Common factors and common breaks in panels : an empirical investigation

Feng, Qu

2020

Feng, Q. (2020). Common factors and common breaks in panels : an empirical investigation. *Economics Letters*, 187, 108897--.
<https://dx.doi.org/10.1016/j.econlet.2019.108897>

<https://hdl.handle.net/10356/147537>

<https://doi.org/10.1016/j.econlet.2019.108897>

© 2019 Elsevier B.V. All rights reserved. This paper was published in *Economics Letters* and is made available with permission of Elsevier B.V.

Downloaded on 26 Apr 2023 18:42:05 SGT

Common Factors and Common Breaks in Panels: An Empirical Investigation

Qu Feng*

July 2019

Abstract

Unobserved common factors and common breaks are two important features in empirical studies using large panels. Recently, Baltagi, Feng and Kao (2016, 2019) extended Pesaran's (2006) common correlated effects (CCE) approach by allowing for common breaks and endogenous regressors in large heterogeneous panels. In this note, we empirically investigate these estimators and compare them with CCE and Bai's (2009) interactive fixed effects (IFE) estimators in the context of China's provincial infrastructure investment.

Keywords: Common Factors, Common Breaks, Infrastructure Investment.

JEL Classification: C23, C33, E22, H54, O40

*School of Social Sciences, Nanyang Technological University, Singapore. HSS-04-48, 48 Nanyang Avenue, Singapore 639818. Email Address: qfeng@ntu.edu.sg.

1 Introduction

Among the recent influential papers in panel data, Pesaran (2006) and Bai (2009) proposed error factor structures to model cross-sectional dependence due to unobservable common factors. Pesaran (2006) develops the common correlated effects (CCE) estimator for large heterogeneous panels with an error factor structure. The idea of the CCE approach is to use cross-sectional averages of the data to proxy for the unobserved factors, thus the slope parameters can be estimated by least squares using augmented data. Bai (2009) takes a different perspective and treats the factor structure as an interactive fixed effects (IFE) in a homogeneous panel data model, and develops an iterative estimation procedure based on a principal components estimator.

Since the advancement of Pesaran’s (2006) CCE approach and Bai’s (2009) IFE method, a multifactor error structure has been widely employed in empirical studies to model cross-sectional dependence and heterogeneous effects of unobserved macro shocks. In the applications of the CCE approach, for example, common factors are used to account for spillover in a study of private returns to R&D (Eberhardt, Helmers and Strauss, 2013), and to control for unobserved heterogeneity when examining the relationship between public debt and long-run growth (Eberhardt and Presbitero, 2015). In Boneva and Linton’s (2017) research on the issuing of a corporate bond, unobserved common shocks such as the global financial crisis are modelled by a factor structure in a discrete-choice model in heterogeneous panels. In addition, heterogeneous responses to aggregate shocks are allowed for by common factors in examining the effect of financial aid on macro outcomes by Temple and Van de Sijpe (2017), also, the reaction in a given US state to capital tax changes in other states by Chirinko and Wilson (2017).

In the applications of the IFE approach, Kim and Oka (2014) examine the effects of unilateral divorce laws on divorce rates in the US. They control for endogeneity due to the correlation between the unobserved heterogeneity and regressors to deal with bias in the resulting estimates of the treatment effects. Similarly, Gobillon and Magnac (2016) use Bai’s (2009) IFE approach to evaluate the effect of an enterprise zone program. Totty (2017), on the other hand, estimates the effect of minimal wage increase on employment in the US with a factor structure to address concerns related to unobserved heterogeneity.¹

¹CCE estimates are also reported in this empirical study.

In these empirical studies using the CCE and IFE approaches, one concern could be structural breaks. To apply these two approaches, long panel data sets are usually required. However, over a long span, parameter instability due to structural breaks is possible. In addition, with the exception of Temple and Van de Sijpe (2017), and Chirinko and Wilson (2017), endogeneity due to the correlation between the regressors and the idiosyncratic errors could bias the resulting estimates. Though an error factor structure can be used to control for the correlation between the regressors and the unobserved factors or loadings, the correlation between the regressors and the idiosyncratic errors could still be present due to reverse causality. This endogeneity is common in empirical studies using aggregate data, for example, the return of public infrastructure as surveyed by Gramlich (1994) and Calderon, Moral-Benito and Serven (2015).

Baltagi, Feng and Kao (BFK hereafter) (2016, 2019) extend Pesaran’s (2006) CCE approach by allowing for unknown common structural changes in slopes and error factor structure and endogenous regressors in large heterogeneous panels.² They find that Pesaran’s CCE approach is still valid when dealing with unobservable common factors in the presence of common breaks in slopes and error factor loadings and endogenous regressors. Given that there are no empirical investigations of the proposed estimators available, this paper aims to compare these estimators to Bai’s (2009) IFE estimator and Pesaran’s (2006) CCE mean group estimator in the context of China’s provincial infrastructure investment covering the period 1996-2015. In this specific empirical context, the trade-offs of allowing for endogeneity and common structural breaks in heterogeneous panels with an error factor structure can be illustrated.

The paper is organized as follows. Section 2 introduces BFK’s (2016, 2019) estimators in heterogeneous panels to deal with common factors and common breaks. Section 3 provides an empirical investigation by using a panel data set of China’s provincial infrastructure investment, and Section 4 concludes.

²Other methods dealing with common factors and common breaks in panels include Kim (2011), Li, Qian and Su (2016), etc. For a detailed review, please refer to BFK (2016).

2 Model

Pesaran (2006) proposes the CCE approach to deal with unobserved common factors in the errors in a large heterogeneous panel data model:

$$y_{it} = x'_{it}\beta_i + e_{it}, e_{it} = \gamma'_i f_t + \varepsilon_{it}, i = 1, \dots, N; t = 1, \dots, T. \quad (1)$$

x_{it} denotes explanatory variables, and e_{it} is cross-sectionally correlated due to unobserved factors f_t with corresponding loading vector γ_i . ε_{it} is the idiosyncratic error. The possibility that x_{it} could be affected by the unobservable common effects f_t renders the inconsistency of OLS.

Model (1) is extended by BFK (2016) by allowing for an unknown common break k_0 in some or all components of β_i ,

$$\begin{aligned} y_{it} &= x'_{it}\beta_i(k_0) + e_{it}, \\ \beta_i(k_0) &= \begin{cases} \beta_{1i}, & t = 1, \dots, k_0, \\ \beta_{2i} \neq \beta_{1i}, & t = k_0 + 1, \dots, T. \end{cases} \end{aligned} \quad (2)$$

BFK (2016) show that k_0 can be consistently estimated by least squares and that the cross-sectional mean of slopes β'_i s can be consistently estimated by the CCE approach in each regime. The resulting CCE mean group estimators in each regime is labelled as CCEMG-b here.

BFK (2019), on the other hand, consider both endogenous regressors and a common break k_0 in the slopes β_i and a different common break $k_1 \neq k_0$ in the error factor loadings. Endogeneity could be due to possible reverse causality or other sources. Thus, the model considered here can accommodate appealing empirical features: slope heterogeneity, cross-sectional dependence modelled by common factors, common breaks and endogeneity. BFK (2019) show that the CCE approach is still valid to deal with observe common factors and the least squares estimator of k_0 after using cross-sectional averages as proxies for unobserved common factors is still consistent in the presence of endogenous regressors and break in error factor loadings. The cross-sectional mean of the slopes β'_i s can be consistently estimated by IV estimation in each regime defined by the OLS estimator of k_0 , as in Hall, Han and Boldea (2012), Perron and Yamamoto (2015) in the time series setup.

Let $b_i = (\beta'_{1i}, \beta'_{2i})'$, $i = 1, \dots, N$. For every i and any possible break date $k = 1, \dots, T-1$, define $X_{1i}(k) = (x_{i1}, \dots, x_{ik})'$, $X_{2i}(k) = (x_{i,k+1}, \dots, x_{iT})'$. Let $Y_i = (y_{i1}, \dots, y_{iT})'$ and

$\varepsilon_i = (\varepsilon_{i1}, \varepsilon_{i2}, \dots, \varepsilon_{iT})'$ denote the stacked data and errors over time. Using the notation $\mathbb{X}_i(k) = \begin{pmatrix} X_{1i}(k) & 0 \\ 0 & X_{2i}(k) \end{pmatrix}$ and $F = (f'_1, f'_2, \dots, f'_T)'$, equation (2) can be written in matrix form as

$$Y_i = \mathbb{X}_i(k_0)b_i + F\gamma_i(k_1) + \varepsilon_i, i = 1, \dots, N. \quad (3)$$

Denote $w_{it} = (y_{it}, x'_{it})'$, $\bar{w}_t = \sum_{i=1}^N w_{it}$, $\bar{W} = (\bar{w}'_1, \bar{w}'_2, \dots, \bar{w}'_T)'$ and $M_w = I_T - \bar{W}(\bar{W}'\bar{W})^{-1}\bar{W}'$. BFK (2019) argue that $\bar{W}$ can be treated as exogenous asymptotically when N is large, and that it can be included as the first-stage regressors along with instruments z_{it} . Similar to the definition of $\mathbb{X}_i(k)$, we define the instrument matrix $\mathbb{Z}_i(k) = \begin{pmatrix} Z_{1i}(k) & 0 \\ 0 & Z_{2i}(k) \end{pmatrix}$ where $Z_{1i}(k) = (z'_{i1}, \dots, z'_{ik})'$ and $Z_{2i}(k) = (z'_{ik+1}, \dots, z'_{iT})'$. Denote $\mathbb{Z}_i^+(k) = (\mathbb{Z}_i(k), \bar{W})$. The predicted value of $\mathbb{X}_i(\tilde{k})$ is $\hat{\mathbb{X}}_i(\tilde{k}) = P_{\mathbb{Z}_i^+(\tilde{k})}\mathbb{X}_i(\tilde{k})$. Given the OLS estimator of the break date, $\tilde{k}$, the IV estimator of b_i is given by $[\hat{\mathbb{X}}_i(\tilde{k})'M_w\hat{\mathbb{X}}_i(\tilde{k})]^{-1}\hat{\mathbb{X}}_i(\tilde{k})'M_wY_i$, $i = 1, \dots, N$, and the mean group estimator of the cross-sectional mean of b_i , $i = 1, \dots, N$, is defined in BFK (2019) by:

$$\frac{1}{N} \sum_{i=1}^N [\hat{\mathbb{X}}_i(\tilde{k})'M_w\hat{\mathbb{X}}_i(\tilde{k})]^{-1}\hat{\mathbb{X}}_i(\tilde{k})'M_wY_i, \quad (4)$$

labelled as CCEMG-IV-b here.

3 Empirical Example: Infrastructure Investment

In this section, we empirically investigate how to deal with common factors and common breaks using the estimators proposed by BFK (2016, 2019) using a panel data set for China's provincial infrastructure investments over the period 1996-2015. This data set is collected from the website of National Bureau of Statistics of China and used in Feng and Wu (2018). The resulting estimates are compared to those obtained using Bai's (2009) IFE estimator, and Pesaran's (2006) CCE mean group estimator.

To address whether public infrastructure investment enhances the growth of the whole economy, we estimate the output elasticity with respect to public infrastructure in an aggregate production function. China is widely considered as an investment-driven economy with infrastructure investment accounting for an average rate of 9.3% of China's GDP over the period 1996-2015 (Feng and Wu, 2018). In this example, we make use of China's institutional context to obtain an instrument to deal with endogeneity issue. The endo-

geneity due to the reverse causality between output and infrastructure has been widely documented in the literature (Gramlich, 1994).

We start with a homogeneous panel data model based on an aggregate production function:

$$g_{it} = \beta_0 + \beta_b b_{it} + \beta_k k_{it} + \mu_i + \lambda_t + \epsilon_{it}, \quad (5)$$

where g_{it} is the logarithm of GDP per labor in province i in year t , and b_{it} is the logarithm of public infrastructure stock per labor, and k_{it} is the logarithm of non-infrastructure capital stock per labor. The public infrastructure stock is constructed from data on the sum of provincial-level investments in the 3 industries: (1) production and supply of electricity, gas and water; (2) transport, storage and post; (3) information transmission, computer services and software. In equation (5), β_b and β_k are the output elasticities of public infrastructure and non-infrastructure capital. μ_i denotes province specific factors, such as location, weather, endowments of raw materials, etc. Time effects λ_t are used to control for national-level macro shocks. ϵ_{it} denotes idiosyncratic error. To estimate the parameter of interest β_b , first-differenced equation form (6) below is used to deal with the non-stationarity of macroeconomic variables g_{it} , b_{it} , k_{it} :

$$\Delta g_{it} = \beta_b \Delta b_{it} + \beta_k \Delta k_{it} + \Delta \lambda_t + \Delta \epsilon_{it}, \quad (6)$$

where Δb_{it} and Δk_{it} are considered as endogenous. Thus, first-differenced (FD) estimation and first-differenced instrumental variable (FDIV) estimation are considered in that paper. Summary statistics of the variables used in the regressions are reported in Table 1. For detailed information of the data construction and variables, see Feng and Wu (2018).

Panel A of Table 2 gives FD estimates assuming that the regressors Δb_{it} and Δk_{it} are exogenous. Besides the full sample estimates in column (1), estimates using subsamples of non-eastern and eastern provinces are reported in columns (2) and (3) to highlight cross-region heterogeneity. Similarly, to allow for structural changes in elasticities, subsample estimates using the periods 1997-2007 and 2008-2015 are presented in columns (4) and (5). Substantial differences in the magnitude of the estimated β_b are observed, indicating that cross-region heterogeneity and structural changes should be accommodated in an empirically more flexible model. Panel B of Table 2 is the counterpart of Panel A assuming that Δb_{it} and Δk_{it} are endogenous. Δenb_{it} , the infrastructure capital per labor in two economically neighboring provinces, and lagged values Δk_{it-2} in differenced form are used

as instruments for Δb_{it} and Δk_{it} . The validity of instruments has been discussed in Feng and Wu (2018). In line with Feng and Wu (2018), after controlling for endogeneity, there is no strong evidence on a large and significant estimate of β_b . In addition, comparisons between columns (7) and (8), and between (9) and (10) also confirm the finding in Panel A of potential cross-region heterogeneity and structural change.

Besides endogeneity, this paper also considers three other empirical features in various cases: slope heterogeneity, common factors and a common break in slopes. Table 3, columns (1)-(4) of Panel A, consider the case of exogenous regressors, including Bai's (2009) IFE estimates in column (1), mean-group (MG) estimates without considering unobserved factors in column (2), Pesaran's (2006) CCE mean-group (CCEMG) estimates in column (3), CCEMG allowing for a common break in slopes (CCEMG-b) in column (4). In column (1) the IFE estimates of β_b and β_k are 0.197 and 0.349, respectively.³ Compared with column (1) of Table 2, the IFE estimate of β_k varies little after controlling for interactive fixed effects, while the IFE estimate of β_b increases from 0.127 to 0.197. Column (2) of Table 3 estimates a heterogeneous model to allow for different elasticities across provinces:

$$\Delta g_{it} = \beta_{b,i} \Delta b_{it} + \beta_{k,i} \Delta k_{it} + \Delta \lambda_t + \Delta \epsilon_{it}. \quad (7)$$

Column (3) of Table 3 assumes a factor structure in the error $\Delta \epsilon_{it} = \gamma'_i f_t + \varepsilon_{it}$ in equation (7) to capture the heterogeneous impact of unobserved macro shocks f_t , and the fact that regressors $\Delta b_{it}, \Delta k_{it}$ can be affected by the unobserved common factors f_t . For this model, Pesaran's (2006) CCEMG can be applied directly. Column (4) extends Pesaran's (2006) approach by allowing for a common break k_0 in the slopes:

$$\Delta g_{it} = \beta_{b,i}(k_0) \Delta b_{it} + \beta_{k,i}(k_0) \Delta k_{it} + \Delta \lambda_t + \Delta \epsilon_{it}, \Delta \epsilon_{it} = \gamma'_i f_t + \varepsilon_{it}. \quad (8)$$

Compared with the usual first-difference estimates in column (1) of Table 2, CCEMG in column (3) of Table (3) accommodates two empirical features: slope heterogeneity and cross-sectional dependence. CCEMG-b in column (4) adds one more feature of parameter structural change to CCEMG in column (3). The CCEMG estimate of β_b and β_k are 0.194 and 0.407, respectively, very close to the corresponding IFE estimates in column (1) of

³A Stata ado file *regife* developed by Gomez (2015) is used to calculate Bai's (2009) interactive fixed effects estimates. Here, 2 factors are assumed. The estimates are quantitatively similar to those with 3 factors in the errors.

Table 3 and both are slightly different from the FD estimates in column (1) of Table 3. In column (4) of Table 3, using the estimation procedure in BFK (2016), the estimated common break 2004 splits β_b and β_k in two regimes of 1997-2004 and 2005-2015.⁴ The CCEMG-b estimates of β_b and β_k deviate moderately from their CCEMG counterparts in column (3) of Table 3 in different directions.

Columns (5)-(7) of Panel B of Table 3 are the IV versions of columns of (2)-(4) of Table 3 assuming that $\Delta b_{it}, \Delta k_{it}$ are endogenous.⁵ CCEMG-IV in column (6) refers to the estimator (4) in Section 2 assuming no break. Column (8) is the IV version of CCEMG with an imposed common break in the slopes in 2007, while in column (7) an estimated common break date is used. As in Panel B of Table 2, the instruments $\Delta enb_{it}, \Delta k_{it-2}$ are used for the endogenous $\Delta b_{it}, \Delta k_{it}$. Compared to the FDIV estimates in column (6) of Table 2, the CCEMG-IV estimates in column (6) of Table 3 show a positive and significant β_k , but weak evidence on the productivity of infrastructure.

In column (7) of Table 3, as suggested by Theorem 1 of BFK (2019), with endogenous regressors the estimated common break date remains the same as 2004 in column (4). Interestingly, CCEMG-IV-b estimates of β_b and β_k in the period 1997-2004 are 0.289, 0.527 and significant, but no longer significant in the period 2005-2015. Compared with the case of exogenous regressors in column (4), the IV estimates in column (7) have much bigger standard errors.

To look at the effect of choosing structural break on coefficient estimates, we also use the imposed break date of 2007 as in the subsample estimates in columns (9) and (10) of Table 2. In this case, the CCEMG-IV-b estimate of β_k becomes 0.765 and significant over the period 2008-2015, but the coefficient of β_b is still insignificant. Comparison

⁴There are empirical studies that provide evidence why 2004 could be a break point. Lu and Xiang (2016) find the overall slowdown of economic efficiency in China after 2003 and link it to a clear policy turning point in the regional allocation of economic resources. Wu et al. (2017) also find that the return rates of infrastructure investment peak at 2003 by matching firm-level data with Chinese provincial infrastructure investment data.

⁵MG-IV in column (5) is the IV version of MG without considering unobserved common factors. In the simplified case without unobserved common factors, $Y_i = \mathbb{X}_i(k_0)b_i + \varepsilon_i, i = 1, \dots, N$, BFK (2019) show that the OLS estimator $\hat{k}, b_i$ can be consistently estimated by the IV estimator

$$\hat{b}_{i,IV}(\hat{k}) = [\mathbb{X}_i(\hat{k})' P_{\mathbb{Z}_i(\hat{k})} \mathbb{X}_i(\hat{k})]^{-1} \mathbb{X}_i(\hat{k})' P_{\mathbb{Z}_i(\hat{k})} Y_i,$$

where the projection matrix $P_{\mathbb{Z}_i(\hat{k})} = \mathbb{Z}_i(\hat{k})[\mathbb{Z}_i(\hat{k})' \mathbb{Z}_i(\hat{k})]^{-1} \mathbb{Z}_i(\hat{k})'$. The cross-sectional mean of b_i can be consistently estimated by a mean group (called MG-IV) estimator $\frac{1}{N} \sum_{i=1}^N \hat{b}_{i,IV}(\hat{k})$.

between columns (7) and (8) also indicate the sensitivity of selecting sample periods on the estimates of elasticities in the case of endogeneity.

This application shows that the proposed panel data model has the advantage of accommodating more empirical features in the data than existing models considered the literature in Panel A of Table 3. However, the trade-off seems also pronounced, especially when the endogeneity issue arises. The estimates in Panel B of Table 3 become less accurate especially when the sample size is not very big. From this point of view, applied researchers have to strike a balance between model flexibility and data constraints.

4 Conclusion

Unobserved common factors and common breaks are two important features in empirical studies using large panels. In this note, using a panel data set of China's provincial infrastructure investment we deal with these features by applying the estimators proposed by BFK (2016, 2019) based on Pesaran's (2006) CCE approach. The empirical results indicate a clear trade-off between model flexibility and data constraints.

References

- [1] Bai, J. (2009), Panel Data Models with Interactive Fixed Effects, *Econometrica*, 77, 1229-1279.
- [2] Baltagi, B.H., Feng, Q., and Kao, C. (2016), Estimation of Heterogeneous Panels with Structural Breaks, *Journal of Econometrics*, 191, 176-195.
- [3] Baltagi, B.H., Feng, Q., and Kao, C. (2019), Structural Breaks in Heterogeneous Panels, *Journal of Applied Econometrics*, forthcoming.
- [4] Boneva, L. and O. Linton, (2017). A discrete-choice model for large heterogeneous panels with interactive fixed effects with an application to the determinants of corporate bond issuance, *Journal of Applied Econometrics*, 32, 1226-1243.
- [5] Chirinko, Robert S. & Wilson, Daniel J., 2017. Tax competition among U.S. states: Racing to the bottom or riding on a seesaw?, *Journal of Public Economics*, 155(C), 147-163.
- [6] Calderon, C., Moral-Benito, E., and Servén, L. (2015). Is Infrastructure Capital Productive? A Dynamic Heterogeneous Approach. *Journal of Applied Econometrics*, 30(2), 177-198.
- [7] Eberhardt, M., C. Helmers and H. Strauss, (2013), Do Spillovers Matter When Estimating Private Returns to R&D?, *Review of Economics and Statistics*, 95, 436-448.

- [8] Eberhardt, M. and A. Presbitero, (2015), Public debt and growth: Heterogeneity and non-linearity, *Journal of International Economics*, 97, 45-58.
- [9] Feng, Q., Wu, G.L., (2018), On the Reverse Causality between Output and Infrastructure: the Case of China, *Economic Modelling*, 74, 97-104.
- [10] Gobillon, L. and T. Magnac, (2016), Regional Policy Evaluation: Interactive Fixed Effects and Synthetic Controls, *Review of Economics and Statistics*, 98, 535-551.
- [11] Gomez, M. (2015). REGIFE: Stata module to estimate linear models with interactive fixed effect. Statistical Software Components.
- [12] Gramlich, E. (1994). Infrastructure investment: A review essay. *Journal of Economic Literature*, 32, 1176-1196.
- [13] Hall, A.R., S. Han and O. Boldea (2012), Inference Regarding Multiple Structural Changes in Linear Models with Endogenous Regressors, *Journal of Econometrics* 170, 281-302.
- [14] Kim, D. (2011), Estimating a Common Deterministic Time Trend Break in Large Panels with Cross Sectional Dependence, *Journal of Econometrics*, 164, 310-330.
- [15] Kim, D., and T. Oka (2014), Divorce Law Reforms and Divorce Rates in the U.S.: An Interactive Fixed Effects Approach, *Journal of Applied Econometrics*, 29(2), 231-245.
- [16] Li, D., J. Qian, and Su, L. (2016), Panel Data Models with Interactive Fixed Effects and Multiple Structural Breaks, *Journal of the American Statistical Association*, 111, 1804-1819.
- [17] Lu, M., and K. Xiang (2016), Great Turning: How Has the Chinese Economy Been Trapped in an Efficiency-and-Balance Tradeoff? *Asian Economic Papers*, 15(1), 25-50.
- [18] Pesaran, M. H. (2006), Estimation and Inference in Large Heterogeneous Panels with Multifactor Error Structure, *Econometrica*, 74, 967-1012.
- [19] Perron, P., and Y. Yamamoto (2015), Using OLS to Estimate and Test for Multiple Structural Changes in Models with Endogenous Regressors, *Journal of Applied Econometrics*, 30(1), 119-144.
- [20] Temple, J. and Van de Sijpe, N. (2017), Foreign Aid and Domestic Absorption, *Journal of International Economics*, 108, 431-443.
- [21] Totty, E. (2017), The Effect of Minimum Wages on Employment: A Factor Model Approach, *Economic Inquiry*, 55, 1712-1737.
- [22] Wu, G.L., Feng, Q., and Wang, Z. (2017), Estimating Productivity of Public Infrastructure Investment, Working Paper, Nanyang Technological University, Singapore.

Table 1 Summary Statistics

Symbol	Definition	Unit	Mean	Std. D.	no. of obs.	Form in regression	Data sources
<i>y</i>	real output per labor	10,000 yuan	2.38	1.79	599	log	China NBS Website
<i>b</i>	real public infrastructure capital per labor	10,000 yuan	1.17	0.92	599	log	China NBS Website
<i>k</i>	real non-infrastructure capital per labor	10,000 yuan	5.21	5.05	599	log	China NBS Website
<i>enb</i>	real infrastructure capital per labor in 2 economically neighboring provinces	10,000 yuan	1.07	0.77	599	log	authors' calculation
<i>G</i>	public infrastructure investment flow	100 million yuan	674	624	599		China NBS Website
<i>L</i>	number of workers	10,000	3080	1847	599		China NBS Website

Notes:

1. All variables are measured at the provincial level during 1996-2015.
2. Units and summary statistics of all variables are reported before taking log.
3. For detailed information on the data and variables, see Feng and Wu (2018).

Table 2 Output Elasticities: Heterogeneity and Structural Change by Subsamples

Dependent variable: Output per labor

Independent variables:	Panel A: FD					Panel B: FDIV				
	Full sample	Heterogeneity		Structural change		Full sample	Heterogeneity		Structural change	
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
Infrastructure per labor	0.127*** (0.025)	0.166*** (0.025)	0.107*** (0.026)	0.144*** (0.031)	0.088** (0.036)	0.077 (0.150)	-0.070 (0.316)	0.033 (0.137)	0.202 (0.199)	-0.202 (0.244)
Non-infrastructure per labor	0.324*** (0.027)	0.346*** (0.024)	0.321*** (0.025)	0.340*** (0.040)	0.315*** (0.024)	0.250*** (0.079)	0.245* (0.133)	0.286*** (0.074)	0.099 (0.126)	0.438*** (0.128)
Regions	All	Non-eastern	Eastern	All	All	All	Non-eastern	Eastern	All	All
Periods	All	All	All	1997-2007	2008-2015	All	All	All	1997-2007	2008-2015
Year effects	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
No. of observations	569	360	209	329	240	509	322	187	269	240
Overall R ²	0.727	0.762	0.758	0.755	0.670	0.704	0.614	0.742	0.687	0.510
Instruments						$\Delta enb_t, \Delta k_{t-2}$				
1st-stage regression coefficient						0.256	0.153	0.287	0.241	0.283
1st-stage <i>t</i> -ratio						(4.61)	(2.02)	(3.92)	(3.15)	(3.45)

Notes:

- Columns (1)-(5) of Panel A report the first-difference (FD) estimates using the full sample, subsamples of non-eastern and eastern provinces, subsamples of periods 1997-2007 and 2008-2015, respectively. Cross-region heterogeneity is highlighted in columns (2)-(3), while structural change is allowed in columns (4)-(5).
- Columns (6)-(10) of Panel B report the first-difference instrumental variable (FDIV) estimates. Similar to Panel A, subsamples are used to allow for cross-region heterogeneity in parameters in columns (7)-(8), and parameter structural change in columns (9)-(10).
- Instruments $\Delta enb_t, \Delta k_{t-2}$ are used for endogenous $\Delta b_t, \Delta k_t$ in the differenced equation.
- Standard errors are reported in parentheses. The stars, *, ** and *** indicate the significance level at 10%, 5% and 1%, respectively.
- Standard errors are adjusted for 30 clusters. Robust standard errors are used in columns (2)-(3) and (7)-(8).
- A depreciation rate of 10% is used to calculate the public infrastructure and the non-public capital stocks.
- For the definition, unit of variables and data sources, please refer to Table 1.

Table 3 Output Elasticities: Common Factors and Common Break Estimates

Dependent variable: Output per labor

Independent variables:	Panel A: exogeneity					Panel B: endogeneity					
	IFE (1)	MG (2)	CCEMG (3)	CCEMG-b (4)		MG-IV (5)	CCEMG-IV (6)	CCEMG-IV-b (7)		(8)	
Public infrastructure per labor	0.197*** (0.017)	0.205*** (0.025)	0.194*** (0.023)	0.252*** (0.044)	0.179*** (0.036)	0.156 (0.132)	0.165 (0.137)	0.289** (0.137)	0.468 (0.418)	0.208* (0.128)	5.854 (6.176)
Non-public capital per labor	0.349*** (0.018)	0.361*** (0.031)	0.407*** (0.037)	0.386*** (0.052)	0.441*** (0.047)	0.231* (0.149)	0.286* (0.174)	0.527*** (0.168)	-0.370 (0.597)	0.341* (0.210)	0.765*** (0.220)
Periods	All	All	All	1997-2004	2005-2015	All	All	1997-2004	2005-2015	1997-2007	2008-2015
Year effects	Yes	Yes	Yes	Yes		Yes	Yes	Yes		Yes	
No. of observations	569	569	569	239	330	509	509	179	330	269	240
Overall R^2		0.65	0.72	0.78							
Empirical features											
slope heterogeneity	No	Yes	Yes	Yes		Yes	Yes	Yes		Yes	
cross-sectional dependence	Yes	No	Yes	Yes		No	Yes	Yes		Yes	
structural change	No	No	No	Yes		No	No	Yes		imposed	
endogeneity	No	No	No	No		Yes	Yes	Yes		Yes	

Notes:

- Columns (1)-(4) of Panel A report estimates assuming exogenous regressors, including Bai's (2009) interactive fixed effects (IFE) estimates in column (1), mean-group (MG) estimates without considering unobserved factors in column (2), Pesaran's (2006) CCE mean-group (CCEMG) estimates in column (3), CCEMG allowing for a common break in slopes (CCEMG-b) in column (4).
- Columns (5)-(7) of Panel B report IV estimates, corresponding to the IV versions of (2), (3) and (4). Column (8) is the IV version of CCEMG with an imposed common break in slopes in year 2007, while in column (7) an estimated common break date is used.
- Instruments Δnb_t , Δk_{t-2} are used for endogenous Δb_t , Δk_t in Panel B of Table 3.