Supplemental Appendix for

# Testing for Unobserved Heterogeneity via *k-means* Clustering

by Andrew J. Patton and Brian M. Weller

15 July, 2019

## S.A.1: Extensions of Theorem 1

To streamline exposition, in this appendix we focus on the case that $d \equiv \dim(Y_{it}) = 1$ and $G = 2$. All of the results below hold for any finite value of $d$ and $G$.

We present a simplified version of Theorem 1 for $d = 1,\ G = 2$. In this instance, it is more natural to consider a $t$-test of the difference in cluster means.

**Corollary 1** *Assume $G = 2$ and $\dim(Y_{it}) = 1$. Let $\hat{\boldsymbol{\gamma}}_{NR}$ be the estimated group assignments based on sample $\mathcal{R}$, and let $\tilde{\boldsymbol{\mu}}_{NP}(\hat{\boldsymbol{\gamma}}_{NR})$ be the estimated group means from sample $\mathcal{P}$ using group assignments $\hat{\gamma}_{NR}$. Define the t-statistic for the differences in the estimated means as*

$$tstat_{NPR} = \frac{\sqrt{NP}\left(\tilde{\mu}_{1,NP}(\hat{\boldsymbol{\gamma}}_{NR}) - \tilde{\mu}_{2,NP}(\hat{\boldsymbol{\gamma}}_{NR})\right)}{\hat{\omega}_{NPR}} \tag{2}$$

$$\text{where} \quad \hat{\omega}^2_{NPR} \equiv \frac{1}{NP}\sum_{i=1}^{N}\sum_{t\in\mathcal{P}}\left(Y_{it} - \bar{Y}_{iP}\right)^2\left(\hat{\pi}^{-2}_{1,NR}\mathbf{1}\left\{\hat{\gamma}_{i,NR} = 1\right\} + \hat{\pi}^{-2}_{2,NR}\mathbf{1}\left\{\hat{\gamma}_{i,NR} = 2\right\}\right) \tag{3}$$

$$\bar{Y}_{iP} \equiv \frac{1}{P}\sum_{t\in\mathcal{P}} Y_{it} \tag{4}$$

$$\hat{\pi}_{g,NR} \equiv \frac{1}{N}\sum_{i=1}^{N}\mathbf{1}\left\{\hat{\gamma}_{i,NR} = g\right\}, \quad \text{for } g = 1,2 \tag{5}$$

*(a) Under Assumptions 1 and 2,*

$$tstat_{NPR} \xrightarrow{d} N(0,1), \quad \text{as } N,P,R \to \infty \tag{6}$$

*(b) Under Assumptions 1 and 2′,*

$$|tstat_{NPR}| \xrightarrow{p} \infty, \quad \text{as } N,P,R \to \infty \tag{7}$$

First, we consider allowing for general time series dependence up to some lag $M$. To do so, we define $\mathcal{G}_t$ as the information set $\sigma\left(\{Y_{is}\}_{i=1}^{N}, s \leq t\right)$, and modify Assumption 1 to:

**Assumption 1$''$:** (a) The data come from $Y_{it} = m_i + \varepsilon_{it}$, for $i = 1, ..., N$, and $t = 1, .., T$, where $m_i \in [\underline{m}, \bar{m}] \subset \mathbb{R}$ and $V[\varepsilon_{it}] \equiv \sigma_i^2 \in [\underline{\sigma}^2, \bar{\sigma}^2] \subset \mathbb{R}_+ \ \forall\ i$, $E[\varepsilon_{it}] = 0$ and $E\left[|\varepsilon_{it}|^{4+\delta}\right] < \infty$ $\forall\ i$ for some $\delta > 0$, (b) $\varepsilon_{it} \perp\!\!\!\perp \varepsilon_{js}\ \forall\ t, s$, for $i \neq j$ (c) $\varepsilon_{it} \perp\!\!\!\perp X$ for all $X \in \mathcal{G}_{t-M}$, for $\forall\ i, t$ and (d) $N, P, R \to \infty$.

Assumption 1$''$(a) allows for cross-sectional heteroskedasticity, and heterogeneity more generally, in the distribution of residuals, subject to them being mean zero and having finite fourth moments. Assumption 1$''$(b) imposes cross-sectional independence, and 1$''$(c) allows for general time series dependence up to lag $M$, but imposes independence beyond $M$ lags. The main change required when allowing for time series dependence is that the formation of subsamples now requires some structure. We suggest using $\mathcal{R} = \{1, 2, ..., R - M\}$ and $\mathcal{P} = \{R + 1, ..., R + P \equiv T\}$.

**Theorem 6** *Let $\hat{\boldsymbol{\gamma}}_{NR}$ be the estimated group assignments based on sample $\mathcal{R}$, and let $\tilde{\boldsymbol{\mu}}_{NP}(\hat{\boldsymbol{\gamma}}_{NR})$ be the estimated group means from sample $\mathcal{P}$ using group assignments $\hat{\boldsymbol{\gamma}}_{NR}$. Define the t-statistic for the differences in the estimated means as*

$$\begin{aligned} tstat_{NPR} &= \frac{\sqrt{NP}\left(\tilde{\mu}_{1,NP}(\hat{\boldsymbol{\gamma}}_{NR}) - \tilde{\mu}_{2,NP}(\hat{\boldsymbol{\gamma}}_{NR})\right)}{\hat{\omega}_{NPR}} \\ \text{where} \quad \hat{\omega}_{NPR}^2 &\equiv \sum_{i=1}^{N} \boldsymbol{\iota}_P' \hat{\boldsymbol{\varepsilon}}_i \hat{\boldsymbol{\varepsilon}}_i' \boldsymbol{\iota}_P \left(\hat{\pi}_{1,NR}^{-2} \mathbf{1}\left\{\hat{\gamma}_{i,NR} = 1\right\} - \hat{\pi}_{2,NR}^{-2} \mathbf{1}\left\{\hat{\gamma}_{i,NR} = 2\right\}\right) \\ \hat{\boldsymbol{\varepsilon}}_{iP} &= \mathbf{Y}_{iP} - \boldsymbol{\iota}_P \bar{Y}_{iP} \\ \text{and} \quad \hat{\pi}_{g,NR} &\equiv \frac{1}{N} \sum_{i=1}^{N} \mathbf{1}\left\{\hat{\gamma}_{i,NR} = g\right\}, \quad \text{for } g = 1, 2 \end{aligned}$$

*and $\mathbf{Y}_{iP} \equiv [Y_{i1}, ..., Y_{iP}]'$ and $\boldsymbol{\iota}_P$ is a $P \times 1$ vector of ones.*

*(a) Under Assumptions 1$''$ and 2,*

$$tstat_{NPR} \xrightarrow{d} N(0, 1), \quad \text{as } N, P \to \infty$$

*(b) Under Assumptions 1$''$ and 2$'$,*

$$|tstat_{NPR}| \xrightarrow{p} \infty, \quad \text{as } N, P, R \to \infty$$

**Proof of Theorem 6.** (a) We first find the limiting distribution of $\sqrt{NP}\left(\tilde{\mu}_{1,NP}\left(\hat{\gamma}_{NR}\right)-\tilde{\mu}_{2,NP}\left(\hat{\gamma}_{NR}\right)\right)$ conditional on $\mathcal{F}_R$. Note that

$$\begin{aligned}\tilde{\mu}_{g,NP}\left(\hat{\gamma}_{NR}\right) &= \frac{1}{\hat{N}_{g,NR}}\sum_{i=1}^{N}\left(\mathbf{1}\left\{\hat{\gamma}_{i,NR}=g\right\}\frac{1}{P}\sum_{t=R+1}^{T}Y_{i,t}\right)\\ &\equiv \frac{1}{NP}\sum_{i=1}^{N}\sum_{t=R+1}^{T}Y_{i,t}\hat{\pi}_{g,NR}^{-1}\mathbf{1}\left\{\hat{\gamma}_{i,NR}=g\right\}\end{aligned}$$

Then

$$\sqrt{NP}\left(\tilde{\mu}_{1,NP}\left(\hat{\gamma}_{NR}\right)-\tilde{\mu}_{2,NP}\left(\hat{\gamma}_{NR}\right)\right)=\frac{1}{\sqrt{NP}}\sum_{i=1}^{N}\sum_{t=R+1}^{T}\hat{Z}_{i,NR}\varepsilon_{it}$$

where $\hat{Z}_{i,NR}\equiv\hat{\pi}_{1,NR}^{-1}\mathbf{1}\left\{\hat{\gamma}_{i,NR}=1\right\}-\hat{\pi}_{2,NR}^{-1}\mathbf{1}\left\{\hat{\gamma}_{i,NR}=2\right\}$. We now verify that we can invoke a CLT for $\sqrt{NP}\frac{1}{NP}\sum_{i=1}^{N}\sum_{t=R+1}^{P}\xi_{it,NR}$, where $\xi_{it,NR}\equiv\hat{Z}_{i,NR}\varepsilon_{it}$. Note that conditional on $\mathcal{F}_R$, the sequence $\left\{\xi_{it,NR}\right\}$ is heterogeneously distributed, and $M$-dependent by Assumption $1''$(c) which immediately implies strong mixing. Also note that conditional on $\mathcal{F}_R$, $\xi_{it,NR}$ is independent of $\xi_{jt,NR}$ $\forall\ i\neq j$. Then note that

$$E\left[\xi_{it,NR}|\mathcal{F}_R\right]=\hat{Z}_{i,NR}E\left[\varepsilon_{it}|\mathcal{F}_R\right]=\hat{Z}_{i,NR}E\left[E\left[\varepsilon_{it}|\mathcal{G}_{t-M}\right]|\mathcal{F}_R\right]=0,\ \text{ for } t\geq R+1$$

Next, let

$$\boldsymbol{\xi}_{i,NPR}'\equiv\left[\xi_{i,R+1,NR},...,\xi_{i,R+P,NR}\right]=\hat{Z}_{i,NR}\left[\varepsilon_{i,R+1},...,\varepsilon_{i,R+P}\right]'\equiv\hat{Z}_{i,NR}\boldsymbol{\varepsilon}_{iP}'$$

and note that

$$E\left[\boldsymbol{\xi}_{i,NPR}\boldsymbol{\xi}_{i,NPR}'\middle|\mathcal{F}_R\right]=\hat{Z}_{i,NR}^2E\left[\boldsymbol{\varepsilon}_{iP}\boldsymbol{\varepsilon}_{iP}'\right]\equiv\hat{Z}_{i,NR}^2\Omega_i$$

Note that by Assumption $1''$(a) and (c), $\Omega_i$ is a Toeplitz matrix, with $\sigma_i^2$ on the main diagonal, $\psi_{i,1}\equiv Cov\left[\varepsilon_{i,t},\varepsilon_{i,t+1}\right]$ on the secondary diagonal, etc. out to $\psi_{i,M}\equiv Cov\left[\varepsilon_{i,t},\varepsilon_{i,t+M}\right]$ on the $(M+1)^{th}$ diagonal, and with zeros elsewhere. This structure simplifies the estimation of $\Omega_i$.

Finally, define

$$\begin{aligned}\bar{\omega}_{NR}^2 &\equiv \lim_{N,P\to\infty}\frac{1}{NP}\sum_{i=1}^{N}\hat{Z}_{i,NR}^2\boldsymbol{\iota}_P'\Omega_i\boldsymbol{\iota}_P\\ &= \lim_{N,P\to\infty}\frac{1}{N}\sum_{i=1}^{N}\hat{Z}_{i,NR}^2\sigma_i^2+\lim_{N,P\to\infty}\frac{2}{N}\sum_{i=1}^{N}\hat{Z}_{i,NR}^2\left(\sum_{k=1}^{M}\left(1-k/P\right)\psi_{i,k}\right)\end{aligned}$$

The general estimator of the asymptotic covariance in Hansen (2007) is given below, which we then simplify based on our $M$-dependence assumption.

$$\begin{aligned} \hat{\omega}_{NPR}^2 &= \frac{1}{N}\sum_{i=1}^{N}\hat{Z}_{i,NR}^2\hat{\psi}_{i,0,P}+\frac{2}{N}\sum_{i=1}^{N}\hat{Z}_{i,NR}^2\left(\sum_{k=1}^{M}(1-k/P)\,\hat{\psi}_{i,k,P}\right) \\ \text{where }\ \hat{\psi}_{i,k,P} &\equiv \frac{1}{P}\sum_{t=R+1}^{T-k}\left(Y_{it}-\bar{Y}_{iP}\right)\left(Y_{i,t+k}-\bar{Y}_{iP}\right),\ \ k=0,1,...,M \end{aligned}$$

Our Assumption $1''$ is sufficient for Assumptions 1, 2 and 3(b) of Hansen (2007, Theorem 3), and thus we have, conditional on $\mathcal{F}_R$,

$$\frac{\sqrt{NP}\frac{1}{NP}\sum_{i=1}^{N}\sum_{t=R+1}^{T}\hat{Z}_{i,NR}\varepsilon_{it}}{\bar{\omega}_{NR}}\xrightarrow{d}N(0,1)\ \text{ and }\ \hat{\omega}_{NPR}^2\xrightarrow{p}\bar{\omega}_{NR}^2 \text{ as } N,P,R\to\infty$$

This implies that the $t$-statistic obeys

$$\frac{\sqrt{NP}\frac{1}{NP}\sum_{i=1}^{N}\sum_{t=R+1}^{T}\hat{Z}_{i,NR}\varepsilon_{it}}{\hat{\omega}_{NPR}}\xrightarrow{d}N(0,1)\,,\ \text{as } N,P,R\to\infty$$

As the limiting distribution of the $t$-statistic does not depend on $\mathcal{F}_R$, its unconditional distribution is also $N(0,1)$, completing the proof.

(b) As in Theorem 1(b), note that $\tilde{\mu}_{NP}(\hat{\gamma}_{NR})-\mu^* = (\hat{\mu}_{NR}-\mu^*)+(\tilde{\mu}_{NP}(\hat{\gamma}_{NR})-\hat{\mu}_{NR})$. Our Assumption $1''$ is sufficient for Assumption 1 of Bonhomme and Manresa (2015), and their Theorem 1 implies that the first term on the RHS is $o_p(1)$, as $N,R\to\infty$. The second term is:

$$\begin{aligned} \tilde{\mu}_{g,NP}(\hat{\gamma}_{NR})-\hat{\mu}_{g,NR} &= \frac{1}{N}\sum_{i=1}^{N}\hat{\pi}_{g,NR}^{-1}\mathbf{1}\left\{\hat{\gamma}_{i,NR}=g\right\}\left(\frac{1}{P}\sum_{t=R+1}^{T}Y_{i,t}-\frac{1}{R-M}\sum_{t=1}^{R-M}Y_{i,t}\right) \\ &= \frac{1}{N}\sum_{i=1}^{N}\hat{\pi}_{g,NR}^{-1}\mathbf{1}\left\{\hat{\gamma}_{i,NR}=g\right\}\left(\frac{1}{P}\sum_{t=R+1}^{T}\varepsilon_{i,t}-\frac{1}{R-M}\sum_{t=1}^{R-M}\varepsilon_{i,t}\right) \\ &\leq \underline{\pi}^{-1}\frac{1}{N}\sum_{i=1}^{N}\left(\frac{1}{P}\sum_{t=R+1}^{T}\varepsilon_{i,t}-\frac{1}{R-M}\sum_{t=1}^{R-M}\varepsilon_{i,t}\right) \\ &= \underline{\pi}^{-1}\left(\frac{1}{NP}\sum_{i=1}^{N}\sum_{t=R+1}^{T}\varepsilon_{i,t}-\frac{1}{N(R-M)}\sum_{i=1}^{N}\sum_{t=1}^{R-M}\varepsilon_{i,t}\right) \\ &= o_p(1)\,,\ \text{as } N,P,R\to\infty \end{aligned}$$

since our Assumption $1''$ is sufficient for Assumptions 1, 2 and 3(b) of Hansen (2007), which implies that $\frac{1}{N(R-M)}\sum_{i=1}^{N}\sum_{t=1}^{R-M}\varepsilon_{i,t}=o_p(1)$ and $\frac{1}{NP}\sum_{i=1}^{N}\sum_{t=R+1}^{T}\varepsilon_{i,t}=o_p(1)$. This holds for $g=1,2,$

and thus $\tilde{\mu}_{NP}(\hat{\gamma}_{NR}) \xrightarrow{p} \mu^*$, as $N, P, R \to \infty$. This implies that $\tilde{\mu}_{1,NP}(\hat{\gamma}_{NR}) - \tilde{\mu}_{2,NP}(\hat{\gamma}_{NR}) \xrightarrow{p} \mu_1^* - \mu_2^* \neq 0$ by Assumption $2'$(b). Thus $|tstat| \xrightarrow{p} \infty$, as $N, P, R \to \infty$. ■

We next consider allowing for general time series and cross-sectional dependence, by adapting Assumption 1 of Bonhomme and Manresa (2015) to our application. Consider the following assumption. Let $K$ denote some finite constant.

**Assumption $1'''$:** (a) The data come from $Y_{it} = m_i + \varepsilon_{it}$, for $i = 1, ..., N$, and $t = 1, .., T$, where $m_i \in [\underline{m}, \bar{m}] \subset \mathbb{R}$ and $V[\varepsilon_{it}] \equiv \sigma_i^2 \in [\underline{\sigma}^2, \bar{\sigma}^2] \subset \mathbb{R}_+ \ \forall\ i$, and $E[\eta_{it}^4] \leq \bar{\kappa} < \infty \ \forall\ i$

(b) $\left| \frac{1}{NT} \sum_{i=1}^N \sum_{t=1}^T \sum_{s=1}^T E[\varepsilon_{it}\varepsilon_{is}] \right| \leq K < \infty$

(c) $\left| \frac{1}{N^2T} \sum_{i=1}^N \sum_{j=1}^N \sum_{t=1}^T \sum_{s=1}^T Cov[\varepsilon_{it}\varepsilon_{jt}, \varepsilon_{is}\varepsilon_{js}] \right| \leq K < \infty$

(d) $\frac{1}{N} \sum_{i=1}^N \sum_{j=1}^N \left| \frac{1}{T} \sum_{t=1}^T E[\varepsilon_{it}\varepsilon_{jt}] \right| \leq K < \infty$

(e) $\sqrt{NT} \frac{1}{NT} \sum_{i=1}^N \sum_{t=1}^T \varepsilon_{it} \xrightarrow{d} N(0, \bar{\upsilon}^2)$ for some $\bar{\upsilon}^2 > 0$, and there exists an estimator $\hat{\upsilon}_{NT}^2$ that is robust to cross-sectional heteroskedasticity in $\{\varepsilon_{it}\}$ and is consistent for $\bar{\upsilon}^2$, as $N, T \to \infty$.

(f) $N, P, R \to \infty$.

Assumption $1'''$(a) allows for cross-sectional heteroskedasticity, and heterogeneity more generally, in the distribution of residuals, subject to them being mean zero and having finite fourth moments. Assumptions $1'''$(b) and (c) imposes restrictions on the amount of time series dependence in the data, and $1'''$(d) limits the amount of cross-sectional dependence. Assumption $1'''$(e) is a high level assumption that a CLT can be invoked for the sample average of $\{\varepsilon_{it}\}$, and that a consistent estimator of the asymptotic variance is available. There are a variety of CLTs and LLNs that can be used in panel applications to satisfy this assumption, see Pesaran (2015) for a recent textbook treatment of this area. The requirement that this estimator is robust to cross-sectional heteroskedasticity is a mild requirement and is satisfied by many estimators in the literature.

**Theorem 7** *Let $\hat{\gamma}_{NR}$ be the estimated group assignments based on sample $\mathcal{R}$, and let $\tilde{\mu}_{NP}(\hat{\gamma}_{NR})$ be the estimated group means from sample $\mathcal{P}$ using group assignments $\hat{\gamma}_{NR}$. Define the t-statistic* *for the differences in the estimated means as*

$$tstat_{NPR} = \frac{\sqrt{NP}\left(\tilde{\mu}_{1,NP}\left(\hat{\boldsymbol{\gamma}}_{NR}\right) - \tilde{\mu}_{2,NP}\left(\hat{\boldsymbol{\gamma}}_{NR}\right)\right)}{\hat{\omega}_{NPR}}$$

*where* $\hat{\omega}^2_{NPR}$ *is an estimator of the asymptotic variance of*

$$\xi_{it,NR} \equiv \left(\hat{\pi}^{-1}_{1,NR}\mathbf{1}\left\{\hat{\gamma}_{i,NR} = 1\right\} - \hat{\pi}^{-1}_{2,NR}\mathbf{1}\left\{\hat{\gamma}_{i,NR} = 2\right\}\right)\varepsilon_{it}$$

*and takes the same functional form as the estimator* $\hat{\upsilon}^2_{NT}$ *in Assumption* $1'''$*(e).*

*(a) Under Assumptions* $1'''$ *and* 2*,*

$$tstat_{NPR} \xrightarrow{d} N(0,1), \quad as\ N, P \to \infty \tag{8}$$

*(b) Under Assumptions* $1'''$ *and* $2'$*,*

$$|tstat_{NPR}| \xrightarrow{p} \infty, \quad as\ N, P, R \to \infty \tag{9}$$

**Proof of Theorem 7.** (a) Note that

$$\begin{aligned}\sqrt{NP}\left(\tilde{\mu}_{1,NP}\left(\hat{\boldsymbol{\gamma}}_{NR}\right) - \tilde{\mu}_{2,NP}\left(\hat{\boldsymbol{\gamma}}_{NR}\right)\right) &= \frac{1}{\sqrt{NP}}\sum_{i=1}^{N}\sum_{t\in\mathcal{P}}\varepsilon_{it}\left(\hat{\pi}^{-1}_{1,NR}\mathbf{1}\left\{\hat{\gamma}_{i,NR} = 1\right\} - \hat{\pi}^{-1}_{2,NR}\mathbf{1}\left\{\hat{\gamma}_{i,NR} = 2\right\}\right) \\ &= \frac{1}{\sqrt{NP}}\sum_{i=1}^{N}\sum_{t\in\mathcal{P}}\hat{Z}_{i,NR}\varepsilon_{it}\end{aligned}$$

where $\hat{Z}_{i,NR} \equiv \hat{\pi}^{-1}_{1,NR}\mathbf{1}\left\{\hat{\gamma}_{i,NR} = 1\right\} - \hat{\pi}^{-1}_{2,NR}\mathbf{1}\left\{\hat{\gamma}_{i,NR} = 2\right\}$. By Assumption $1'''$(e) we know that $\frac{1}{\sqrt{NP}}\sum_{i=1}^{N}\sum_{t=1}^{P}\varepsilon_{it} \xrightarrow{d} N\left(0, \bar{\upsilon}^2\right)$, so

$$\bar{\upsilon}^2 = \lim_{N,P\to\infty} V\left[\frac{1}{\sqrt{NP}}\sum_{i=1}^{N}\sum_{t=1}^{P}\varepsilon_{it}\right] = \lim_{N,P\to\infty}\frac{1}{NP}\sum_{i=1}^{N}\sum_{j=1}^{N}\sum_{t=1}^{P}\sum_{s=1}^{P}E\left[\varepsilon_{it}\varepsilon_{js}\right]$$

Conditional on $\mathcal{F}_R$, the weights, $\hat{Z}_{i,NR}$, on $\varepsilon_{it}$ are known, and are bounded since $\underline{\pi} > 0$. Define the variable $\xi_{it,NR} \equiv \hat{Z}_{i,NR}\varepsilon_{it}$, and note that we have:

$$E\left[\xi_{it,NR}\middle|\mathcal{F}_R\right] = \hat{Z}_{i,NR}E\left[\varepsilon_{it}\right] = 0$$

Moreover,

$$\begin{aligned}
E\left[\left|\xi_{it,NR}\right|^q \middle| \mathcal{F}_R\right] &= \left|\hat{Z}_{i,NR}\right|^q E\left[\left|\varepsilon_{it}\right|^q\right], \text{ for } q \text{ s.t. } E\left[\left|\varepsilon_{it}\right|^q\right] < \infty \\
E\left[\xi_{it,NR}\xi_{is,NR} \middle| \mathcal{F}_R\right] &= \hat{Z}_{i,NR}^2 E\left[\varepsilon_{it}\varepsilon_{is}\right] \ \forall\ i,t,s \\
E\left[\xi_{it,NR}\xi_{jt,NR} \middle| \mathcal{F}_R\right] &= \hat{Z}_{i,NR}\hat{Z}_{j,NR} E\left[\varepsilon_{it}\varepsilon_{jt}\right] \ \forall\ i,j,t \\
Cov\left[\xi_{it,NR}\xi_{jt,NR}, \xi_{jt,NR}\xi_{js,NR} \middle| \mathcal{F}_R\right] &= \hat{Z}_{i,NR}^2 \hat{Z}_{j,NR}^2 Cov\left[\varepsilon_{it}\varepsilon_{jt}, \varepsilon_{is}\varepsilon_{js}\right] \ \forall\ i,j,t,s
\end{aligned}$$

and so the moment and memory properties of $\left\{\xi_{it,NR}\right\}$ are completely determined by the moment and memory properties of $\{\varepsilon_{it}\}$. Thus any CLT that applies to $\{\varepsilon_{it}\}$, and which allows for cross-sectional heteroskedasticity, will also apply to $\left\{\xi_{it,NR}\right\}$, conditional on $\mathcal{F}_R$. This implies that

$$\frac{1}{\sqrt{NP}} \sum_{i=1}^{N} \sum_{t=1}^{P} \xi_{it,NR} \xrightarrow{d} N\left(0, \bar{\omega}^2\right)$$

$$\text{where} \quad \bar{\omega}^2 = \lim_{N,P\to\infty} V\left[\frac{1}{\sqrt{NP}} \sum_{i=1}^{N} \sum_{t=1}^{P} \xi_{it,NR}\right]$$

By Assumption $1'''$(e) we know that there exists an estimator $\hat{\upsilon}_{NP}^2$ such that $\hat{\upsilon}_{NP}^2 \xrightarrow{p} \bar{\upsilon}^2$, as $N, P \to \infty$. As $\hat{Z}_{i,NR}$ is non-zero and finite, any estimator $\hat{\upsilon}_{NP}^2$ that is consistent for $\bar{\upsilon}^2$, and robust to cross-sectional heteroskedasticity, can also be applied to $\xi_{it,NR}$, yielding an estimator $\hat{\omega}_{NPR}^2$ that is consistent for $\bar{\omega}^2$. This implies that the $t$-statistic obeys:

$$tstat = \frac{\sqrt{NP}\left(\tilde{\mu}_{1,NP}\left(\hat{\gamma}_{NR}\right) - \tilde{\mu}_{2,NP}\left(\hat{\gamma}_{NR}\right)\right)}{\hat{\omega}_{NPR}} \xrightarrow{d} N\left(0,1\right) \quad \text{as } N, P, R \to \infty$$

As the limiting distribution of the $t$-statistic does not depend on $\mathcal{F}_R$, its unconditional distribution is also $N\left(0,1\right)$, completing the proof.

(b) Note that $\tilde{\mu}_{NP}\left(\hat{\gamma}_{NR}\right) - \mu^* = \left(\hat{\mu}_{NR} - \mu^*\right) + \left(\tilde{\mu}_{NP}\left(\hat{\gamma}_{NR}\right) - \hat{\mu}_{NR}\right)$. Our Assumption $1'''$ is sufficient for Assumption 1 of Bonhomme and Manresa (2015), and their Theorem 1 implies that the first term on the RHS is $o_p(1)$, as $N, R \to \infty$. The second term is:

$$\begin{aligned}
\tilde{\mu}_{g,NP}(\hat{\gamma}_{NR}) - \hat{\mu}_{g,NR} &= \frac{1}{N}\sum_{i=1}^{N} \hat{\pi}_{g,NR}^{-1} \mathbf{1}\{\hat{\gamma}_{i,NR} = g\} \left( \frac{1}{P}\sum_{t\in\mathcal{P}} Y_{i,t} - \frac{1}{R}\sum_{t\in\mathcal{R}} Y_{i,t} \right) \\
&= \frac{1}{N}\sum_{i=1}^{N} \hat{\pi}_{g,NR}^{-1} \mathbf{1}\{\hat{\gamma}_{i,NR} = g\} \left( \frac{1}{P}\sum_{t\in\mathcal{P}} \varepsilon_{i,t} - \frac{1}{R}\sum_{t\in\mathcal{R}} \varepsilon_{i,t} \right) \\
&\leq \underline{\pi}^{-1} \left( \frac{1}{NP}\sum_{i=1}^{N}\sum_{t\in\mathcal{P}} \varepsilon_{i,t} - \frac{1}{NR}\sum_{i=1}^{N}\sum_{t\in\mathcal{R}} \varepsilon_{i,t} \right) \\
&= o_p(1), \text{ as } N, P, R \to \infty
\end{aligned}$$

since $\underline{\pi} > 0$ and using a LLN for $\frac{1}{NP}\sum_{i=1}^{N}\sum_{t\in\mathcal{P}} \varepsilon_{i,t}$ and $\frac{1}{NR}\sum_{i=1}^{N}\sum_{t\in\mathcal{R}} \varepsilon_{i,t}$ which follows from Theorem 1 of Bonhomme and Manresa (2015). This holds for $g = 1, 2$, and thus $\tilde{\mu}_{NP}(\hat{\gamma}_{NR}) \xrightarrow{p} \mu^*$, as $N, P, R \to \infty$. This implies that $\tilde{\mu}_{1,NP}(\hat{\gamma}_{NR}) - \tilde{\mu}_{2,NP}(\hat{\gamma}_{NR}) \xrightarrow{p} \mu_1^* - \mu_2^* \neq 0$ by Assumption 2′(b). Thus $|tstat| \xrightarrow{p} \infty$, as $N, P, R \to \infty$. ■

## S.A.2: Additional proofs

**Proof of Lemma 1.** We know that the limit of the objective function of the correctly specified model is minimized at $(\boldsymbol{\mu}^*, \boldsymbol{\gamma}^*)$, and the MSE at that point is

$$\begin{aligned}
MSE^*(\boldsymbol{\mu}^*, \boldsymbol{\gamma}^*) &= \lim_{N,T\to\infty} \frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T}\sum_{g=1}^{G} \left(Y_{it} - \mu_g^*\right)^2 \mathbf{1}\{\gamma_i^* = g\} \\
&= \lim_{N,T\to\infty} \frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T} \varepsilon_{it}^2 \\
&= \frac{1}{N}\sum_{i=1}^{N} \sigma_i^2 \\
&\equiv \bar{\sigma}^2
\end{aligned}$$

Let $\boldsymbol{\gamma}^\star$ be such that, for all $i, j \in \{1, ..., N\}$, $\gamma_i^\star = \gamma_j^\star \Rightarrow \gamma_i^* = \gamma_j^*$. That is, all clusters defined by $\boldsymbol{\gamma}^\star$ can be generated by taking the correct set of clusters (given by $\boldsymbol{\gamma}^*$) and then splitting some $\boldsymbol{\gamma}^*$-clusters into two or more clusters. This implies that $\boldsymbol{\mu}_g^\star = \boldsymbol{\mu}_{g'}^*$ for some $g'$, for all $g$. For any such $\left(\boldsymbol{\mu}^{\bigstar},\gamma^{\bigstar}\right)$ the limit of the objective function is

$$\begin{aligned} MSE^{*}\left(\boldsymbol{\mu}^{\bigstar},\boldsymbol{\gamma}^{\bigstar}\right) &= \lim_{N,T\to\infty}\frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T}\sum_{g=1}^{\tilde{G}}\left(Y_{it}-\mu_{g}^{\bigstar}\right)^{2}\mathbf{1}\left\{\gamma_{i}^{\bigstar}=g\right\}=\lim_{N,T\to\infty}\frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T}\varepsilon_{it}^{2} \\ &= \frac{1}{N}\sum_{i=1}^{N}\sigma_{i}^{2}\equiv\bar{\sigma}^{2}\leq MSE^{*}\left(\boldsymbol{\mu},\boldsymbol{\gamma}\right) \end{aligned}$$

and so $\left(\boldsymbol{\mu}^{\bigstar},\gamma^{\bigstar}\right)$ is a solution to the population $\tilde{G}$-cluster estimation problem. ■

**Lemma 3** *For $d=1$, Assumption $2'$(b) implies Assumption $3''$(b).*

**Proof of Lemma 3.** Consider the case that $G=3$ and $\tilde{G}=2$ for simplicity, and assume $\mu_1^*<\mu_2^*<\mu_3^*$. Every element of a group has the same mean (by Assumption $2'$) and so if it is optimal for one member of a given group to be assigned to a specific group in the $\tilde{G}$-cluster model then it is optimal for *all* members of that true group. This implies that there are no split true groups between the $\tilde{G}$-cluster model groups. There are then three possible groupings for the $\tilde{G}=2$ model, in terms of the true group assignments: $\{1,(2,3)\}$, $\{(1,2),3\}$, $\{(1,3),2\}$. The latter allocation can be easily shown to be suboptimal since $\mu_1^*<\mu_2^*<\mu_3^*$, so we need only consider the first two cases.

In the first case, we have $\mu_1^{\bigstar}=\mu_1^*$, since that group comprises all the true group one variables. The other location parameter will be a convex combination of $\mu_2^*$ and $\mu_3^*$:

$$\mu_2^{\bigstar}=\frac{\pi_2}{\pi_2+\pi_3}\mu_2^*+\frac{\pi_3}{\pi_2+\pi_3}\mu_3^*$$

Then note that $\left|\mu_1^{\bigstar}-\mu_2^{\bigstar}\right|=\left|\mu_1^{*}-\mu_2^{\bigstar}\right|>\left|\mu_1^*-\mu_2^*\right|>c$, where the first inequality holds since $\mu_2^{\bigstar}\in\left(\mu_2^*,\mu_3^*\right)$ and the second inequality holds by Assumption $2'$(b). A similar inequality holds if we consider the other allocation: $\left|\mu_1^{\bigstar}-\mu_2^{\bigstar}\right|=\left|\mu_1^{\bigstar}-\mu_3^{*}\right|>\left|\mu_2^*-\mu_3^*\right|>c$ since in this case we have $\mu_2^{\bigstar}=\mu_3^*$ and $\mu_1^{\bigstar}\in\left(\mu_1^*,\mu_2^*\right)$. The extension to the general case $G>\tilde{G}\geq 2$ is proven similarly.

Next we provide an example where this implication fails for $d>1$. Consider $d=2$, $G=3$ and $\tilde{G}=2$. Assume $\boldsymbol{\mu}_1^*=[0,0]$, $\boldsymbol{\mu}_2^*=[2,0]$ and $\boldsymbol{\mu}_3^*=\left[1,\sqrt{3}\right]$, i.e., these points form an equilateral triangle on $\mathbb{R}^2$ with side lengths equal to two. Assume that $\pi_1 = \pi_2 \geq \underline{\pi} > 0$ and $\pi_3 > 1/3$, leading to the optimal $\tilde{G} = 2$ group assignment being $\{(1,2), 3\}$. Thus $\boldsymbol{\mu}_2^{\bigstar} = \boldsymbol{\mu}_3^*$ and

$$\boldsymbol{\mu}_1^{\bigstar} = \frac{1}{2}(\boldsymbol{\mu}_1^* + \boldsymbol{\mu}_2^*) = [1, 0]$$

In this case we find $\left\|\boldsymbol{\mu}_1^{\bigstar} - \boldsymbol{\mu}_2^{\bigstar}\right\| = \sqrt{3} < \min_{g \neq g'} \left\|\boldsymbol{\mu}_g^* - \boldsymbol{\mu}_{g'}^*\right\|$. Thus the $\tilde{G} = 2$ model has optimal clusters that are closer together than the clusters in the DGP. ■

**Proof of Theorem 2.** (a) This case is identical to the case considered in Theorem 1(a): a model with $\tilde{G}$ clusters is estimated, but the null of only a single cluster is true. Thus we obtain

$$Fstat \xrightarrow{d} \chi^2_{\tilde{G}-1}, \text{ as } N, P, R \to \infty.$$

(b) Now we consider a $\tilde{G}$-cluster model when the DGP has $G \in \left\{2, ..., \tilde{G} - 1\right\}$ clusters, and so the $\tilde{G}$-cluster model is too large. Note that

$$\tilde{\boldsymbol{\mu}}_{NP}(\hat{\boldsymbol{\gamma}}_{NR}) - \boldsymbol{\mu}^{\bigstar} = \left(\hat{\boldsymbol{\mu}}_{NR} - \boldsymbol{\mu}^{\bigstar}\right) + \left(\tilde{\boldsymbol{\mu}}_{NP}(\hat{\boldsymbol{\gamma}}_{NR}) - \hat{\boldsymbol{\mu}}_{NR}\right)$$

The first term on the RHS is $o_p(1)$ as $N, R \to \infty$ by Assumption 3′(a). The second term is treated as in Theorem 1(b) and is $o_p(1)$ as $N, P, R \to \infty$.

By Lemma 1, $\boldsymbol{\mu}^{\bigstar}$ is a re-ordering of $[\boldsymbol{\mu}^{*\prime}, \boldsymbol{\varphi}^{*\prime}]'$, where $\boldsymbol{\varphi}^*$ is a $\left(\tilde{G} - G\right)$ vector with elements drawn with replacement from $\boldsymbol{\mu}^*$. The well-separatedness assumption on the DGP (Assumption 2'(b)) implies that all of the $G(G-1)/2$ pairwise differences of elements of $\boldsymbol{\mu}^*$ are non-zero, i.e., $\left|\mu_g^* - \mu_{g'}^*\right| > c > 0 \;\forall\; g \neq g'$. Combining this with Lemma 1 we have:

$$\begin{aligned} \sum_{g=1}^{\tilde{G}-1} \sum_{g'=g+1}^{\tilde{G}} \mathbf{1}\left\{\left|\mu_g^{\bigstar} - \mu_{g'}^{\bigstar}\right| = 0\right\} &\leq \left(\tilde{G} - G + 1\right)\left(\tilde{G} - G\right)/2 \\ \text{and so} \sum_{g=1}^{\tilde{G}-1} \sum_{g'=g+1}^{\tilde{G}} \mathbf{1}\left\{\left|\mu_g^{\bigstar} - \mu_{g'}^{\bigstar}\right| > c\right\} &\geq \left(4\tilde{G} - 3G\right)(G-1)/2 \end{aligned}$$

Thus, while not all of the pairwise differences in $\boldsymbol{\mu}_g^{\bigstar}$ will be non-zero, there will be at least $\left(4\tilde{G} - 3G\right)(G-1)/2$ non-zero pairwise differences. This implies that

$$\tilde{\boldsymbol{\mu}}'_{NP}(\hat{\boldsymbol{\gamma}}_{NR}) A'_{1,\tilde{G}} \left(A_{1,\tilde{G}} \hat{\Omega}_{NPR} A'_{1,\tilde{G}}\right)^{-1} A_{1,\tilde{G}} \tilde{\boldsymbol{\mu}}_{NP}(\hat{\boldsymbol{\gamma}}_{NR}) \xrightarrow{p} \boldsymbol{\mu}^{\bigstar\prime} A'_{1,\tilde{G}} \left(A_{1,\tilde{G}} \bar{\Omega}_{NR} A'_{1,\tilde{G}}\right)^{-1} A_{\tilde{G}} \boldsymbol{\mu}^{\bigstar} > 0$$

by the positive definiteness of $\bar{\Omega}_{NR}$ and the full row rank of $A_G$. And thus

$$Fstat = NP\tilde{\boldsymbol{\mu}}_{NP}'\left(\hat{\boldsymbol{\gamma}}_{NR}\right)A_{1,\tilde{G}}'\left(A_{1,\tilde{G}}\hat{\Omega}_{NPR}A_{1,\tilde{G}}'\right)^{-1}A_{1,\tilde{G}}\tilde{\boldsymbol{\mu}}_{NP}\left(\hat{\boldsymbol{\gamma}}_{NR}\right)\xrightarrow{p}\infty, \text{ as } N,P,R\to\infty$$

completing the proof.

(c) Now we consider a $\tilde{G}$-cluster model when the DGP has $G>\tilde{G}$ clusters, and so the $\tilde{G}$-cluster model is misspecified. Note that

$$\tilde{\boldsymbol{\mu}}_{NP}\left(\hat{\boldsymbol{\gamma}}_{NR}\right)-\boldsymbol{\mu}^{\bigstar}=\left(\hat{\boldsymbol{\mu}}_{NR}-\boldsymbol{\mu}^{\bigstar}\right)+\left(\tilde{\boldsymbol{\mu}}_{NP}\left(\hat{\boldsymbol{\gamma}}_{NR}\right)-\hat{\boldsymbol{\mu}}_{NR}\right)$$

The first term on the RHS is $o_p(1)$ as $N,R\to\infty$ by Assumption $3''$(a). The second term is treated as in Theorem 1(b) and is $o_p(1)$ as $N,P,R\to\infty$. This implies that

$$\tilde{\boldsymbol{\mu}}_{NP}'\left(\hat{\boldsymbol{\gamma}}_{NR}\right)A_{1,\tilde{G}}'\left(A_{1,\tilde{G}}\hat{\Omega}_{NPR}A_{1,\tilde{G}}'\right)^{-1}A_{1,\tilde{G}}\tilde{\boldsymbol{\mu}}_{NP}\left(\hat{\boldsymbol{\gamma}}_{NR}\right)\xrightarrow{p}\boldsymbol{\mu}^{\bigstar\prime}A_{1,\tilde{G}}'\left(A_{1,\tilde{G}}\Omega_{NR}A_{1,\tilde{G}}'\right)^{-1}A_{1,\tilde{G}}\boldsymbol{\mu}^{\bigstar}>0$$

by Assumption $3''$(b), the positive definiteness of $\bar{\Omega}_{NP}$ and the full row rank of $A_G$. Thus

$$Fstat = NP\tilde{\boldsymbol{\mu}}_{NP}'\left(\hat{\boldsymbol{\gamma}}_{NR}\right)A_{1,\tilde{G}}'\left(A_{1,\tilde{G}}\hat{\Omega}_{NPR}A_{1,\tilde{G}}'\right)^{-1}A_{1,\tilde{G}}\tilde{\boldsymbol{\mu}}_{NP}\left(\hat{\boldsymbol{\gamma}}_{NR}\right)\xrightarrow{p}\infty, \text{ as } N,P,R\to\infty$$

completing the proof. ■

# References


[1] Bonhomme, S., and Manresa, E., 2015, Grouped patterns of heterogeneity in panel data, *Econometrica*, 83(3), 1147-1184.

[2] Hansen, C.B., 2007, Asymptotic properties of a robust variance matrix estimator for panel data when $T$ is large, *Journal of Econometrics*, 141, 597-620.

[3] Pesaran, M.H., 2015, *Time Series and Panel Data Econometrics*, Oxford University Press, Oxford.

**Table SA.1: Finite sample rejection rates, no sample splitting**

| $d$ | $G$ | $N=30$ $T=50$ | 30 250 | 30 1000 | 150 50 | 150 250 | 150 1000 | 600 50 | 600 250 | 600 1000 |
|---|---|---|---|---|---|---|---|---|---|---|
| | | *Normal data* | | | | | | | | |
| 1 | 2 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 |
| 1 | 3 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 |
| 1 | 4 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 |
| 1 | 5 | 1.000 | **0.999** | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 |
| 1 | Bonf. | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 |
| 2 | 2 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 |
| 2 | 3 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 |
| 2 | 4 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 |
| 2 | 5 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 |
| 2 | Bonf. | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 |
| 5 | 2 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 |
| 5 | 3 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 |
| 5 | 4 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 |
| 5 | 5 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 |
| 5 | Bonf. | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 |
| | | *Heterogeneous data* | | | | | | | | |
| 1 | 2 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 |
| 1 | 3 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 |
| 1 | 4 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 |
| 1 | 5 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 |
| 1 | Bonf. | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 |
| 2 | 2 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 |
| 2 | 3 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 |
| 2 | 4 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 |
| 2 | 5 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 |
| 2 | Bonf. | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 |
| 5 | 2 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 |
| 5 | 3 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 |
| 5 | 4 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 |
| 5 | 5 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 |
| 5 | Bonf. | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 |

Notes: This table presents the proportion of simulations in which the null of a single cluster is rejected in favor of multiple clusters, using the test proposed in Theorem 1 but *without* sample splitting, at a 0.05 significance level. The top panel presents results for *iid* Normal data; the lower panel presents results when the distribution is randomly drawn from one of $N(0,1)$, $Exp(2)$, $Unif(-3,3)$, $\chi^2(4)$ or $t(5)$, each standardized to have zero mean and unit variance. The dimension of the variables is denoted $d$, the number of groups considered under the alternative is denoted $G$, the number of variables is denoted $N$, and the number of time series observations is denoted $T$. Rows labeled "Bonf." use tests with a Bonferroni correction to consider $G \in \{2,3,4,5\}$ under the alternative. The number of simulations is 1000.