}_{z\tilde{e},i,h}\right)
\end{aligned} \tag{53}
$$

where by the FOCs (51) all terms inside the parentheses are zero.

As for the first term, let

$$
\phi^{*d}_{i,h} = \underset{\phi_{i,h}}{\operatorname{argmin}} \frac{1}{N}\sum_h\sum_i \left[\bar{d}_{zx,i}\beta^d_{g_i^0,h} + \bar{d}_{zc,i}\phi_{i,h}\right]' \hat{\Omega}_{i,h} \left[\bar{d}_{zx,i}\beta^d_{g_i^0,h} + \bar{d}_{zc,i}\phi_{i,h}\right] \ . \tag{54}
$$

Then the first term is lower bounded by

$$
\begin{aligned}
& \tfrac{1}{N}\sum_h \sum_i \left[\bar{d}_{zx,i}\beta^d_{g_i^0,h} + \bar{d}_{zc,i}\phi^d_{i,h}\right]' \hat{\Omega}_{i,h} \left[\bar{d}_{zx,i}\beta^d_{g_i^0,h} + \bar{d}_{zc,i}\phi^d_{i,h}\right] \\
\geqslant\ & \tfrac{1}{N}\sum_h \sum_i \left[\bar{d}_{zx,i}\beta^d_{g_i^0,h} + \bar{d}_{zc,i}\phi^{*d}_{i,h}\right]' \hat{\Omega}_{i,h} \left[\bar{d}_{zx,i}\beta^d_{g_i^0,h} + \bar{d}_{zc,i}\phi^{*d}_{i,h}\right] \\
=\ & \tfrac{1}{N}\sum_h \sum_i \beta^{d\prime}_{g_i^0,h}\left(\bar{d}'_{zx,i} M_{zc,i,h} \bar{d}_{zx,i}\right)\beta^d_{g_i^0,h} \\
=\ & \sum_h \sum_g \beta^{d\prime}_{g,h}\left(\tfrac{1}{N}\sum_{i\in\mathcal{S}_g^0} \bar{d}'_{zx,i} M_{zc,i,h} \bar{d}_{zx,i}\right)\beta^d_{g,h} \\
\geqslant\ & \sum_h \sum_g \|\tilde{\beta}_{g,h} - \hat{\beta}_{g,h}\|^2 \rho_{g,h} \geqslant \max_{g,h} \|\tilde{\beta}_{g,h} - \hat{\beta}_{g,h}\|^2 \rho_{g,h}
\end{aligned}
\tag{55}
$$

where $\rho_{g,h}$ is some positive finite number. Here, the first inequality is obtained by construction. The second last inequality follows Assumption 1.D and 1.F, under which

$$
\operatorname*{plim}_{N,T\to\infty} \frac{1}{N}\sum_{i\in\mathcal{S}_g^0} \bar{d}'_{zx,i}\overline{M}_{zc,i,h}\bar{d}_{zx,i} = \operatorname*{plim}_{N\to\infty} \frac{1}{N}\sum_{i\in\mathcal{S}_g^0} d'_{zx,i} M_{zc,i,h} d_{zx,i} \tag{56}
$$

which as Lemma 3 shows is positive definite and thus lower bounded by its minimum eigenvalue $\infty > \rho_{g,h} > 0$. The last inequality follows from the fact that the summands $\|\tilde{\beta}_{g,h} - \hat{\beta}_{g,h}\|^2 \rho_{g,h}$ are non-negative. Combining the two terms, we have

$$
\hat{Q}_{NT}(\hat{\boldsymbol{\beta}}, \hat{\boldsymbol{\phi}}, \gamma^0) - \hat{Q}_{NT}(\tilde{\boldsymbol{\beta}}, \tilde{\phi}, \gamma^0) \geq \max_{g,h} \|\tilde{\beta}_{g,h} - \hat{\beta}_{g,h}\|^2 \rho_{g,h} \tag{57}
$$

and so it suffices to show that the LHS is $o_p(1)$.

Notice that by the Lemma 3, we have $\mathbb{P}\left(\boldsymbol{\beta} \in \mathcal{N}_\eta\right) \xrightarrow{p} 1$, under which Lemma 5 shows that the misclassification error is $o_p(1)$. Therefore, we have

$$
\sup_{(\boldsymbol{\beta},\boldsymbol{\phi})\in\mathcal{N}_\eta} |\,\hat{Q}_{NT}(\boldsymbol{\beta}, \boldsymbol{\phi}, \hat{\gamma}) - \hat{Q}_{NT}(\boldsymbol{\beta}, \boldsymbol{\phi}, \gamma^0)| = o_p(NT^{-\delta})\ . \tag{58}
$$

It follows that

$$
\begin{aligned}
0 \leq\ & \hat{Q}_{NT}(\hat{\boldsymbol{\beta}}, \hat{\boldsymbol{\phi}}, \gamma^0) - \hat{Q}_{NT}(\tilde{\boldsymbol{\beta}}, \tilde{\boldsymbol{\phi}}, \gamma^0) \\
=\ & \hat{Q}_{NT}(\hat{\boldsymbol{\beta}}, \hat{\boldsymbol{\phi}}, \hat{\gamma}) - \hat{Q}_{NT}(\tilde{\boldsymbol{\beta}}, \tilde{\boldsymbol{\phi}}, \hat{\gamma}) + o_p(NT^{-\delta}) \quad . \\
\leq\ & o_p(NT^{-\delta})
\end{aligned}
\tag{59}
$$

Here, the first and the last inequality are obtained by construction: given on the true grouping γ^0, the infeasible estimator $\tilde{\boldsymbol{\beta}}$ is the minimizer, whereas given the estimated groups $\hat{\gamma}$ (by GLP), $\hat{\boldsymbol{\beta}}$ is the minimizer.

To sum up, we have

$$o_p(NT^{-\delta}) \geqslant \max_{g,h} \|\tilde{\beta}_{g,h} - \hat{\beta}_{g,h}\|^2 \rho_{g,h} \tag{60}$$

for some positive finite number $\rho_{g,h}$, which implies that $\|\tilde{\beta}_{g,h} - \hat{\beta}_{g,h}\| = o_p(NT^{-\delta})$ for all g and h.

■

S4.3 Theorem 2

Proof. To derive the asymptotic distribution, I follow Fernández-Val and Lee (2013) to expand the FOC to higher order. For later use, denote the first order conditions with known group structure by

$$\zeta(\beta_{j,h}, \phi(\beta_{j,h})) = \frac{1}{N_j} \sum_{i \in S_j^0} \bar{d}'_{zx,i} \hat{\Omega}_{i,h} \left(\bar{d}_{zy,i,h} - \bar{d}_{zx,i} \beta_{j,h} - \bar{d}_{zc,i} \phi(\beta_{j,h}) \right) = 0 \ . \tag{61}$$

and

$$\xi(\beta_{j,h}, \phi(\beta_{j,h})) = \bar{d}'_{zc,i} \hat{\Omega}_{i,h} \left(\bar{d}_{zy,i,h} - \bar{d}_{zx,i} \beta_{j,h} - \bar{d}_{zc,i} \phi(\beta_{j,h}) \right) = 0 \ . \tag{62}$$

Expand (61) around parameters $\beta^0_{j,h}$ and $\tilde{\phi}(\beta^0_{j,h})$, where $\tilde{}$ indicates that $\tilde{\phi}(\beta^0_{j,h})$ solves (62) given $\beta^0_{j,h}$. We have

$$0 = \zeta(\beta^0_{j,h}, \tilde{\phi}(\beta^0_{j,h})) + \frac{d\zeta(\beta^0_{j,h}, \tilde{\phi}(\beta^0_{j,h}))}{d\beta'}|_{\overline{\beta}} \times (\tilde{\beta}_{j,h} - \beta^0_{j,h}) \tag{63}$$

for some $\overline{\beta}$ between $\beta^0_{j,h}$ and $\tilde{\beta}_{j,h}$. Multiply both sides by $N_j T$, we have

$$\frac{d\zeta(\beta^0_{j,h}, \tilde{\phi}(\beta^0_{j,h}))}{d\beta'}|_{\overline{\beta}} \sqrt{N_j T} (\tilde{\beta}_{j,h} - \beta^0_{j,h}) = -\sqrt{N_j T} \zeta(\beta^0_{j,h}, \tilde{\phi}(\beta^0_{j,h})) \tag{64}$$

Therefore, the proof proceeds in three steps: First, we need to show that

$$\frac{d\zeta(\beta^0_{j,h}, \tilde{\phi}(\beta^0_{j,h}))}{d\beta'}|_{\overline{\beta}} \xrightarrow{p} -\Sigma_{j,h} \tag{65}$$

Second, we need to show that

$$\sqrt{N_j T} \zeta(\beta^0_{j,h}, \tilde{\phi}(\beta^0_{j,h})) \xrightarrow{d} N(\kappa \mathcal{B}, \Psi_{j,h}) \tag{66}$$

The asymptotic normality of the infeasible estimator follows immediately from the first two results. Finally, I show that the GLP estimator is asymptotically equivalent to the infeasible counterpart.

Part I. Taking the derivatives of (61), we have

$$\frac{d\zeta(\beta^0_{j,h},\tilde{\phi}(\beta^0_{j,h}))}{d\beta}|_{\overline{\beta}} = -\frac{1}{N_j}\sum_{i\in S^0_j}\overline{d}'_{zx,i}\hat{\Omega}_{i,h}\overline{d}_{zx,i} - \frac{1}{N_j}\sum_{i\in S^0_j}\bar{d}'_{zx,i}\hat{\Omega}_{i,h}\bar{d}'_{zc,i}\frac{\partial\phi(\beta)}{\partial\beta}|_{\overline{\beta}}. \tag{67}$$

Similarly, taking derivatives of (62) and evaluating at $\overline{\beta}$, we have

$$-\bar{d}'_{zc,i}\hat{\Omega}_{i,h}\bar{d}_{zx,i} - \bar{d}'_{zc,i}\hat{\Omega}_{i,h}\bar{d}_{zc,i}\frac{\partial\phi(\beta)}{\partial\beta}|_{\overline{\beta}} = 0 \tag{68}$$

Next by Lemma 6, we have (uniformly over i),

$$\begin{aligned}
0 &= -\,d'_{zc,i}\Omega_{i,h}d_{zx,i} - \bar{d}'_{zc,i}\hat{\Omega}_{i,h}\bar{d}_{zc,i}\frac{\partial\phi(\beta)}{\partial\beta}|_{\overline{\beta}}\\
&= -\,d'_{zc,i}\Omega_{i,h}d_{zx,i} + o_p(1) - (o_p(1) + d'_{zc,i}\Omega_{i,h}d_{zc,i})\frac{\partial\phi(\beta)}{\partial\beta}|_{\overline{\beta}}\\
\Longrightarrow\ \frac{\partial\phi(\beta)}{\partial\beta}|_{\overline{\beta}} &= -\left(d'_{zc,i}\Omega_{i,h}d_{zc,i}\right)^{-1}d'_{zc,i}\Omega_{i,h}d_{zx,i} + o_p(1)
\end{aligned} \tag{69}$$

Substitute it back to (67), we have

$$\begin{aligned}
&\frac{d\zeta(\beta^0_{j,h},\tilde{\phi}(\beta^0_{j,h}))}{d\beta}|_{\overline{\beta}}\\
=&-\frac{1}{N_j}\sum_{i\in S^0_j}\overline{d}'_{zx,i}\hat{\Omega}_{i,h}\overline{d}_{zx,i} - \frac{1}{N_j}\sum_{i\in S^0_j}\bar{d}'_{zx,i}\hat{\Omega}_{i,h}\bar{d}'_{zc,i}\frac{\partial\phi(\beta)}{\partial\beta}|_{\overline{\beta}}.\\
=&-\frac{1}{N_j}\sum_{i\in S^0_j}d'_{zx,i}\Omega_{i,h}d_{zx,i} + \frac{1}{N_j}\sum_{i\in S^0_j}d'_{zx,i}\Omega_{i,h}d'_{zc,i}\left(d'_{zc,i}\Omega_{i,h}d_{zc,i}\right)^{-1}d'_{zc,i}\Omega_{i,h}d_{zx,i} + o_p(1)\\
=&-\frac{1}{N_j}\sum_{i\in S^0_j}d'_{zx,i}M_{zc,i,h}d_{zx,i} + o_p(1)
\end{aligned} \tag{70}$$

which gives Part I. Specifically, in the above derivation, the $o_p(1)$ terms remain bounded because Lemma 6 holds uniformly over i. For instance, we have

$$\begin{aligned}
& -\tfrac{1}{N_j}\textstyle\sum_{i\in S_j^0}\overline{d}'_{zx,i}\hat{\Omega}_{i,h}\overline{d}_{zx,i} \\
= & -\tfrac{1}{N_j}\textstyle\sum_{i\in S_j^0} d'_{zx,i}\Omega_{i,h}d_{zx,i} + \left(\tfrac{1}{N_j}\textstyle\sum_{i\in S_j^0} d'_{zx,i}\Omega_{i,h}d_{zx,i} - \tfrac{1}{N_j}\textstyle\sum_{i\in S_j^0}\overline{d}'_{zx,i}\hat{\Omega}_{i,h}\overline{d}_{zx,i}\right) \\
\leq & -\tfrac{1}{N_j}\textstyle\sum_{i\in S_j^0} d'_{zx,i}\Omega_{i,h}d_{zx,i} + \sup_i\left(d'_{zx,i}\Omega_{i,h}d_{zx,i} - \overline{d}'_{zx,i}\hat{\Omega}_{i,h}\overline{d}_{zx,i}\right) \\
\leq & -\tfrac{1}{N_j}\textstyle\sum_{i\in S_j^0} d'_{zx,i}\Omega_{i,h}d_{zx,i} + o_p(1)
\end{aligned} \tag{71}$$

Part II. The FOC for the infeasible estimator evaluated at $\beta^0_{j,h}, \tilde{\phi}(\beta^0_{j,h})$ gives

$$\begin{aligned}
\zeta(\beta^0_{j,h}, \tilde{\phi}(\beta^0_{j,h})) =& \frac{1}{N_j}\sum_{i\in S_j^0}\bar{d}'_{zx,i}\hat{\Omega}_{i,h}\left(\bar{d}_{zy,i,h} - \bar{d}_{zx,i}\beta^0_{j,h} - \bar{d}_{zc,i}\tilde{\phi}(\beta^0_{j,h})\right) \\
=& \frac{1}{N_j}\sum_{i\in S_j^0}\bar{d}'_{zx,i}\hat{\Omega}_{i,h}\bar{d}_{zc,i}\left(\phi^0_{i,h} - \tilde{\phi}(\beta^0_{j,h})\right) + \frac{1}{N_j}\sum_{i\in S_j^0}\bar{d}'_{zx,i}\hat{\Omega}_{i,h}\bar{d}_{z\epsilon,i,h}
\end{aligned} \tag{72}$$

To derive the stochastic expansion of $\zeta(\beta^0_{j,h}, \tilde{\phi}(\beta^0_{j,h}))$ we need to expand $\phi^0_{i,h} - \tilde{\phi}(\beta^0_{j,h})$. First, from the FOC of individual parameters (62), we have

$$\begin{aligned}
0 = \xi(\beta^0_{j,h}, \tilde{\phi}(\beta^0_{j,h})) =& \bar{d}'_{zc,i}\hat{\Omega}_{i,h}\left(\bar{d}_{zy,i,h} - \bar{d}_{zx,i}\beta^0_{j,h} - \bar{d}_{zc,i}\tilde{\phi}(\beta^0_{j,h})\right) \\
=& \bar{d}'_{zc,i}\hat{\Omega}_{i,h}\bar{d}_{z\epsilon,i,h} + \bar{d}'_{zc,i}\hat{\Omega}_{i,h}\bar{d}_{zc,i}\left(\phi^0_{i,h} - \tilde{\phi}(\beta^0_{j,h})\right)
\end{aligned} \tag{73}$$

which gives the individual estimator

$$\left(\phi^0_{i,h} - \tilde{\phi}(\beta^0_{j,h})\right) = -\left(\bar{d}'_{zc,i}\hat{\Omega}_{i,h}\bar{d}_{zc,i}\right)^{-1}\bar{d}'_{zc,i}\hat{\Omega}_{i,h}\bar{d}_{z\epsilon,i,h} \tag{74}$$

Denote $\mathcal{O}_{zc,i,h} = \bar{d}'_{zc,i}\hat{\Omega}_{i,h}\bar{d}_{zc,i}$. By add and subtract, we have

$$
\begin{aligned}
&\tilde{\mathcal{O}}^{-1}_{zc,i,h}\bar{d}'_{zc,i}\hat{\Omega}_{i,h}\bar{d}_{z\epsilon,i,h} \\
=&\left[\tilde{\mathcal{O}}^{-1}_{zc,i,h} - \mathcal{O}^{-1}_{zc,i,h} + \mathcal{O}^{-1}_{zc,i,h}\right] \times \left[\bar{d}_{zc,i} - d_{zc,i} + d_{zc,i}\right]' \times \left[\hat{\Omega}_{i,h} - \Omega_{i,h} + \Omega_{i,h}\right] \bar{d}_{z\epsilon,i,h} \\
=&\left[\tilde{\mathcal{O}}^{-1}_{zc,i,h} - \mathcal{O}^{-1}_{zc,i,h} + \mathcal{O}^{-1}_{zc,i,h}\right] \times \left[\left(\bar{d}_{zc,i} - d_{zc,i}\right)'\left(\hat{\Omega}_{i,h} - \Omega_{i,h}\right) + \left(\bar{d}_{zc,i} - d_{zc,i}\right)'\Omega_{i,h}\right. \\
&\qquad\qquad\qquad\qquad\qquad\left. + \, d_{zc,i}\left(\hat{\Omega}_{i,h} - \Omega_{i,h}\right) + d_{zc,i}\Omega_{i,h}\right]\bar{d}_{z\epsilon,i,h} \\
=&\left[\tilde{\mathcal{O}}^{-1}_{zc,i,h} - \mathcal{O}^{-1}_{zc,i,h} + \mathcal{O}^{-1}_{zc,i,h}\right] \times \left[\left(\bar{d}_{zc,i} - d_{zc,i}\right)'\Omega_{i,h}\right. \\
&\qquad\qquad\qquad\qquad\qquad\left. + \, d_{zc,i}\left(\hat{\Omega}_{i,h} - \Omega_{i,h}\right) + d_{zc,i}\Omega_{i,h}\right] + o_p(1) \\
=&\mathcal{O}^{-1}_{zc,i,h}\left[\left(\bar{d}_{zc,i} - d_{zc,i}\right)'\Omega_{i,h} + d_{zc,i}\left(\hat{\Omega}_{i,h} - \Omega_{i,h}\right) + d_{zc,i}\Omega_{i,h}\right]\bar{d}_{z\epsilon,i,h} + o_p(1) \\
=&[A_1 + A_2 + A_3]\bar{d}_{z\epsilon,i,h} + o_p(1)
\end{aligned}
\tag{75}
$$

Consider the first term in the last equality. Multiply it by $\bar{d}_{z\epsilon,i,h}$ we have

$$
\begin{aligned}
&\mathcal{O}^{-1}_{zc,i,h}\left(\bar{d}_{zc,i} - d_{zc,i}\right)'\Omega_{i,h}\bar{d}_{z\epsilon,i,h} \\
=&\mathcal{O}^{-1}_{zc,i,h}\left(\frac{1}{T}\sum_t z_{i,t}c'_{i,t} - \mathbb{E}[z_{i,t}c'_{i,t}]\right)'\Omega_{i,h}\left(\frac{1}{T}\sum_t z_{i,t}\epsilon_{i,t+h}\right) \\
=&\frac{1}{\sqrt{T}}\mathcal{O}^{-1}_{zc,i,h}\left(\frac{1}{\sqrt{T}}\sum_t z_{i,t}c'_{i,t} - \mathbb{E}[z_{i,t}c'_{i,t}]\right)'\Omega_{i,h}\left(\frac{1}{\sqrt{T}}\sum_t z_{i,t}\epsilon_{i,t+h}\right)
\end{aligned}
\tag{76}
$$

where we have

$$
\frac{1}{\sqrt{T}}\sum_t z_{i,t}c'_{i,t} - \mathbb{E}[z_{i,t}c'_{i,t}] = O_p(1), \quad \frac{1}{\sqrt{T}}\sum_t z_{i,t}\epsilon_{i,t+h} = O_p(1) \tag{77}
$$

and thus this term is $O_p(\frac{1}{\sqrt{T}})$, we will formally show this later, since we need to sum over i.

Now substitute $A_1 + A_2 + A_3 + o_p(1)$ back to the group-parameter FOC, and as before, in each step we "drop" terms involving two (or more) estimation errors:

$$\begin{aligned}
&\zeta(\beta^0_{j,h}, \tilde{\phi}(\beta^0_{j,h})) \\
=&\frac{1}{N_j}\sum_{i\in S^0_j} \bar{d}'_{zx,i}\hat{\Omega}_{i,h}\bar{d}_{zc,i}\left(\phi^0_{i,h} - \tilde{\phi}(\beta^0_{j,h})\right) + \frac{1}{N_j}\sum_{i\in S^0_j} \bar{d}'_{zx,i}\hat{\Omega}_{i,h}\bar{d}_{z\epsilon,i,h} \\
=&-\frac{1}{N_j}\sum_{i\in S^0_j} \bar{d}'_{zx,i}\hat{\Omega}_{i,h}\bar{d}_{zc,i}\left(A_1 + A_2 + A_3\right)\bar{d}_{z\epsilon,i,h} + \frac{1}{N_j}\sum_{i\in S^0_j} \bar{d}'_{zx,i}\hat{\Omega}_{i,h}\bar{d}_{z\epsilon,i,h} + o_p(1) \\
=&\frac{1}{N_j}\sum_{i\in S^0_j} \bar{d}'_{zx,i}\hat{\Omega}_{i,h}\left[I - \bar{d}_{zc,i}\left(A_1 + A_2 + A_3\right)\right]\bar{d}_{z\epsilon,i,h} + o_p(1)
\end{aligned} \tag{78}$$

First consider

$$\begin{aligned}
\bar{d}'_{zx,i}\hat{\Omega}_{i,h} &= \left[\bar{d}_{zx,i} - d_{zx,i} + d_{zx,i}\right]'\left[\hat{\Omega}_{i,h} - \Omega_{i,h} + \Omega_{i,h}\right] \\
&= \left(\bar{d}_{zx,i} - d_{zx,i}\right)'\left(\hat{\Omega}_{i,h} - \Omega_{i,h}\right) + \left(\bar{d}_{zx,i} - d_{zx,i}\right)'\Omega_{i,h} \\
&\quad + d'_{zx,i}\left(\hat{\Omega}_{i,h} - \Omega_{i,h}\right) + d'_{zx,i}\Omega_{i,h} + o_p(1) \\
&= \left(\bar{d}_{zx,i} - d_{zx,i}\right)'\Omega_{i,h} + d'_{zx,i}\left(\hat{\Omega}_{i,h} - \Omega_{i,h}\right) + d'_{zx,i}\Omega_{i,h} + o_p(1) \\
&= B_1 + B_2 + B_3 + o_p(1)
\end{aligned} \tag{79}$$

Next, consider

$$\begin{aligned}
&\bar{d}'_{zx,i}\hat{\Omega}_{i,h}\bar{d}_{zc,i}[A_1 + A_2 + A_3] \\
=&(B_1 + B_2 + B_3)(\bar{d}_{zc,i} - d_{zc,i} + d_{zc,i})[A_1 + A_2 + A_3] + o_p(1) \\
=&(B_1 + B_2 + B_3)(\bar{d}_{zc,i} - d_{zc,i} + d_{zc,i})[A_1 + A_2 + A_3] + o_p(1) \\
=&\left(B_1 d_{zc,i} + B_2 d_{zc,i} + B_3(\bar{d}_{zc,i} - d_{zc,i}) + B_3 d_{zc,i}\right)[A_1 + A_2 + A_3] + o_p(1) \\
=&B_1 d_{zc,i} A_3 + B_2 d_{zc,i} A_3 \\
&\quad + B_3(\bar{d}_{zc,i} - d_{zc,i})A_3 + B_3 d_{zc,i} A_1 + B_3 d_{zc,i} A_2 + B_3 d_{zc,i} A_3 + o_p(1)
\end{aligned} \tag{80}$$

In total, we keep nine terms, listed below:

$$
\begin{aligned}
&B_1\overline{d}_{z\epsilon,i,h} &&= \left(\bar{d}_{zx,i} - d_{zx,i}\right)' \Omega_{i,h}\overline{d}_{z\epsilon,i,h} \\
&B_1 d_{zc,i} A_3 \overline{d}_{z\epsilon,i,h} &&= \left(\bar{d}_{zx,i} - d_{zx,i}\right)' \Omega_{i,h} d_{zc,i}\mathcal{O}^{-1}_{zc,i,h} d_{zc,i}\Omega_{i,h}\overline{d}_{z\epsilon,i,h} \\
&B_2\overline{d}_{z\epsilon,i,h} &&= d'_{zx,i}\left(\hat{\Omega}_{i,h} - \Omega_{i,h}\right)\overline{d}_{z\epsilon,i,h} \\
&B_2 d_{zc,i} A_3 \overline{d}_{z\epsilon,i,h} &&= d'_{zx,i}\left(\hat{\Omega}_{i,h} - \Omega_{i,h}\right) d_{zc,i}\mathcal{O}^{-1}_{zc,i,h} d_{zc,i}\Omega_{i,h}\overline{d}_{z\epsilon,i,h} \\
&B_3 d_{zc,i} A_2 \overline{d}_{z\epsilon,i,h} &&= d'_{zx,i}\Omega_{i,h} d_{zc,i}\mathcal{O}^{-1}_{zc,i,h} d_{zc,i}\left(\hat{\Omega}_{i,h} - \Omega_{i,h}\right)\overline{d}_{z\epsilon,i,h} \\
&B_3(\bar{d}_{zc,i} - d_{zc,i}) A_3 \overline{d}_{z\epsilon,i,h} &&= d'_{zx,i}\Omega_{i,h}(\bar{d}_{zc,i} - d_{zc,i})\mathcal{O}^{-1}_{zc,i,h} d_{zc,i}\Omega_{i,h}\overline{d}_{z\epsilon,i,h} \\
&B_3 d_{zc,i} A_1 \overline{d}_{z\epsilon,i,h} &&= d'_{zx,i}\Omega_{i,h} d_{zc,i}\mathcal{O}^{-1}_{zc,i,h}\left(\bar{d}_{zc,i} - d_{zc,i}\right)' \Omega_{i,h}\overline{d}_{z\epsilon,i,h} \\
&B_3\overline{d}_{z\epsilon,i,h} &&= d'_{zx,i}\Omega_{i,h}\overline{d}_{z\epsilon,i,h} \\
&B_3 d_{zc,i} A_3 \overline{d}_{z\epsilon,i,h} &&= d'_{zx,i}\Omega_{i,h} d_{zc,i}\mathcal{O}^{-1}_{zc,i,h} d_{zc,i}\Omega_{i,h}\overline{d}_{z\epsilon,i,h}
\end{aligned} \tag{81}
$$

To finish the proof, we need to derive the asymptotic distribution of the above terms. We first collect the terms by defining:

$$
\begin{aligned}
\mathcal{B}_1 =& \frac{1}{N_j}\sum_{i\in\mathcal{S}^0_j} B_1\overline{d}_{z\epsilon,i,h} - B_1 d_{zc,i} A_3\overline{d}_{z\epsilon,i,h} \\
=& \frac{1}{N_j}\sum_{i\in\mathcal{S}^0_j} \left(\bar{d}_{zx,i} - d_{zx,i}\right)' \left[\Omega_{i,h} - \Omega_{i,h} d_{zc,i}\mathcal{O}^{-1}_{zc,i,h} d_{zc,i}\Omega_{i,h}\right]\overline{d}_{z\epsilon,i,h} \\
=& \frac{1}{N_j T^2}\sum_{i\in\mathcal{S}^0_j}\sum_t\sum_s \left(z_{i,t}x'_{i,t} - \mathbb{E}z_{i,t}x'_{i,t}\right)' M_{zc,i,h} z_{i,s}\epsilon_{i,s+h} \qquad (82) \\
\mathcal{B}_2 =& \frac{1}{N_j}\sum_{i\in\mathcal{S}^0_j} B_2\overline{d}_{z\epsilon,i,h} - B_2 d_{zc,i} A_3\overline{d}_{z\epsilon,i,h} - B_3 d_{zc,i} A_2\overline{d}_{z\epsilon,i,h} \\
=& \frac{1}{N_j}\sum_{i\in\mathcal{S}^0_j} d'_{zx,i}\left(\hat{\Omega}_{i,h} - \Omega_{i,h}\right)\left[I - d_{zc,i}\overline{d}_{z\epsilon,i,h}\mathcal{O}^{-1}_{zc,i,h} d_{zc,i}\Omega_{i,h}\right]\overline{d}_{z\epsilon,i,h} \\
&- \frac{1}{N_j}\sum_{i\in\mathcal{S}^0_j} d'_{zx,i}\Omega_{i,h} d_{zc,i}\mathcal{O}^{-1}_{zc,i,h} d_{zc,i}\left(\hat{\Omega}_{i,h} - \Omega_{i,h}\right)\overline{d}_{z\epsilon,i,h} \\
=& \frac{1}{N_j}\sum_{i\in\mathcal{S}^0_j} d'_{zx,i}\left(\hat{\Omega}_{i,h} - \Omega_{i,h}\right)\Omega^{-1}_{i,h} M_{zc,i,h}\overline{d}_{z\epsilon,i,h}
\end{aligned}
$$

$$
-\frac{1}{N_j}\sum_{i\in\mathcal{S}_j^0} d'_{zx,i}\Omega_{i,h}d_{zc,i}\mathcal{O}^{-1}_{zc,i,h}d_{zc,i}\left(\hat{\Omega}_{i,h}-\Omega_{i,h}\right)\overline{d}_{z\epsilon,i,h} \tag{83}
$$

$$
\begin{aligned}
\mathcal{B}_3 =& -\frac{1}{N_j}\sum_{i\in\mathcal{S}_j^0} B_3(\overline{d}_{zc,i}-d_{zc,i})A_3\overline{d}_{z\epsilon,i,h} \\
=&\frac{1}{N_j}\sum_{i\in\mathcal{S}_j^0} d'_{zx,i}\Omega_{i,h}(\overline{d}_{zc,i}-d_{zc,i})\mathcal{O}^{-1}_{zc,i,h}d_{zc,i}\Omega_{i,h}\overline{d}_{z\epsilon,i,h} \qquad (84)\\
\mathcal{B}_4 =& -\frac{1}{N_j}\sum_{i\in\mathcal{S}_j^0} B_3 d_{zc,i}A_1\overline{d}_{z\epsilon,i,h} \\
=&\frac{1}{N_j}\sum_{i\in\mathcal{S}_j^0} d'_{zx,i}\Omega_{i,h}d_{zc,i}\mathcal{O}^{-1}_{zc,i,h}\left(\bar{d}_{zc,i}-d_{zc,i}\right)'\Omega_{i,h}\overline{d}_{z\epsilon,i,h} \qquad (85)
\end{aligned}
$$

We now show that these terms are $O(\frac{1}{T})$. The proofs follow similar strategy and I derive here for $\mathcal{B}_1$. Multiplied by $\sqrt{N_jT}$, we have

$$
\sqrt{N_jT}\mathcal{B}_1 = \frac{1}{N_j^{1/2}T^{3/2}}\sum_{i\in\mathcal{S}_j^0}\sum_t\sum_s\left(x_{i,t}z'_{i,t}-\mathbb{E}x_{i,t}z'_{i,t}\right)M_{zc,i,h}z_{i,s}\epsilon_{i,s+h} \tag{86}
$$

For notational simplicity, we define $\omega_{i,t} = \left(x_{i,t}z'_{i,t}-\mathbb{E}x_{i,t}z'_{i,t}\right)M_{zc,i,h}$ and $\eta_{i,s} = z_{i,s}\epsilon_{i,s+h}$. Notice that since $M_{zc,i,h}$ is nonstochastic, we have

$$
\mathbb{E}\omega_{i,t} = \mathbb{E}\left(x_{i,t}z'_{i,t}-\mathbb{E}x_{i,t}z'_{i,t}\right)M_{zc,i,h} = 0\ . \tag{87}
$$

We then rewrite $\mathcal{B}_1$ as

$$
\sqrt{N_jT}\mathcal{B}_1 = \frac{1}{N_j^{1/2}T^{3/2}}\sum_{i\in\mathcal{S}_j^0}\sum_t\sum_s\mathbb{E}\omega_{i,t}\eta_{i,s} + \frac{1}{N_j^{1/2}T^{3/2}}\sum_{i\in\mathcal{S}_j^0}\sum_t\sum_s\left(\omega_{i,t}\eta_{i,s}-\mathbb{E}\omega_{i,t}\eta_{i,s}\right) \tag{88}
$$

We then bound the RHS of the above equation. First, denote by ι_1 some arbitrary nonrandom vector with unit norm, we have

$$
\frac{1}{N_j^{1/2}T^{3/2}}\sum_{i\in S_j^0}\sum_t\sum_s\mathbb{E}\left[\iota_1 x_{i,t}z'_{i,t}M_{zc,i,h}z_{i,s}\epsilon_{i,s+h}\right]
$$

$$\leqslant \frac{1}{N_j^{1/2}T^{3/2}} \sum_{i\in S_j^0}\sum_t\sum_s \left|\mathbb{E}\left[\iota_1 x_{i,t}z_{i,t}' M_{zc,i,h} z_{i,s}\epsilon_{i,s+h}\right]\right|$$
$$= \sqrt{\frac{N_j}{T}}\frac{1}{N_jT} \sum_{i\in S_j^0}\sum_t\sum_s \left|\mathbb{E}\left[\iota_1 x_{i,t}z_{i,t}' M_{zc,i,h} z_{i,s}\epsilon_{i,s+h}\right] - \mathbb{E}\left[\iota_1 x_{i,t}z_{i,t}' M_{zc,i,h}\right]\mathbb{E}\left[z_{i,s}\epsilon_{i,s+h}\right]\right|$$
$$\leqslant \sqrt{\frac{N_j}{T}}\frac{1}{N_jT} \sum_{i\in S_j^0}\sum_t\sum_s 8\left\|\iota_1 x_{i,t}z_{i,t}' M_{zc,i,h}\right\|^q \left\|z_{i,s}\epsilon_{i,s+h}\right\|^q \alpha(|t-s|)^{1-2/q} \tag{89}$$

where the last inequality is obtained by Davydov inequality. By Assumption 1.D and 1.E we can set $q = 4(1+\delta)$. Moreover, recall that $M_{zc,i,h}$ is nonstochastic and finite so by Cauchy-Schwarz inequality, only the moment bound for $x_{i,t}z_{i,t}$ is needed. Combined with Assumption 1.B, leads to $\sum_t\sum_s \alpha(|t-s|)^{1-1/2(1+\delta)} = O(T)$ and thus

$$\frac{1}{N_j^{1/2}T^{3/2}} \sum_{i\in S_j^0}\sum_t\sum_s \mathbb{E}\left[\iota_1 x_{i,t}z_{i,t}' M_{zc,i,h} z_{i,s}\epsilon_{i,s+h}\right] \leqslant O_p\left(\sqrt{\frac{N_j}{T}}\right) . \tag{90}$$

Second, using a similar strategy, we have

$$\mathbb{E}\left[\frac{1}{N_j^{1/2}T^{3/2}} \sum_{i\in S_j^0}\sum_t\sum_s \left(\omega_{i,t}\eta_{i,s} - \mathbb{E}\left[\omega_{i,t}\eta_{i,s}\right]\right)\right]^2$$
$$= \frac{1}{N_jT^3} \sum_{i\in S_j^0}\sum_t\sum_s\sum_m\sum_r \mathbb{E}\left(\omega_{i,t}\eta_{i,s} - \mathbb{E}\left[\omega_{i,t}\eta_{i,s}\right]\right)\left(\omega_{i,m}\eta_{i,r} - \mathbb{E}\left[\omega_{i,m}\eta_{i,r}\right]\right)$$
$$= \frac{1}{N_jT^3} \sum_{i\in S_j^0}\sum_t\sum_s\sum_m\sum_r \mathbb{E}\left(\omega_{i,t}\eta_{i,s}\omega_{i,m}\eta_{i,r}\right) - \mathbb{E}\left[\omega_{i,t}\eta_{i,s}\right]\mathbb{E}\left[\omega_{i,m}\eta_{i,r}\right]$$
$$\leqslant \frac{1}{N_jT^3} \sum_{i\in S_j^0}\sum_t\sum_s\sum_m\sum_r 8\left\|\omega_{i,t}\eta_{i,s}\right\|^q \left\|\omega_{i,m}\eta_{i,r}\right\|^q \alpha(|t-s-m-r|)^{1-2/q} . \tag{91}$$

We can then similarly obtain and upper bound by Assumption 1.B, 1.D and 1.E that

$$\mathbb{E}\left[\frac{1}{N_j^{1/2}T^{3/2}} \sum_{i\in S_j^0}\sum_t\sum_s \left(\eta_{i,t}\xi_{i,s} - \mathbb{E}\left[\eta_{i,t}\xi_{i,s}\right]\right)\right]^2 \leqslant O_p(\frac{1}{T}) \tag{92}$$

and thus

$$\frac{1}{N_j^{1/2}T^{3/2}}\sum_{i\in S_j^0}\sum_t\sum_s\left(\eta_{i,t}\xi_{i,s}-\mathbb{E}\left[\eta_{i,t}\xi_{i,s}\right]\right)=O_p(T^{-1/2})=o_p(1)\ . \tag{93}$$

In conclusion, the bias term $\mathcal{B}_1$ is dominated by the first term

$$\mathbb{E}\left[\left(x_{i,t}z'_{i,t}-\mathbb{E}x_{i,t}z'_{i,t}\right)M_{zc,i,h}z_{i,s}\epsilon_{i,s+h}\right]$$

which leads to asymptotic bias. The remaining terms $\mathcal{B}_2,\mathcal{B}_3$ and $\mathcal{B}_4$ are bounded similarly.

Having established the asymptotic properties of the bias terms, it remains to show the asymptotic distribution. Specifically, we have

$$\begin{aligned}\mathcal{V}=&\frac{1}{N_j}\sum_{i\in\mathcal{S}_j^0}B_3\overline{d}_{z\epsilon,i,h}-B_3d_{zc,i}A_3\overline{d}_{z\epsilon,i,h}\\=&\frac{1}{N_j}\sum_{i\in\mathcal{S}_j^0}d'_{zx,i}\left[\Omega_{i,h}-\Omega_{i,h}d_{zc,i}\mathcal{O}_{zc,i,h}^{-1}d_{zc,i}\Omega_{i,h}\right]\overline{d}_{z\epsilon,i,h}\\=&\frac{1}{N_jT}\sum_{i\in\mathcal{S}_j^0}\sum_t d'_{zx,i}M_{zc,i,h}z_{i,t}\epsilon_{i,t+h}\end{aligned}\tag{94}$$

Assumptions 1.B, 1.D, 1.E and 3.A guarantee that we have

$$\frac{1}{\sqrt{N_jT}}\sum_{i\in S_j^0}\sum_s\mathbb{E}\left[x_{i,t}z'_{i,t}\right]M_{zc,i,h}z_{i,s}\epsilon_{i,s+h}\xrightarrow{d}N(0,\Psi_{j,h}) \tag{95}$$

by standard central limit theorem, e.g. Lemma 3 in (Hahn and Kuersteiner, 2011).

We are now ready to show the asymptotic distribution. Using (72), we have

$$\begin{aligned}&\sqrt{N_jT}\zeta(\beta_{j,h}^0,\tilde{\phi}(\beta_{j,h}^0))\\=&\sqrt{N_jT}\left(\mathcal{V}+\mathcal{B}_1+\mathcal{B}_2+\mathcal{B}_3+\mathcal{B}_4\right)+o_p(1)\end{aligned}\tag{96}$$

with $\sqrt{N_jT}\mathcal{V}\xrightarrow{d}N(0,\Psi_{j,h})$, and $\sqrt{N_jT}\left(\mathcal{B}_1+\mathcal{B}_2+\mathcal{B}_3+\mathcal{B}_4\right)=O(\sqrt{N_j/T})$. Combined with results from part I, the asymptotic distribution of the infeasible estimator follows.

Part III. By Theorem 1, we have

$$\sqrt{N_j T}(\hat{\beta}_{j,h} - \tilde{\beta}_{j,h}) = o_p\left(\frac{\sqrt{N_j} N T^{\frac{1}{2}}}{T^{\delta}}\right). \tag{97}$$

Under the assumption that $N_j/T = \kappa_j$ and $N_j/N = \pi_j$ with $\kappa_j \in [0, \infty)$ and $\pi_j \in (0, 1)$, we have

$$\frac{\sqrt{N_j} N T^{\frac{1}{2}}}{T^{\delta}} = \frac{\kappa_j^2}{\pi_j} T^{2-\delta} \tag{98}$$

which converges to zero if we set $\delta > 2$. This completes the proof. ■

S4.4 Proposition 1

Proof. We introduce some new notations in this section to discuss the group number selection. Specifically, we denote the objective function in the main text (4) by the following

$$\{\hat{\boldsymbol{\beta}}^G, \hat{\boldsymbol{\phi}}^G, \hat{\gamma}^G\} = \operatorname*{argmin}_{\boldsymbol{\beta}\in\Theta_G, \boldsymbol{\phi}\in\Phi_N, \gamma\in\mathcal{G}(G)} \hat{Q}_{NT}(\boldsymbol{\beta}, \boldsymbol{\phi}, \gamma) \ , \tag{99}$$

where the subscript G explicitly indicates the dependence on the number of groups. The minimized objective is $\hat{Q}_{NT,G} = \min_{\boldsymbol{\beta}\in\Theta_G, \boldsymbol{\phi}\in\Phi_N, \gamma\in\mathcal{G}_G} \hat{Q}_{NT}(\boldsymbol{\beta}, \boldsymbol{\phi}, \gamma)$. That is, we assume that we can obtain the global minimizer.

To prove the proposition, let us consider two cases.

Case I: Under-select number of groups $\hat{G} < G^0$. By the definition of the proposed criterion (15), we want to show that

$$\mathbb{P}\left(\min_{G<G^0} \hat{Q}_{NT,G} - \hat{Q}_{NT,G^0} + \underbrace{\varrho_{N,T}\hat{Q}_{NT,G_{max}}(G-G^0)(H+1)}_{\xrightarrow{p} 0_-, \text{ by } 4.B} > 0\right) \to 1 \ . \tag{100}$$

Note that by Theorem 1 and Lemma 4, we have $\hat{\boldsymbol{\beta}}^G \xrightarrow{p} \hat{\boldsymbol{\beta}}^0, \hat{\boldsymbol{\phi}}^G \xrightarrow{p} \hat{\boldsymbol{\phi}}^0$. Then by the continuous mapping theorem, we have $\hat{Q}_{NT,G^0} \xrightarrow{p} Q^0$.

Furthermore, by Assumption 4.A we have

$$\min_{1\le G<G^0} \inf_{\gamma\in\mathcal{G}(G)} \hat{Q}_{NT,G} \xrightarrow{p} \underline{Q} > Q^0 \tag{101}$$

and thus

$$\begin{aligned}
&\min_{G<G^0} \hat{Q}_{NT,G} - \hat{Q}_{NT,G^0} + \varrho_{N,T}\hat{Q}_{NT,G_{max}}(G-G^0)(H+1) \\
\xrightarrow{p} &\underline{Q} - Q^0 + o_p(1) > 0
\end{aligned}$$

and thus (100) holds.

Case II: Over-select number of groups $\hat{G} > G^0$**.** By definition, we want to show that

$$\mathbb{P}\left(\min_{G^0<G\leq G_{max}} IC(G) > IC(G^0)\right) \to 1 \tag{102}$$

It is equivalent to show that there exists some $a > 0$ such that

$$\mathbb{P}\left(\min_{G>G^0} T^a(\hat{Q}_{NT,G} - \hat{Q}_{NT,G^0}) + \underbrace{T^a \varrho_{N,T} \hat{Q}_{NT,G_{max}}(G-G^0)(H+1)}_{\xrightarrow{p}\infty_+,\text{ by}4.B} > 0\right) \to 1\ . \tag{103}$$

Given that the second component converges to (positive) infinity, a sufficient condition for the above is that

$$T^a(\hat{Q}_{NT,G} - \hat{Q}_{NT,G^0}) = O_p(1)\ . \tag{104}$$

However, it is generally difficult to compare GLP estimates with different number of groups. Notice that by add and subtract, we can rewrite the above as

$$T^a(\hat{Q}_{NT,G} - \hat{Q}_{NT}(\boldsymbol{\beta}^0, \boldsymbol{\phi}^0, \gamma^0) + \hat{Q}_{NT}(\boldsymbol{\beta}^0, \boldsymbol{\phi}^0, \gamma^0) - \hat{Q}_{NT,G^0}) = O_p(1)\ . \tag{105}$$

Therefore, it is enough to show that there exists some $a > 0$ such that we instead show that

$$T^a\left|\hat{Q}_{NT,G}(\boldsymbol{\beta}, \boldsymbol{\phi}, \gamma) - \hat{Q}_{NT}(\boldsymbol{\beta}^0, \boldsymbol{\phi}^0, \gamma^0)\right| \leq O_p(1)\ . \tag{106}$$

By Lemma 2 (notice that we have not used the assumption of $G = G^0$ in the lemma):

$$\begin{aligned} & Q_{NT,G}(\hat{\boldsymbol{\beta}}, \hat{\boldsymbol{\phi}}, \hat{\gamma}) = \hat{Q}_{NT,G}(\hat{\boldsymbol{\beta}}, \hat{\boldsymbol{\phi}}, \hat{\gamma}) + o_p(NT^{-\delta}) \\ \leq\ & \hat{Q}_{NT}(\boldsymbol{\beta}^0, \boldsymbol{\phi}^0, \gamma^0) + o_p(NT^{-\delta}) = Q_{NT}(\boldsymbol{\beta}^0, \boldsymbol{\phi}^0, \gamma^0) + o_p(NT^{-\delta}) \end{aligned} \tag{107}$$

and thus

$$\left|\hat{Q}_{NT,G}(\boldsymbol{\beta}, \boldsymbol{\phi}, \gamma) - \hat{Q}_{NT}(\boldsymbol{\beta}^0, \boldsymbol{\phi}^0, \gamma^0)\right| = o_p(NT^{-\delta}) \tag{108}$$

■

S5 Proofs of Lemmas

S5.1 Lemma 1

Proof. The proof for the uniform convergence of $z_{i,t}w'_{i,t}$ and $z_{i,t}\epsilon_{i,t+h}$ are almost identical, and hence we focus on the former here. Denote by $\overline{A}_i = \frac{1}{T}\sum_t A_{i,t} = \frac{1}{T}\sum_t z_{i,t}w'_{i,t} - \mathbb{E}[z_{i,t}w'_{i,t}]$, an $L \times K$ matrix. By Boole's inequality, we have

$$\begin{aligned}
&\mathbb{P}\left(\sup_i \left\|\overline{A}_i\right\| \geq \tilde{\nu}\right) \\
\leq& N \sup_i \mathbb{P}\left(\left\|\overline{A}_i\right\| \geq \tilde{\nu}\right) \\
\leq& N \sup_i \mathbb{P}\left(\sum_l \sum_k \left\|\overline{A}_{i,l,k}\right\| \geq \tilde{\nu}\right) \\
\leq& NLK \sup_i \sup_{l,k} \mathbb{P}\left(\left|\frac{1}{T}\sum_t A_{it,l,k}\right| \geq \tilde{\nu}/\sqrt{LK}\right)
\end{aligned} \tag{109}$$

We would like to evaluate the RHS using Lemma 7. To do so, let us verify conditions (i) and (ii) of the Lemma. First notice that $A_{it,l,k} = z_{i,t,l}w_{i,t,k} - \mathbb{E}z_{i,t,l}w_{i,t,k}$ is mean-zero, stationary strong mixing with desired mixing coefficients by Assumption 1.B. So condition (i) follows. By Assumption 1.C, the tail property (ii) is also satisfied. Therefore, Lemma 7 gives (where $X_t = A_{it,l,k}$, $4\lambda = T\tilde{\nu}/\sqrt{LK}$ and $r = T^{1/2}$)

$$\begin{aligned}
&NLK \sup_i \sup_{l,k} \mathbb{P}\left(\left|\frac{1}{T}\sum_t A_{it,l,k}\right| \geq \tilde{\nu}/\sqrt{LK}\right) \\
=& NLK \sup_i \sup_{l,k} \mathbb{P}\left(\left|\sum_t A_{it,l,k}\right| \geq T\tilde{\nu}/\sqrt{LK}\right) \\
\leq& NLK \left(4\left(1 + \frac{\sqrt{T}\tilde{\nu}^2}{16LKM}\right)^{-\sqrt{T}/2} + \frac{16C\sqrt{LK}}{\tilde{\nu}} \exp\left[-c_1\left(\frac{\sqrt{T}\tilde{\nu}}{4c_3\sqrt{LK}}\right)^c\right]\right)
\end{aligned} \tag{110}$$

where $M = \sup_{t>0}\left(\mathbb{E}(A^2_{it,l,k}) + 2\sum_{s>t}|\mathbb{E}(A_{it,l,k}A_{is,l,k})|\right)$.

It remains to show the convergence rate of the RHS. We start by showing $M < \infty$ using the Davydov inequality from Lemma 8. Specifically, for $s > t$ and let $p = q = 4(1+\delta)$,

$$E\left|A_{it,l,k}A_{is,l,k}\right| \leq 8\left(\alpha(s-t)\right)^{1-\frac{1}{2(1+\delta)}}\left(\mathbb{E}\left[\left|A_{i1,l,k}\right|^{4(1+\delta)}\right]\right)^{\frac{1}{2(1+\delta)}} \ . \tag{111}$$

Then under Assumption 1.D (or Assumption 1.E when we consider $A_{it,l} = z_{it,l}\epsilon_{i,t+h}$), there exists some finite positive number C such that

$$E\left|A_{i1,l,k}\right|^2 + 2\sum_{s>t}\left|\mathbb{E}A_{it,l,k}A_{is,l,k}\right| \leq C\sum_{s>0}\left(\alpha(s)\right)^{1-\frac{1}{2(1+\delta)}} < \infty \tag{112}$$

and thus $M < \infty$.

Next we aim to show the following: for any $\tilde{\nu} > 0$, we have for all $\delta > 0$,

$$\begin{array}{c} T^{\delta}\left(1+\frac{\sqrt{T}\tilde{\nu}^2}{16LKM}\right)^{-\sqrt{T}/2} \to 0 \quad \text{as } T \to \infty \\ T^{\delta}\frac{1}{\tilde{\nu}}\exp\left[-c_1\left(\frac{\sqrt{T}\tilde{\nu}}{4c_3\sqrt{LK}}\right)^c\right] \to 0 \quad \text{as } T \to \infty \end{array} \ . \tag{113}$$

Consider for example the first part. Denote $\nu = \tilde{\nu}^2/(16LKM)$. It is equivalent to show that $\delta \ln T - \frac{1}{2}\sqrt{T}\ln(1+\nu T^{1/2})$ converges to $-\infty$ as $T \to \infty$. Given an arbitrary $\nu > 0$, we have $\ln(1+\nu T^{1/2}) > 1$ for sufficiently large T. Therefore, for sufficiently large T,

$$\delta \ln T - \frac{1}{2}\sqrt{T}\ln(1+\nu T^{1/2}) < \delta \ln T - \frac{1}{2}\sqrt{T} \tag{114}$$

and the RHS converges to $-\infty$ as $T \to \infty$. Similarly for the second part, it is equivalent to show that $\frac{1}{\tilde{\nu}}\exp\left[\delta \ln T - T^{c/2}\nu\right] \to 0$ with some arbitrary positive $\tilde{\nu}, \nu$ and c, which trivially holds as $\delta \ln T - T^{c/2}\nu$ converges to $-\infty$. Therefore, we have established that for any $\tilde{\nu} > 0$, we have for all $\delta > 0$,

$$4\left(1+\frac{\sqrt{T}\tilde{\nu}^2}{16LKM}\right)^{-\sqrt{T}/2} + \frac{16C\sqrt{LK}}{\tilde{\nu}}\exp\left[-c_1\left(\frac{\sqrt{T}\tilde{\nu}}{4c_3\sqrt{LK}}\right)^c\right] = o(T^{-\delta}) \tag{115}$$

and thus

$$\mathbb{P}\left(\sup_i \left\|\overline{A}_i\right\| \geq \tilde{\nu}\right) = o(NT^{-\delta})\ . \tag{116}$$

$\blacksquare$

S5.2 Lemma 2

Proof. By definition, we have for any $\beta \in \Theta,\ \phi \in \Phi$

$$
\begin{aligned}
&\left|\hat{Q}_{iTh}(\beta,\phi) - Q_{iTh}(\beta,\phi)\right| \\
=&\left|\hat{m}'_{i,h}\hat{\Omega}_{i,h}\hat{m}_{i,h} - m_{i,h}\Omega_{i,h}m_{i,h}\right| \\
=&\left|\hat{m}'_{i,h}\hat{\Omega}_{i,h}\hat{m}_{i,h} - \hat{m}'_{i,h}\Omega_{i,h}\hat{m}_{i,h} + \hat{m}'_{i,h}\Omega_{i,h}\hat{m}_{i,h} - m_{i,h}\Omega_{i,h}m_{i,h}\right| \\
\leq&\left|\hat{m}'_{i,h}(\hat{\Omega}_{i,h} - \Omega_{i,h})\hat{m}_{i,h}\right| + \left|\hat{m}'_{i,h}\Omega_{i,h}\hat{m}_{i,h} - m_{i,h}\Omega_{i,h}m_{i,h}\right| \\
=&\left|\hat{m}'_{i,h}(\hat{\Omega}_{i,h} - \Omega_{i,h})\hat{m}_{i,h} - m'_{i,h}(\hat{\Omega}_{i,h} - \Omega_{i,h})m_{i,h} + m'_{i,h}(\hat{\Omega}_{i,h} - \Omega_{i,h})m_{i,h}\right| \\
&+ \left|(\hat{m}_{i,h} - m_{i,h})'\Omega_{i,h}(\hat{m}_{i,h} - m_{i,h}) + 2m'_{i,h}\Omega_{i,h}(\hat{m}_{i,h} - m_{i,h})\right| \\
\leq&\left|(\hat{m}_{i,h} - m_{i,h})'(\hat{\Omega}_{i,h} - \Omega_{i,h})(\hat{m}_{i,h} - m_{i,h}) + 2m'_{i,h}(\hat{\Omega}_{i,h} - \Omega_{i,h})(\hat{m}_{i,h} - m_{i,h})\right| \\
&+ \left|m'_{i,h}(\hat{\Omega}_{i,h} - \Omega_{i,h})m_{i,h}\right| \\
&+ \left|(\hat{m}_{i,h} - m_{i,h})'\Omega_{i,h}(\hat{m}_{i,h} - m_{i,h})\right| + 2\left|m'_{i,h}\Omega_{i,h}(\hat{m}_{i,h} - m_{i,h})\right| \\
\leq&\|\hat{m}_{i,h} - m_{i,h}\|^2\left[\|\Omega_{i,h}\| + \|\hat{\Omega}_{i,h} - \Omega_{i,h}\|\right] \\
&+ 2\|m_{i,h}\|\|\hat{m}_{i,h} - m_{i,h}\|\left[\|\Omega_{i,h}\| + \|\hat{\Omega}_{i,h} - \Omega_{i,h}\|\right] \\
&+ \|m_{i,h}\|^2\|\hat{\Omega}_{i,h} - \Omega_{i,h}\|
\end{aligned}
\tag{117}
$$

where we have repeatedly used the triangle inequality, the matrix identity (29) and the matrix inequalities (27).

Next, we take the supremum oevr i and $\beta \in \Theta, \phi \in \Phi$. Since both sides are positive, it boils down to taking the supremum over for each individual terms. Let us examine each terms separately. First, consider $m_{i,h}$. Plug in the definition, we have[3]

$$
\begin{aligned}
m_{i,h}(\beta,\phi) =& \ \mathbb{E}[\hat{m}_{i,h}(\beta,\phi)] = \mathbb{E}[d_{zy,i,h} - d_{zx,i}\beta - d_{zc,i}\phi] \\
=& \ \mathbb{E}\left[z_{it}x'_{it}\left(\beta^0_{g_i^0,h} - \beta\right) + z_{it}c'_{it}\left(\phi^0_{i,h} - \phi\right) + z_{it}\epsilon_{it+h}\right] \\
=& \ d_{zx,i}\left(\beta^0_{g_i^0,h} - \beta\right) + d_{zc,i}\left(\phi^0_{i,h} - \phi\right) = d_{zw,i}(\theta^0_{i,h} - \theta)
\end{aligned}
\tag{118}
$$

where with a slight abuse of notation, I write $\theta = (\beta', \phi')'$. Taking the supremum gives, for

[3] I use here $m_{i,h}(\beta,\phi)$ to explicitly indicate that the moment function is a function of β and ϕ.

all $h = 0, \ldots, H$,

$$
\begin{aligned}
&\sup_i \sup_{\beta\in\Theta,\phi\in\Phi} \|m_{i,h}(\beta,\phi)\| \\
\leq& \sup_i \|d_{zw,i}\| \sup_i \sup_{\beta\in\Theta,\phi\in\Phi} \|\theta^0_{i,h} - \theta\| \qquad\qquad . \\
\leq& \left(\sup_i \mathbb{E}\|z_{i,t}w'_{i,t}\|\right) \left(\sup_i \sup_{\beta\in\Theta,\phi\in\Phi} \|\theta^0_{i,h} - \theta\|\right) < \infty
\end{aligned}
\tag{119}
$$

The first inequality follows from Cauchy-Schwarz inequality (26); the second inequality applies Jensen's inequality to the expectation operator; the last inequality comes from Assumption 1.D and 1.A. In particular, Assumption 1.A states that the parameter space Θ and Φ are compact. Therefore, there exists some finite constants $\infty > C_1 > 0$ such that $\mathrm{Diam}(\Theta), \mathrm{Diam}(\Phi) \leq C_1$. Given (119), it follows immediately that

$$
\sup_i \sup_{\beta\in\Theta,\phi\in\Phi} \|m_{i,h}(\beta,\phi)\|^2 < \infty \; . \tag{120}
$$

Second, by Assumption 1.F the weighting matrix $\Omega_{i,h}$ is finite positive definite, then there exists some finite positive constant C_3 such that $\sup_i \|\Omega_{i,h}\| < C_3 < \infty$. Moreover, we also assume that $\sup_i \|\hat{\Omega}_{i,h} - \Omega_{i,h}\| = o_p(1)$.

Third, expanding $\hat{m}_{i,h} - m_{i,h}$ gives

$$
\hat{m}_{i,h}(\beta,\phi) - m_{i,h}(\beta,\phi) = (\overline{d}_{zw,i} - d_{zw,i})\left(\theta^0_{i,h} - \theta\right) + d_{z\epsilon,i,h} \tag{121}
$$

Taking the supremum gives

$$
\begin{aligned}
&\sup_i \sup_{\beta\in\Theta,\phi\in\Phi} \|(\hat{m}_{i,h} - m_{i,h})(\beta,\phi)\| \\
\leq& \sup_i \|\overline{d}_{zw,i} - d_{zw,i}\| \sup_i \sup_{\beta\in\Theta,\phi\in\Phi} \|\theta^0_{i,h} - \theta\| + \sup_i \|d_{z\epsilon,i,h}\| = o_p(NT^{-\delta})
\end{aligned}
\tag{122}
$$

where the last equality follows from Lemma 1, and we again use the assumption that the parameter space is compact. Given (122), it follows immediately that

$$
\begin{aligned}
& \sup_i \sup_{\beta\in\Theta,\phi\in\Phi} \|(\hat{m}_{i,h} - m_{i,h})(\beta,\phi)\|^2 \\
= \; & \sup_i \sup_{\beta\in\Theta,\phi\in\Phi} \|(\overline{d}_{zw,i} - d_{zw,i})(\theta^0_{i,h} - \theta_{i,h}) + d_{z\epsilon,i,h}\|^2 \\
\leq \; & 2\sup_i \|\overline{d}_{zw,i} - d_{zw,i}\|^2 \sup_i \sup_{\beta\in\Theta,\phi\in\Phi} \|\theta^0_{i,h} - \theta\|^2 + 2\|d_{z\epsilon,i,h}\|^2 = o_p(NT^{-\delta})
\end{aligned} \quad . \tag{123}
$$

Now combine the results above, and taking supremum over i and the parameter space $\beta \in \Theta,\ \phi \in \Phi$ of (168), we have

$$
\begin{aligned}
& \sup_i \sup_{\beta\in\Theta,\ \phi\in\Phi} \left|\hat{Q}_{iTh}(\beta_{g_i,h},\phi_{i,h}) - Q_{iTh}(\beta_{g_i,h},\phi_{i,h})\right| \\
\leq \; & \sup_i \sup_{\beta\in\Theta,\ \phi\in\Phi} \|\hat{m}_{i,h} - m_{i,h}\|^2 \left[\|\Omega_{i,h}\| + \|\hat{\Omega}_{i,h} - \Omega_{i,h}\|\right] \\
& + 2\sup_i \sup_{\beta\in\Theta,\ \phi\in\Phi} \|m_{i,h}\|\|\hat{m}_{i,h} - m_{i,h}\| \left[\|\Omega_{i,h}\| + \|\hat{\Omega}_{i,h} - \Omega_{i,h}\|\right] \\
& + \sup_i \sup_{\beta\in\Theta,\ \phi\in\Phi} \|m_{i,h}\|^2 \|\hat{\Omega}_{i,h} - \Omega_{i,h}\| \\
\leq \; & C_1 \sup_i \|\overline{d}_{zw,i} - d_{zw,i}\| + C_2 \sup_i \|d_{z\epsilon,i,h}\| + C_3 \sup_i \|\hat{\Omega}_{i,h} - \Omega_{i,h}\| \\
\leq \; & (C_1 + C_2 + C_3) o(NT^{-\delta}) = o_p(NT^{-\delta})
\end{aligned} \tag{124}
$$

where C_1, C_2, C_3 are some finite positive constant and the last inequality follows from Lemma 1.

Part II follows immediately:

$$
\begin{aligned}
& \sup_{\boldsymbol{\beta}\in\Theta_G,\boldsymbol{\phi}\in\Phi_N,\gamma\in\mathcal{G}} \left|\hat{Q}_{NT}(\boldsymbol{\beta},\boldsymbol{\phi},\gamma) - Q_{NT}(\boldsymbol{\beta},\boldsymbol{\phi},\gamma)\right| \\
\leq & \sup_{\boldsymbol{\beta}\in\Theta_G,\boldsymbol{\phi}\in\Phi_N,\gamma\in\mathcal{G}} \left|\frac{1}{N}\sum_i \sum_h \left[\hat{Q}_{iTh}(\beta_{g_i,h},\phi_{i,h}) - Q_{iTh}(\beta_{g_i,h},\phi_{i,h})\right]\right| \\
\leq & H \max_h \sup_i \sup_{\beta\in\Theta,\ \phi\in\Phi} \left|\hat{Q}_{iTh}(\beta,\phi) - Q_{iTh}(\beta,\phi)\right| \leq H o_p(1) = o_p(1)\ .
\end{aligned} \tag{125}
$$

■

S5.3 Lemma 3

Proof. For notational simplicity, denote $\beta^d_{i,h} = \beta^0_{g^0_i,h} - \beta_{g_i,h}$ and $\phi^d_{i,h} = \phi^0_{i,h} - \phi_{i,h}$. By definition (35), we have

$$\begin{aligned}&Q_{NT}(\boldsymbol{\beta}, \boldsymbol{\phi}, \gamma) - Q_{NT}(\boldsymbol{\beta}^0, \boldsymbol{\phi}^0, \gamma^0)\\ =&\frac{1}{N}\sum_h \sum_i \left[d_{zx,i}\beta^d_{i,h} + d_{zc,i}\phi^d_{i,h}\right]' \Omega_{i,h} \left[d_{zx,i}\beta^d_{i,h} + d_{zc,i}\phi^d_{i,h}\right] \ .\end{aligned} \tag{126}$$

Define $\phi^{d*}_{i,h}$ to be the minimizer of the above objective function, i.e.

$$\phi^{d*}_{i,h} = \arg\min_{\phi^d_{i,h}} \frac{1}{N}\sum_h \sum_i \left[d_{zx,i}\beta^d_{i,h} + d_{zc,i}\phi^d_{i,h}\right]' \Omega_{i,h} \left[d_{zx,i}\beta^d_{i,h} + d_{zc,i}\phi^d_{i,h}\right] \tag{127}$$

First order condition for this minimization problem gives

$$d'_{zc,i}\Omega_{i,h}\left(d_{zx,i}\beta^d_{i,h} + d_{zc,i}\phi^{d*}_{i,h}\right) = 0 \tag{128}$$

and thus

$$\phi^{d*}_{i,h} = -\left[d'_{zc,i}\Omega_{i,h}d_{zc,i}\right]^{-1}\left[d'_{zc,i}\Omega_{i,h}d_{zx,i}\right]\beta^d_{i,h} \ . \tag{129}$$

Using the definition of $\phi^{d*}_{i,h}$ and the first order condition, we have

$$\begin{aligned}&Q_{NT}(\boldsymbol{\beta}, \boldsymbol{\phi}, \gamma) - Q_{NT}(\boldsymbol{\beta}^0, \boldsymbol{\phi}^0, \gamma^0)\\ \geqslant&\frac{1}{N}\sum_h \sum_i \left[d_{zx,i}\beta^d_{i,h} + d_{zc,i}\phi^{d*}_{i,h}\right]' \Omega_{i,h} \left[d_{zx,i}\beta^d_{i,h} + d_{zc,i}\phi^{d*}_{i,h}\right]\\ =&\frac{1}{N}\sum_h \sum_i (\beta^0_{g^0_i,h} - \beta_{g_i,h})' \left(d'_{zx,i}M_{zc,i,h}d_{zx,i}\right)(\beta^0_{g^0_i,h} - \beta_{g_i,h})\\ =&\frac{1}{N}\sum_h \sum_i \sum_g \sum_{\tilde{g}} \mathbf{1}\{g^0_i = g\}\mathbf{1}\{g_i = \tilde{g}\}(\beta^0_{g,h} - \beta_{\tilde{g},h})' \left(d'_{zx,i}M_{zc,i,h}d_{zx,i}\right)(\beta^0_{g,h} - \beta_{\tilde{g},h})\\ =&\sum_h \sum_g \sum_{\tilde{g}} (\beta^0_{g,h} - \beta_{\tilde{g},h})' \left[\frac{1}{N}\sum_i \mathbf{1}\{g^0_i = g, g_i = \tilde{g}\}\left(d'_{zx,i}M_{zc,i,h}d_{zx,i}\right)\right](\beta^0_{g,h} - \beta_{\tilde{g},h})\\ \geqslant&\sum_h \sum_g \sum_{\tilde{g}} \rho_{min,g\tilde{g},h}\|\beta^0_{g,h} - \beta_{\tilde{g},h}\|^2\end{aligned}$$

where $M_{zc,i,h} = \Omega_{i,h} - \Omega_{i,h} d_{zc,i}(d'_{zc,i}\Omega_{i,h}d_{zc,i})^{-1}d'_{zc,i}\Omega_{i,h}$ is defined as in Section 3 in the main text, and $\rho_{min,g\tilde{g},h}$ is the minimal eigenvalues of $\frac{1}{N}\sum_i \mathbf{1}\{g_i^0 = g, g_i = \tilde{g}\}\left(d'_{zx,i}M_{zc,i,h}d_{zx,i}\right)$.

Next I show that $\rho_{min,g\tilde{g},h} > \underline{\rho}$ for some $\underline{\rho}$, by showing that $d'_{zx,i}M_{zc,i,h}d_{zx,i}$ is positive definite. Notice that by Cholesky decomposition $\Omega_{i,h} = L_{i,h}L'_{i,h}$, we can write it as

$$
\begin{aligned}
d'_{zx,i}M_{zc,i,h}d_{zx,i} = &\; d'_{zx,i}\left[\Omega_{i,h} - \Omega_{i,h}d_{zc,i}(d'_{zc,i}\Omega_{i,h}d_{zc,i})^{-1}d'_{zc,i}\Omega_{i,h}\right]d_{zx,i} \\
= &\; d'_{zx,i}L_{i,h}\left[I - L'_{i,h}d_{zc,i}(d'_{zc,i}\Omega_{i,h}d_{zc,i})^{-1}d'_{zc,i}L_{i,h}\right]L'_{i,h}d_{zx,i} \\
= &\; \left(L'_{i,h}d_{zx,i}\right)'\widetilde{M}_{zc,i,h}\left(L'_{i,h}d_{zx,i}\right) \\
= &\; \left(\widetilde{M}_{zc,i,h}L'_{i,h}d_{zx,i}\right)'\left(\widetilde{M}_{zc,i,h}L'_{i,h}d_{zx,i}\right)
\end{aligned} \tag{130}
$$

where the last equality comes from the fact that $\widetilde{M}_{zc,i,h}$ is idempotent. Specifically, we have

$$
\begin{aligned}
&\widetilde{M}'_{zc,i,h}\widetilde{M}_{zc,i,h} \\
= &\left[I - L'_{i,h}d_{zc,i}(d'_{zc,i}\Omega_{i,h}d_{zc,i})^{-1}d'_{zc,i}L_{i,h}\right]'\left[I - L'_{i,h}d_{zc,i}(d'_{zc,i}\Omega_{i,h}d_{zc,i})^{-1}d'_{zc,i}L_{i,h}\right] \\
= &\; I - L'_{i,h}d_{zc,i}(d'_{zc,i}\Omega_{i,h}d_{zc,i})^{-1}d'_{zc,i}L_{i,h} - L'_{i,h}d_{zc,i}(d'_{zc,i}\Omega_{i,h}d_{zc,i})^{-1}d'_{zc,i}L_{i,h} \\
&+ L'_{i,h}d_{zc,i}(d'_{zc,i}\Omega_{i,h}d_{zc,i})^{-1}d'_{zc,i}L_{i,h}L'_{i,h}d_{zc,i}(d'_{zc,i}\Omega_{i,h}d_{zc,i})^{-1}d'_{zc,i}L_{i,h} \\
= &\; I - L'_{i,h}d_{zc,i}(d'_{zc,i}\Omega_{i,h}d_{zc,i})^{-1}d'_{zc,i}L_{i,h}
\end{aligned} .
$$

Therefore, to show positive definiteness is to show that

$$
\begin{aligned}
\widetilde{M}_{zc,i,h}L'_{i,h}d_{zx,i} = &\; \left[I - L'_{i,h}d_{zc,i}(d'_{zc,i}\Omega_{i,h}d_{zc,i})^{-1}d'_{zc,i}L_{i,h}\right]L'_{i,h}d_{zx,i} \\
= &\; L'_{i,h}d_{zx,i} - L'_{i,h}d_{zc,i}(d'_{zc,i}\Omega_{i,h}d_{zc,i})^{-1}d'_{zc,i}\Omega_{i,h}d_{zx,i}
\end{aligned} \tag{131}
$$

is of full column rank.

We prove this by contradiction. Assume that (131) is does not have full column rank, then there exists a vector $\alpha \neq 0$ such that

$$
L'_{i,h}d_{zx,i}\alpha = L'_{i,h}d_{zc,i}(d'_{zc,i}\Omega_{i,h}d_{zc,i})^{-1}d'_{zc,i}\Omega_{i,h}d_{zx,i}\alpha \tag{132}
$$

Observe that by Assumptions 1.D and 1.F, $d'_{zc,i}\Omega_{i,h}d_{zc,i}$ has full rank and invertible, and

$d'_{zc,i}\Omega_{i,h}d_{zx,i}$ is of full column rank. Then $(d'_{zc,i}\Omega_{i,h}d_{zc,i})^{-1}d'_{zc,i}\Omega_{i,h}d_{zx,i}\alpha \neq 0$. As a result, there exists some $\tau \neq 0$ such that

$$L'_{i,h}d_{zx,i}\alpha = L'_{i,h}d_{zc,i}\tau \ . \tag{133}$$

It follows that $d_{zx,i}$ and $d_{zc,i}$ are linearly dependent, violating our rank condition 1.D. Therefore, (131) is of full column rank and $d'_{zx,i}M_{zc,i,h}d_{zx,i}$ is positive definite.

Notice that the above holds for any i, and thus under any partition over $i = 1, \ldots, N$ we have $\rho_{min,g\tilde{g},h} \geqslant \underline{\rho} > 0$. Then we have

$$\begin{aligned}
& Q_{NT}(\boldsymbol{\beta}, \boldsymbol{\phi}, \gamma) - Q_{NT}(\boldsymbol{\beta}^0, \boldsymbol{\phi}^0, \gamma^0) \\
\geqslant & \sum_g \sum_{\tilde{g}} \min_h \rho_{min,g\tilde{g},h} \left(\sum_h \|\beta^0_{g,h} - \beta_{\tilde{g},h}\|^2 \right) \\
\geqslant & \sum_g \left[\sum_{\tilde{g}} \min_h \rho_{min,g\tilde{g},h} \left(\min_{\tilde{g}\in\{1,\ldots,G^0\}} \|\beta^0_g - \beta_{\tilde{g}}\|^2 \right) \right] \\
\geqslant & \sum_g \left[\max_{\tilde{g}\in\{1,\ldots,G^0\}} \min_h \rho_{min,g\tilde{g},h} \right] \left[\min_{\tilde{g}\in\{1,\ldots,G^0\}} \|\beta^0_g - \beta_{\tilde{g}}\|^2 \right] \\
\geqslant & \sum_g \left[\min_{\gamma\in\mathcal{G}} \max_{\tilde{g}\in\{1,\ldots,G^0\}} \min_h \rho_{min,g\tilde{g},h} \right] \left[\min_{\tilde{g}\in\{1,\ldots,G^0\}} \|\beta^0_g - \beta_{\tilde{g}}\|^2 \right] \\
\geqslant & \max_{g\in\{1,\ldots,G^0\}} \left[\min_{\gamma\in\mathcal{G}} \max_{\tilde{g}\in\{1,\ldots,G^0\}} \min_h \rho_{min,g\tilde{g},h} \right] \left[\min_{\tilde{g}\in\{1,\ldots,G^0\}} \|\beta^0_g - \beta_{\tilde{g}}\|^2 \right] \ .
\end{aligned} \tag{134}$$

There are five inequalities in (134). The first, the second and the fourth are obtained by construction, i.e.,

$$\begin{aligned}
\textstyle\sum_h \rho_{min,g\tilde{g},h}\|\beta^0_{g,h} - \beta_{\tilde{g},h}\|^2 &\geqslant \textstyle\sum_h \left(\min_h \rho_{min,g\tilde{g},h}\right) \|\beta^0_{g,h} - \beta_{\tilde{g},h}\|^2 \\
\|\beta^0_g - \beta_{\tilde{g}}\|^2 &\geqslant \min_{\tilde{g}\in\{1,\ldots,G^0\}} \|\beta^0_g - \beta_{\tilde{g}}\|^2 \\
\max_{\tilde{g}\in\{1,\ldots,G^0\}} \min_h \rho_{min,g\tilde{g},h} &\geqslant \min_{\gamma\in\mathcal{G}} \max_{\tilde{g}\in\{1,\ldots,G^0\}} \min_h \rho_{min,g\tilde{g},h}
\end{aligned} \tag{135}$$

The third and the last inequalities follow the same logic: since each single elements in the summation is positive, the summation is larger than any single element, including the maximal one, i.e., $\sum_i a_i \geqslant \max_i a_i$ as long as $a_i \geqslant 0$. Equation (134) implies that the auxiliary function $\tilde{Q}(\boldsymbol{\beta}, \boldsymbol{\phi}, \gamma)$ is uniquely minimized at the true parameter values, since the RHS is bounded away from zero.

The next step is to bound the LHS using Lemma 2. Specifically, we have

$$\begin{aligned} & Q_{NT}(\hat{\boldsymbol{\beta}}, \hat{\boldsymbol{\phi}}, \hat{\gamma}) = \hat{Q}_{NT}(\hat{\boldsymbol{\beta}}, \hat{\boldsymbol{\phi}}, \hat{\gamma}) + o_p(1) \\ \leq \ & \hat{Q}_{NT}(\boldsymbol{\beta}^0, \boldsymbol{\phi}^0, \gamma^0) + o_p(1) = Q_{NT}(\boldsymbol{\beta}^0, \boldsymbol{\phi}^0, \gamma^0) + o_p(1) \end{aligned} \tag{136}$$

where $\hat{Q}$ is by definition minimized at $(\hat{\boldsymbol{\beta}}, \hat{\boldsymbol{\phi}}, \hat{\gamma})$. Rearrange terms,

$$o_p(1) \geqslant Q_{NT}(\hat{\boldsymbol{\beta}}, \hat{\boldsymbol{\phi}}, \hat{\gamma}) - Q_{NT}(\boldsymbol{\beta}^0, \boldsymbol{\phi}^0, \gamma^0)\ . \tag{137}$$

Combining (134) and (137):

$$\begin{aligned} o_p(1) \geqslant & Q_{NT}(\hat{\boldsymbol{\beta}}, \hat{\boldsymbol{\phi}}, \hat{\gamma}) - Q_{NT}(\boldsymbol{\beta}^0, \boldsymbol{\phi}^0, \gamma^0) \\ \geqslant & \max_{g \in \{1,\ldots,G^0\}} \left[\min_{\gamma \in \mathcal{G}} \max_{\tilde{g} \in \{1,\ldots,G^0\}} \min_h \rho_{min, g\tilde{g}, h} \right] \left[\min_{\tilde{g} \in \{1,\ldots,G^0\}} \|\beta_g^0 - \beta_{\tilde{g}}\|^2 \right] \geqslant 0. \end{aligned} \tag{138}$$

Written compactly, we have

$$o_p(1) \geqslant \max_{g \in \{1,\ldots,G^0\}} \left[\min_{\gamma \in \mathcal{G}} \max_{\tilde{g} \in \{1,\ldots,G^0\}} \min_h \rho_{min, g\tilde{g}, h} \right] \left[\min_{\tilde{g} \in \{1,\ldots,G^0\}} \|\beta_g^0 - \hat{\beta}_{\tilde{g}}\|^2 \right] \geqslant 0$$

Since $\min_{\gamma \in \mathcal{G}} \max_{\tilde{g} \in \{1,\ldots,G^0\}} \min_h \rho_{min, g\tilde{g}, h}$ is strictly positive and *finite*, the above inequality implies

$$o_p(1) = \max_{g \in \{1,\ldots,G^0\}} \left[\min_{\tilde{g} \in \{1,\ldots,G^0\}} \|\beta_g^0 - \hat{\beta}_{\tilde{g}}\|^2 \right]. \tag{139}$$

which proves half of Lemma 3.

Finally, we prove the remaining half of the lemma. First define for any given grouping g the partition that minimizes the distance of IRs as

$$\sigma(g) = \underset{\tilde{g} \in \{1,\ldots,G^0\}}{\arg\min} \|\beta_g^0 - \hat{\beta}_{\tilde{g}}\|^2. \tag{140}$$

We then want to show that the mapping $\sigma(g):\{1,\ldots,G^0\}\to\{1,\ldots,G^0\}$ is one-to-one with probability approaching 1 so that $\sigma(g)^{-1}$ is well-defined. Notice that

$$\beta_g^0-\beta_{\tilde{g}}^0=(\beta_g^0-\hat{\beta}_{\sigma(g)})+(\hat{\beta}_{\sigma(g)}-\hat{\beta}_{\sigma(\tilde{g})})+(\hat{\beta}_{\sigma(\tilde{g})}-\beta_{\tilde{g}}^0). \tag{141}$$

By triangle inequality, we have

$$\left(\|\beta_g^0-\beta_{\tilde{g}}^0\|^2\right)^{\frac{1}{2}}\leq\underbrace{\left(\|\beta_g^0-\hat{\beta}_{\sigma(g)}\|^2\right)^{\frac{1}{2}}}_{o_p(1)}+\left(\|\hat{\beta}_{\sigma(g)}-\hat{\beta}_{\sigma(\tilde{g})}\|^2\right)^{\frac{1}{2}}+\underbrace{\left(\|\hat{\beta}_{\sigma(\tilde{g})}-\beta_{\tilde{g}}^0\|^2\right)^{\frac{1}{2}}}_{o_p(1)} \tag{142}$$

where the first and the third term on the RHS are $o_p(1)$ by (139) and (140). Moreover, the LHS is strictly positive as long as $g\neq\tilde{g}$ by Assumption 2.B. Therefore, the above inequality states that for any $g\neq l$, we have

$$\|\hat{\beta}_{\sigma(g)}-\hat{\beta}_{\sigma(l)}\|^2>0\ .$$

That is, with probability approaching to 1, $\sigma(g)\neq\sigma(\tilde{g})$ for all $g\neq\tilde{g}$. Hence, the mapping $\sigma(g)$ is one-to-one with probability approaching to 1.

Now we are ready to look at the second part of the Hausdorff distance between β^0 and $\hat{\beta}$. We have for all $\tilde{g}\in\{1,\ldots,G^0\}$,

$$\min_{g\in\{1,\ldots,G^0\}}\|\beta_g^0-\hat{\beta}_{\tilde{g}}\|^2\leq\underbrace{\|\beta^0_{\sigma^{-1}(\tilde{g})}-\hat{\beta}_{\tilde{g}}\|^2}_{\text{by construction}}=\underbrace{\min_{l\in\{1,\ldots,G^0\}}\|\beta^0_{\sigma^{-1}(\tilde{g})}-\hat{\beta}_l\|^2}_{\text{by definition of }\sigma}=\underbrace{o_p(1)}_{\text{by (139)}}\ . \tag{143}$$

The above inequality holds for all possible partition $\tilde{g}$, and thus

$$\max_{\tilde{g}\in\{1,\ldots,G^0\}}\left[\min_{g\in\{1,\ldots,G^0\}}\|\beta_g^0-\hat{\beta}_{\tilde{g}}\|^2\right]=o_p(1) \tag{144}$$

Given (139) and (144), we have $d_H(\hat{\boldsymbol{\beta}},\boldsymbol{\beta}^0)\xrightarrow{p}0$ by definition. ■

S5.4 Lemma 4

Proof. By definition, we can write $\hat{\phi}_{i,h}(\beta_{g_i,h})$ as

$$\begin{aligned}\hat{\phi}_{i,h}(\beta_{g_i,h}) = \quad & \phi^0_{i,h} + \left(\bar{d}'_{zc,i}\hat{\Omega}_{i,h}\bar{d}_{zc,i}\right)^{-1}\bar{d}'_{zc,i}\hat{\Omega}_{i,h}\bar{d}_{zx,i}\left[\beta^0_{g^0_i,h} - \beta_{g_i}\right] \\ & + \left(\bar{d}'_{zc,i}\hat{\Omega}_{i,h}\bar{d}_{zc,i}\right)^{-1}\bar{d}'_{zc,i}\hat{\Omega}_{i,h}\bar{d}_{z\epsilon,i,h}\end{aligned} \tag{145}$$

Rearranging terms gives

$$\begin{aligned}\|\hat{\phi}_{i,h}(\beta_{g_i,h}) - \phi^0_{i,h}\| \leq \quad & \left\| \left(\bar{d}'_{zc,i}\hat{\Omega}_{i,h}\bar{d}_{zc,i}\right)^{-1}\bar{d}'_{zc,i}\hat{\Omega}_{i,h}\bar{d}_{zx,i}\left[\beta^0_{g^0_i,h} - \beta_{g_i}\right] \right\| \\ & + \left\| \left(\bar{d}'_{zc,i}\hat{\Omega}_{i,h}\bar{d}_{zc,i}\right)^{-1}\bar{d}'_{zc,i}\hat{\Omega}_{i,h}\bar{d}_{z\epsilon,i,h} \right\| \\ \leq \quad & \left\| \left(\bar{d}'_{zc,i}\hat{\Omega}_{i,h}\bar{d}_{zc,i}\right)^{-1} \right\| \times \left\| \bar{d}'_{zc,i}\hat{\Omega}_{i,h}\bar{d}_{zx,i} \right\| \times \left\| \beta^0_{g^0_i,h} - \beta_{g_i} \right\| \\ & + \left\| \left(\bar{d}'_{zc,i}\hat{\Omega}_{i,h}\bar{d}_{zc,i}\right)^{-1} \right\| \times \left\| \bar{d}'_{zc,i}\hat{\Omega}_{i,h}\bar{d}_{z\epsilon,i,h} \right\|\end{aligned} \tag{146}$$

Consider for example the first term, we want to show that

$$\left\| \bar{d}'_{zc,i}\hat{\Omega}_{i,h}\bar{d}_{zc,i} - d_{zc,i}\Omega_{i,h}d_{zc,i} \right\| \leq o_p(1) \tag{147}$$

By add and subtract, we have

$$\begin{aligned} & \left\| \bar{d}'_{zc,i}\hat{\Omega}_{i,h}\bar{d}_{zc,i} - d_{zc,i}\Omega_{i,h}d_{zc,i} \right\| \\ = & \left\| \bar{d}'_{zc,i}(\hat{\Omega}_{i,h} - \Omega_{i,h})\bar{d}_{zc,i} + \bar{d}'_{zc,i}\Omega_{i,h}(\bar{d}_{zc,i} - d_{zc,i}) + (\bar{d}_{zc,i} - d_{zc,i})'\Omega_{i,h}d_{zc,i} \right\| \\ \leq & \left\| \bar{d}'_{zc,i}(\hat{\Omega}_{i,h} - \Omega_{i,h})\bar{d}_{zc,i} \right\| + \left\| \bar{d}'_{zc,i}\Omega_{i,h}(\bar{d}_{zc,i} - d_{zc,i}) \right\| + \left\| (\bar{d}_{zc,i} - d_{zc,i})'\Omega_{i,h}d_{zc,i} \right\| \\ \leq & \left\| \bar{d}_{zc,i} \right\|^2 \left\| \hat{\Omega}_{i,h} - \Omega_{i,h} \right\| + \left\| \bar{d}'_{zc,i}\Omega_{i,h} \right\| \left\| \bar{d}_{zc,i} - d_{zc,i} \right\| + \left\| \bar{d}_{zc,i} - d_{zc,i} \right\| \left\| \Omega_{i,h}d_{zc,i} \right\| \end{aligned} \tag{148}$$

Taking the supremum over i on both sides, by Lemma 1 and the assumption on the weighting matrix 1.F, we have

$$\sup_i \left\| \bar{d}'_{zc,i}\hat{\Omega}_{i,h}\bar{d}_{zc,i} - d_{zc,i}\Omega_{i,h}d_{zc,i} \right\| = o_p(1)\ . \tag{149}$$

Therefore, we have $\sup_i \left\| \bar{d}'_{zc,i} \hat{\Omega}_{i,h} \bar{d}_{zc,i} \right\| = O_p(1)$. Results for other terms similarly follows. Taken together, we have

$$\sup_i \|\hat{\phi}_{i,h}(\beta_{g_i,h}) - \phi^0_{i,h}\| \leq O_p(1) \sup_i \left\| \beta^0_{g^0_i,h} - \beta_{g_i} \right\| + o_p(1) = o_p(1) \tag{150}$$

where the last inequality comes from the assumption that $\sup_i \left\| \beta^0_{g^0_i,h} - \beta_{g_i} \right\| = o_p(1)$. ■

S5.5 Lemma 5

Proof. By construction, we have

$$\hat{g}_i(\boldsymbol{\beta}) = g \quad \Longrightarrow \quad \sum_h \hat{Q}_{iTh}(\beta_{g,h}, \hat{\phi}_{i,h}(\beta_{g,h})) \leq \sum_h \hat{Q}_{iTh}(\beta_{\tilde{g},h}, \hat{\phi}_{i,h}(\beta_{\tilde{g},h})) \tag{151}$$

for arbitrary $\tilde{g}$, which implies

$$\mathbf{1}\left\{\hat{g}_i(\boldsymbol{\beta}) = g\right\} \leq \mathbf{1}\left\{\sum_h \hat{Q}_{iTh}(\beta_{g,h}, \hat{\phi}_{i,h}(\beta_{g,h})) \leq \sum_h \hat{Q}_{iTh}(\beta_{\tilde{g},h}, \hat{\phi}_{i,h}(\beta_{\tilde{g},h}))\right\}. \tag{152}$$

As a result, we have

$$\begin{aligned}
&\mathbf{1}\{\hat{g}_i(\boldsymbol{\beta}) \neq g_i^0\} \\
=& \sum_g \mathbf{1}\{g_i^0 \neq g\}\mathbf{1}\{\hat{g}_i(\boldsymbol{\beta}) = g\} \\
=& \sum_g \sum_{\tilde{g}\neq g} \mathbf{1}\{g_i^0 = \tilde{g}\}\mathbf{1}\{\hat{g}_i(\boldsymbol{\beta}) = g\} \\
\leq& \sum_g \max_{\tilde{g}\neq g} \mathbf{1}\{g_i^0 = \tilde{g}\}\mathbf{1}\left\{\hat{g}_i(\boldsymbol{\beta}) = g\right\} \\
\leq& \sum_g \max_{\tilde{g}\in\{1,\ldots,G^0\}} \mathbf{1}\left\{g_i^0 = \tilde{g}, \tilde{g} \neq g\right\}\mathbf{1}\left\{\sum_h \hat{Q}_{iTh}(\beta_{g,h}, \hat{\phi}_{i,h}(\beta_{g,h})) \leq \sum_h \hat{Q}_{iTh}(\beta_{\tilde{g},h}, \hat{\phi}_{i,h}(\beta_{\tilde{g},h}))\right\}
\end{aligned}$$

Now let us focus on the last indicator function. The idea is to use Lemma 2 that

$$\sup_i \sup_{\beta\in\Theta,\ \phi\in\Phi} \left|\hat{Q}_{iTh}(\beta, \phi) - Q_{iTh}(\beta, \phi)\right| = o_p(NT^{-\delta}) \tag{153}$$

and the fact that the population objective function is uniquely minimized at the true parameter values. In order to do this, suppose there exists some $\Delta > 0$ such that

$$\Delta - \sum_h Q_{iTh}(\beta_{g,h}, \hat{\phi}_{i,h}(\beta_{g,h})) + \sum_h Q_{iTh}(\beta_{\tilde{g},h}, \hat{\phi}_{i,h}(\beta_{\tilde{g},h})) \leq 0 \tag{154}$$

The by probability algebra, we have :

$$
\begin{aligned}
&\mathbf{1}\left\{\sum_h \hat{Q}_{iTh}(\beta_{g,h},\hat{\phi}_{i,h}(\beta_{g,h})) \leq \sum_h \hat{Q}_{iTh}(\beta_{\tilde{g},h},\hat{\phi}_{i,h}(\beta_{\tilde{g},h}))\right\} \\
=&\mathbf{1}\left\{0 \leq \sum_h \hat{Q}_{iTh}(\beta_{\tilde{g},h},\hat{\phi}_{i,h}(\beta_{\tilde{g},h})) - \sum_h \hat{Q}_{iTh}(\beta_{g,h},\hat{\phi}_{i,h}(\beta_{g,h}))\right\} \\
\leq&\mathbf{1}\left\{\Delta - \sum_h Q_{iTh}(\beta_{g,h},\hat{\phi}_{i,h}(\beta_{g,h})) + \sum_h Q_{iTh}(\beta_{\tilde{g},h},\hat{\phi}_{i,h}(\beta_{\tilde{g},h}))\right. \\
&\qquad\qquad \left.\leq \sum_h \hat{Q}_{iTh}(\beta_{\tilde{g},h},\hat{\phi}_{i,h}(\beta_{\tilde{g},h})) - \sum_h \hat{Q}_{iTh}(\beta_{g,h},\hat{\phi}_{i,h}(\beta_{g,h}))\right\} \\
=&\mathbf{1}\left\{\Delta \leq \left(\sum_h \hat{Q}_{iTh}(\beta_{\tilde{g},h},\hat{\phi}_{i,h}(\beta_{\tilde{g},h})) - \sum_h Q_{iTh}(\beta_{\tilde{g},h},\hat{\phi}_{i,h}(\beta_{\tilde{g},h}))\right)\right. \\
&\qquad\qquad \left. - \left(\sum_h Q_{iTh}(\beta_{g,h},\hat{\phi}_{i,h}(\beta_{g,h})) - \sum_h \hat{Q}_{iTh}(\beta_{g,h},\hat{\phi}_{i,h}(\beta_{g,h}))\right)\right\} \\
\leq&\mathbf{1}\left\{\Delta \leq \sum_h \left|\hat{Q}_{iTh}(\beta_{\tilde{g},h},\hat{\phi}_{i,h}(\beta_{\tilde{g},h})) - Q_{iTh}(\beta_{\tilde{g},h},\hat{\phi}_{i,h}(\beta_{\tilde{g},h}))\right|\right. \\
&\qquad\qquad \left. + \sum_h \left|Q_{iTh}(\beta_{g,h},\hat{\phi}_{i,h}(\beta_{g,h})) - \hat{Q}_{iTh}(\beta_{g,h},\hat{\phi}_{i,h}(\beta_{g,h}))\right|\right\} \\
\leq&\mathbf{1}\left\{\Delta \leq 2\sum_h \sup_{\beta\in\Theta,\phi\in\Phi} \left|\hat{Q}_{iTh}(\beta,\phi) - Q_{iTh}(\beta,\phi)\right|\right\}
\end{aligned}
$$

Next we show that Δ does exists. Rewrite (154), we would like to find $\Delta > 0$ such that

$$
\sum_h Q_{iTh}(\beta_{g,h},\hat{\phi}_{i,h}(\beta_{g,h})) - \sum_h Q_{iTh}(\beta_{\tilde{g},h},\hat{\phi}_{i,h}(\beta_{\tilde{g},h})) \geq \Delta\ . \tag{155}
$$

Add and subtract $\sum_h Q_{iTh}(\beta^0_{\tilde{g},h},\phi^0_{i,h})$ to the LHS, the condition becomes

$$
\begin{aligned}
&\left(\sum_h Q_{iTh}(\beta_{g,h},\hat{\phi}_{i,h}(\beta_{g,h})) - \sum_h Q_{iTh}(\beta^0_{\tilde{g},h},\phi^0_{i,h})\right) \\
&\quad + \left(\sum_h Q_{iTh}(\beta^0_{\tilde{g},h},\phi^0_{i,h}) - \sum_h Q_{iTh}(\beta_{\tilde{g},h},\hat{\phi}_{i,h}(\beta_{\tilde{g},h}))\right) \geq \Delta\ .
\end{aligned}
\tag{156}
$$

Therefore, the goal is to derive lower bounds for the LHS terms. Suppose $\boldsymbol{\beta} \in \mathcal{N}_\eta$. Since

$g_i^0 = \tilde{g}$ and $g \neq \tilde{g}$, we have

$$
\begin{aligned}
& \sum_h \left\|\beta^0_{\tilde{g},h} - \beta_{g,h}\right\| \\
= \ & \sum_h \left\|\beta^0_{\tilde{g},h} - \beta^0_{g,h} + \beta^0_{g,h} - \beta_{g,h}\right\| \\
\geq \ & \sum_h \left\|\beta^0_{\tilde{g},h} - \beta^0_{g,h}\right\| - \sum_h \left\|\beta^0_{g,h} - \beta_{g,h}\right\| \\
\geq \ & \sum_h \left\|\beta^0_{\tilde{g},h} - \beta^0_{g,h}\right\| - \eta \geq c_\beta - \eta
\end{aligned}
\tag{157}
$$

where $c_\beta \triangleq \min_{j \neq k} \sum_h \|\beta^0_{j,h} - \beta^0_{k,h}\| > 0$ by Assumption 2.B. For sufficiently small η, we have $\sum_h \|\beta^0_{\tilde{g},h} - \beta_{g,h}\| \geq c_\beta - \eta > 0$.

To apply the above result, notice that for individual level objective function $Q_{iTh}(\beta, \phi)$, we have

$$
\begin{aligned}
Q_{iTh}(\beta, \phi) = \ & \mathbb{E}[z_{i,t}(y_{i,t+h} - x'_{i,t}\beta - c'_{i,t}\phi)]'\Omega_{i,h}\mathbb{E}[z_{i,t}(y_{i,t+h} - x'_{i,t}\beta - c'_{i,t}\phi)] \\
= \ & (\theta^0_{i,h} - \theta)'\mathbb{E}[z_{i,t}w'_{i,t}]'\Omega_{i,h}\mathbb{E}[z_{i,t}w'_{i,t}](\theta^0_{i,h} - \theta) \\
\geq \ & \rho_{\min,i,h}\|\theta^0_{i,h} - \theta\|^2 \geq \rho_{\min,i,h}\|\beta^0_{g_i^0,h} - \beta\|^2
\end{aligned}
\tag{158}
$$

for generic $\beta \in \Theta, \phi \in \Phi$. Therefore, we have

$$
\begin{aligned}
& \sum_h Q_{iTh}\left(\beta_{g,h}, \hat{\phi}_{i,h}(\beta_{g,h})\right) - \sum_h Q_{iTh}\left(\beta^0_{\tilde{g},h}, \phi^0_{i,h}\right) \\
= & \sum_h Q_{iTh}\left(\beta_{g,h}, \hat{\phi}_{i,h}(\beta_{g,h})\right) - 0 \\
\geq & \sum_h \rho_{min,i,h}\left\|\beta^0_{\tilde{g},h} - \beta_{g,h}\right\| \geq \left(\min_h \rho_{min,i,h}\right) \sum_h \left\|\beta^0_{\tilde{g},h} - \beta_{g,h}\right\| \\
\geq & \left(\min_h \rho_{min,i,h}\right)(c_\beta - \eta) \ .
\end{aligned}
\tag{159}
$$

Similarly, for generic $\beta, \tilde{\beta} \in \Theta, \phi, \tilde{\phi} \in \Phi$:

$$
\begin{aligned}
& Q_{iTh}(\beta,\phi) - Q_{iTh}(\tilde{\beta},\tilde{\phi}) \\
=& (\theta^0_{i,h} - \theta)'\mathbb{E}[z_{i,t}w'_{i,t}]'\Omega_{i,h}\mathbb{E}[z_{i,t}w'_{i,t}](\theta^0_{i,h} - \theta) \\
& \qquad - (\theta^0_{i,h} - \tilde{\theta})'\mathbb{E}[z_{i,t}w'_{i,t}]'\Omega_{i,h}\mathbb{E}[z_{i,t}w'_{i,t}](\theta^0_{i,h} - \tilde{\theta}) \\
=& (2\theta^0_{i,h} - \theta - \tilde{\theta})'\mathbb{E}[z_{i,t}w'_{i,t}]'\Omega_{i,h}\mathbb{E}[z_{i,t}w'_{i,t}](\tilde{\theta} - \theta) \\
\leq& \left|(2\theta^0_{i,h} - \theta - \tilde{\theta})'\mathbb{E}[z_{i,t}w'_{i,t}]'\Omega_{i,h}\mathbb{E}[z_{i,t}w'_{i,t}](\tilde{\theta} - \theta)\right| \\
\leq& \left\|2\theta^0_{i,h} - \theta - \tilde{\theta}\right\| \times \left\|\mathbb{E}[z_{i,t}w'_{i,t}]'\Omega_{i,h}\mathbb{E}[z_{i,t}w'_{i,t}]\right\| \times \left[\|\beta - \tilde{\beta}\|^2 + \|\phi - \tilde{\phi}\|^2\right]^{1/2} \\
\leq& C_2\left[\|\beta - \tilde{\beta}\| + \|\phi - \tilde{\phi}\|\right]
\end{aligned} \tag{160}
$$

where $C_2 < \infty$ is some finite constant. To see this, observe that by Assumption 1.A the parameter space is compact and thus bounded; moreover, by Assumption 1.D and 1.F, $\mathbb{E}[z_{i,t}w'_{i,t}]'\Omega_{i,h}\mathbb{E}[z_{i,t}w'_{i,t}]$ is finite positive definite and thus its norm is bounded; these conditions holds uniformly over i and the parameter space.

Then by (160), we can derive the upper bound:

$$
\begin{aligned}
0 \leq& \sum_h Q_{iTh}(\beta_{\tilde{g},h}, \hat{\phi}_{i,h}(\beta_{\tilde{g},h})) - \sum_h Q_{iTh}\left(\beta^0_{\tilde{g},h}, \phi^0_{i,h}\right) \\
\leq& C_2 \sum_h \left[\|\beta^0_{\tilde{g},h} - \beta_{\tilde{g},h}\| + \|\hat{\phi}_{i,h}(\beta_{\tilde{g},h}) - \phi^0_{i,h}\|\right] \\
\leq& C_2 \sum_h \sup_i \|\hat{\phi}_{i,h}(\beta_{\tilde{g},h}) - \phi^0_{i,h}\| + C_2 H\eta
\end{aligned} \tag{161}
$$

where the first inequality comes from the fact that $Q_{iTh}(\beta,\phi)$ is uniquely minimized at $(\beta^0_{\tilde{g},h}, \phi^0_{i,h})$, and the last inequality is obtained by the condition that $\boldsymbol{\beta} \in \mathcal{N}_\eta$. Therefore, we have

$$
\sum_h Q_{iTh}(\beta^0_{\tilde{g},h}, \phi^0_{i,h}) - \sum_h Q_{iTh}(\beta_{\tilde{g},h}, \hat{\phi}_{i,h}(\beta_{\tilde{g},h})) \geq -C_2 \sum_h \sup_i \|\hat{\phi}_{i,h}(\beta_{\tilde{g},h}) - \phi^0_{i,h}\| - C_2 H\eta\ . \tag{162}
$$

Taken together, we have

$$
\begin{aligned}
&\left(\sum_h Q_{iTh}(\beta_{g,h}, \hat{\phi}_{i,h}(\beta_{g,h})) - \sum_h Q_{iTh}(\beta^0_{\tilde{g},h}, \phi^0_{i,h})\right) \\
&\quad + \left(\sum_h Q_{iTh}(\beta^0_{\tilde{g},h}, \phi^0_{i,h}) - \sum_h Q_{iTh}(\beta_{\tilde{g},h}, \hat{\phi}_{i,h}(\beta_{\tilde{g},h}))\right) \\
&\geq \left(\min_h \rho_{min,i,h}\right)(c_\beta - \eta) - C_2 \sum_h \sup_i \|\hat{\phi}_{i,h}(\beta_{\tilde{g},h}) - \phi^0_{i,h}\| - C_2 H\eta
\end{aligned}
\tag{163}
$$

It remains to show that the RHS is asymptotically bounded away from zero. Notice that since we are conditioning on $g^0_i = \tilde{g}$ and $\boldsymbol{\beta} \in \mathcal{N}_\eta$, we have $\max_{\tilde{g}} \|\beta_{\tilde{g},h} - \beta^0_{\tilde{g},h}\| = o_p(1)$, which satisfies the condition of Lemma 4. Therefore, we have for all h,

$$
\sup_i \|\hat{\phi}_{i,h}(\beta_{\tilde{g},h}) - \phi^0_{i,h}\| = o_p(1)\ . \tag{164}
$$

Denote the event $A_\eta = \left\{\sup_i \|\hat{\phi}_{i,h}(\beta_{\tilde{g},h}) - \phi^0_{i,h}\| \leq \eta\right\}$ for any given $\eta > 0$. We know that $\mathbb{P}(A_\eta) = 1$. Then conditional on A_η, we have

$$
\begin{aligned}
&\left(\sum_h Q_{iTh}(\beta_{g,h}, \hat{\phi}_{i,h}(\beta_{g,h})) - \sum_h Q_{iTh}(\beta^0_{\tilde{g},h}, \phi^0_{i,h})\right) \\
&\quad + \left(\sum_h Q_{iTh}(\beta^0_{\tilde{g},h}, \phi^0_{i,h}) - \sum_h Q_{iTh}(\beta_{\tilde{g},h}, \hat{\phi}_{i,h}(\beta_{\tilde{g},h}))\right) \\
&\geq \left(\min_h \rho_{min,i,h}\right) c_\beta - \eta\left(\min_h \rho_{min,i,h} + 2C_2 H\right) \triangleq \Delta > 0
\end{aligned}
\tag{165}
$$

where we use the fact that $0 < \min_h \rho_{min,i,h}, C_2 < \infty$ by Assumptions 1.A, 1.D and 1.F, $H < \infty$ by Assumption 1.G. The last inequality holds if we choose a sufficiently small η.

Having established the expression for Δ under the event A_η, we have

$$
\begin{aligned}
&\mathbf{1}\left\{\sum_h \hat{Q}_{iTh}(\beta_{g,h},\hat{\phi}_{i,h}(\beta_{g,h})) \le \sum_h \hat{Q}_{iTh}(\beta_{\tilde{g},h},\hat{\phi}_{i,h}(\beta_{\tilde{g},h}))\right\} \\
\le&\mathbf{1}\left\{\sum_h \hat{Q}_{iTh}(\beta_{g,h},\hat{\phi}_{i,h}(\beta_{g,h})) \le \sum_h \hat{Q}_{iTh}(\beta_{\tilde{g},h},\hat{\phi}_{i,h}(\beta_{\tilde{g},h}))\right\}\mathbf{1}\left\{A_\eta\right\} \\
&+\mathbf{1}\left\{\sum_h \hat{Q}_{iTh}(\beta_{g,h},\hat{\phi}_{i,h}(\beta_{g,h})) \le \sum_h \hat{Q}_{iTh}(\beta_{\tilde{g},h},\hat{\phi}_{i,h}(\beta_{\tilde{g},h}))\right\}\mathbf{1}\left\{A_\eta^c\right\}. \qquad (166)\\
\le&\mathbf{1}\left\{\Delta \le 2\sum_h \sup_{\beta\in\Theta,\phi\in\Phi}\left|\hat{Q}_{iTh}(\beta,\phi)-Q_{iTh}(\beta,\phi)\right|\right\}\mathbf{1}\left\{A_\eta\right\}+\mathbf{1}\left\{A_\eta^c\right\} \\
\le&2\mathbf{1}\left\{\sum_h \sup_{\beta\in\Theta,\phi\in\Phi}\left|\hat{Q}_{iTh}(\beta,\phi)-Q_{iTh}(\beta,\phi)\right| \ge \Delta/2\right\}+\mathbf{1}\left\{A_\eta^c\right\}
\end{aligned}
$$

Plug this back into the misclassification indicator, we have

$$
\begin{aligned}
&\mathbb{P}\left(\sup_{\boldsymbol{\beta}\in\mathcal{N}_\eta}\sup_i \mathbf{1}\left\{\hat{g}_i(\boldsymbol{\beta})\neq g_i^0\right\}>0\right) \\
\le&\mathbb{P}\left(\sup_{\boldsymbol{\beta}\in\mathcal{N}_\eta}\sup_i \sum_g \max_{\tilde{g}\neq g}\mathbf{1}\{g_i^0=\tilde{g}\}\mathbf{1}\left\{\hat{g}_i(\boldsymbol{\beta})=g\right\}>0\right) \\
\le&\mathbb{P}\left(\sup_{\boldsymbol{\beta}\in\mathcal{N}_\eta}\sup_i \sum_g \max_{\tilde{g}\neq g}\mathbf{1}\{g_i^0=\tilde{g}\}\mathbf{1}\left\{\sum_h \hat{Q}_{iTh}(\beta_{g,h},\hat{\phi}_{i,h}(\beta_{g,h}))\right.\right. \\
&\qquad\qquad\left.\left.\le \sum_h \hat{Q}_{iTh}(\beta_{\tilde{g},h},\hat{\phi}_{i,h}(\beta_{\tilde{g},h}))\right\}>0\right) \\
\le&G^0(G^0-1)\max_g\max_{\tilde{g}\neq g}\mathbb{P}\left(\sup_{\boldsymbol{\beta}\in\mathcal{N}_\eta}\sup_i 2\mathbf{1}\left\{\sum_h \sup_{\beta\in\Theta,\phi\in\Phi}\left|\hat{Q}_{iTh}(\beta,\phi)-Q_{iTh}(\beta,\phi)\right|\ge\Delta/2\right\}\right. \\
&\qquad\qquad\left.+\sup_{\boldsymbol{\beta}\in\mathcal{N}_\eta}\sup_i\mathbf{1}\left\{A_\eta^c\right\}>0\right) \\
\le&2G^0(G^0-1)\max_g\max_{\tilde{g}\neq g}\mathbb{P}\left(\sum_h\sup_i\sup_{\beta\in\Theta,\phi\in\Phi}\left|\hat{Q}_{iTh}(\beta,\phi)-Q_{iTh}(\beta,\phi)\right|\ge\Delta/2\right) \\
&+2G^0(G^0-1)\max_g\max_{\tilde{g}\neq g}\mathbb{P}\left(\sup_i\|\hat{\phi}_{i,h}(\beta_{\tilde{g},h})-\phi_{i,h}^0\|>\eta\right) \\
\le&2G^0(G^0-1)\max_g\max_{\tilde{g}\neq g}[o_p(NT^{-\delta})+o_p(1)]=o_p(1)
\end{aligned}
\tag{167}
$$

where the third and the fourth inequality comes from the Boole's inequality, and the last inequality follows from Lemma 2, Lemma 4, and the Assumption 2.C that the number of groups is finite. ■

S5.6 Lemma 6

Proof. The proofs of each terms follow the same logic, and hence here we show the proof for the first term only. By simple manipulations, we have

$$\begin{aligned}
&\left|\bar{d}'_{zx,i}\hat{\Omega}_{i,h}\bar{d}_{zx,i} - d'_{zx,i}\Omega_{i,h}d_{zx,i}\right| \\
=&\left|\bar{d}'_{zx,i}\hat{\Omega}_{i,h}\bar{d}_{zx,i} - \bar{d}'_{zx,i}\Omega_{i,h}\bar{d}_{zx,i} + \bar{d}'_{zx,i}\Omega_{i,h}\bar{d}_{zx,i} - d'_{zx,i}\Omega_{i,h}d_{zx,i}\right| \\
\leq&\left|\bar{d}'_{zx,i}(\hat{\Omega}_{i,h} - \Omega_{i,h})\bar{d}_{zx,i}\right| + \left|\bar{d}'_{zx,i}\Omega_{i,h}\bar{d}_{zx,i} - d_{zx,i}\Omega_{i,h}d_{zx,i}\right| \\
=&\left|\bar{d}'_{zx,i}(\hat{\Omega}_{i,h} - \Omega_{i,h})\bar{d}_{zx,i} - d'_{zx,i}(\hat{\Omega}_{i,h} - \Omega_{i,h})d_{zx,i} + d'_{zx,i}(\hat{\Omega}_{i,h} - \Omega_{i,h})d_{zx,i}\right| \\
&+ \left|(\bar{d}_{zx,i} - d_{zx,i})'\Omega_{i,h}(\bar{d}_{zx,i} - d_{zx,i}) + 2d'_{zx,i}\Omega_{i,h}(\bar{d}_{zx,i} - d_{zx,i})\right| \\
\leq&\left|(\bar{d}_{zx,i} - d_{zx,i})'(\hat{\Omega}_{i,h} - \Omega_{i,h})(\bar{d}_{zx,i} - d_{zx,i}) + 2d'_{zx,i}(\hat{\Omega}_{i,h} - \Omega_{i,h})(\bar{d}_{zx,i} - d_{zx,i})\right| \qquad (168) \\
&+ \left|d'_{zx,i}(\hat{\Omega}_{i,h} - \Omega_{i,h})d_{zx,i}\right| \\
&+ \left|(\bar{d}_{zx,i} - d_{zx,i})'\Omega_{i,h}(\bar{d}_{zx,i} - d_{zx,i})\right| + 2\left|d'_{zx,i}\Omega_{i,h}(\bar{d}_{zx,i} - d_{zx,i})\right| \\
\leq&\|\bar{d}_{zx,i} - d_{zx,i}\|^2 \left[\|\Omega_{i,h}\| + \|\hat{\Omega}_{i,h} - \Omega_{i,h}\|\right] \\
&+ 2\|d_{zx,i}\|\|\bar{d}_{zx,i} - d_{zx,i}\| \left[\|\Omega_{i,h}\| + \|\hat{\Omega}_{i,h} - \Omega_{i,h}\|\right] \\
&+ \|d_{zx,i}\|^2\|\hat{\Omega}_{i,h} - \Omega_{i,h}\|
\end{aligned}$$

We have three terms on the RHS. Next we take the supremum on both sides and bound the three terms separately.

Term 1.

$$\begin{aligned}
& \sup_i \|\bar{d}_{zx,i} - d_{zx,i}\|^2 \left[\|\Omega_{i,h}\| + \|\hat{\Omega}_{i,h} - \Omega_{i,h}\|\right] \\
\leq \;& \sup_i \|\bar{d}_{zx,i} - d_{zx,i}\|^2 \left[\sup_i \|\Omega_{i,h}\| + \sup_i \|\hat{\Omega}_{i,h} - \Omega_{i,h}\|\right] \qquad (169) \\
\leq \;& o_p(NT^{-\delta}) \left(O_p(1) + o_p(NT^{-\delta})\right)
\end{aligned}$$

by Lemma 1 and Assumption 1.F.

Term 2.

$$\begin{aligned} & \sup_i \|d_{zx,i}\|\|\bar{d}_{zx,i} - d_{zx,i}\| \left[\|\Omega_{i,h}\| + \|\hat{\Omega}_{i,h} - \Omega_{i,h}\|\right] \\ \leq \; & \sup_i \|d_{zx,i}\| \sup_i \|\bar{d}_{zx,i} - d_{zx,i}\| \left[\|\Omega_{i,h}\| + \|\hat{\Omega}_{i,h} - \Omega_{i,h}\|\right] = O_p(1)o_p(1) = o_p(1) \end{aligned} \tag{170}$$

where we use the full rank condition 1.D and result from Term 1.

Term 3.

$$\begin{aligned} & \sup_i \|d_{zx,i}\|^2\|\hat{\Omega}_{i,h} - \Omega_{i,h}\| \\ \leq \; & \sup_i \|d_{zx,i}\|^2\| \sup_i \|\hat{\Omega}_{i,h} - \Omega_{i,h}\| = O_p(1)o_p(NT^{-\delta}) \end{aligned} \tag{171}$$

where we use Assumptions 1.D and 1.F.

Combining the three terms, we

$$\begin{aligned} & \sup_i \left(\bar{d}'_{zx,i}\hat{\Omega}_{i,h}\bar{d}_{zx,i} - d'_{zx,i}\Omega_{i,h}d_{zx,i}\right) \\ \leq \; & \sup_i \left|\bar{d}'_{zx,i}\hat{\Omega}_{i,h}\bar{d}_{zx,i} - d'_{zx,i}\Omega_{i,h}d_{zx,i}\right| = o_p(1) \end{aligned} \tag{172}$$

as desired. The proof for the other terms are almost identical. ■

References

Ahn, M., Batty, M., and Meisenzahl, R. R. (2018). Household Debt-to-Income Ratios in the Enhanced Financial Accounts. FEDS Notes, Board of Governors of the Federal Reserve System, Washington.

Anderson, T. and Hsiao, C. (1982). Formulation and Estimation of Dynamic Models Using Panel Data. *Journal of Econometrics*, 18(1):47–82.

Barnichon, R. and Brownlees, C. (2019). Impulse Response Estimation by Smooth Local Projections. *Review of Economics and Statistics*, 101(3):522–530.

Bonhomme, S. and Manresa, E. (2015). Grouped Patterns of Heterogeneity in Panel Data. *Econometrica*, 83(3):1147–1184.

Cameron, A. C. and Miller, D. L. (2015). A Practitioner's Guide to Cluster-Robust Inference. *Journal of Human Resources*, 50(2):317–372.

Dhaene, G. and Jochmans, K. (2015). Split-panel Jackknife Estimation of Fixed-effect Models. *The Review of Economic Studies*, 82(3):991–1030.

Fernández-Val, I. and Lee, J. (2013). Panel Data Models with Nonadditive Unobserved Heterogeneity: Estimation and Inference. *Quantitative Economics*, 4(3):453–481.

Gao, J. (2007). *Nonlinear Time Series: Semiparametric and Nonparametric Methods.* Number 108 in Monographs on Statistics and Applied Probability. Chapman & Hall/CRC, Boca Raton.

Hahn, J. and Kuersteiner, G. (2011). BIAS REDUCTION FOR DYNAMIC NONLINEAR PANEL MODELS WITH FIXED EFFECTS. *Econometric Theory*, 27(6):1152–1191.

Hall, P. and Heyde, C. C. (1980). *Martingale Limit Theory and Its Application.* Academic Press, New York.

Kilian, L. and Lütkepohl, H. (2017). *Structural Vector Autoregressive Analysis.* Cambridge University Press, 1 edition.

Lewis, D. J., Melcangi, D., and Pilossoph, L. (2019). Latent Heterogeneity in the Marginal Propensity to Consume. Federal Reserve Bank of New York Staff Report 902.

Merlevède, F., Peligrad, M., and Rio, E. (2011). A Bernstein type inequality and moderate deviations for weakly dependent sequences. *Probability Theory and Related Fields*, 151(3-4):435–474.

Nagar, A. L. (1959). The Bias and Moment Matrix of the General k-Class Estimators of the Parameters in Simultaneous Equations. *Econometrica*, 27(4):575.

Saiz, A. (2010). The Geographic Determinants of Housing Supply. *Quarterly Journal of Economics*, 125(3):1253–1296.

Su, L., Shi, Z., and Phillips, P. C. B. (2016). Identifying Latent Structures in Panel Data. *Econometrica*, 84(6):2215–2264.

Wang, W., Phillips, P. C. B., and Su, L. (2018). Homogeneity Pursuit in Panel Data Models: Theory and Application. *Journal of Applied Econometrics*, 33(6):797–815.

Tables and figures

Table S1: Infeasible GLP Coverage Rates (%)

			$G^0 = 2$							$G^0 = 3$						
Design	N	T	h=0	h=1	h=2	h=3	h=4	h=5	h=6	h=0	h=1	h=2	h=3	h=4	h=5	h=6
1	100	100	94.3	91.9	91.0	91.1	88.5	89.6	90.3	92.8	91.8	89.4	89.0	87.7	87.9	86.9
	100	200	94.3	93.8	92.9	92.1	91.9	92.1	91.2	93.6	93.8	92.3	91.9	90.5	90.6	90.0
	100	300	94.6	92.6	93.0	93.0	92.7	93.0	93.0	93.8	93.1	92.0	92.3	92.0	92.0	91.3
	200	100	95.3	91.3	88.2	87.9	86.6	85.3	84.5	94.0	91.4	89.5	87.4	85.1	84.1	82.8
	200	200	95.1	92.8	92.5	91.8	91.3	91.0	90.6	93.9	92.7	92.8	91.7	91.4	90.7	89.9
	200	300	93.9	92.9	93.6	93.8	92.4	92.8	90.6	93.8	93.2	93.2	93.0	91.8	91.1	91.1
	300	100	94.6	89.3	85.4	84.1	83.8	82.5	81.0	94.1	90.0	86.8	84.1	81.5	79.0	78.1
	300	200	95.8	91.7	89.9	89.0	88.9	89.0	86.6	94.8	92.3	91.3	90.2	89.9	87.4	87.1
	300	300	94.6	93.6	92.8	91.7	91.6	90.4	89.6	95.3	93.3	92.6	92.3	91.5	90.5	90.0
2	100	100	93.6	92.9	91.4	90.7	91.3	91.6	91.6	94.1	92.7	91.6	91.2	91.0	92.0	90.5
	100	200	93.9	93.8	92.8	92.9	92.3	92.2	93.1	93.2	93.2	92.1	92.0	92.4	91.9	91.9
	100	300	94.2	94.0	93.3	93.1	93.1	93.1	93.8	93.9	93.1	93.1	93.5	92.6	92.4	92.6
	200	100	95.0	91.5	89.7	88.0	86.9	85.3	86.1	94.5	92.7	90.5	89.6	89.1	88.7	88.8
	200	200	94.8	93.8	92.4	91.6	91.1	90.4	91.5	94.9	93.9	92.9	91.7	91.9	91.9	91.7
	200	300	94.9	94.7	94.0	93.3	93.2	93.1	92.7	94.1	94.2	93.5	93.1	92.5	92.4	91.6
	300	100	95.6	91.5	87.1	83.9	83.1	84.2	83.7	94.3	91.4	89.3	87.2	86.5	86.6	86.8
	300	200	94.9	92.7	92.2	90.9	91.0	89.6	89.2	93.9	92.7	91.3	90.9	90.8	89.3	89.8
	300	300	95.5	94.0	92.7	91.7	91.3	90.8	91.3	94.3	94.4	92.8	92.1	91.7	92.2	92.1

Note: Standard errors are clustered at individual level (Cameron and Miller, 2015).

Tables and figures

Table S2: Infeasible GLP Jackknife Coverage Rates (%)

			$G^0 = 2$							$G^0 = 3$						
Design	N	T	h=0	h=1	h=2	h=3	h=4	h=5	h=6	h=0	h=1	h=2	h=3	h=4	h=5	h=6
1	100	100	94.0	93.3	94.8	93.3	92.8	92.5	92.3	93.7	93.4	93.2	93.0	91.4	91.6	91.2
	100	200	93.3	94.8	93.6	93.9	93.9	94.1	95.1	93.9	94.9	93.8	93.6	94.4	93.9	94.1
	100	300	94.6	93.9	94.4	94.3	94.3	94.5	94.8	94.2	94.2	94.2	94.1	94.3	94.2	94.4
	200	100	93.5	94.0	93.3	93.4	92.7	92.6	92.5	93.5	93.9	93.1	92.8	91.6	91.4	91.1
	200	200	94.2	94.9	94.3	94.4	94.3	94.0	93.4	94.7	94.3	94.3	94.2	94.0	93.2	92.8
	200	300	95.2	94.9	93.8	94.7	94.4	94.7	95.0	94.9	94.7	94.7	93.9	94.2	94.4	93.6
	300	100	93.1	93.4	93.3	92.5	91.5	92.2	92.3	93.3	93.1	92.2	92.6	91.5	91.2	91.1
	300	200	94.1	94.7	94.3	93.6	94.0	93.6	93.3	94.2	93.9	93.8	93.7	93.7	92.9	93.3
	300	300	94.6	93.9	93.9	92.9	93.5	93.7	94.2	94.5	94.4	93.9	93.4	93.4	93.1	93.5
2	100	100	93.7	93.4	94.0	92.8	92.8	92.2	92.3	93.6	93.5	93.9	92.9	92.4	92.6	92.3
	100	200	93.3	94.7	94.1	94.1	94.5	94.7	94.9	93.9	94.9	94.0	94.2	94.5	94.1	94.3
	100	300	94.5	94.2	94.1	94.3	93.9	94.3	95.2	94.2	94.7	94.3	94.6	93.8	94.9	94.9
	200	100	93.4	93.9	93.8	93.3	92.9	92.2	92.5	93.7	93.8	93.2	92.6	92.3	92.3	92.0
	200	200	94.4	94.4	94.2	94.4	93.7	92.8	93.5	94.7	94.6	94.2	94.9	93.8	92.8	93.0
	200	300	95.4	94.9	94.0	94.7	94.9	95.0	94.6	95.0	94.2	94.6	94.5	94.8	94.6	94.3
	300	100	93.3	93.3	92.7	92.5	91.1	91.5	92.3	93.7	93.4	92.2	93.1	92.0	92.0	92.8
	300	200	94.2	95.0	93.7	93.7	94.1	93.8	93.8	94.5	93.8	94.0	93.7	94.4	93.8	94.0
	300	300	94.8	93.7	93.5	92.5	93.5	94.5	94.1	94.6	94.5	93.7	93.4	93.7	93.8	94.3

Note: Standard errors are clustered at individual level (Cameron and Miller, 2015).

Tables and figures

Table S3: GLP Jackknife Coverage Rates (%)

			$G^0 = 2$							$G^0 = 3$						
Design	N	T	h=0	h=1	h=2	h=3	h=4	h=5	h=6	h=0	h=1	h=2	h=3	h=4	h=5	h=6
1	100	100	92.2	80.3	77.7	80.0	82.8	86.7	89.7	91.3	72.3	69.4	71.5	74.9	78.3	81.1
	100	200	93.3	87.6	89.5	89.7	90.9	92.6	93.5	93.7	84.6	82.7	84.3	85.3	87.0	89.6
	100	300	94.1	92.2	93.8	93.2	94.0	93.6	94.2	94.4	90.7	89.5	89.5	90.5	91.2	92.1
	200	100	90.5	66.4	62.8	64.8	74.0	79.6	87.2	90.5	53.3	49.1	53.3	62.5	66.7	72.3
	200	200	93.5	81.5	85.2	87.3	89.0	90.6	92.2	93.3	74.3	70.6	73.7	77.0	80.2	82.9
	200	300	95.7	88.0	90.9	92.9	93.1	94.2	94.5	94.4	84.9	83.7	84.0	86.0	88.0	89.4
	300	100	88.5	56.6	53.0	53.9	64.1	78.4	86.8	88.5	42.0	37.1	42.8	51.7	60.6	65.7
	300	200	92.7	76.4	79.1	81.8	85.4	87.7	92.0	93.1	64.7	61.1	63.3	67.8	73.6	77.2
	300	300	94.2	84.5	88.1	89.9	91.2	93.0	93.4	94.7	80.3	76.5	78.8	80.4	83.3	86.4
2	100	100	93.6	93.5	94.0	92.7	92.8	92.0	92.3	94.2	93.5	93.9	93.0	92.4	92.6	92.2
	100	200	93.3	94.7	94.1	94.1	94.5	94.7	94.9	93.9	94.9	94.0	94.2	94.5	94.1	94.3
	100	300	94.5	94.2	94.1	94.3	93.9	94.3	95.2	94.2	94.7	94.3	94.6	93.8	94.9	94.9
	200	100	93.2	93.9	94.0	93.2	92.8	92.4	92.6	93.5	93.8	93.3	92.8	92.2	92.5	92.2
	200	200	94.5	94.4	94.2	94.4	93.7	92.8	93.5	94.8	94.5	94.2	94.9	93.8	92.8	93.0
	200	300	95.4	94.9	94.0	94.7	94.9	95.0	94.6	95.0	94.2	94.6	94.5	94.8	94.6	94.3
	300	100	93.0	93.1	92.8	92.4	91.3	91.6	92.3	93.4	93.4	92.3	92.9	92.0	92.1	92.6
	300	200	94.1	95.0	93.7	93.6	94.1	93.8	93.8	94.5	93.8	94.1	93.7	94.4	93.8	94.0
	300	300	94.8	93.7	93.5	92.5	93.5	94.5	94.1	94.6	94.5	93.7	93.4	93.7	93.8	94.3

Note: Standard errors are clustered at individual level (Cameron and Miller, 2015).

Table S4: GLP Performance ($T = 50$)

Design	G^0	N	AC (%)	BR (%)		RMSE (×100)			
				GLP	IGLP	GLP	PAN	IND	IGLP
1	2	100	74.0	17.8	17.7	43.1	35.4	107.3	16.5
		200	74.3	12.7	12.6	42.0	34.8	123.5	14.2
		300	74.2	10.4	10.4	41.9	34.7	115.1	13.2
	3	100	64.5	23.7	22.1	78.1	83.9	138.1	27.2
		200	64.3	16.9	15.8	77.2	83.4	179.0	23.6
		300	64.4	13.9	13.0	76.6	83.2	220.7	22.1
2	2	100	97.0	18.2	17.8	12.5	23.2	62.6	9.7
		200	97.0	12.9	12.7	11.2	22.9	69.5	8.0
		300	97.1	10.5	10.4	10.7	22.8	65.7	7.4
	3	100	96.2	22.2	21.5	17.4	37.8	81.6	14.7
		200	96.2	15.8	15.4	15.0	37.4	135.5	11.9
		300	96.1	12.9	12.7	14.1	37.3	97.8	10.8

Table S5: GLP Performance ($T = 200$)

Design	G^0	N	AC (%)	BR (%)		RMSE (×100)			
				GLP	IGLP	GLP	PAN	IND	IGLP
1	2	300	92.8	8.7	8.6	18.2	32.8	43.4	4.2
		400	92.6	7.5	7.5	18.3	32.7	43.4	3.8
		500	92.7	6.7	6.7	18.2	32.7	43.3	3.6
	3	300	90.2	10.9	10.6	33.6	81.0	52.6	6.4
		400	90.2	9.4	9.2	33.4	81.0	52.6	5.6
		500	90.2	8.4	8.2	33.5	81.0	52.6	5.3
2	2	300	100.0	8.7	8.6	2.5	21.9	26.2	2.4
		400	100.0	7.5	7.5	2.3	21.9	26.2	2.3
		500	100.0	6.7	6.7	2.2	21.9	26.2	2.1
	3	300	100.0	10.6	10.5	3.9	36.4	34.5	3.7
		400	100.0	9.2	9.1	3.5	36.3	34.4	3.4
		500	100.0	8.2	8.2	3.3	36.3	34.4	3.1

Table S6: GLP Performance ($T = 300$)

				BR (%)		RMSE (×100)			
Design	G^0	N	AC (%)	GLP	IGLP	GLP	PAN	IND	IGLP
1	2	500	96.3	6.6	6.6	12.8	32.7	35.0	2.7
		1000	96.3	4.7	4.7	12.7	32.6	35.0	2.2
		1500	96.3	3.8	3.8	12.7	32.6	35.0	2.0
	3	500	95.1	8.2	8.1	23.9	80.9	42.3	3.9
		1000	95.1	5.8	5.7	23.9	80.9	42.3	3.3
		1500	95.1	4.7	4.7	23.8	80.9	42.3	3.0
2	2	500	100.0	6.6	6.6	1.6	21.9	21.1	1.6
		1000	100.0	4.7	4.7	1.3	21.9	21.1	1.3
		1500	100.0	3.8	3.8	1.2	21.8	21.1	1.2
	3	500	100.0	8.1	8.0	2.4	36.3	27.8	2.4
		1000	100.0	5.7	5.7	2.0	36.3	27.8	1.9
		1500	100.0	4.7	4.6	1.8	36.3	27.8	1.7

Table S7: GLP Coverage Rates (%, $T = 50$)

Design	G	N	GLP							IGLP						
			h=0	h=1	h=2	h=3	h=4	h=5	h=6	h=0	h=1	h=2	h=3	h=4	h=5	h=6
1	2	100	89.6	81.6	71.2	73.0	80.3	83.8	86.6	94.2	90.3	86.1	84.2	84.9	85.5	84.7
		200	85.9	66.1	44.9	53.8	65.1	75.1	78.6	93.8	86.0	79.0	76.9	75.9	75.0	75.8
		300	82.3	52.6	28.6	40.6	56.9	66.7	69.5	95.1	80.6	73.1	69.3	66.7	65.8	65.2
	3	100	90.9	81.9	83.1	84.8	86.9	88.3	90.4	93.8	89.6	86.1	83.4	81.5	78.9	78.3
		200	88.9	65.9	69.5	74.7	78.4	80.7	83.8	94.6	87.2	82.3	76.7	72.8	69.8	67.4
		300	85.7	54.3	52.7	62.3	67.7	72.1	79.8	94.1	83.0	75.5	68.5	63.6	60.3	58.0
2	2	100	89.7	91.5	87.5	86.5	85.1	86.6	86.0	95.5	91.2	88.3	86.6	86.2	86.1	84.6
		200	83.6	88.4	84.9	78.5	76.2	77.6	76.8	95.0	88.7	85.3	79.8	77.7	77.8	76.8
		300	79.7	85.2	76.1	71.2	67.5	71.0	72.5	94.8	85.8	78.3	72.7	69.2	69.6	71.4
	3	100	90.0	91.9	90.4	88.9	87.3	88.1	87.8	93.8	91.5	89.0	88.3	86.9	87.4	86.7
		200	86.8	90.1	85.4	82.2	82.0	82.7	83.0	94.7	90.2	86.7	83.6	82.7	81.6	82.3
		300	82.0	88.3	82.2	77.7	75.9	77.5	76.4	94.5	88.1	83.5	79.5	76.6	77.4	75.6

Note: This table reports the coverage probability of the large T inference as in Theorem 2.

Table S8: GLP Coverage Rates (%, $T = 200$)

Design	G	N	GLP h=0	GLP h=1	GLP h=2	GLP h=3	GLP h=4	GLP h=5	GLP h=6	IGLP h=0	IGLP h=1	IGLP h=2	IGLP h=3	IGLP h=4	IGLP h=5	IGLP h=6
1	2	300	92.7	89.1	92.2	89.8	88.7	89.2	89.0	95.1	92.8	91.0	89.2	88.9	88.3	88.0
		400	89.7	87.1	91.0	88.3	86.0	86.7	87.2	94.9	91.5	89.7	87.4	87.2	86.3	86.4
		500	89.3	86.6	88.8	85.5	85.3	86.6	85.1	95.1	91.6	88.1	86.0	84.1	83.8	83.4
	3	300	92.3	80.0	80.3	85.4	89.2	92.8	94.4	94.8	93.5	90.4	89.0	88.5	88.0	86.0
		400	92.1	77.4	76.0	81.8	87.0	90.8	92.8	95.0	92.3	90.8	89.0	87.6	86.3	86.0
		500	91.4	71.6	73.5	78.1	85.5	89.0	92.1	94.8	91.4	89.6	86.9	85.6	85.1	82.2
2	2	300	93.6	92.6	89.5	89.5	90.0	88.9	90.4	94.6	93.2	90.8	90.0	90.2	89.9	89.6
		400	92.6	91.7	90.4	86.9	86.9	88.1	87.1	94.3	93.2	90.5	88.1	87.9	88.1	87.6
		500	91.3	90.0	88.2	86.8	85.9	86.4	86.0	94.3	91.7	89.6	87.8	87.4	87.1	85.8
	3	300	93.5	92.5	91.8	90.3	90.6	90.8	91.4	95.1	93.3	92.5	90.6	91.0	90.7	90.7
		400	93.0	92.3	90.0	88.5	89.6	88.8	88.3	94.6	92.8	90.7	89.6	89.8	89.3	88.1
		500	93.2	91.6	88.9	87.1	87.5	88.9	87.7	94.4	92.4	90.4	88.5	87.5	89.1	88.2

Note: This table reports the coverage probability of the large T inference as in Theorem 2.

Table S9: GLP Coverage Rates (%, $T = 300$)

Design	G	N	GLP h=0	GLP h=1	GLP h=2	GLP h=3	GLP h=4	GLP h=5	GLP h=6	IGLP h=0	IGLP h=1	IGLP h=2	IGLP h=3	IGLP h=4	IGLP h=5	IGLP h=6
1	2	500	91.9	91.0	95.5	91.8	90.6	90.5	89.5	94.0	91.2	89.5	88.7	87.4	89.1	87.6
		1000	89.1	87.2	93.7	89.9	87.1	84.2	82.5	95.2	90.4	85.9	82.9	80.8	80.2	78.6
		1500	87.0	80.7	92.4	86.4	81.1	79.2	77.4	95.4	86.3	81.4	78.1	75.7	74.3	72.9
	3	500	93.2	84.4	85.1	88.3	91.0	92.6	93.6	94.1	93.7	91.9	89.7	88.8	87.8	86.5
		1000	91.6	75.3	76.3	80.2	85.8	88.7	90.4	95.4	90.9	88.2	84.4	82.3	80.3	78.4
		1500	89.5	68.2	70.1	74.1	80.9	87.1	88.8	95.1	89.7	84.5	81.1	76.6	74.0	71.3
2	2	500	92.1	92.5	89.4	88.0	88.2	88.5	88.5	94.7	93.8	90.3	89.5	88.6	88.8	88.6
		1000	91.7	88.8	84.8	83.3	82.7	81.4	82.6	94.9	90.7	87.6	85.4	84.4	82.0	82.8
		1500	89.7	85.4	79.5	76.9	77.2	75.1	77.3	94.7	88.3	83.3	79.8	79.5	76.5	77.7
	3	500	93.2	92.7	90.3	90.2	90.9	90.4	90.5	94.3	93.4	91.6	90.9	90.6	90.3	90.6
		1000	92.5	90.9	87.7	85.9	85.3	85.6	84.7	94.2	92.3	89.2	87.7	86.3	86.2	85.0
		1500	91.6	86.5	84.9	82.1	81.2	81.1	82.0	94.7	89.7	87.8	84.4	82.2	82.0	82.7

Note: This table reports the coverage probability of the large T inference as in Theorem 2.

Table S10: GLP with Unknown Group Number (RMSE $\times 100$)

		$G^0 = 2$					$G^0 = 3$				
	N	100	200	100	200	400	100	200	100	200	400
Design	T	50	50	400	400	400	50	50	400	400	400
1	PAN	**35.5**	**34.9**	32.7	32.7	32.6	83.9	83.4	81.0	81.0	80.9
	$\hat{G} = 2$	46.3	44.3	**10.5**	**10.4**	**9.8**	**79.1**	**78.0**	29.1	29.2	29.0
	$\hat{G} = 3$	53.6	51.8	15.4	15.1	14.8	82.7	81.4	**19.3**	**19.3**	**19.1**
	$\hat{G} = 4$	57.3	55.0	17.2	16.6	16.2	85.8	84.3	24.0	23.9	23.5
	$\hat{G} = 5$	60.5	57.6	18.3	17.7	17.2	88.5	86.3	25.4	25.3	24.9
	$\hat{G} = 6$	62.9	59.9	19.2	18.5	18.0	90.4	87.9	26.2	26.0	25.7
	$\hat{G} = 7$	65.0	61.4	19.9	19.1	18.5	91.9	89.3	26.9	26.5	26.1
	$\hat{G} = 8$	66.6	63.0	20.5	19.6	19.0	93.5	90.2	27.3	26.9	26.5
	IND	101.3	103.1	30.1	30.2	30.1	126.0	126.6	36.3	36.4	36.4
	IC	1.0	1.0	2.0	2.0	2.0	2.0	2.0	3.1	3.0	3.1
2	PAN	23.2	22.9	21.9	21.9	21.9	37.7	37.4	36.3	36.3	36.3
	$\hat{G} = 2$	**12.9**	**11.5**	**2.7**	**2.0**	**1.5**	**23.8**	**22.6**	17.4	17.1	17.0
	$\hat{G} = 3$	24.9	23.2	8.2	7.8	7.6	18.4	16.1	**4.1**	**3.1**	**2.2**
	$\hat{G} = 4$	28.6	26.7	9.5	9.0	8.8	32.6	30.0	10.6	9.9	9.5
	$\hat{G} = 5$	31.3	29.2	10.4	9.9	9.6	37.5	34.6	12.4	11.6	11.1
	$\hat{G} = 6$	33.2	30.9	11.0	10.5	10.2	41.1	37.7	13.4	12.6	12.1
	$\hat{G} = 7$	34.9	32.5	11.5	10.9	10.6	43.5	39.9	14.2	13.4	12.8
	$\hat{G} = 8$	36.2	33.8	12.0	11.3	11.0	45.7	41.7	14.9	13.9	13.4
	IND	61.9	64.8	18.1	18.2	18.2	89.8	81.7	23.9	23.9	23.9
	IC	2.0	2.0	2.0	2.0	2.0	2.1	2.1	3.0	3.0	3.0

Note: This table reports the RMSE of the GLP with different supplied group number. Cells chosen by the information criterion are in bold.

Table S11: Compare Weighting Matrix: Classification Accuracy (%)

Design	N	T	$G^0 = 2$					$G^0 = 3$				
			UH	H	U	2SLS	IV	UH	H	U	2SLS	IV
1	100	100	83.5	**84.1**	83.0	83.0	83.1	78.2	**78.6**	77.1	76.9	77.2
	100	200	92.6	**92.7**	91.8	91.7	91.8	90.0	**90.3**	88.8	88.6	88.8
	100	300	96.3	**96.3**	95.6	95.6	95.6	95.0	**95.1**	94.0	93.9	94.0
	200	100	83.5	**84.0**	82.9	82.8	83.0	77.9	**78.4**	76.8	76.7	77.0
	200	200	92.4	**92.6**	91.7	91.6	91.7	90.1	**90.3**	89.0	88.8	89.0
	200	300	96.3	**96.3**	95.6	95.6	95.6	94.9	**95.0**	94.0	93.9	94.0
	300	100	83.5	**84.0**	82.9	82.9	83.0	77.9	**78.3**	76.9	76.7	77.0
	300	200	92.6	**92.7**	91.8	91.8	91.8	90.1	**90.3**	88.9	88.7	88.9
	300	300	96.3	**96.3**	95.6	95.6	95.6	94.9	**95.1**	94.0	93.9	94.0
2	100	100	99.7	**99.7**	96.6	96.5	96.6	99.6	**99.7**	90.3	90.1	91.7
	100	200	100.0	**100.0**	99.6	99.5	99.5	**100.0**	100.0	98.0	98.0	98.1
	100	300	**100.0**	**100.0**	99.9	99.9	99.9	**100.0**	**100.0**	99.4	99.4	99.4
	200	100	99.7	**99.7**	96.7	96.6	96.7	99.6	**99.6**	90.1	90.0	91.7
	200	200	100.0	**100.0**	99.5	99.5	99.5	100.0	**100.0**	98.1	98.0	98.1
	200	300	**100.0**	**100.0**	99.9	99.9	99.9	**100.0**	**100.0**	99.4	99.4	99.4
	300	100	99.7	**99.7**	96.6	96.5	96.6	99.6	**99.6**	89.9	89.8	91.7
	300	200	100.0	**100.0**	99.5	99.5	99.5	100.0	**100.0**	98.1	98.0	98.1
	300	300	**100.0**	**100.0**	99.9	99.9	99.9	**100.0**	**100.0**	99.5	99.5	99.5

Note: This table reports the classification accuracy of the GLP estimator with different weighting schemes. UH, H, U, 2SLS, IV indicate unit-and-horizon specific weights, horizon-specific weights, unit-specific weights, two stage least squares weights and IV weights respectively. Cells with the highest accuracy are in bold.

Table S12: Compare weighting matrix: RMSE (×100)

$\hat{G}$	Design	N	T	UH	H	U	2SLS	IV	PAN	IND	IGLP
2	1	100	100	30.0	**29.6**	30.4	30.7	30.5	33.5	63.7	10.0
		100	200	19.0	**18.9**	19.9	19.9	19.9	32.9	43.3	6.4
		100	300	13.6	**13.4**	14.5	14.5	14.5	32.8	35.0	5.1
		200	100	29.4	**29.1**	30.0	30.2	30.1	33.2	63.8	7.8
		200	200	18.8	**18.5**	19.6	19.7	19.6	32.8	43.4	4.9
		200	300	13.2	**13.1**	14.2	14.2	14.2	32.7	35.0	3.8
		300	100	29.2	**28.9**	29.8	30.1	29.9	33.1	63.8	7.0
		300	200	18.4	**18.3**	19.3	19.4	19.3	32.7	43.3	4.2
		300	300	13.0	**12.9**	14.0	14.0	13.9	32.7	35.0	3.2
	2	100	100	7.1	**6.5**	10.2	10.2	10.1	22.3	38.6	5.8
		100	200	4.2	**3.9**	5.1	4.9	4.9	22.0	26.2	3.8
		100	300	3.2	**3.1**	3.5	3.3	3.3	21.9	21.1	3.1
		200	100	5.9	**5.3**	9.4	9.3	9.2	22.1	38.5	4.6
		200	200	3.3	**3.0**	4.5	4.3	4.3	21.9	26.2	2.9
		200	300	2.4	**2.2**	2.8	2.6	2.6	21.9	21.1	2.2
		300	100	5.5	**4.8**	9.3	9.2	9.1	22.1	38.6	4.1
		300	200	2.9	**2.6**	4.2	4.0	4.0	21.9	26.1	2.5
		300	300	2.1	**1.9**	2.5	2.3	2.3	21.9	21.1	1.9
3	1	100	100	**52.0**	52.6	53.2	53.6	53.5	81.6	77.5	15.0
		100	200	34.3	**34.1**	35.9	35.8	35.6	81.2	52.5	9.5
		100	300	24.8	**24.5**	26.6	26.4	26.3	81.0	42.3	7.5
		200	100	**51.9**	52.2	52.9	53.3	53.2	81.5	77.5	12.0
		200	200	33.7	**33.5**	35.1	35.1	34.9	81.0	52.4	7.0
		200	300	24.6	**24.3**	26.3	26.1	26.1	81.0	42.3	5.6
		300	100	**51.6**	51.9	52.7	53.0	52.9	81.4	77.5	10.8
		300	200	33.6	**33.4**	35.1	35.1	34.9	81.0	52.5	6.1
		300	300	24.4	**24.2**	26.2	26.0	25.9	81.0	42.3	4.7
	2	100	100	10.3	**9.6**	17.7	17.8	16.2	36.8	50.6	9.0
		100	200	6.4	**6.1**	8.8	8.7	8.6	36.5	34.4	6.0
		100	300	5.0	**4.8**	6.0	5.8	5.8	36.4	27.8	4.8
		200	100	8.6	**7.8**	16.8	16.9	15.1	36.6	50.7	7.0
		200	200	4.9	**4.5**	7.8	7.7	7.6	36.4	34.4	4.4
		200	300	3.8	**3.6**	5.0	4.9	4.8	36.3	27.8	3.5
		300	100	7.8	**6.9**	16.5	16.7	14.7	36.6	50.6	6.2
		300	200	4.3	**3.9**	7.4	7.4	7.2	36.3	34.4	3.7
		300	300	3.3	**3.0**	4.6	4.4	4.4	36.3	27.8	2.9

Note: This table reports the RMSE of the GLP estimator with different weighting schemes. UH, H, U, 2SLS, IV indicate unit-and-horizon specific weights, horizon-specific weights, unit-specific weights, two stage least squares weights and IV weights respectively. Cells with the lowest RMSE are in bold.

Table S13: Compare weighting matrix: Band Ratios (%)

			$G^0=2$						$G^0=3$					
Design	N	T	UH	H	U	2SLS	IV	IGLP	UH	H	U	2SLS	IV	IGLP
1	100	100	**14.5**	15.9	15.3	15.6	15.8	15.7	**17.7**	20.3	18.6	19.1	19.5	19.3
	100	200	**14.3**	15.0	14.7	14.9	15.0	14.8	**17.5**	18.8	18.0	18.3	18.4	18.1
	100	300	**14.2**	14.7	14.5	14.6	14.7	14.5	**17.4**	18.2	17.7	18.0	18.1	17.7
	200	100	**10.3**	11.2	10.8	11.0	11.2	11.2	**12.5**	14.4	13.2	13.6	13.8	13.8
	200	200	**10.1**	10.6	10.4	10.5	10.6	10.6	**12.3**	13.3	12.7	13.0	13.1	12.9
	200	300	**10.1**	10.4	10.3	10.3	10.4	10.3	**12.3**	12.9	12.6	12.7	12.8	12.6
	300	100	**8.4**	9.2	8.8	9.0	9.1	9.2	**10.2**	11.8	10.8	11.1	11.3	11.3
	300	200	**8.3**	8.7	8.5	8.6	8.7	8.6	**10.1**	10.9	10.4	10.6	10.7	10.6
	300	300	**8.2**	8.5	8.4	8.5	8.5	8.5	**10.0**	10.5	10.3	10.4	10.4	10.4
2	100	100	**14.6**	15.9	15.3	15.8	16.0	15.7	**17.9**	19.5	19.0	19.6	19.8	19.0
	100	200	**14.3**	15.0	14.7	14.9	15.0	14.8	**17.5**	18.3	18.0	18.3	18.4	17.9
	100	300	**14.2**	14.7	14.5	14.6	14.7	14.5	**17.4**	18.0	17.7	17.9	18.0	17.6
	200	100	**10.4**	11.3	10.9	11.2	11.3	11.2	**12.7**	13.8	13.5	13.9	14.1	13.7
	200	200	**10.1**	10.6	10.4	10.5	10.6	10.6	**12.4**	13.0	12.7	13.0	13.0	12.8
	200	300	**10.1**	10.4	10.3	10.3	10.4	10.3	**12.3**	12.7	12.5	12.7	12.7	12.6
	300	100	**8.5**	9.2	8.9	9.1	9.3	9.2	**10.4**	11.3	11.0	11.4	11.5	11.2
	300	200	**8.3**	8.7	8.5	8.6	8.7	8.6	**10.1**	10.6	10.4	10.6	10.7	10.5
	300	300	**8.2**	8.5	8.4	8.5	8.5	8.5	**10.0**	10.4	10.2	10.3	10.4	10.3

Note: This table reports the band ratios of the GLP estimator with different weighting schemes. UH, H, U, 2SLS, IV indicate unit-and-horizon specific weights, horizon-specific weights, unit-specific weights, two stage least squares weights and IV weights respectively. Cells with the lowest band ratios are in bold.

Table S14: COMPARE WEIGHTING MATRIX: COVERAGE RATES (%, DESIGN 1, $G^0 = 2$)

N	T	Weight	h=0	h=1	h=2	h=3	h=4	h=5	h=6
100	100	UH	-14.0	-2.3	-5.8	-8.5	-6.0	-4.7	-2.6
		H	**-2.6**	**-1.3**	**-3.2**	**-5.1**	**-2.4**	**-0.9**	**0.7**
		U	-1.1	3.1	-4.1	-8.6	-7.6	-4.7	-1.4
		2SLS	-2.2	-1.2	-5.4	-10.5	-8.5	-4.2	-0.4
		IV	-2.3	-0.3	-4.6	-7.9	-6.3	-2.5	0.9
		IGLP	94.9	92.1	92.1	90.0	89.7	89.2	89.4
	200	UH	-6.8	-0.9	0.1	0.4	-1.2	-1.0	-0.5
		H	**0.0**	**0.1**	**1.9**	**1.5**	**0.6**	**0.6**	**1.4**
		U	1.0	0.4	1.3	0.6	-0.3	-0.5	0.5
		2SLS	0.4	-1.5	0.4	-0.3	-0.1	-0.5	1.4
		IV	0.5	-1.3	1.1	0.3	-0.1	-0.4	1.3
		IGLP	94.1	93.2	93.2	92.5	92.2	92.8	92.4
	300	UH	-2.0	-0.5	0.4	1.4	-0.1	-0.5	0.7
		H	**0.5**	**0.2**	**1.8**	**2.8**	**1.3**	**1.3**	**2.0**
		U	0.4	0.0	1.7	2.4	0.9	0.4	0.9
		2SLS	0.0	-2.1	2.0	2.3	1.1	0.4	1.8
		IV	0.0	-1.6	2.0	2.5	1.1	0.9	2.0
		IGLP	93.7	93.5	93.0	92.3	93.3	92.7	92.6
200	100	UH	-26.2	-6.1	-11.3	-13.6	-10.4	-7.5	-4.4
		H	**-4.0**	**-4.5**	**-7.4**	**-9.8**	**-6.9**	**-2.5**	**0.9**
		U	-1.7	2.7	-8.9	-17.5	-13.3	-7.2	-3.7
		2SLS	-4.2	-4.8	-13.2	-21.1	-15.7	-8.1	-2.3
		IV	-4.2	-3.3	-11.0	-18.0	-13.2	-6.6	-1.1
		IGLP	94.0	90.4	87.9	86.6	86.0	86.6	86.7
	200	UH	-14.2	-1.4	2.0	-1.2	-0.9	-1.8	-1.1
		H	**-2.6**	**-0.9**	**2.3**	**0.8**	**0.6**	**1.0**	**0.5**
		U	-0.7	0.1	2.8	-1.0	-1.4	-0.2	-1.1
		2SLS	-2.6	-6.4	0.7	-2.7	-2.0	-0.1	0.0
		IV	-2.4	-5.7	0.8	-2.1	-1.5	0.0	0.0
		IGLP	95.1	92.8	90.9	91.1	90.0	89.6	90.9
	300	UH	-8.5	-1.1	1.4	0.9	-1.2	-1.3	-1.1
		H	**-1.3**	**0.6**	**2.1**	**2.6**	**1.5**	**0.8**	**1.2**
		U	-0.3	0.1	2.7	2.1	0.6	-0.3	-0.1
		2SLS	-0.9	-3.4	1.3	1.8	0.6	0.7	1.1
		IV	-1.4	-3.3	1.2	2.1	0.4	0.9	1.2
		IGLP	94.9	92.8	92.2	91.4	92.2	91.4	91.3
	100	UH	-35.8	-10.8	-18.2	-19.7	-12.5	-8.5	-3.8
		H	**-8.5**	**-10.5**	**-16.6**	**-17.2**	**-9.4**	**-3.0**	**2.6**
		U	-3.9	3.3	-17.3	-28.4	-18.7	-10.3	-3.8
		2SLS	-8.0	-8.7	-24.2	-35.2	-22.5	-12.0	-2.3

	IV	-7.7	-7.4	-21.5	-31.0	-18.6	-9.0	0.0
	IGLP	95.0	89.6	86.7	83.9	81.2	81.6	79.8
200	UH	-19.2	-4.8	1.8	-1.2	-2.9	-2.6	-2.1
	H	**-2.6**	**-4.3**	**1.8**	**1.7**	**0.3**	**0.2**	**1.1**
	U	-1.5	-1.0	2.5	-1.6	-4.0	-3.3	-1.8
	2SLS	-2.9	-12.0	-0.8	-3.7	-4.1	-3.6	0.1
	IV	-2.9	-11.0	-0.7	-3.3	-3.1	-3.1	0.8
	IGLP	95.9	93.1	89.8	89.8	88.9	88.5	88.7
300	UH	-11.1	-1.3	1.7	-0.2	-1.3	-1.2	-1.2
	H	**-1.2**	**-0.8**	**3.1**	**2.2**	**1.9**	**0.9**	**1.6**
	U	-0.2	-0.8	3.2	1.6	-0.4	-0.7	-0.3
	2SLS	-1.4	-7.0	2.4	1.3	0.1	0.0	1.3
	IV	-1.0	-5.9	2.4	1.7	0.6	0.0	1.2
	IGLP	94.9	93.6	92.9	92.5	91.1	91.4	89.2

Note: This table provides the coverage rates for the infeasible GLP and the differences between IGLP and the GLP.

Table S16: Compare weighting matrix: Coverage Rates (%, Design 2, $G^0 = 2$)

N	T	Weight	h=0	h=1	h=2	h=3	h=4	h=5	h=6
		UH	-14.0	-4.8	-9.6	-7.9	-6.3	-4.7	-3.8
		H	**-0.7**	**-0.3**	**-0.2**	**-0.7**	**0.1**	**0.1**	**0.1**
		U	-6.8	-4.6	-5.7	-4.0	-3.8	-2.6	-1.8
	100	2SLS	-20.6	-1.6	-4.0	-1.9	-1.9	-0.5	-0.2
		IV	-18.8	-1.0	-3.2	-2.0	-1.1	-0.4	-0.4
		IGLP	94.9	92.1	92.1	90.0	89.7	89.2	89.4
		UH	-6.2	-1.7	-3.8	-4.8	-2.3	-1.9	-0.5
		H	**0.0**	**0.4**	**0.1**	**-0.3**	**0.4**	**0.4**	**1.1**
		U	-1.3	-2.3	-3.8	-3.5	-1.5	-0.9	-0.2
100	200	2SLS	-3.8	0.5	-0.4	-0.8	0.6	0.3	1.3
		IV	-3.3	0.8	-0.7	-0.7	0.3	0.3	1.4
		IGLP	94.1	93.2	93.2	92.5	92.2	92.8	92.4
		UH	-4.2	-1.4	-3.0	-1.3	-1.2	-1.3	-0.6
		H	**0.2**	**0.3**	**-0.1**	**0.7**	**0.8**	**0.6**	**0.8**
		U	1.0	-2.5	-2.3	-0.8	-0.6	0.0	-0.2
	300	2SLS	-0.1	0.2	-0.1	0.5	0.7	0.5	0.4
		IV	-0.3	0.2	0.0	0.6	0.8	0.8	0.7
		IGLP	93.7	93.5	93.0	92.3	93.3	92.7	92.6
		UH	-26.6	-5.9	-15.4	-15.8	-9.5	-7.4	-5.7
		H	**-4.0**	**-0.6**	**-0.6**	**-1.5**	**0.1**	**-0.6**	**0.8**
		U	-9.3	-10.3	-12.8	-10.5	-6.0	-5.4	-3.8
	100	2SLS	-33.7	-3.7	-8.5	-6.3	-2.9	-1.7	-0.5
		IV	-32.1	-2.6	-6.9	-5.5	-1.7	-1.5	0.6
		IGLP	94.0	90.4	87.9	86.6	86.0	86.6	86.7
		UH	-11.9	-2.6	-8.2	-8.9	-5.8	-4.2	-2.8
		H	**-0.5**	**0.4**	**0.3**	**-0.5**	**-0.6**	**-0.3**	**0.0**
		U	-0.1	-4.8	-7.4	-6.4	-4.9	-3.8	-2.9
200	200	2SLS	-7.1	0.4	-0.6	-1.3	-1.5	-1.1	0.1
		IV	-6.6	0.3	0.0	-1.2	-1.0	-1.3	0.2
		IGLP	95.1	92.8	90.9	91.1	90.0	89.6	90.9
		UH	-6.4	-1.9	-5.5	-5.6	-3.2	-2.2	-1.9
		H	**-0.2**	**0.2**	**-0.9**	**0.0**	**0.6**	**-0.1**	**0.2**
		U	-0.1	-5.0	-5.6	-4.0	-2.8	-1.8	-1.4
	300	2SLS	-1.6	0.0	-1.1	-0.5	0.1	-0.2	0.1
		IV	-1.4	0.1	-1.2	-0.5	0.1	-0.1	0.1
		IGLP	94.9	92.8	92.2	91.4	92.2	91.4	91.3
		UH	-37.1	-9.3	-21.6	-19.7	-13.9	-9.0	-6.3
		H	**-4.2**	**-0.5**	**-2.0**	**-1.8**	**-1.4**	**-1.6**	**-0.5**
		U	-13.2	-15.3	-18.5	-13.3	-9.8	-8.4	-6.0
	100	2SLS	-44.6	-7.8	-12.7	-9.7	-4.4	-3.0	-1.7

	IV	-42.8	-6.1	-10.4	-8.1	-4.3	-2.9	-1.1
	IGLP	95.0	89.6	86.7	83.9	81.2	81.6	79.8
200	UH	-16.4	-4.2	-11.4	-10.7	-8.3	-5.1	-3.2
	H	**-1.3**	**-1.2**	**-0.1**	**-0.7**	**-0.7**	**-0.1**	**-1.5**
	U	-0.9	-9.9	-10.8	-8.6	-7.1	-4.7	-3.9
	2SLS	-9.7	-0.5	-1.5	-2.5	-1.9	-0.9	-1.2
	IV	-9.4	-0.1	-1.5	-2.1	-1.7	-0.8	-1.0
	IGLP	95.9	93.1	89.8	89.8	88.9	88.5	88.7
300	UH	-11.6	-3.1	-7.1	-5.2	-5.1	-2.6	-2.2
	H	**-0.6**	**-0.6**	**-0.2**	**0.1**	**-0.5**	**0.1**	**0.1**
	U	-0.6	-8.0	-7.7	-4.7	-4.8	-2.7	-2.8
	2SLS	-2.6	-0.4	-0.6	-0.9	-0.7	0.0	-0.4
	IV	-2.5	-0.6	-0.4	-0.6	-0.7	0.0	-0.2
	IGLP	94.9	93.6	92.9	92.5	91.1	91.4	89.2

Table S17: Compare weighting matrix: Coverage Rates (%, Design 1, $G^0 = 3$)

N	T	Weight	h=0	h=1	h=2	h=3	h=4	h=5	h=6
100	100	UH	-8.1	-6.4	-2.8	-0.1	1.5	2.6	1.3
		H	**0.0**	**-4.5**	**-2.1**	**0.8**	**3.4**	**5.6**	**6.9**
		U	0.9	-1.1	-2.5	-1.7	1.4	4.3	4.3
		2SLS	0.1	-8.9	-9.2	-6.0	-1.1	2.5	4.8
		IV	0.5	-5.9	-6.5	-3.1	0.5	3.8	5.6
		IGLP	94.9	92.1	92.1	90.0	89.7	89.2	89.4
	200	UH	-4.6	-3.4	-1.0	1.2	1.6	2.6	0.7
		H	**-0.1**	**-2.3**	**-1.4**	**0.4**	**1.9**	**4.4**	**4.7**
		U	0.7	-1.6	-1.4	0.1	1.2	3.7	3.4
		2SLS	0.7	-8.1	-7.1	-4.1	-0.9	2.9	3.4
		IV	0.5	-6.1	-5.6	-2.8	0.0	3.6	3.8
		IGLP	94.1	93.2	93.2	92.5	92.2	92.8	92.4
	300	UH	-2.1	-1.5	1.5	1.2	0.2	0.3	0.0
		H	**0.5**	**-0.2**	**1.2**	**1.7**	**2.1**	**2.5**	**3.2**
		U	0.8	-1.0	1.2	1.0	0.5	1.4	2.1
		2SLS	0.5	-4.8	-2.3	-1.2	0.1	1.8	3.0
		IV	0.4	-4.0	-1.4	-0.4	0.5	2.4	3.2
		IGLP	93.7	93.5	93.0	92.3	93.3	92.7	92.6
200	100	UH	-16.3	-15.0	-9.7	-2.3	3.1	4.5	3.6
		H	**-1.9**	**-14.8**	**-13.0**	**-4.5**	**3.0**	**7.8**	**11.2**
		U	-0.1	-4.9	-11.4	-6.9	-0.8	5.2	7.1
		2SLS	-1.7	-20.3	-24.4	-19.4	-7.4	0.1	6.8
		IV	-1.8	-16.1	-20.5	-14.5	-4.3	2.0	7.7
		IGLP	94.0	90.4	87.9	86.6	86.0	86.6	86.7
	200	UH	-9.3	-8.5	-3.5	-0.8	1.0	1.4	-0.9
		H	**-1.1**	**-8.0**	**-7.3**	**-2.5**	**1.3**	**3.4**	**4.5**
		U	-0.6	-4.1	-5.1	-3.3	0.4	2.0	3.1
		2SLS	-1.4	-17.0	-16.5	-11.3	-4.8	-0.4	2.3
		IV	-1.3	-14.1	-13.4	-9.4	-2.7	1.1	3.8
		IGLP	95.1	92.8	90.9	91.1	90.0	89.6	90.9
	300	UH	-4.5	-4.4	0.0	0.6	0.0	-0.4	-1.8
		H	**0.0**	**-2.5**	**-1.4**	**0.9**	**1.7**	**3.9**	**3.5**
		U	1.0	-1.9	-0.6	0.2	0.4	1.9	1.9
		2SLS	0.1	-10.3	-7.8	-4.1	-1.6	2.1	2.6
		IV	0.1	-8.3	-6.6	-2.8	-0.7	2.8	3.1
		IGLP	94.9	92.8	92.2	91.4	92.2	91.4	91.3
	100	UH	-25.0	-21.8	-15.6	-5.3	1.6	6.2	2.6
		H	**-4.7**	**-23.0**	**-20.2**	**-8.6**	**0.5**	**8.1**	**13.5**
		U	-1.6	-7.9	-17.3	-11.0	-3.7	4.9	8.4
		2SLS	-4.3	-26.4	-32.3	-29.7	-15.9	-3.0	6.0

	IV	-3.9	-22.2	-28.0	-23.3	-10.9	-0.1	7.6
	IGLP	95.0	89.6	86.7	83.9	81.2	81.6	79.8
200	UH	-13.6	-11.8	-4.9	-0.2	0.9	1.2	0.0
	H	**-2.9**	**-10.7**	**-9.9**	**-4.1**	**0.2**	**3.7**	**7.0**
	U	-1.0	-5.6	-7.4	-4.4	-1.0	2.7	4.9
	2SLS	-2.6	-23.0	-21.8	-17.5	-8.5	-1.2	4.4
	IV	-2.7	-19.7	-19.3	-14.4	-6.2	0.4	5.4
	IGLP	95.9	93.1	89.8	89.8	88.9	88.5	88.7
300	UH	-9.2	-5.7	-0.9	0.0	0.0	-1.4	-4.6
	H	**-1.7**	**-4.5**	**-3.9**	**-0.5**	**1.4**	**3.0**	**4.3**
	U	-0.8	-2.8	-3.2	-1.5	0.5	1.8	0.8
	2SLS	-1.9	-13.6	-11.8	-7.4	-3.0	1.2	2.8
	IV	-1.8	-12.1	-10.2	-6.2	-1.9	1.9	3.5
	IGLP	94.9	93.6	92.9	92.5	91.1	91.4	89.2

Table S18: Compare weighting matrix: Coverage Rates (%, Design 2, $G^0 = 3$)

N	T	Weight	h=0	h=1	h=2	h=3	h=4	h=5	h=6
		UH	-10.1	-3.6	-7.7	-6.4	-4.3	-3.8	-3.5
		H	**-0.8**	**1.1**	**0.8**	**0.4**	**0.8**	**0.7**	**0.7**
		U	-32.9	-5.3	-11.4	-10.9	-8.1	-6.3	-4.6
	100	2SLS	-42.6	-4.3	-11.6	-10.7	-7.1	-4.4	-2.9
		IV	-38.4	-2.2	-6.8	-5.9	-2.5	-1.3	0.0
		IGLP	94.9	92.1	92.1	90.0	89.7	89.2	89.4
		UH	-4.0	-0.8	-3.1	-2.6	-1.8	-1.1	-0.6
		H	**0.6**	**1.1**	**1.1**	**1.2**	**1.1**	**1.2**	**1.0**
		U	-8.5	-0.6	-1.8	-1.8	-1.2	-0.4	0.0
100	200	2SLS	-16.3	1.1	-0.3	-0.6	-0.2	0.7	0.7
		IV	-14.5	1.4	0.2	0.3	0.6	1.1	0.5
		IGLP	94.1	93.2	93.2	92.5	92.2	92.8	92.4
		UH	-1.2	-0.1	-1.4	-1.3	-0.9	0.2	0.0
		H	**1.3**	**1.0**	**1.0**	**1.3**	**1.0**	**1.5**	**1.1**
		U	-1.8	0.0	-0.9	-0.5	-0.3	0.5	0.5
	300	2SLS	-5.3	1.4	0.4	0.8	0.4	1.7	1.4
		IV	-4.7	1.6	0.4	1.0	0.7	1.5	1.2
		IGLP	93.7	93.5	93.0	92.3	93.3	92.7	92.6
		UH	-18.3	-4.8	-12.9	-11.8	-8.8	-6.7	-4.5
		H	**-1.8**	**0.6**	**-0.6**	**-0.8**	**-0.6**	**0.1**	**1.1**
		U	-42.9	-12.0	-22.4	-18.3	-14.5	-10.4	-7.8
	100	2SLS	-51.8	-11.4	-25.4	-19.3	-13.2	-7.3	-3.4
		IV	-47.6	-7.2	-16.5	-11.2	-6.8	-2.6	-0.7
		IGLP	94.0	90.4	87.9	86.6	86.0	86.6	86.7
		UH	-7.5	-2.2	-6.3	-5.9	-4.3	-2.3	-3.0
		H	**0.2**	**0.2**	**-0.4**	**-0.1**	**0.0**	**0.7**	**0.2**
		U	-13.4	-2.9	-5.0	-3.5	-2.7	-2.0	-1.8
200	200	2SLS	-23.7	-0.2	-2.5	-2.4	-1.3	-0.5	-0.8
		IV	-22.6	0.2	-1.7	-1.3	-0.6	0.2	-0.3
		IGLP	95.1	92.8	90.9	91.1	90.0	89.6	90.9
		UH	-5.4	-1.4	-2.9	-3.3	-3.1	-1.5	-1.2
		H	**-0.2**	**0.0**	**-0.1**	**0.2**	**0.3**	**0.5**	**0.7**
		U	-4.2	-3.0	-3.0	-2.5	-2.4	-1.0	-1.0
	300	2SLS	-10.2	0.7	0.2	-0.6	-0.4	-0.1	0.5
		IV	-10.0	0.8	-0.1	-0.3	-0.1	0.2	0.8
		IGLP	94.9	92.8	92.2	91.4	92.2	91.4	91.3
		UH	-26.8	-7.7	-16.8	-16.2	-12.1	-8.4	-6.2
		H	**-3.1**	**-0.8**	**-0.8**	**-0.9**	**-0.6**	**-0.4**	**-0.4**
		U	-48.9	-20.7	-29.0	-23.1	-17.3	-13.0	-9.8
	100	2SLS	-59.2	-19.3	-31.9	-24.6	-17.4	-11.2	-6.5

	IV	-52.8	-14.1	-20.8	-14.9	-8.4	-4.0	-2.1
	IGLP	95.0	89.6	86.7	83.9	81.2	81.6	79.8
200	UH	-12.0	-3.4	-9.6	-7.6	-6.2	-4.2	-3.1
	H	**-0.8**	**-0.6**	**-0.9**	**-1.0**	**0.1**	**-0.4**	**-0.1**
	U	-17.9	-6.5	-9.1	-6.3	-4.6	-3.1	-2.5
	2SLS	-30.0	-1.5	-4.6	-3.7	-1.9	-0.9	-1.0
	IV	-28.7	-1.0	-4.3	-2.8	-1.5	-0.7	-0.4
	IGLP	95.9	93.1	89.8	89.8	88.9	88.5	88.7
300	UH	-7.0	-2.5	-6.2	-5.0	-4.5	-2.6	-1.3
	H	**-0.4**	**-0.3**	**-0.6**	**-0.1**	**-0.3**	**0.5**	**0.3**
	U	-4.5	-5.3	-5.8	-4.1	-3.3	-2.2	-1.7
	2SLS	-11.7	-0.2	-1.5	-0.8	-1.1	0.0	-0.1
	IV	-11.0	-0.2	-1.3	-0.4	-0.8	0.5	0.0
	IGLP	94.9	93.6	92.9	92.5	91.1	91.4	89.2

Table S19: Compare Inference Methods: Band Ratios (%)

Design	N	T	$G^0 = 2$ Large T	$G^0 = 2$ Small T	$G^0 = 3$ Large T	$G^0 = 3$ Small T
1	100	100	15.8	16.1	20.3	20.9
	100	200	15.0	15.1	18.8	19.0
	100	300	14.7	14.8	18.2	18.4
	200	100	11.2	11.4	14.4	14.8
	200	200	10.6	10.7	13.3	13.4
	200	300	10.4	10.5	12.9	13.0
	300	100	9.2	9.3	11.8	12.1
	300	200	8.7	8.7	10.9	11.0
	300	300	8.5	8.5	10.5	10.6
2	100	100	15.9	16.1	19.5	19.7
	100	200	15.0	15.1	18.3	18.5
	100	300	14.7	14.8	18.0	18.1
	200	100	11.3	11.4	13.8	14.0
	200	200	10.6	10.7	13.0	13.1
	200	300	10.4	10.4	12.7	12.8
	300	100	9.2	9.3	11.3	11.4
	300	200	8.7	8.7	10.6	10.7
	300	300	8.5	8.5	10.4	10.4

Note: This table reports the band ratios for Large T inference (Theorem 2) and small T adjustment (Proposition ??) respectively.

Table S20: Large T Inference Coverage Rates (%)

Design	N	T	$G^0 = 2$							$G^0 = 3$						
			h=0	h=1	h=2	h=3	h=4	h=5	h=6	h=0	h=1	h=2	h=3	h=4	h=5	h=6
1	100	100	92.0	91.2	88.3	86.6	88.2	89.2	89.6	93.5	86.9	88.2	90.7	92.4	93.8	94.1
	100	200	92.2	92.9	93.8	94.0	93.8	92.1	91.7	94.1	90.4	91.0	93.7	94.2	94.9	94.7
	100	300	94.4	93.6	95.4	94.9	94.3	93.2	94.3	94.8	93.9	93.4	93.9	94.4	95.3	94.7
	200	100	89.8	86.0	80.2	76.7	80.0	84.5	86.1	91.7	76.4	76.6	83.9	87.2	90.3	93.6
	200	200	93.3	91.7	93.1	92.1	92.4	91.0	90.6	93.5	86.1	86.6	89.5	92.2	94.0	94.7
	200	300	94.1	92.7	95.1	94.3	92.9	92.2	92.9	95.1	90.2	91.5	91.5	92.3	94.2	94.2
	300	100	88.1	80.7	71.2	70.1	73.8	80.3	81.5	90.1	65.6	66.8	75.9	83.8	88.1	91.7
	300	200	91.7	88.9	92.3	89.8	88.4	88.7	87.7	92.5	81.2	81.7	86.2	89.3	92.5	93.4
	300	300	92.9	92.0	95.9	94.3	92.3	90.8	90.8	94.0	89.9	88.9	90.8	92.1	92.9	93.0
2	100	100	93.6	93.2	92.3	90.6	90.9	90.9	91.4	93.5	93.8	91.8	90.6	91.9	91.0	92.1
	100	200	94.7	94.0	93.3	92.2	92.9	93.9	93.9	93.9	94.1	93.8	93.5	92.9	92.6	93.3
	100	300	94.9	94.2	94.5	93.3	93.7	93.7	94.4	94.2	94.0	94.2	94.3	93.3	94.6	93.6
	200	100	91.4	91.2	88.6	87.6	88.2	87.5	87.8	92.7	91.4	90.2	89.2	88.6	88.0	89.0
	200	200	94.2	93.5	92.9	91.4	92.2	92.1	90.7	94.1	93.4	92.5	93.3	93.1	91.7	91.5
	200	300	94.6	92.6	92.4	93.7	93.5	93.3	93.4	94.4	93.8	93.5	92.4	93.1	92.6	92.5
	300	100	89.0	90.8	85.7	83.4	82.8	83.5	83.6	92.1	89.5	87.7	86.2	86.9	87.3	86.6
	300	200	92.4	92.4	90.7	89.0	89.1	89.0	90.1	93.7	92.7	91.4	90.6	91.2	91.0	91.0
	300	300	93.1	93.4	91.4	91.9	90.7	91.0	91.0	94.1	93.3	93.6	92.8	91.8	91.1	90.9

Note: This table reports the coverage probability of the large T inference as in Theorem 2.

Table S21: Small T Adjustment Coverage Rates (%)

Design	N	T	$G^0 = 2$ h=0	h=1	h=2	h=3	h=4	h=5	h=6	$G^0 = 3$ h=0	h=1	h=2	h=3	h=4	h=5	h=6
1	100	100	92.2	91.7	88.9	87.6	88.8	89.9	90.4	93.7	88.1	89.5	91.9	92.9	94.6	95.0
	100	200	92.3	93.2	93.9	94.1	94.1	92.5	92.6	94.2	90.9	91.6	94.1	94.7	95.3	95.2
	100	300	94.6	93.8	95.7	95.1	94.5	93.4	94.2	94.8	94.1	93.6	94.0	94.6	95.4	95.1
	200	100	89.8	86.5	81.3	77.8	80.8	85.0	86.4	91.7	77.3	78.2	85.0	87.8	91.2	94.1
	200	200	93.4	91.9	93.4	92.1	92.7	91.1	91.0	93.6	86.6	87.2	90.0	92.7	94.3	94.9
	200	300	94.2	92.7	95.2	94.5	93.2	92.5	92.9	95.2	90.5	91.8	91.8	92.6	94.5	94.3
	300	100	87.9	81.7	72.3	71.4	74.8	80.9	82.3	90.1	66.6	68.6	77.3	84.4	88.7	92.2
	300	200	91.8	89.0	92.5	90.2	88.8	88.9	87.9	92.1	81.7	82.1	86.9	89.9	93.0	93.9
	300	300	92.8	92.1	95.8	94.4	92.3	91.2	91.2	94.0	90.1	89.4	91.1	92.5	93.0	93.1
2	100	100	93.7	93.6	92.5	91.0	91.5	91.5	91.7	93.4	93.9	92.4	91.3	92.6	91.6	92.4
	100	200	94.7	94.1	93.5	92.6	93.0	94.0	94.3	93.9	94.3	94.1	93.7	93.2	92.8	93.5
	100	300	94.9	94.3	94.5	93.4	94.0	93.8	94.5	94.4	94.2	94.3	94.5	93.5	94.9	93.8
	200	100	91.3	91.3	89.1	87.9	88.8	88.2	88.3	92.7	91.3	90.7	89.8	89.2	88.5	89.9
	200	200	94.1	93.6	93.1	91.8	92.2	92.2	90.9	94.3	93.5	93.1	93.5	93.1	92.1	91.7
	200	300	94.6	92.7	92.5	93.9	93.8	93.5	93.5	94.4	93.8	93.5	92.5	93.3	92.9	92.8
	300	100	88.9	90.8	86.0	84.1	83.5	84.5	84.3	92.1	89.4	87.8	86.7	87.5	87.9	87.6
	300	200	92.4	92.2	90.9	89.1	89.3	89.1	90.3	93.6	92.6	91.7	90.7	91.4	91.4	91.1
	300	300	93.2	93.4	91.5	92.0	90.7	91.1	91.2	94.2	93.4	93.7	92.8	92.1	91.3	91.0

Note: This table reports the coverage probability of the small T adjustment as in Proposition **??**.

Table S22: GLP Performance (First-Differenced)

			$G^0 = 2$							$G^0 = 3$						
				BR (%)		RMSE ($\times$100)					BR (%)		RMSE ($\times$100)			
Design	N	T	AC (%)	GLP	IGLP	GLP	PAN	IND	IGLP	AC (%)	GLP	IGLP	GLP	PAN	IND	IGLP
1	100	100	84.6	15.3	13.8	29.9	33.8	84.6	10.9	77.5	21.6	14.2	60.8	82.8	2003.9	16.7
	100	200	93.2	14.7	13.3	18.8	33.3	54.7	7.4	89.6	20.1	13.4	37.5	82.5	108.8	10.9
	100	300	96.7	14.5	13.1	13.2	33.2	43.7	6.0	94.7	19.5	13.1	26.7	82.4	53.2	8.6
	200	100	84.8	10.8	9.8	28.8	33.4	87.0	7.6	77.1	15.5	10.2	60.1	82.6	67124.6	11.8
	200	200	93.3	10.4	9.5	17.9	33.1	54.6	5.3	89.6	14.3	9.6	36.9	82.4	6292.9	7.6
	200	300	96.7	10.2	9.3	12.6	33.1	43.7	4.3	94.6	13.9	9.4	26.3	82.3	53.1	6.1
	300	100	84.7	8.8	8.1	28.6	33.2	85.3	6.2	77.0	12.7	8.3	59.9	82.5	4156.9	9.7
	300	200	93.2	8.5	7.7	17.9	33.1	54.7	4.4	89.7	11.7	7.9	36.5	82.4	250.7	6.3
	300	300	96.7	8.4	7.6	12.4	33.0	43.7	3.5	94.7	11.3	7.7	26.1	82.3	54.3	5.0
2	100	100	98.5	15.3	13.8	8.5	22.3	50.1	6.2	97.8	18.8	16.6	12.2	36.8	65.3	9.8
	100	200	99.9	14.7	13.3	4.6	22.0	31.4	4.3	99.8	18.0	16.0	7.1	36.5	41.8	6.7
	100	300	100.0	14.5	13.2	3.6	22.0	25.1	3.5	100.0	17.7	15.9	5.6	36.4	33.4	5.4
	200	100	98.4	10.9	9.8	7.4	22.0	70.6	4.4	97.7	13.4	11.9	10.1	36.5	64.4	7.0
	200	200	99.9	10.4	9.5	3.4	21.9	31.4	3.0	99.8	12.7	11.5	5.2	36.4	41.8	4.8
	200	300	100.0	10.2	9.4	2.6	21.9	25.1	2.5	100.0	12.5	11.4	4.0	36.3	33.4	3.8
	300	100	98.4	8.9	8.0	6.8	22.0	123.0	3.6	97.7	10.9	9.8	9.2	36.4	70.6	5.7
	300	200	99.9	8.5	7.8	3.0	21.9	31.4	2.5	99.8	10.4	9.4	4.5	36.3	41.8	3.9
	300	300	100.0	8.4	7.7	2.1	21.9	25.0	2.0	100.0	10.2	9.3	3.3	36.3	33.4	3.2

Note: This table reports the classification accuracy (AC), the confidence bands ratios between the GLP and the individual LP-IV (BR), and the RMSE of the GLP. GLP, PAN, IND and IGLP stand for the GLP, panel LP-IV, individual LP-IV and the infeasible GLP respectively. The infeasible GLP is group-by-group using standard panel LP-IV where we know the true group structure beforehand. Classification accuracy and band ratios are in percentage terms, and RMSE are multiplied by 100.

Table S23: GLP Coverage Rates (%, First-Differenced)

			$G^0 = 2$							$G^0 = 3$						
Design	N	T	h=0	h=1	h=2	h=3	h=4	h=5	h=6	h=0	h=1	h=2	h=3	h=4	h=5	h=6
1	100	100	93.1	91.3	91.5	95.4	94.7	94.3	94.8	97.9	96.2	97.4	98.4	98.6	98.5	98.3
	100	200	94.8	96.3	95.7	97.0	96.5	97.0	96.0	97.9	97.3	98.3	98.4	98.2	98.3	98.1
	100	300	96.1	97.1	97.3	96.8	96.4	95.6	96.8	98.0	97.4	98.2	98.1	97.9	97.8	98.4
	200	100	91.4	85.1	84.6	93.9	95.9	94.3	94.5	97.7	91.7	95.2	97.8	98.4	97.9	98.5
	200	200	94.0	94.8	94.7	96.0	96.8	96.2	95.5	97.9	96.5	96.4	98.0	97.8	98.2	98.5
	200	300	95.4	96.6	96.2	97.2	95.8	97.3	97.0	98.3	97.0	96.7	97.3	97.5	98.0	98.2
	300	100	88.6	79.2	75.8	93.1	96.1	94.1	93.3	97.1	87.5	92.8	97.6	97.9	98.0	98.6
	300	200	93.1	94.0	92.5	97.1	95.9	95.1	95.0	97.3	95.1	95.1	97.2	98.3	98.1	98.1
	300	300	94.7	96.0	96.3	97.1	95.9	95.8	95.6	97.6	96.3	95.5	97.2	98.3	98.3	98.3
2	100	100	95.0	93.8	94.8	95.8	95.6	96.8	96.9	95.0	95.2	95.3	95.8	95.7	96.5	96.5
	100	200	95.0	94.7	95.3	95.9	96.2	96.3	96.5	96.3	95.4	95.9	96.6	96.8	96.9	97.2
	100	300	96.3	95.7	95.7	96.3	96.0	95.8	96.5	96.2	95.6	96.3	96.5	96.2	97.3	97.3
	200	100	92.7	91.3	93.8	95.7	95.6	97.1	96.3	93.2	92.9	94.8	95.2	96.1	96.6	95.9
	200	200	95.8	94.7	95.4	95.8	96.8	96.6	96.8	96.1	94.5	94.8	95.6	96.7	96.4	96.4
	200	300	95.9	94.2	95.4	96.2	96.1	96.1	96.7	96.3	95.6	96.0	97.0	96.2	96.4	96.3
	300	100	91.9	89.1	93.1	94.1	96.2	96.4	97.1	92.5	91.7	93.5	95.6	95.7	96.3	96.2
	300	200	94.8	92.8	93.8	95.8	96.8	96.1	96.5	95.2	93.3	94.6	95.5	96.3	95.8	96.1
	300	300	94.7	94.0	94.9	96.3	96.6	96.1	96.9	96.5	94.2	95.9	96.4	96.1	96.6	96.5

Note: This table reports the coverage probability of the large T inference as in Theorem 2.

Table S24: Infeasible GLP Coverage Rates (%, First-Differenced)

			$G^0 = 2$							$G^0 = 3$						
Design	N	T	h=0	h=1	h=2	h=3	h=4	h=5	h=6	h=0	h=1	h=2	h=3	h=4	h=5	h=6
1	100	100	94.0	93.4	94.6	93.5	92.8	95.2	94.5	93.0	94.4	95.0	94.2	94.0	94.1	94.1
	100	200	93.4	93.9	94.2	95.1	94.4	94.0	93.9	93.7	93.6	94.0	93.0	93.8	93.2	93.9
	100	300	95.4	94.9	95.0	93.8	93.4	92.9	93.9	93.9	93.2	93.5	94.0	93.2	93.6	94.2
	200	100	93.2	94.1	94.7	94.0	95.2	95.1	94.8	94.1	95.1	94.9	94.1	95.2	95.0	94.5
	200	200	94.1	94.5	94.1	94.5	94.7	95.4	93.4	94.9	94.4	94.6	94.7	94.7	95.3	94.2
	200	300	94.3	95.3	94.7	95.2	93.9	95.7	95.1	94.6	95.2	93.7	94.4	93.8	93.7	94.2
	300	100	94.6	94.1	95.2	95.2	95.2	94.9	95.0	94.6	95.2	94.4	94.5	94.6	94.4	94.5
	300	200	94.9	94.3	94.0	95.4	94.3	95.0	94.0	94.7	95.4	94.7	94.6	94.4	94.6	94.7
	300	300	95.2	94.8	94.6	94.1	93.9	93.6	94.0	94.5	94.5	95.1	94.5	94.6	94.6	94.6
2	100	100	95.1	93.5	93.9	94.3	93.8	94.4	94.7	94.1	94.1	93.8	92.9	93.2	93.8	93.8
	100	200	93.9	93.8	94.1	93.8	93.0	94.8	94.1	93.8	93.4	94.1	94.0	93.7	93.9	94.6
	100	300	94.8	94.5	93.7	94.0	93.7	93.7	93.7	93.3	93.4	93.7	94.0	93.5	94.2	93.9
	200	100	95.5	94.4	95.5	95.5	94.7	95.0	94.8	94.2	94.3	94.2	93.7	94.6	94.6	93.8
	200	200	95.4	94.0	94.7	94.5	95.0	94.7	94.4	95.2	93.5	93.8	94.1	94.5	94.9	94.6
	200	300	95.1	94.6	94.6	94.5	94.5	94.8	95.3	94.5	94.8	95.3	94.6	94.0	94.3	94.0
	300	100	94.1	95.2	94.9	94.6	95.2	94.9	95.3	94.7	94.0	94.7	94.6	94.6	94.4	94.1
	300	200	94.3	94.7	93.9	95.4	95.4	94.3	94.8	93.9	94.4	93.8	94.6	94.9	93.5	94.0
	300	300	94.0	95.4	94.8	94.7	95.1	94.0	94.2	95.1	94.1	94.9	95.0	95.1	94.9	94.4

Note: Standard errors are clustered at individual level (Cameron and Miller, 2015).

Table S25: Compare GMM Criteria

			$G^0 = 2$				$G^0 = 3$			
			BR (%)		RMSE ($\times$100)		BR (%)		RMSE ($\times$100)	
Design	N	T	POOL	BASE	POOL	BASE	POOL	BASE	POOL	BASE
1	100	100	15.7	15.7	10.0	10.3	19.3	19.3	15.2	17.2
	100	200	14.8	14.9	6.5	6.6	18.1	18.2	9.5	10.4
	100	300	14.5	14.6	5.2	5.3	17.6	17.9	7.5	8.0
	200	100	11.2	11.1	7.9	8.4	13.8	13.6	12.3	14.8
	200	200	10.5	10.5	4.8	5.0	12.9	12.9	7.2	8.3
	200	300	10.3	10.4	3.8	4.0	12.6	12.7	5.5	6.1
	300	100	9.2	9.1	7.0	7.5	11.3	11.2	10.9	13.6
	300	200	8.6	8.6	4.2	4.4	10.6	10.5	6.3	7.5
	300	300	8.5	8.5	3.2	3.3	10.4	10.4	4.7	5.5
2	100	100	15.7	15.6	5.9	6.0	19.0	19.1	9.1	9.3
	100	200	14.8	14.9	3.9	4.0	17.9	18.2	5.9	6.0
	100	300	14.5	14.6	3.1	3.1	17.6	17.9	4.7	4.8
	200	100	11.2	11.1	4.5	4.7	13.6	13.5	7.0	7.2
	200	200	10.5	10.5	2.9	3.0	12.8	12.9	4.4	4.5
	200	300	10.3	10.3	2.3	2.3	12.6	12.6	3.4	3.5
	300	100	9.2	9.0	4.0	4.2	11.2	11.1	6.1	6.4
	300	200	8.6	8.6	2.5	2.6	10.5	10.5	3.8	3.9
	300	300	8.5	8.4	1.9	2.0	10.3	10.3	2.9	3.0

Note: Column POOL indicates the fully-pooled GMM criterion, Column BASE indicates the baseline GLP criterion, and Column DIFF shows the difference.

Table S26: Coverage Rates (%): Fully-Pooled GMM

Design	N	T	$G^0=2$ h=0	h=1	h=2	h=3	h=4	h=5	h=6	$G^0=3$ h=0	h=1	h=2	h=3	h=4	h=5	h=6
1	100	100	94.4	92.5	90.5	90.2	90.8	89.0	89.9	93.4	92.4	90.0	89.8	87.9	87.1	87.5
	100	200	93.4	93.4	91.9	92.2	92.2	92.2	92.3	93.8	93.5	93.1	92.6	90.9	91.1	89.8
	100	300	92.6	93.3	93.7	93.2	93.2	92.7	92.4	93.9	93.7	93.8	92.5	92.0	91.6	92.1
	200	100	94.2	91.1	88.4	86.9	85.6	85.2	84.1	93.9	91.1	89.1	86.7	84.3	83.8	82.2
	200	200	93.7	92.6	92.2	91.3	90.8	90.3	90.4	94.0	92.7	91.4	91.4	90.8	89.9	89.0
	200	300	94.5	93.2	92.5	91.6	91.9	92.0	90.8	94.2	93.7	93.2	93.1	92.5	91.1	91.5
	300	100	94.1	89.1	84.8	84.0	81.4	81.8	81.9	94.7	91.6	87.4	84.2	80.6	78.9	77.1
	300	200	93.8	92.6	91.4	89.3	89.0	88.7	88.9	94.2	91.5	90.9	89.9	88.2	87.3	86.4
	300	300	94.2	94.6	91.9	91.6	90.2	91.2	91.0	94.3	93.6	92.4	91.8	91.6	90.6	89.9
2	100	100	94.4	92.9	91.7	90.3	90.3	91.1	90.4	93.5	92.5	90.9	90.6	90.3	90.5	90.4
	100	200	93.8	92.7	93.2	92.9	93.0	91.3	92.8	93.8	93.2	92.4	92.0	92.5	92.7	92.9
	100	300	94.1	94.4	94.2	93.3	93.4	92.7	92.5	93.9	93.9	93.0	93.0	93.2	93.5	92.8
	200	100	95.0	92.2	90.5	89.1	88.0	88.9	87.1	94.8	93.3	90.6	88.6	89.8	88.1	88.2
	200	200	94.4	93.7	92.8	91.2	90.8	89.9	91.2	95.0	94.2	93.0	92.3	91.8	91.3	92.2
	200	300	93.6	93.9	93.4	92.5	91.3	92.2	92.2	93.7	93.6	92.4	93.1	93.3	93.4	92.7
	300	100	94.3	91.4	86.9	85.4	85.0	83.1	84.6	95.1	90.8	88.6	87.2	85.6	85.9	86.1
	300	200	94.3	93.5	91.0	89.9	89.6	89.6	89.9	94.2	93.5	92.4	91.1	90.7	90.5	91.1
	300	300	94.1	93.3	91.7	91.4	91.8	91.4	91.8	94.5	92.9	92.6	92.5	92.2	92.1	92.1

Note: Standard errors are clustered at individual level (Cameron and Miller, 2015).

Table S27: Coverage Rates (%): Baseline GLP

Design	N	T	$G^0 = 2$ h=0	h=1	h=2	h=3	h=4	h=5	h=6	$G^0 = 3$ h=0	h=1	h=2	h=3	h=4	h=5	h=6
1	100	100	93.5	91.8	89.8	89.6	90.0	88.7	89.5	94.5	91.1	88.2	87.1	84.3	83.8	83.2
	100	200	94.2	93.9	92.6	92.1	92.3	92.7	92.5	94.5	94.2	92.3	91.8	90.1	89.0	88.0
	100	300	93.7	93.7	94.1	93.2	93.2	93.2	93.1	95.0	94.6	93.9	92.6	91.6	90.9	91.2
	200	100	93.8	89.5	84.7	84.2	83.4	82.8	83.5	93.3	87.9	83.8	79.3	76.7	75.1	73.8
	200	200	92.8	91.7	90.6	89.7	90.2	90.1	90.7	94.0	92.0	89.3	88.3	87.3	85.7	84.0
	200	300	93.3	93.0	91.7	91.0	91.2	92.0	90.4	94.0	93.6	92.0	91.2	89.8	88.3	88.9
	300	100	90.6	85.6	80.8	78.7	78.1	79.0	81.3	92.5	87.5	79.4	73.6	69.7	67.9	67.0
	300	200	92.6	90.3	89.8	87.2	87.6	87.1	87.0	93.7	90.0	86.6	84.4	82.2	81.1	79.8
	300	300	93.7	93.2	90.6	90.4	89.6	90.0	90.3	94.0	92.5	89.6	88.1	87.1	86.2	85.8
2	100	100	94.1	92.1	90.7	89.8	90.3	91.0	90.4	93.4	92.8	90.9	90.4	90.8	91.3	91.6
	100	200	93.4	93.3	93.2	92.0	92.6	92.3	93.1	95.2	93.9	92.7	92.8	93.2	93.6	94.1
	100	300	94.4	94.6	94.3	93.8	93.6	92.9	93.1	95.0	94.9	93.8	93.5	93.7	94.5	94.1
	200	100	93.1	91.1	87.1	86.8	87.1	87.8	87.2	93.2	92.0	88.9	87.4	88.4	88.6	88.7
	200	200	93.8	93.5	91.7	90.8	90.3	90.3	90.7	94.6	94.2	92.1	91.8	91.7	91.2	92.5
	200	300	93.3	94.1	93.0	92.1	91.4	92.2	92.1	94.0	93.4	92.5	93.4	93.3	93.2	93.0
	300	100	91.4	89.5	82.9	82.0	83.0	82.6	83.6	92.0	89.5	86.0	84.4	84.1	85.7	85.1
	300	200	92.7	92.1	88.6	88.3	88.3	89.3	89.8	94.1	93.1	91.0	89.2	89.7	90.2	90.9
	300	300	93.9	92.7	90.6	90.9	91.1	91.1	92.0	94.1	92.7	92.1	91.9	92.1	92.1	92.2

Note: Standard errors are computed as in Theorem 2.

Table S28: GLP Performance (Bias-Corrected)

			$G^0 = 2$							$G^0 = 3$						
				BR (%)		RMSE (×100)					BR (%)		RMSE (×100)			
Design	N	T	AC (%)	GLP	IGLP	GLP	PAN	IND	IGLP	AC (%)	GLP	IGLP	GLP	PAN	IND	IGLP
1	100	100	84.1	15.8	15.7	29.7	33.6	63.8	10.1	78.4	20.3	19.3	52.7	81.6	77.5	14.9
	100	200	92.8	15.0	14.8	18.9	33.0	43.4	6.5	90.2	18.8	18.1	34.3	81.2	52.6	9.5
	100	300	96.3	14.7	14.5	13.4	32.8	35.0	5.1	94.9	18.2	17.7	25.0	81.1	42.4	7.5
	200	100	84.1	11.2	11.2	29.0	33.2	63.8	8.0	78.3	14.4	13.8	52.2	81.5	77.6	12.0
	200	200	92.8	10.6	10.5	18.4	32.8	43.4	4.9	90.2	13.3	12.9	33.8	81.1	52.5	7.2
	200	300	96.3	10.4	10.3	13.0	32.7	34.9	3.8	95.0	12.9	12.6	24.3	81.0	42.3	5.5
	300	100	84.1	9.2	9.2	28.8	33.2	63.9	7.2	78.3	11.8	11.3	52.1	81.4	77.7	10.8
	300	200	92.7	8.7	8.6	18.3	32.8	43.4	4.1	90.0	10.9	10.6	33.9	81.0	52.6	6.3
	300	300	96.4	8.5	8.5	12.9	32.7	35.0	3.2	95.1	10.5	10.4	24.1	81.0	42.3	4.8
2	100	100	99.7	15.9	15.7	6.7	22.3	38.5	6.0	99.7	19.5	19.0	9.6	36.8	50.5	8.9
	100	200	100.0	15.0	14.8	4.0	22.0	26.2	3.9	100.0	18.3	17.9	6.1	36.5	34.4	6.0
	100	300	100.0	14.7	14.5	3.1	21.9	21.1	3.1	100.0	18.0	17.6	4.8	36.4	27.8	4.7
	200	100	99.7	11.3	11.2	5.3	22.1	38.5	4.6	99.6	13.8	13.6	7.6	36.6	50.6	6.8
	200	200	100.0	10.6	10.5	3.0	21.9	26.2	2.9	100.0	13.0	12.8	4.6	36.4	34.4	4.5
	200	300	100.0	10.4	10.3	2.3	21.9	21.1	2.3	100.0	12.7	12.6	3.5	36.3	27.7	3.5
	300	100	99.7	9.2	9.2	4.8	22.1	38.5	4.0	99.6	11.3	11.2	6.8	36.5	50.6	6.1
	300	200	100.0	8.7	8.6	2.6	21.9	26.2	2.5	100.0	10.6	10.5	3.9	36.4	34.4	3.7
	300	300	100.0	8.5	8.5	2.0	21.9	21.1	1.9	100.0	10.4	10.3	3.0	36.3	27.7	2.9

Note: This table reports the classification accuracy (AC), the confidence bands ratios between the GLP and the individual LP-IV (BR), and the RMSE of the GLP. GLP, PAN, IND and IGLP stand for the GLP, panel LP-IV, individual LP-IV and the infeasible GLP respectively. The infeasible GLP is group-by-group using standard panel LP-IV where we know the true group structure beforehand. Classification accuracy and band ratios are in percentage terms, and RMSE are multiplied by 100.

Table S29: GLP Coverage Rates (%, Bias-Corrected)

			$G^0 = 2$							$G^0 = 3$						
Design	N	T	h=0	h=1	h=2	h=3	h=4	h=5	h=6	h=0	h=1	h=2	h=3	h=4	h=5	h=6
1	100	100	92.3	91.2	89.4	86.2	86.5	87.5	88.0	93.7	86.9	87.8	90.8	91.7	93.4	94.3
	100	200	94.4	93.7	94.2	93.2	93.2	92.8	93.1	93.7	91.0	91.5	93.0	94.2	95.0	95.4
	100	300	95.5	94.0	94.1	94.8	94.2	94.1	93.4	94.2	92.9	93.3	94.2	94.8	95.1	94.8
	200	100	90.6	85.4	79.0	75.9	79.0	83.6	85.4	92.8	77.8	79.0	83.2	87.5	89.5	92.9
	200	200	93.4	92.4	93.4	92.6	90.3	90.8	90.0	94.0	85.9	87.0	89.3	91.3	93.5	94.5
	200	300	93.9	93.3	95.7	93.9	93.9	94.4	93.1	94.7	90.9	91.2	92.6	93.9	94.5	94.0
	300	100	91.9	80.9	68.4	67.2	71.6	76.8	81.0	92.2	68.8	68.2	73.9	80.9	86.7	90.4
	300	200	94.5	89.7	91.8	88.7	89.2	89.0	88.4	93.9	81.7	82.3	84.4	89.8	92.0	92.9
	300	300	93.9	93.4	95.1	93.4	93.1	92.6	91.3	94.5	88.1	89.4	91.9	93.0	94.3	93.7
2	100	100	94.7	92.6	90.6	90.9	90.1	89.0	89.5	94.6	91.8	91.4	92.0	91.9	91.9	91.8
	100	200	95.0	93.7	93.2	92.5	91.9	92.4	93.2	94.8	93.8	93.0	92.9	92.7	93.4	93.5
	100	300	95.7	94.5	93.7	93.0	93.9	94.3	92.6	94.6	93.8	94.4	94.4	94.4	93.5	94.4
	200	100	94.7	92.2	87.5	87.4	86.5	86.2	87.3	93.3	92.0	89.6	90.0	89.1	88.4	89.4
	200	200	94.8	92.1	91.5	91.3	91.8	90.4	92.5	94.4	93.0	92.1	92.1	91.8	91.5	91.0
	200	300	94.5	93.9	92.2	92.3	92.4	93.2	91.9	94.2	93.3	93.3	93.1	93.3	92.9	93.0
	300	100	94.2	89.0	85.5	84.0	84.0	84.5	84.1	94.7	90.3	88.0	87.0	86.4	85.3	86.9
	300	200	94.6	91.2	90.4	89.5	89.1	88.2	89.0	95.1	91.8	90.5	90.1	91.2	91.7	90.4
	300	300	94.2	92.4	90.4	91.5	90.8	90.8	91.1	94.8	93.4	92.5	92.2	92.2	92.2	92.2

Note: This table reports the coverage probability of the bias-corrected GLP. The coverage rates are computed using large T inference as in Theorem 2.

Table S30: GLP Performance: Horizon-by-Horizon Grouping ($G^0 = 2$)

			AC (%)		BR (%)			RMSE (×100)				
Design	N	T	BASE	HBH	BASE	HBH	IGLP	BASE	HBH	PAN	IND	IGLP
1	100	100	84.1	64.3	15.9	15.9	15.7	29.6	50.9	33.5	63.9	10.1
	100	200	92.7	69.2	15.0	15.0	14.8	18.9	35.7	32.9	43.4	6.4
	100	300	96.3	72.2	14.7	14.7	14.5	13.5	28.5	32.8	35.0	5.2
	200	100	84.1	63.9	11.2	11.2	11.2	28.9	50.8	33.2	63.8	7.9
	200	200	92.7	69.0	10.6	10.6	10.5	18.5	35.5	32.8	43.4	4.9
	200	300	96.3	72.1	10.4	10.4	10.3	13.1	28.5	32.7	35.0	3.8
	300	100	84.0	63.8	9.2	9.2	9.2	28.8	50.7	33.1	63.8	7.0
	300	200	92.6	68.9	8.7	8.7	8.6	18.4	35.5	32.8	43.4	4.2
	300	300	96.3	72.0	8.5	8.5	8.5	13.0	28.4	32.7	35.1	3.2
2	100	100	99.7	65.5	15.9	17.0	15.7	6.4	28.7	22.3	38.5	5.9
	100	200	100.0	68.1	15.0	16.0	14.8	3.9	20.2	22.0	26.2	3.9
	100	300	100.0	69.7	14.7	15.6	14.5	3.1	16.5	21.9	21.1	3.1
	200	100	99.7	65.0	11.3	12.1	11.2	5.3	28.6	22.1	38.5	4.5
	200	200	100.0	67.6	10.6	11.3	10.5	3.0	20.1	21.9	26.2	2.9
	200	300	100.0	69.4	10.4	11.1	10.4	2.3	16.4	21.9	21.1	2.2
	300	100	99.7	64.8	9.2	9.9	9.2	4.8	28.6	22.1	38.6	4.0
	300	200	100.0	67.5	8.7	9.2	8.6	2.6	20.1	21.9	26.1	2.5
	300	300	100.0	69.2	8.5	9.0	8.5	2.0	16.4	21.9	21.1	1.9

Note: This table reports the classification accuracy (AC), the confidence bands ratios between the GLP and the individual LP-IV (BR), and the RMSE of the GLP. BASE stands for baseline GLP that groups all horizons together, HBH stands for GLP that groups IRs horizon-by-horizon. PAN, IND and IGLP stand for the panel LP-IV, individual LP-IV and the infeasible GLP respectively. Classification accuracy and band ratios are in percentage terms, and RMSE are multiplied by 100.

Table S31: GLP Performance: Horizon-by-Horizon Grouping ($G^0 = 3$)

			AC (%)		BR (%)			RMSE (×100)				
Design	N	T	BASE	HBH	BASE	HBH	IGLP	BASE	HBH	PAN	IND	IGLP
1	100	100	78.3	60.1	20.3	20.6	19.3	52.8	71.0	81.7	77.6	15.3
	100	200	90.2	69.0	18.8	18.9	18.0	34.2	48.2	81.2	52.5	9.5
	100	300	95.0	73.8	18.2	18.3	17.6	24.7	37.6	81.1	42.2	7.6
	200	100	78.4	59.7	14.4	14.6	13.8	52.1	70.7	81.5	77.6	12.0
	200	200	90.3	68.7	13.3	13.4	12.9	33.6	48.1	81.1	52.6	7.1
	200	300	95.1	73.6	12.9	13.0	12.6	24.1	37.5	81.0	42.2	5.5
	300	100	78.4	59.6	11.8	11.9	11.3	51.8	70.6	81.4	77.5	10.9
	300	200	90.3	68.6	10.9	10.9	10.6	33.5	48.0	81.0	52.5	6.2
	300	300	95.0	73.4	10.5	10.6	10.4	24.2	37.7	81.0	42.3	4.8
2	100	100	99.6	54.1	19.5	22.0	19.1	9.6	41.6	36.8	50.6	8.9
	100	200	100.0	57.0	18.3	20.5	17.9	6.1	29.1	36.5	34.3	6.0
	100	300	100.0	59.1	18.0	20.0	17.6	4.8	23.8	36.4	27.8	4.8
	200	100	99.6	53.3	13.8	15.6	13.6	7.7	41.3	36.6	50.6	7.0
	200	200	100.0	56.4	13.0	14.6	12.8	4.6	29.0	36.4	34.4	4.5
	200	300	100.0	58.4	12.7	14.2	12.6	3.5	23.7	36.3	27.8	3.5
	300	100	99.6	53.0	11.3	12.8	11.2	6.9	41.2	36.5	50.6	6.1
	300	200	100.0	56.1	10.6	11.9	10.5	3.9	29.0	36.3	34.5	3.7
	300	300	100.0	58.1	10.4	11.6	10.3	3.0	23.7	36.3	27.8	3.0

Note: This table reports the classification accuracy (AC), the confidence bands ratios between the GLP and the individual LP-IV (BR), and the RMSE of the GLP. BASE stands for baseline GLP that groups all horizons together, HBH stands for GLP that groups IRs horizon-by-horizon. PAN, IND and IGLP stand for the panel LP-IV, individual LP-IV and the infeasible GLP respectively. Classification accuracy and band ratios are in percentage terms, and RMSE are multiplied by 100.

Table S32: Differences in Coverage Rates (%, HBH-BASE)

			$G^0 = 2$							$G^0 = 3$						
Design	N	T	h=0	h=1	h=2	h=3	h=4	h=5	h=6	h=0	h=1	h=2	h=3	h=4	h=5	h=6
1	100	100	-83.4	-6.2	-29.0	-71.6	-86.3	-88.6	-89.8	-79.5	1.0	-0.9	-22.7	-49.8	-60.8	-64.2
	100	200	-85.7	1.5	-6.3	-45.8	-89.1	-92.7	-92.9	-80.3	4.1	1.2	-17.1	-51.3	-64.4	-67.2
	100	300	-85.6	0.5	-2.2	-20.1	-79.6	-93.6	-93.8	-81.7	3.1	0.9	-10.9	-41.9	-61.1	-65.2
	200	100	-86.5	-18.2	-56.4	-77.2	-80.8	-84.6	-86.1	-81.3	-2.4	-6.7	-49.2	-68.3	-75.8	-80.7
	200	200	-89.6	-1.3	-17.5	-80.3	-90.9	-90.8	-90.5	-84.5	3.4	2.2	-39.2	-64.6	-70.3	-76.1
	200	300	-91.9	1.9	-4.8	-56.1	-93.2	-92.9	-93.8	-85.0	3.4	0.4	-29.1	-61.3	-65.6	-68.3
	300	100	-86.5	-27.5	-61.2	-67.6	-72.4	-79.3	-82.7	-81.3	-5.3	-12.5	-58.2	-71.4	-79.5	-85.9
	300	200	-90.0	0.0	-27.6	-87.8	-89.6	-89.1	-89.5	-85.5	5.5	-1.9	-52.4	-66.7	-73.3	-81.3
	300	300	-92.8	0.5	-9.3	-78.7	-92.4	-93.4	-91.7	-85.2	3.6	-1.2	-38.7	-63.7	-68.1	-73.2
2	100	100	0.2	-37.0	-90.4	-90.5	-91.7	-90.6	-89.5	0.2	-22.8	-63.8	-60.1	-58.9	-59.5	-63.0
	100	200	0.0	-11.7	-83.8	-92.1	-92.7	-93.0	-93.1	-0.1	-12.6	-52.5	-65.2	-61.2	-61.2	-61.1
	100	300	0.0	-7.9	-65.1	-92.3	-93.4	-93.6	-93.1	-0.1	-9.0	-38.7	-68.1	-61.9	-60.6	-61.9
	200	100	-0.1	-72.1	-89.9	-86.8	-88.6	-86.5	-88.2	0.2	-36.1	-59.3	-60.2	-55.7	-55.6	-56.9
	200	200	0.0	-27.4	-92.6	-90.4	-91.2	-92.1	-91.3	0.0	-23.8	-58.1	-66.4	-61.1	-58.7	-58.7
	200	300	-0.1	-19.6	-91.0	-92.1	-92.3	-91.9	-92.5	0.2	-19.9	-53.8	-64.4	-67.8	-60.2	-59.9
	300	100	-0.2	-83.6	-85.4	-84.5	-84.0	-83.9	-85.1	0.1	-42.2	-54.7	-61.6	-53.3	-54.0	-53.7
	300	200	-0.1	-40.6	-90.3	-89.3	-88.9	-89.1	-89.6	0.0	-25.6	-58.0	-64.6	-64.4	-57.7	-57.4
	300	300	0.1	-26.8	-91.1	-90.4	-90.9	-91.7	-91.8	0.1	-24.7	-57.7	-61.7	-70.1	-60.5	-58.9

Note: This table reports the differences in coverage probabilities of the horizon-by-horizon GLP and the baseline GLP. Confidence intervals are computed using Theorem 2.

Table S33: Explaining Group Membership ($\hat{G} = 3$)

	(1)	(2)	(3)	(4)	(5)
2					
RGDP	-0.986				-1.667
	(-1.30)				(-1.33)
Employment		-0.366*			
		(-2.21)			
Debt			-1.971***	-2.007**	-2.436***
			(-4.41)	(-3.15)	(-3.30)
Elasticity				0.777	0.479
				(1.88)	(1.15)
3					
RGDP	-2.386**				-3.214*
	(-2.97)				(-2.46)
Employment		-0.425*			
		(-2.50)			
Debt			-1.539***	-1.623*	-2.441**
			(-3.47)	(-2.56)	(-3.25)
Elasticity				0.796	0.414
				(1.93)	(0.99)
N	382	382	381	253	253

t statistics in parentheses

* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

The coefficients for constant terms are omitted.

Table S34: Explaining Group Membership ($\hat{G} = 4$)

	(1)	(2)	(3)	(4)	(5)
2					
RGDP	-0.687				-1.720
	(-0.85)				(-1.32)
Employment		-0.297			
		(-1.67)			
Debt			-1.715***	-1.890**	-2.344**
			(-3.63)	(-2.86)	(-3.06)
Elasticity				0.445	0.154
				(1.03)	(0.35)
3					
RGDP	-1.488				-2.113
	(-1.91)				(-1.62)
Employment		-0.427*			
		(-2.52)			
Debt			-1.941***	-1.784**	-2.342**
			(-4.31)	(-2.75)	(-3.10)
Elasticity				0.998*	0.664
				(2.36)	(1.57)
4					
RGDP	-2.579**				-3.784**
	(-3.06)				(-2.75)
Employment		-0.422*			
		(-2.36)			
Debt			-1.549***	-1.787**	-2.772***
			(-3.37)	(-2.68)	(-3.50)
Elasticity				0.829	0.410
				(1.95)	(0.95)
N	382	382	381	253	253

t statistics in parentheses

* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

The coefficients for constant terms are omitted.

Table S35: MSA Group Economic Profile (FE')

			GDP		Income		Population		Employment		Regulation		Elasticity		Debt (Low)		Debt (High)	
$\hat{G}$	g	count	mean	std	mean	std	mean	std	mean	std	mean	std	mean	std	mean	std	mean	std
2	1	81	52091	21252	48206	13569	932	1864	581	1212	0.230	0.996	2.092	1.010	1.523	0.491	1.682	0.478
	2	301	45678	11313	45301	8752	676	1564	418	1019	-0.215	0.749	2.725	1.516	1.403	0.480	1.562	0.447
3	1	31	54014	20465	48827	13319	1327	2479	800	1592	0.429	0.929	1.747	0.985	1.720	0.458	1.826	0.448
	2	181	48592	15244	46558	9694	688	1816	434	1186	-0.176	0.865	2.750	1.634	1.337	0.498	1.503	0.435
	3	170	44112	10702	44704	9554	666	1158	409	759	-0.165	0.735	2.584	1.263	1.472	0.449	1.637	0.459
4	1	12	56275	10566	51268	8833	2029	3717	1292	2428	0.466	0.525	1.232	0.893	1.899	0.431	2.064	0.470
	2	157	49273	17550	47505	12012	774	1943	480	1260	0.022	0.918	2.374	1.169	1.444	0.501	1.585	0.449
	3	110	46300	10642	45030	7990	648	1092	409	742	-0.252	0.762	2.813	1.705	1.398	0.433	1.582	0.453
	4	103	43343	10960	43820	8070	599	1095	359	688	-0.236	0.724	2.780	1.452	1.382	0.492	1.548	0.445

Note: GDP and income are per capita measured in dollars. Population and employment are measured in thousands of person and thousands of jobs respectively. Regulation stands for the Wharton residential land use regulatory index, and Elasticity is the house supply elasticity. Debt (Low) and Debt (High) correspond to the lower bound and upper bound of the debt to income ratio.

Table S36: MSA Group Economic Profile (Horizon-by-Horizon)

			GDP		Income		Population		Employment		Regulation		Elasticity		Debt (Low)		Debt (High)	
H	g	count	mean	std	mean	std	mean	std	mean	std	mean	std	mean	std	mean	std	mean	std
1	1	88	48936	15843	47241	10238	1040	2613	644	1703	0.272	0.760	1.971	0.814	1.562	0.501	1.724	0.487
	2	168	48399	15056	46585	10848	611	1176	381	751	-0.291	0.831	2.881	1.685	1.326	0.438	1.498	0.413
	3	126	43898	11168	44101	8405	673	1201	415	799	-0.177	0.784	2.651	1.344	1.473	0.506	1.617	0.465
6	1	30	52836	21679	48810	13915	1449	2591	888	1692	0.460	0.579	1.619	0.884	1.784	0.448	1.916	0.457
	2	244	47653	13691	45884	9203	768	1743	483	1136	-0.166	0.829	2.666	1.413	1.347	0.438	1.509	0.406
	3	108	44039	12190	45188	10476	445	743	264	474	-0.181	0.824	2.690	1.558	1.512	0.537	1.686	0.508
12	1	42	53631	18970	49158	12791	1202	2251	743	1470	0.443	0.838	1.741	0.886	1.683	0.494	1.850	0.491
	2	240	47948	13953	46127	9992	733	1704	461	1112	-0.216	0.807	2.687	1.430	1.349	0.428	1.519	0.407
	3	100	42085	10726	44052	8335	525	1012	311	643	-0.122	0.779	2.717	1.576	1.510	0.556	1.652	0.510
18	1	20	53007	24374	48565	14018	1592	2968	948	1912	0.260	0.631	1.836	1.043	1.777	0.485	1.882	0.482
	2	90	49726	14294	47234	10539	955	2312	605	1512	0.171	0.833	1.963	0.814	1.449	0.469	1.611	0.444
	3	272	45710	13005	45286	9455	592	1135	366	740	-0.248	0.808	2.859	1.550	1.396	0.481	1.560	0.452
24	1	46	51229	19036	47611	12017	1173	2227	716	1444	0.453	0.821	1.825	0.932	1.640	0.560	1.757	0.500
	2	185	47221	11778	46084	9129	752	1840	472	1205	-0.215	0.810	2.656	1.489	1.405	0.459	1.584	0.462
	3	151	45537	15091	45195	10392	568	1024	349	658	-0.178	0.780	2.758	1.454	1.392	0.477	1.544	0.425

Figure S1: Information Criterion

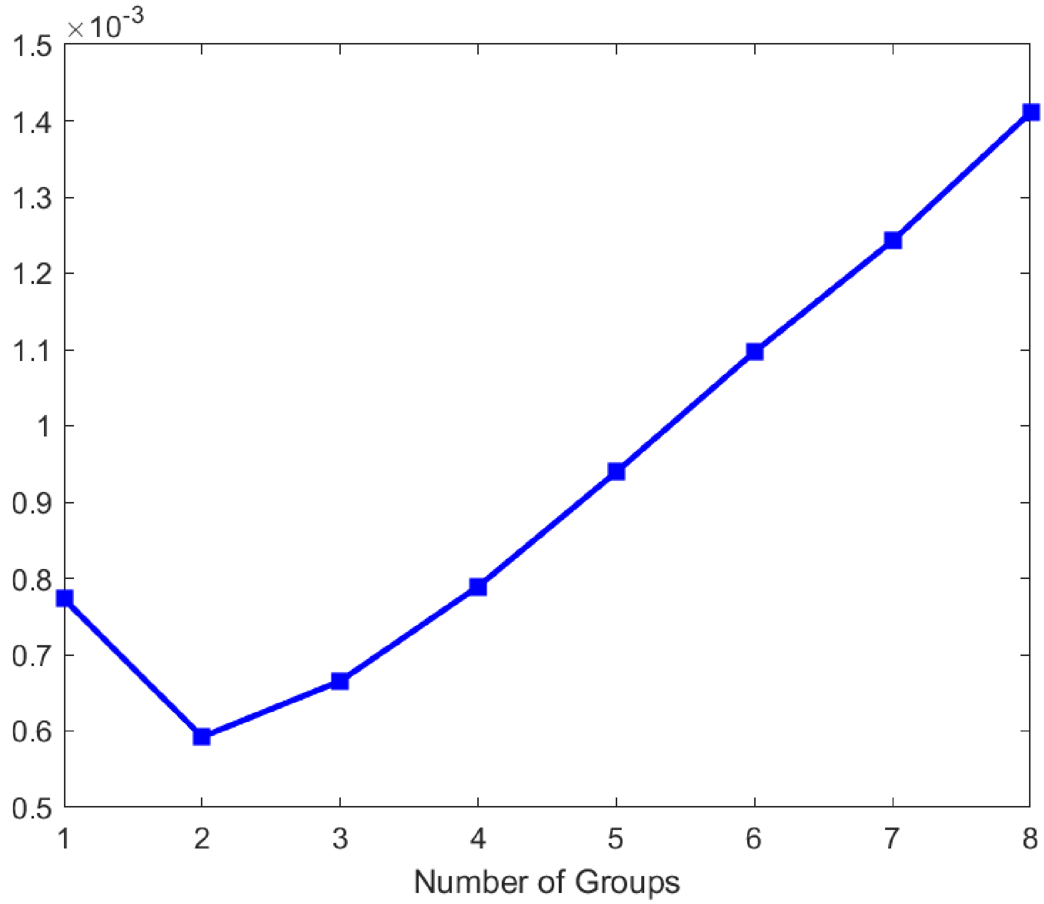


Figure S2: Group Impulse Responses ($\hat{G} = 2$)

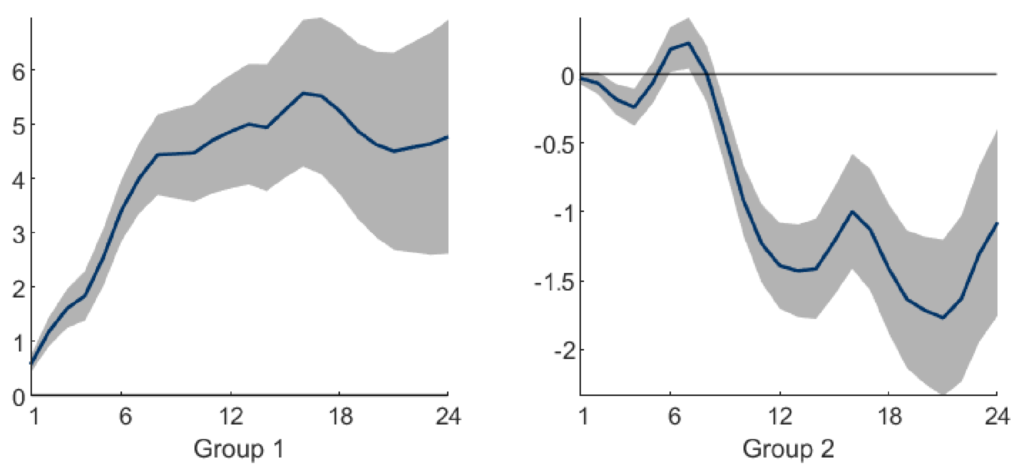


Figure S3: Group Impulse Responses ($\hat{G} = 4$)

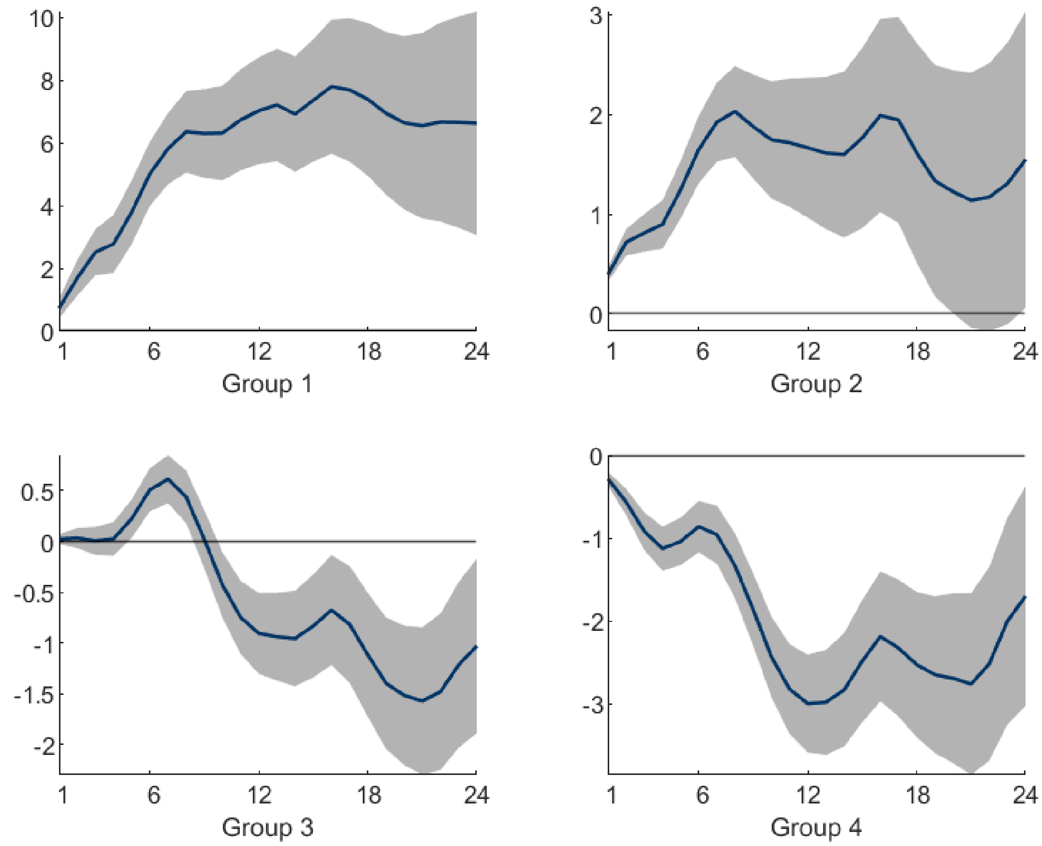


Figure S4: Geographical Distribution ($\hat{G} = 4$, FE))

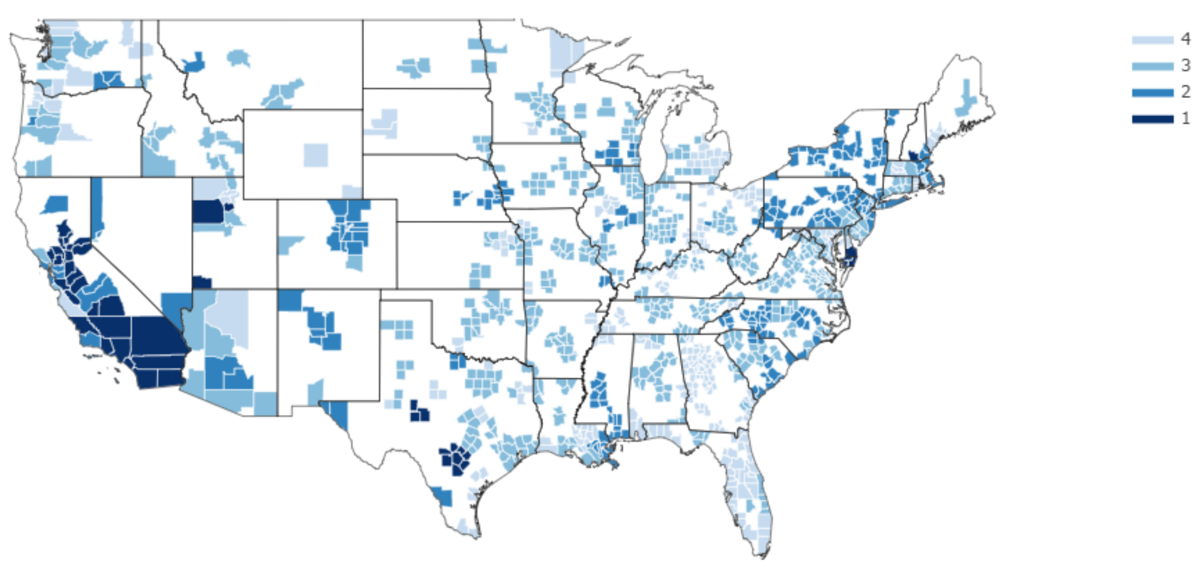


Figure S5: Information Criterion (FE')

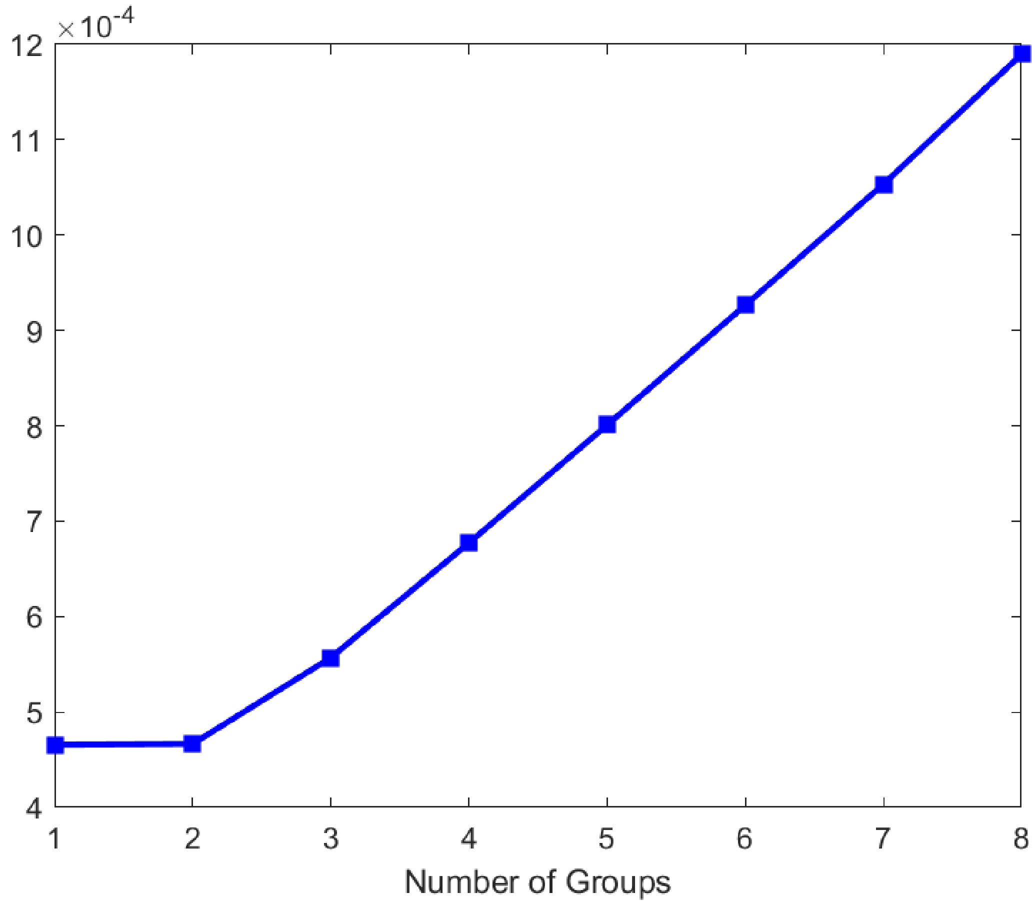


Figure S6: Impulse Responses ($\hat{G} = 2$, FE')

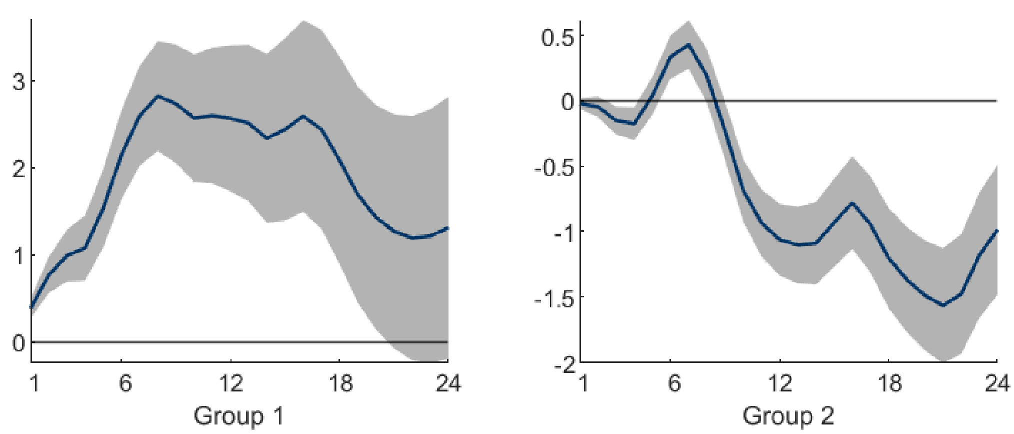


Figure S7: Impulse Responses ($\hat{G} = 3$, FE')

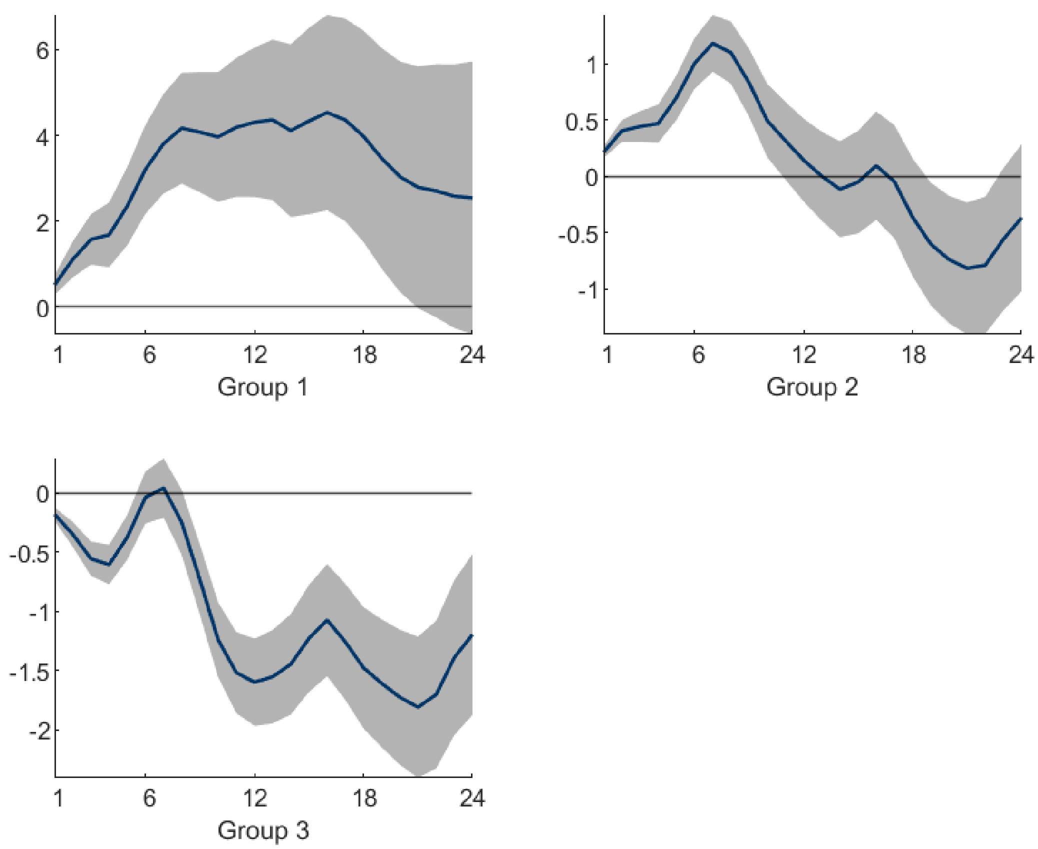


Figure S8: Impulse Responses ($\hat{G} = 4$, FE')

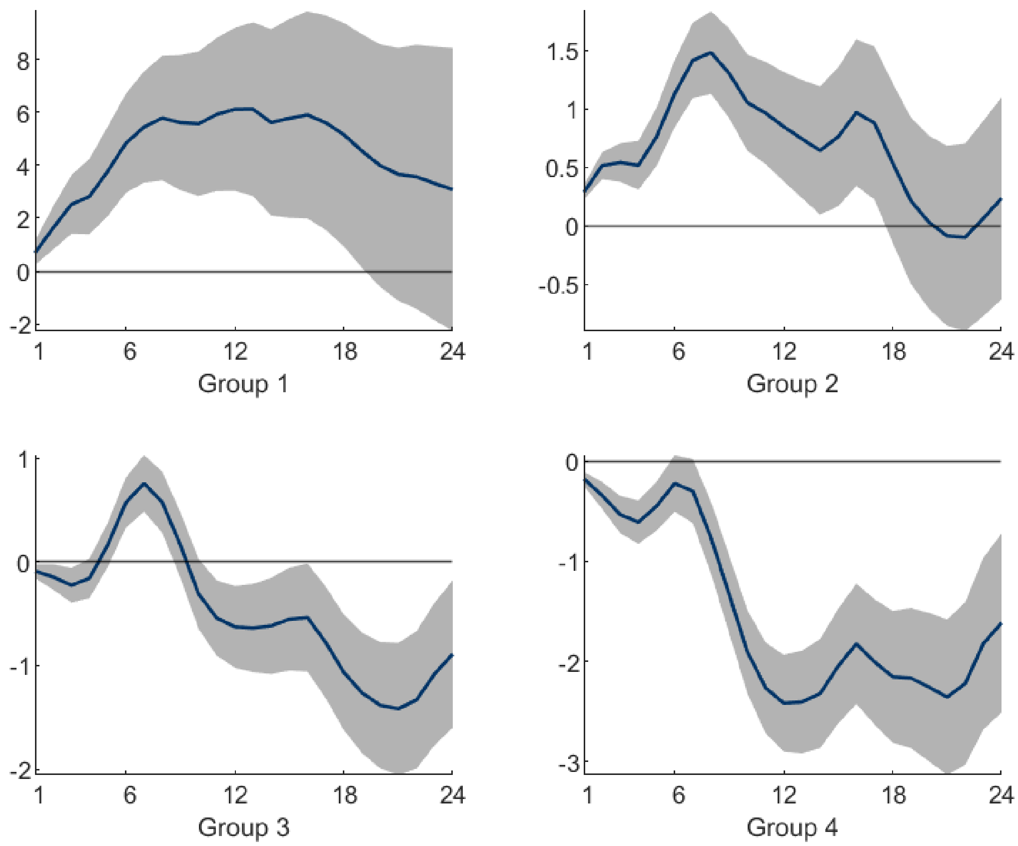


Figure S9: Geographical Distribution ($\hat{G} = 3$, FE')

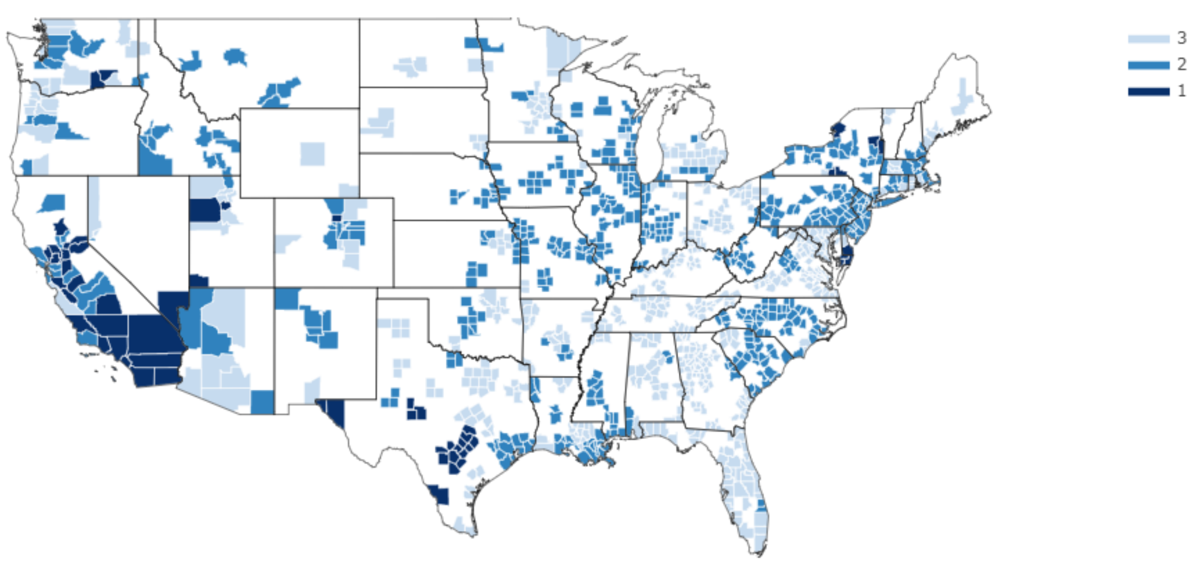


Figure S10: Geographical Distribution ($\hat{G} = 4$, FE')

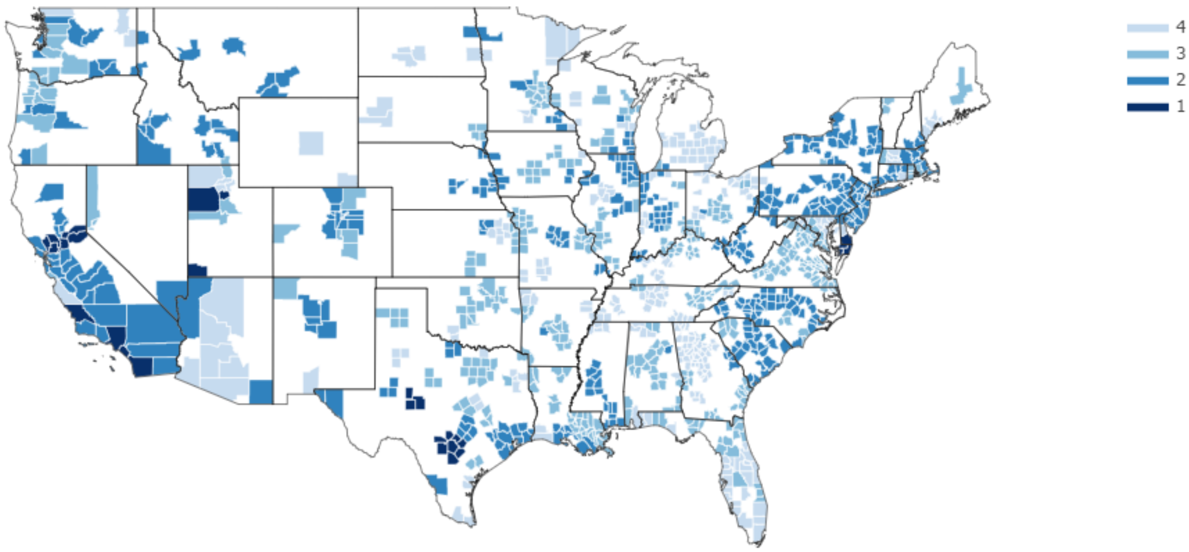
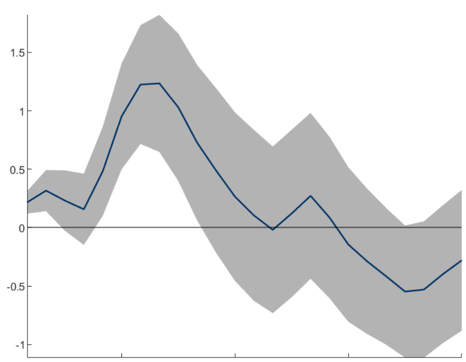
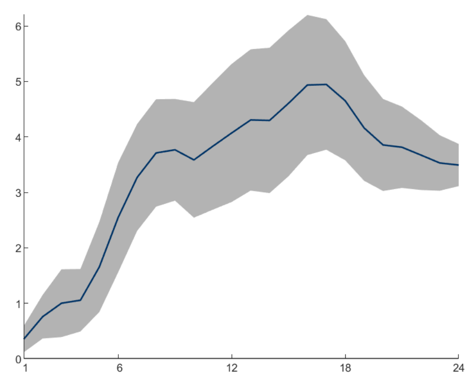


Figure S11: Alternative Criterion ($\hat{G} = 3$, FE')



(a) The Richest 10% MSAs



(b) The Poorest 10% MSAs in Group 1

Figure S12: Impulse Responses (Horizon-by-Horizon)

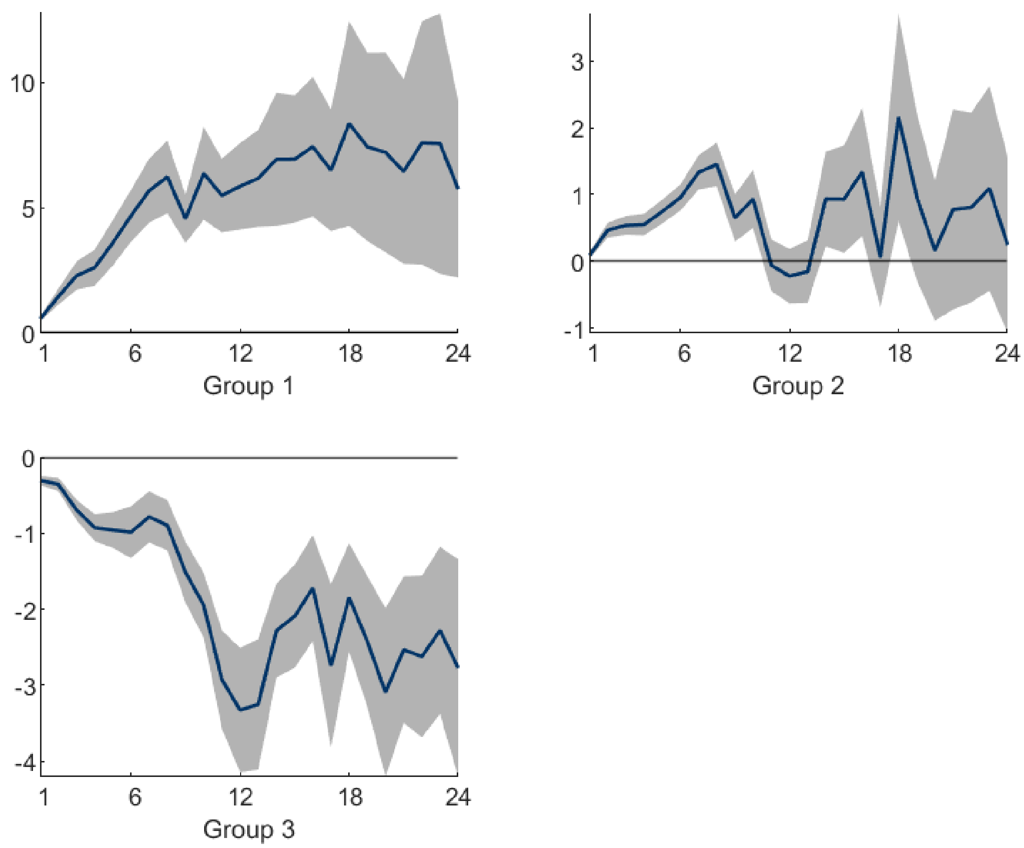
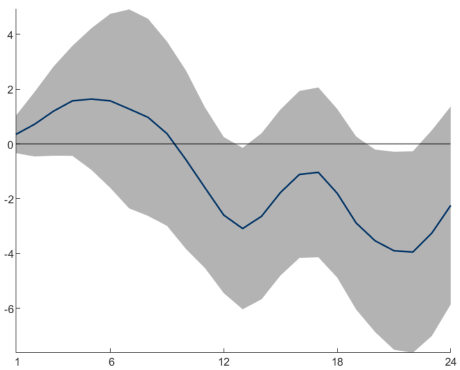
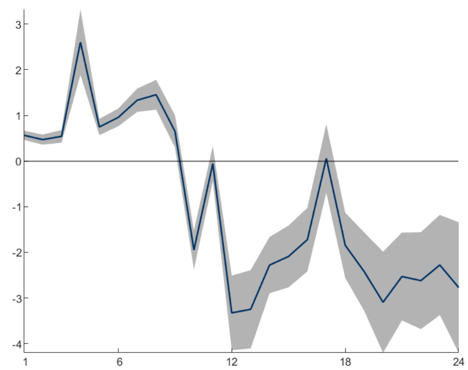


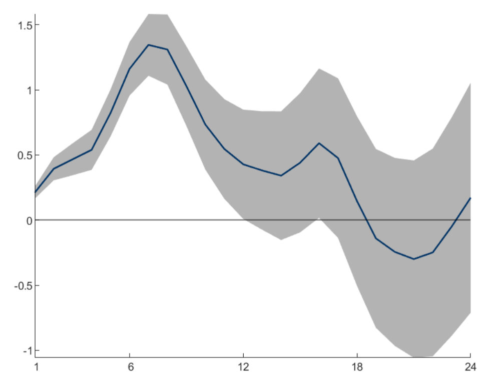
Figure S13: House Prices Impulse Responses (Charleston, WV)



(a) Individual LP-IV



(b) Horizon-by-Horizon GLP



(c) Baseline GLP

Note: All IRs are measured in percentage and correspond to a one percentage increase in the FFR. The shaded areas indicate 95% confidence bands.

Figure S14: Impulse Responses (HBH, $h = 1$)

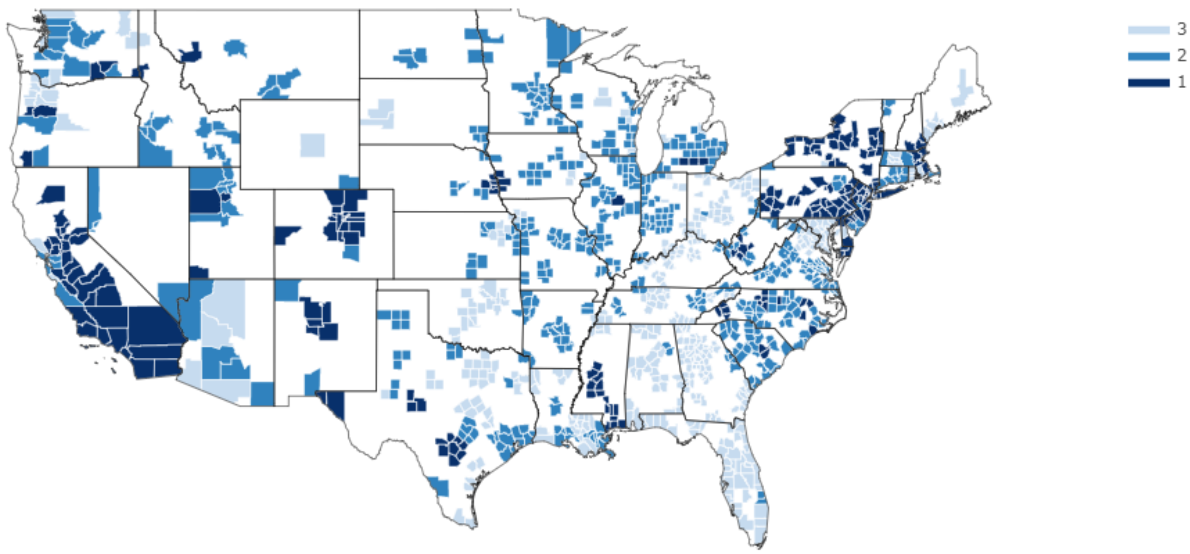


Figure S15: Impulse Responses (HBH, $h = 6$)

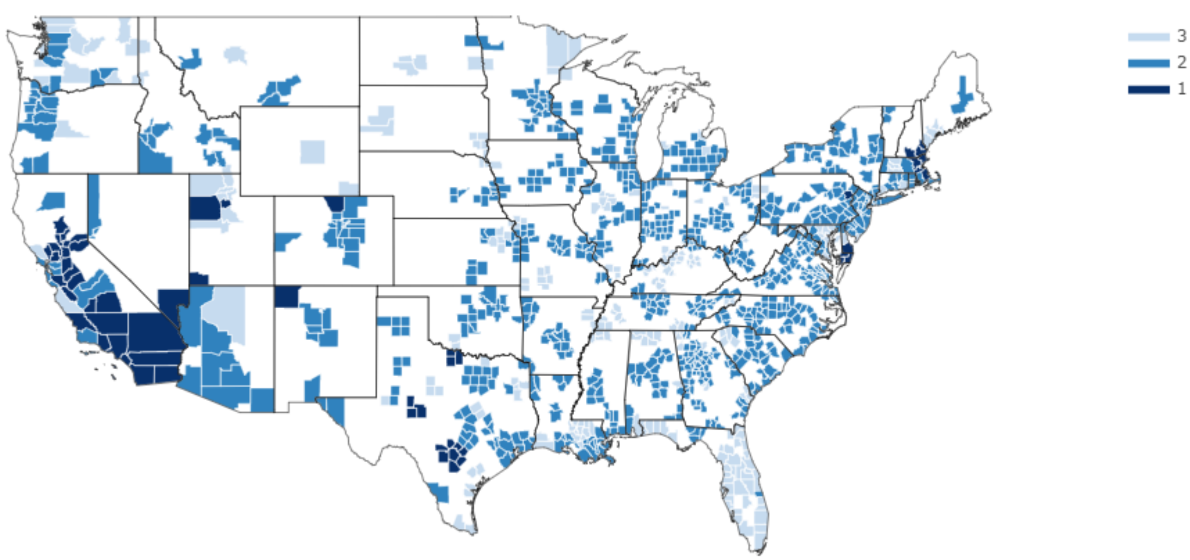


Figure S16: Impulse Responses (HBH, $h = 12$)

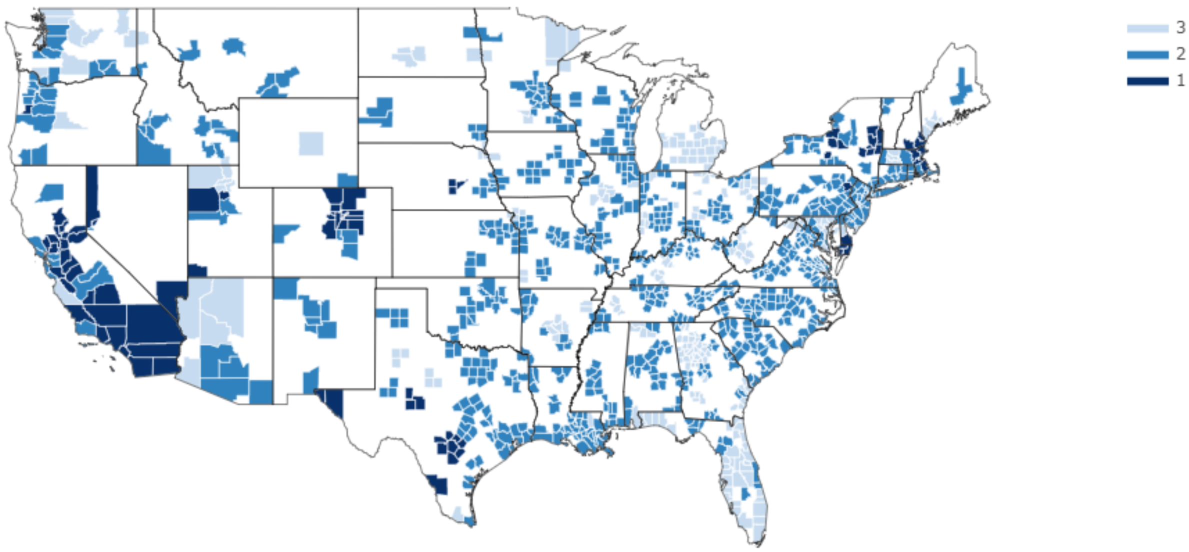


Figure S17: Impulse Responses (HBH, $h = 18$)

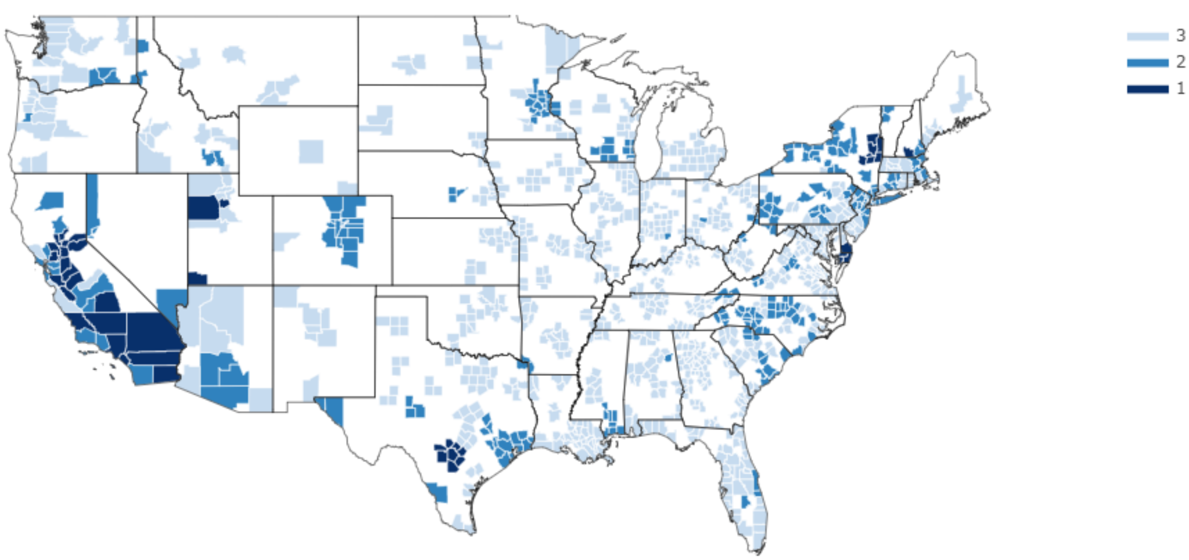


Figure S18: Impulse Responses (HBH, $h = 24$)