Online Supplement to ‘Large System of Seemingly Unrelated Regressions: A Penalized Quasi-Maximum Likelihood Estimation Perspective’

Qingliang Fan Xiao Han Bibo Jiang Guangming Pan

May 7, 2019

Abstract

This appendix presents an empirical study of public capital returns and additional simulation results of Fan et al. (2019).

1 Applications for returns of public capital

In the main text, we discussed the finite sample performance of the PQMLE when n is comparable to T. Here, we apply our proposed method to study how public expenditure affects local economic growth. Public infrastructure, with components such as state highways, bridges, water utilities, etc., contributes to long-run economic growth and competitiveness. However, how much public funds should be spent on the development of public infrastructure and what is the potential crowding-out effect are always popular topics in public policy debate. The validity of policy evaluations relies on a good estimation of the public capital productivity.

Following the study of Munnell (1990), the classic Cobb-Douglas productivity function we consider is:

$$gsp_{it} = \alpha_i + \beta_{1i} prc_{it} + \beta_{2i} hwy_{it} + \beta_{3i} water_{it} + \beta_{4i} util_{it} + \beta_{5i} emp_{it} + \beta_{6i} ump_{it} + \epsilon_{it} \qquad (1.1)$$

where gsp_{it} is the gross state product of state i at time t, prc_{it} is the stock of state private capital[1] after adjustment for capital depreciation, hwy_{it}, $water_{it}$ and $util_{it}$ are the three components of the stock of state public capital, which are highway capital, water utility capital and other utility capital, respectively, emp_{it} is the total state employment and ump_{it} is the state unemployment rate. All these variables used in the regressions are the natural logarithms of the original variables. Assuming that ϵ_t is i.i.d. with mean 0 and covariance matrix $\mathbf{\Sigma}_0$, we use the state-level panel data to estimate both the model coefficients $\boldsymbol{\beta}$ and the covariance matrix $\mathbf{\Sigma}_0$.

We use 42 annual observations of the 48 U.S. continental states from 1970 to 2011, which is an extended version of the data set used in Munnell (1990). Gross state product annual data and state private capital data are from the Bureau of Economic Analysis. Employment data are from the Bureau of Labor Statistics. Capital outlay, which details the items of state public capital, is from the U.S. Census Bureau (government division) data of state government finances. State private capital by sector, specifically, non-farm manufacturing and farm sector data, are from the Bureau

[1] State private capital is collected from agriculture, manufacturing and non-manufacturing sectors.

of Labor Statistics and the Bureau of Economic Analysis. The detailed methodology of proxies for the variables in the production function can be found in Munnell (1990).

Since $T = 42$ and $n = 48$, conventional estimation methods are not feasible for estimating the SUR system. Aggregation of individuals for the SUR model or single-equation estimation are often applied.[2] However, there are some serious disadvantages of regional aggregation in this study. First, interactions between states are not captured if we aggregate the data to study the productivity of public capital. By aggregating the data, we also miss the connections of state-level economy through channels other than geographical association, such as economic and legislative associations. Second, when using the aggregated state-level data, we cannot explore the specific characteristics of each individual state's public capital productivity. Even though aggregation of macroeconomic data is common in empirical studies, the homogeneity assumption is often invalid, which raises concerns about internal validity study.

Using the newly proposed method in Fan et al. (2018), we are able to estimate the productivity of public capital in each state while accounting for contemporaneous correlations. It is important that the estimation and predictions are based on individual states, which are more relevant in state-level legislation and policy evaluation. Additionally, we can explore the correlation structure of the unobserved errors in 48 production functions directly through both geographical associations and other possible channels. The proposed PQMLE fits well in this empirical study because some states might have correlations with other states in the random disturbances of Equation (1.1), whereas some states might not have correlations with any other state. Moreover, we need not assume any a priori covariance structure. In our empirical findings, the sparsity rate of the correlations is 64.98%.

For comparison, we also report the SUR model regression results using aggregated state-level data. The FGLS results and selected [3] PQMLE results are shown in Table 1 and Table 2, respectively. Several meaningful differences require attention. Taking the gulf region as one example, the water utility spending β_3 is not significant. However, looking at each state in the gulf region,

[2] The nine regions GF, MW, MA, MT, NE, SO, SW, CN, and WC are geographical regions, as discussed in detail in Munnell (1990).

[3] This is mainly due to the size of the original table. The whole table is available upon request.

specifically, Alabama, Florida, Louisiana and Mississippi, the coefficients of water utility are quite different for each state. The PQMLE for Alabama's water utility is positively correlated with local economic growth and is significant at the 5% level. For another example, in the south region, the other utility capital β_4 is negatively correlated with growth and is not significant. This result raises concerns that in this region, the productivity of other utility spending is not significantly associated with regional economic growth. However, looking at each state individually, North Carolina, for example, has the same sign but is much more significant. Granted that no causal relationship can be implied here, one possible policy suggestion is that the state of North Carolina could reduce spending on this component. In the west coast region, which includes California, Oregon and Washington, private capital shows the crowd-out effect. But this does not hold for California alone, where the crowd-out effect is not significant in our regression results. Considering the different economic growth patterns of the three states, it is not surprising that some coefficients differ in sign and significance. We also use the PQMLE to estimate the covariance matrix. The selected large correlations of the estimated sparse correlation matrix are shown in Table 3.

In summary, we find that, generally, states that invest more in public infrastructure tend to have greater output. This result agrees with the main findings of Munnell (1990) and Greene (2010). Our results are more efficient than the single equation OLS method, which ignores the cross-equation correlations. We also find that some individual states behave quite differently if aggregated into regional-level data. This result confirms our concern that individual characteristics are not well captured when using aggregated data.

In Table 3, we observe geographical correlation, such as that between New York and Rhode Island, Connecticut and Massachusetts, and North Carolina and Virginia, in addition to other correlation (possibly economic), such as that between Colorado and Texas. We find some correlation at the state level that is not observable in the regional aggregated data.

Table 1: FGLS regression results for the public capital productivity SUR model

	α_i	β_{1i}	β_{2i}	β_{3i}	β_{4i}	β_{5i}	β_{6i}
GF	-2.0921	0.1444	0.3754	0.0582	0.0138	0.9191	-0.0470
	(0.5858)	(0.0615)	(0.0755)	(0.0998)	(0.0635)	(0.1160)	(0.0289)
MW	-3.5254	0.4695	0.2361	0.3382	-0.0230	0.4928	-0.1391
	(0.8626)	(0.0758)	(0.0838)	(0.1250)	(0.0535)	(0.1701)	(0.0248)
MA	-7.6833	0.4849	0.1146	-0.1672	0.0468	1.4887	-0.0046
	(1.3524)	(0.0807)	(0.0569)	(0.0411)	(0.0475)	(0.2229)	(0.0297)
MT	-2.4793	0.7687	0.2309	0.0116	0.0008	0.3583	-0.1663
	(0.4579)	(0.1112)	(0.0616)	(0.0581)	(0.0382)	(0.1188)	(0.0286)
NE	-0.0098	0.8756	-0.0591	-0.0948	0.0228	0.3628	-0.0328
	(1.5097)	(0.0482)	(0.0383)	(0.1278)	(0.0276)	(0.2855)	(0.0368)
SO	-1.6778	1.1376	0.0220	-0.0706	-0.0251	0.0855	-0.1003
	(0.3052)	(0.0645)	(0.0333)	(0.0676)	(0.0265)	(0.0932)	(0.0156)
SW	-2.0714	1.0867	-0.0636	0.0438	0.0727	0.0493	-0.1260
	(0.4638)	(0.0974)	(0.0650)	(0.1002)	(0.0442)	(0.1031)	(0.0238)
CN	-1.1654	1.0717	0.0427	0.1615	-0.0507	-0.1129	-0.1837
	(0.6440)	(0.1124)	(0.0787)	(0.0758)	(0.0395)	(0.1866)	(0.0253)
WC	-4.6350	-0.1022	0.3725	0.4335	-0.0867	1.2325	-0.0609
	(0.5908)	(0.0590)	(0.0527)	(0.1541)	(0.0412)	(0.1567)	(0.0406)

The standard errors are in parentheses.

Table 2: Selected regression results for the PQMLE

	α_i	β_{1i}	β_{2i}	β_{3i}	β_{4i}	β_{5i}	β_{6i}
AL	-1.3286	0.5748	0.1331	0.2066	0.0255	0.3919	-0.0674
	(0.4786)	(0.1624)	(0.0883)	(0.1190)	(0.0353)	(0.2020)	(0.0199)
CA	-3.2155	-0.0127	0.2457	0.4014	-0.0358	1.0994	-0.0432
	(0.9026)	(0.0508)	(0.0647)	(0.1745)	(0.0166)	(0.1577)	(0.0430)
MI	-4.1030	0.1646	0.4012	0.3557	0.0082	0.8327	-0.0763
	(0.9891)	(0.0498)	(0.0645)	(0.1615)	(0.0434)	(0.2526)	(0.0423)
ME	-2.5852	0.8337	0.0256	-0.0117	0.0048	0.6743	-0.0300
	(0.8491)	(0.1023)	(0.0664)	(0.0727)	(0.0159)	(0.2599)	(0.0324)
NC	-0.4457	0.9875	0.1491	0.0030	-0.0509	-0.0168	-0.0816
	(0.5277)	(0.0998)	(0.0473)	(0.0774)	(0.0325)	(0.1678)	(0.0265)
NJ	1.2664	0.8976	0.0089	-0.0553	0.0136	0.0729	-0.0495
	(1.0111)	(0.0412)	(0.0351)	(0.0955)	(0.0224)	(0.1985)	(0.0329)
NY	-17.5926	0.0006	0.2757	-0.1000	0.1794	2.9105	0.1440
	(1.8745)	(0.0733)	(0.0762)	(0.0158)	(0.0563)	(0.2243)	(0.0475)

The standard errors are in parentheses.

Table 3: Selected large estimated correlations of unobserved errors of the production function between states

State	State	Correlation
CO	TX	0.2325
GA	VA	0.1998
NY	RI	0.1859
OR	TX	0.1808
KS	SD	0.1745
CO	OK	0.1663
WY	NV	0.1616
IA	SD	0.1549
AR	KY	0.1535
CT	MA	0.1444
NC	VA	0.1266

2 Additional simulation studies

The results in the main text suggest that the PQMLE has good estimation and inference properties due to the better estimation of the true covariance matrix $\boldsymbol{\Sigma}_0$. The simulation results show that in the model setting of commonly encountered cases, our method performs better than contemporary high-dimensional methods. In this section, we compare the finite sample performance of our proposed estimators in SUR models to that of more conventional methods, complementing the results in the main text. Specifically, we compare the estimator for both the regression coefficient $\boldsymbol{\beta}_0$ and covariance matrix $\boldsymbol{\Sigma}_0$ with that of FGLS, MLE and MD.

The simulation study is based on the data generating process of model (2.1). For simplicity, we set $k_i = 2$ for all equations $i = 1, \dots, n$; hence, $\boldsymbol{x}_{it} = (1, x_{it,1})$ are the regressors for equation i and $\boldsymbol{\beta}_{0i} = (\beta_{01i}, \beta_{02i})$ are the corresponding model coefficients. $x_{it,1}$ is generated as an AR(1) process with autocorrelation coefficient $\rho = 0.6$ and i.i.d. $N(0, 1)$ innovations for each $i = 1, ..., n$. In this simulation study, we consider $n = 20$ and $n = 50$, and the true model coefficients of the two SUR systems are reported in Table 4 and 5 in the Appendix. $(\boldsymbol{U_t})$ is generated as i.i.d. Gaussian[4]

[4]Since Gaussian errors are used, the estimators are indeed the PMLE and MLE instead of the PQMLE and QMLE. We expect that the relative performance of the penalized estimator and other aforementioned estimators is similar if non-Gaussian errors are used.

with mean 0 and $E(\boldsymbol{U_t U_t'}) = \boldsymbol{\Sigma}_0$. To investigate the performance of the PQMLE for models with $\boldsymbol{\Sigma}_0$ having different sparsity structures, we consider the following three cases. Without loss of generality, we set all the diagonal elements of $\boldsymbol{\Sigma}_0$ to 1.

Case 1: $\boldsymbol{\Sigma}_0$ is a band covariance matrix where the first off-diagonal elements are 0.5 and all other off-diagonal elements are zero.

Case 2: Set the first off-diagonal elements to 0.5, the second off-diagonal elements to 0.05 and all other elements to 0. This case is used to check whether the estimation procedure can catch small but nonzero covariance elements.

Case 3: Construct $\boldsymbol{\Sigma}_0$ by setting the m^{th} off-diagonal elements to 0.8^m for $m = 1, \ldots, n/2$. This case is used to check the performance of the PQMLE on models with less sparse error covariance matrices. The sparsity rate is only 23.68% and 24.48%, respectively, for $n = 20$ and $n = 50$. On the other hand, the off-diagonal elements of $\boldsymbol{\Sigma}_0$ decrease gradually and become increasingly close to 0 as they get further away from the diagonal elements. Indeed, when $m = 25$, $0.8^m = 0.0038$ is a tiny nonzero number. By setting $\boldsymbol{\Sigma}_0$ in this way, we check the identification ability of the PQMLE in the neighborhood of 0.

We consider six combinations of n and T. In particular, $T = 15, 25$ and 50 are considered for $n = 20$ and $T = 30, 100$ and 200 are considered for $n = 50$. We are particularly interested in the performance of the estimators when $T < n$, which is becoming increasingly popular in practice due to the availability of individual-level data. Although the asymptotic theories are established for cases where T diverges faster than n, the finite sample performance of the PQMLE shows some merit in cases where the restrictions are not as stringent.

Remark 2.1 *Clearly, when $n > T$, the sample covariance matrix $\boldsymbol{S}$ of the residuals cannot be full rank, and the solution of (2.5) in Fan et al. (2018) will be degenerate. In the calculation, we set $\boldsymbol{S} = \boldsymbol{S} + \iota \boldsymbol{I}_n$, where ι is a very small number, e.g., 10^{-6}. This process is equivalent to augmenting the data with points that do not lie perfectly in the span of the observed data. Our algorithm is designed to handle this common issue in real data. The algorithm established here converges with the same rate as that in Bien and Tibshirani (2011). Notably, when $n \approx T$, the regular QMLE of* *the covariance matrix does not work and is not recommended for empirical researchers. For a fair comparison, we add a small ι to the diagonal for the other aforementioned conventional methods. The proposed PQMLE is shown to have the best performance.*

The error covariance matrix $\mathbf{\Sigma}_0$ and regression coefficient $\boldsymbol{\beta}_0$ are estimated using the simulated data via the proposed PQMLE, MLE, FGLS and MD methods. We compare these four estimators based on the following three statistics.

1. *cov. RMSE*, the overall root mean square error of the covariance matrix estimator $\widehat{\mathbf{\Sigma}}$, i.e., its empirical Frobenius norm, which is given as *cov. RMSE* $= \sqrt{\sum_{ij}(\hat{\sigma}_{ij} - \sigma^0_{ij})^2} = \|\widehat{\mathbf{\Sigma}} - \mathbf{\Sigma}_0\|_F$.

2. *MAD*, the overall median absolute deviation of coefficient matrix $\boldsymbol{\beta}_0$, with $MAD = \text{median}_{n_{sim}}\left(\frac{\sum_{i=1}^n\left(|\hat{\beta}_{1i}-\beta_{01i}|+|\hat{\beta}_{2i}-\beta_{02i}|\right)}{2n}\right)$, where n_{sim} is the number of simulations.

3. *MSE*, the overall mean squared error of coefficient matrix $\boldsymbol{\beta}_0$, with $MSE = \text{median}_{n_{sim}}\left(\frac{\sum_{i=1}^n\left((\hat{\beta}_{1i}-\beta_{01i})^2+\widehat{var}(\hat{\beta}_{1i})+(\hat{\beta}_{2i}-\beta_{02i})^2+\widehat{var}(\hat{\beta}_{2i})\right)}{2n}\right)$, where $\widehat{var}(\hat{\beta}_{1i})$ and $\widehat{var}(\hat{\beta}_{2i})$ of the PQMLE of $\boldsymbol{\beta}_0$ are obtained from the estimated asymptotic covariance matrix according to Theorem 3, where the matrix $\boldsymbol{M}$ is estimated as $\widehat{\boldsymbol{M}} = \frac{1}{nT}\sum_{t=1}^T \boldsymbol{X}_t'\widehat{\mathbf{\Sigma}}\boldsymbol{X}_t$.

Finally, since the PQMLE enjoys model selection consistency, i.e., it can estimate the zero elements in $\mathbf{\Sigma}_0$ as 0 and the nonzero elements as nonzero, we report the following statistic for the PQMLE.

4. *Sp rate*, the sparsity recognition rate, which is calculated as $n_s/n(n-1)$, where n_s is the number of successfully recognized off-diagonal elements of the estimated covariance matrix. If the PQMLE estimates a nonzero off-diagonal element as nonzero or a zero off-diagonal element as zero, we count it as one successful event. n_s is the total number of such successful events.

Note that this value is calculated only for PQMLE since this statistic is not relevant to the other conventional estimators.

Tables 6–8 present the simulation results for $n = 20$. The results in the three tables suggest at least two features. First, the proposed PQMLE produces a more accurate estimation of the covariance matrix in terms of the empirical Frobenius norm and sparsity rate. Second, as a result of the better estimation of the covariance matrix, the PQMLE performs better than the alternative estimators in terms of the median absolute deviation and MSE of the estimator of $\boldsymbol{\beta}_0$. Specifically, for $n = 20$, the proposed PQMLE has the best MAD and MSE. This result is robust for Cases 1, 2 and 3 and for different sample sizes. Furthermore, for $n = 20$ and $T = 15$, the PQMLE is much better than the other methods in terms of the coefficient estimator MSE. The sparsity recognition rate increases with sample size T in general. The simulation results for $n = 50$ are reported in Tables 9–11. Here, we observe similar patterns, i.e., the bias of the PQMLE is better than that of the other methods. Additionally, the correct covariance structure recognition rate increases as T increases. Cases 1 and 2 have, in general, better results than difficult Case 3, which has less satisfactory results in terms of the model selection correction rate. As shown in Table 10, the model selection consistency property holds for the proposed PQMLE when the sparsity is relatively high. Additionally, when n is comparable to T and the sparsity is not severe, it is very difficult to distinguish the small elements from 0 for finite samples, as shown in Table 11.

We conclude that for a large SUR system with a sparse error covariance matrix, the PQMLE generally outperforms the alternative estimators. The simulation results indicate that our method is better than FGLS, MLE and MD in terms of: 1) the RMSE of σ's, 2) estimation bias of the regression coefficients $\boldsymbol{\beta}$ and 3) MSE of coefficients $\boldsymbol{\beta}$. The proposed PQMLE can also recognize the sparsity structure of the covariance matrix as the sample size increases. In the finite sample case, where the cross-sectional dimension n and total number of model parameters are relatively large compared to the time dimension T and the covariance matrix is sparse, our method achieves better performance. In practice, since we usually do not know the covariance matrix sparsity structures of cross-sectional units, our proposed method provides a better solution than the alternatives to estimate the general covariance matrix and regression coefficients.

3 Definition of $\hat{J}_\alpha$ test

In this section, we formally define the $\hat{J}_\alpha$ test of Pesaran and Yamagata (2017), which is compared with our proposed estimator in the second simulation setting. The factor model was already introduced in the simulation section 4.2 of Fan et al. (2018), so we will not repeat it here. Define the diagonal elements of $\boldsymbol{\Sigma}$, namely, the $n \times n$ diagonal matrix, $\boldsymbol{D} = diag\left(\sigma_{11}, \sigma_{22}, ..., \sigma_{nn}\right)$, with $\sigma_{ii} = E\left(u_{it}^2\right)$. Consider the statistic

$$W_d = \left(\boldsymbol{\tau}_T' \boldsymbol{M}_F \boldsymbol{\tau}_T\right) \hat{\boldsymbol{\alpha}}' \boldsymbol{D}^{-1} \hat{\boldsymbol{\alpha}} = \left(\boldsymbol{\tau}_T' \boldsymbol{M}_F \boldsymbol{\tau}_T\right) \sum_{i=1}^{n} \left(\frac{\hat{\alpha}_i^2}{\sigma_{ii}}\right) \tag{3.1}$$

where

$$\hat{\alpha}_i = \left(\alpha_i \boldsymbol{\tau}_T' + \boldsymbol{\beta}_i' \boldsymbol{F}' + \boldsymbol{u}_i'\right) \left(\frac{\boldsymbol{M}_F \boldsymbol{\tau}_T}{\boldsymbol{\tau}_T' \boldsymbol{M}_F \boldsymbol{\tau}_T}\right) = \alpha_i + \boldsymbol{u}_i' \boldsymbol{c}, \boldsymbol{c} = \left(\frac{\boldsymbol{M}_F \boldsymbol{\tau}_T}{\boldsymbol{\tau}_T' \boldsymbol{M}_F \boldsymbol{\tau}_T}\right).$$

is an efficient estimator of α_i. Additionally, $\boldsymbol{\tau}_T = (1, 1, ..., 1)'$, $\boldsymbol{f}_t = (f_{1t}, f_{2t}, ..., f_{mt})'$ is the $m \times 1$ vector of factors, $\boldsymbol{F} = (\boldsymbol{f_1}, \boldsymbol{f_2}, \boldsymbol{f_3}, \ldots, \boldsymbol{f_T})'$, $\boldsymbol{M}_F = \boldsymbol{I}_T - \boldsymbol{F}\left(\boldsymbol{F}'\boldsymbol{F}\right)^{-1} \boldsymbol{F}'$. Its feasible counterpart is given by

$$\hat{W}_d = \left(\boldsymbol{\tau}_T' \boldsymbol{M}_F \boldsymbol{\tau}_T\right) \hat{\boldsymbol{\alpha}}' \hat{\boldsymbol{D}}_v^{-1} \hat{\boldsymbol{\alpha}} = \left(\frac{\boldsymbol{\tau}_T' \boldsymbol{M}_F \boldsymbol{\tau}_T}{v^{-1} T}\right) \sum_{i=1}^{N} \left(\frac{\hat{\alpha}_i^2}{\hat{\sigma}_{ii}}\right) \tag{3.2}$$

where $\widehat{\boldsymbol{\Sigma}}_{ii} = \hat{\boldsymbol{u}}_i' \hat{\boldsymbol{u}}_i / T$, and the degrees of freedom $v = T - m - 1$. In the simulations, we use a three-factor model, such that $m = 3$. The infeasible statistic, W_d, can be written as

$$W_d = \sum_{i=1}^{N} z_i^2 \tag{3.3}$$

where

$$z_i^2 = \hat{\alpha}_i^2 \left(\boldsymbol{\tau}_T' \boldsymbol{M}_F \boldsymbol{\tau}_T\right) / \sigma_{ii} \tag{3.4}$$

It is then easily seen that

$$\hat{W}_d = \sum_{i=1}^{N} t_i^2 \tag{3.5}$$

where t_i denotes the standard t-ratio of α_i in the OLS regression of y_{it} on an intercept and $\boldsymbol{f}_t$, namely,

$$t_i^2 = \frac{\hat{\alpha}_i^2 \left(\boldsymbol{\tau}_T' \boldsymbol{M}_F \boldsymbol{\tau}_T\right)}{v^{-1} T \hat{\sigma}_{ii}} \tag{3.6}$$

A standardized version of $\hat{W}_d$, defined by (3.2), is

$$\frac{n^{-1/2}\left[\hat{W}_d - E\left(\hat{W}_d\right)\right]}{\sqrt{Var(\hat{W}_d)}} \tag{3.7}$$

where

$$n^{-1}E(\hat{W}_d) = E(t_i^2) \tag{3.8}$$

$$n^{-1}Var(\hat{W}_d) = n^{-1}Var(\sum_{i=1}^{n} t_i^2) = N^{-1}\sum_{i=1}^{n} Var(t_i^2) + \frac{2}{n}\sum_{i=2}^{n}\sum_{j=1}^{i-1} Cov(t_i^2, t_j^2) \tag{3.9}$$

Under Gaussianity, the individual t_i statistics are identically distributed as Student's t with v degrees of freedom, and we have (assuming $v = T - m - 1 > 4$)

$$E(t_i^2) = \frac{v}{v-2} \quad \text{and} \quad Var(t_i^2) = \left(\frac{v}{v-2}\right)^2 \frac{2(v-1)}{v-4} \tag{3.10}$$

Using (3.8)–(3.10), the standardized statistic (3.7) can be written as

$$J_\alpha(\theta_n^2) = \frac{n^{-1/2}\left[\hat{W}_d - E\left(\hat{W}_d\right)\right]}{\sqrt{Var(\hat{W}_d)}} = \frac{n^{-1/2}\sum_{i=1}^{n}(t_i^2 - \frac{v}{v-2})}{\sqrt{\left(\frac{v}{v-2}\right)^2 \frac{2(v-1)}{v-4}(1+\theta_n^2)}} \tag{3.11}$$

where

$$\theta_n^2 = n^{-1}\sum_{i=2}^{n}\sum_{j=1}^{i-1} Corr(t_i^2, t_j^2)$$

and

$$Corr(t_i^2, t_j^2) = Cov(t_i^2, t_j^2)/[Var(t_i^2)Var(t_j^2)]^{1/2}$$

To make the J_α test operational, one must provide a large n consistent estimator of θ_n^2. The J_α test is standardized assuming t_i has a standard t distribution; the test will continue to have satisfactory small sample performance even if such an assumption does not hold due to the non-Gaussianity of the underlying errors.

REFERENCES

Fan, Q., Han, X., Jiang, B. and Pan, G. (2019) Large system of seemingly unrelated regressions: a penalized quasi-maximum likelihood estimation perspective. *Econometric Theory*, forthcoming.

Greene, W. (2010) *Econometric Analysis.* 7^{th} edition, Prentice Hall.

Munell, A. (1990) Why has productivity declined? Productivity and public investment. *New England Economic Review*, 3–22.

Pesaran, M. H. and Yamagata, T. (2017) Testing for alpha in linear factor pricing models with a large number of securities. working paper.

Table 4: True slope coefficients $\boldsymbol{\beta}_0$ for $n = 20$

i	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20
β_{01i}	0.12	-0.27	-0.05	0.37	-0.08	0.21	-0.23	-0.08	0.16	-0.15	0.28	-0.27	-0.08	0.29	-0.10	0.31	-0.33	-0.07	0.26	-0.17
β_{02i}	0.63	-0.14	0.14	-0.12	-0.11	-0.17	-0.13	0.52	-0.32	-0.26	0.33	-0.14	0.19	-0.12	-0.39	0.17	-0.13	0.44	-0.22	-0.36

Table 5: True slope coefficients $\boldsymbol{\beta}_0$ for $n = 50$

i	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20
β_{01i}	0.12	-0.27	-0.05	0.37	-0.08	0.21	-0.23	-0.08	0.16	-0.15	0.28	-0.27	-0.08	0.29	-0.10	0.31	-0.33	-0.07	0.26	-0.17
β_{02i}	0.63	-0.14	0.14	-0.12	-0.11	-0.17	-0.13	0.52	-0.32	-0.26	0.33	-0.14	0.19	-0.12	-0.39	0.17	-0.13	0.44	-0.22	-0.36
i	21	22	23	24	25	26	27	28	29	30	31	32	33	34	35	36	37	38	39	40
β_{01i}	0.13	0.33	-0.21	-0.10	0.09	0.13	0.42	-0.06	0.41	0.34	0.50	0.45	0.11	0.26	0.29	0.33	0.14	0.17	-0.15	0.18
β_{02i}	0.20	0.19	0.15	0.15	0.40	0.32	0.33	-0.12	0.29	-0.11	0.24	0.30	-0.08	0.18	-0.21	0.11	0.31	0.25	0.06	-0.23
i	41	42	43	44	45	46	47	48	49	50										
β_{01i}	0.34	0.37	0.42	0.23	-0.21	0.22	0.61	0.17	0.27	0.22										
β_{02i}	0.24	0.23	0.25	0.18	-0.09	0.32	0.23	0.18	0.15	0.16										

Table 6: Simulation results for Case 1, $n = 20$

	$T = 15$				$T = 25$				$T = 50$			
	PQMLE	FGLS	QMLE	MD	PQMLE	FGLS	QMLE	MD	PQMLE	FGLS	QMLE	MD
Cov. RMSE	0.181	0.256	0.256	0.256	0.144	0.206	0.195	0.202	0.118	0.147	0.142	0.145
MAD	0.211	0.211	0.222	0.222	0.145	0.148	0.160	0.160	0.082	0.089	0.105	0.105
MSE	0.089	0.103	0.147	0.120	0.047	0.056	0.080	0.070	0.019	0.024	0.041	0.039
Sp Rate	0.331	-	-	-	0.568	-	-	-	0.767	-	-	-

Table 7: Simulation results for Case 2, $n = 20$

	$T = 15$				$T = 25$				$T = 50$			
	PQMLE	FGLS	QMLE	MD	PQMLE	FGLS	QMLE	MD	PQMLE	FGLS	QMLE	MD
Cov. RMSE	0.184	0.256	0.256	0.256	0.152	0.208	0.196	0.203	0.128	0.149	0.143	0.145
MAD	0.213	0.220	0.221	0.221	0.161	0.165	0.166	0.166	0.095	0.107	0.117	0.117
MSE	0.098	0.111	0.167	0.133	0.054	0.064	0.087	0.075	0.025	0.031	0.036	0.041
Sp Rate	0.385	-	-	-	0.575	-	-	-	0.833	-	-	-

Table 8: Simulation results for Case 3, $n = 20$

	$T = 15$				$T = 25$				$T = 50$			
	PQMLE	FGLS	QMLE	MD	PQMLE	FGLS	QMLE	MD	PQMLE	FGLS	QMLE	MD
Cov. RMSE	0.270	0.280	0.281	0.281	0.200	0.209	0.200	0.206	0.150	0.153	0.150	0.152
MAD	0.213	0.214	0.217	0.217	0.149	0.148	0.169	0.169	0.084	0.086	0.106	0.106
MSE	0.089	0.105	0.157	0.127	0.044	0.055	0.085	0.075	0.019	0.023	0.039	0.037
Sp Rate	0.623	-	-	-	0.674	-	-	-	0.728	-	-	-

The PQMLE is the proposed estimator. FGLS, QMLE, MD are the feasible generalized least squares, quasi-maximum likelihood and minimum distance estimators, respectively. The statistics are defined in the paper. The results of the case where $n > T$ are computed by adding $\iota = 10^{-6}$ to the diagonal. The sparsity recognition rate is reported for PQMLE only.

Table 9: Simulation results for Case 1, $n = 50$

	$T = 30$				$T = 100$				$T = 200$			
	PQMLE	FGLS	MLE	MD	PQMLE	FGLS	MLE	MD	PQMLE	FGLS	MLE	MD
Cov. RMSE	0.117	0.182	0.182	0.182	0.046	0.102	0.101	0.101	0.046	0.072	0.072	0.072
MAD	0.146	0.147	0.149	0.149	0.053	0.056	0.073	0.076	0.035	0.038	0.051	0.051
MSE	0.129	0.137	0.156	0.152	0.095	0.097	0.106	0.105	0.089	0.091	0.095	0.095
Sp Rate	0.217	-	-	-	0.846	-	-	-	0.858	-	-	-

Table 10: Simulation results for Case 2, $n = 50$

	$T = 30$				$T = 100$				$T = 200$			
	PQMLE	FGLS	MLE	MD	PQMLE	FGLS	MLE	MD	PQMLE	FGLS	MLE	MD
Cov. RMSE	0.119	0.182	0.182	0.182	0.078	0.102	0.101	0.101	0.061	0.072	0.072	0.072
MAD	0.139	0.140	0.140	0.140	0.065	0.068	0.073	0.073	0.044	0.046	0.052	0.052
MSE	0.130	0.138	0.158	0.152	0.100	0.102	0.116	0.115	0.091	0.093	0.095	0.095
Sp Rate	0.242	-	-	-	0.919	-	-	-	0.946	-	-	-

Table 11: Simulation results for Case 3, $n = 50$

	$T = 30$				$T = 100$				$T = 200$			
	PQMLE	FGLS	MLE	MD	PQMLE	FGLS	MLE	MD	PQMLE	FGLS	MLE	MD
Cov. RMSE	0.185	0.187	0.187	0.187	0.100	0.105	0.104	0.104	0.073	0.074	0.074	0.074
MAD	0.140	0.141	0.143	0.143	0.057	0.060	0.072	0.072	0.040	0.041	0.051	0.051
MSE	0.129	0.137	0.154	0.148	0.094	0.098	0.106	0.105	0.090	0.091	0.095	0.095
Sp Rate	0.410	-	-	-	0.455	-	-	-	0.689	-	-	-

The PQMLE is the proposed estimator. FGLS, QMLE, MD are the feasible generalized least squares, quasi-maximum likelihood and minimum distance estimators, respectively. The statistics are defined in the paper. The results for the case of $n > T$ are computed by adding $\iota = 10^{-6}$ to the diagonal. The sparsity recognition rate is reported for PQMLE only.