Does China Invest Too Much in Infrastructure?

Daxin Dong

March 21, 2016

Abstract

This paper works on the debate about possible overinvestment of infrastructure (or more generally, public capital) in China. In order to evaluate the optimal allocation between infrastructure and non-infrastructure capital, much literature compares their marginal products by implicitly assuming that optimality holds when the marginal products are equal. We show in an endogenous growth model that this criterion is misused as long as the depreciation rates of different capitals differ. We derive a simple benchmark criterion for policy evaluation. Empirically, this paper employs a semiparametric varying coefficient model to calculate the marginal products of infrastructure capital and private capital in a panel of 30 Chinese provinces during 2004-2013. By considering the depreciation rate difference, it is found that generally there is still an underinvestment of infrastructure in Chinese provinces. We find even in the western provinces with relatively high infrastructure capital ratio and were believed to have “too much infrastructure” according to Shi and Huang (2014) among others, the optimal match between infrastructure and non-infrastructure capital was roughly maintained after 2010.

Keywords: Infrastructure, Public Capital, Marginal Product, Overinvestment, China, Semiparametric Varying Coefficient Model

JEL Classification: H54, O11, O18, O25

*Institute for International Economic Policy, University of Bonn. Address: Lennéstraße 37, 53113 Bonn, Germany. E-mail: daxindong@uni-bonn.de. The author thanks the participants at various conferences and seminars for their valuable comments.
1 Introduction

The accumulation of infrastructure played an important role in China’s rapid economic development over the past decades. However, nowadays an overinvestment of infrastructure might exist and be detrimental in China according to the concerns of some economists and policy makers. Figure 1 shows the gross capital formation rates as percentage of output in 9 countries during 1978-2014.

Figure 1: Gross capital formation rates (% of GDP) in 9 countries, 1978-2014

Compared to other countries, it is obvious that the investment rate in China is unusually high. In recent years its capital formation rate reached a level of over 45%. In the past three decades, only Singapore has ever reached this level in early 1980s. A great part of China’s capital investment was distributed into infrastructure industries. Since 1990, the ratio of infrastructure investment which is indicated by the line “China (infrastructure)” gradually increased and is now around 15% of annual GDP. This ratio of infrastructure investment is even close to the gross capital formation rates over all sectors in Japan, Brazil, America and Germany. This phenomenon induces a natural

1Typical arguments for overinvestment (or no overinvestment) of public capital or gross capital as a whole in China can be grouped into 7 types. Unfortunately they are all not competent to answer our research question. Appendix A.1 provides a brief summary on these arguments.

2Data source: “China (by NBS)” is the “Capital Formation Rate” reported by National Bureau of Statistics (NBS) of China. “China (infrastructure)”, which is also reported by the NBS, refers the ratio of only infrastructure-relevant industrial investment based on author’s calculation as the sum of “Total Investment in Fixed Assets” across 4 industrial sectors: (i) production and supply of electricity, gas and water, (ii) transport, storage and post, (iii) information transmission, computer services and software, (iv) management of water conservancy, environment and public facilities. Other data is directly from World Development Indicators (WDI).

Because of data availability several lines end at year 2013, and “China (infrastructure)” starts at 1990 while “Russia” starts at 1989, respectively. The gross capital formation rate of China reported by the NBS is slightly different from the data by WDI. But the difference is usually small and less than 2% in most years.
question: does China really need so much infrastructure? In November 2008, the Chinese central government started an unprecedented 4 trillion RMB (586 billion US$) fiscal stimulus package to deal with the global financial crisis. The money was mainly invested into fields of infrastructure. The warning about possible overinvestment of infrastructure is especially impressive in recent years when people found the public debt of some local governments, which is often related to radical infrastructure construction, had risen to a frightful level. But how much infrastructure is too much, and how to evaluate the proper level of infrastructure capital stock in aggregate economy? Reading literature does not really encourage us to answer these both theoretical and empirical questions with confidence.

The list of literature on the benefit and functioning of infrastructure (or similarly, public capital or government capital) can be extremely long while the attained consensus is not so much. Since it is extremely complicated to determine an “optimal” level of infrastructure (which involves the exact knowledge about the utility function, production technique, financing channels, household wealth distribution, demographic and geographic situation and so forth), we instead focus on the efficient allocation between infrastructure and non-infrastructure investment within a country. Climbing out of literature, the basic idea of our paper is quite simple. (i) We first employ a theoretical endogenous growth model to derive a criterion for evaluating the overinvestment and underinvestment of infrastructure capital. (ii) Then we empirically estimate the effects of public and private capital on Chinese provincial level output. (iii) By comparing our estimation result with the criterion from theoretical model, we are able to answer the question whether China invests too much in infrastructure. In this paper we focus on the spending aspect of government investment with respect to aggregate production and hence, some other issues like financing the construction and poverty reduction by infrastructure development are not within our concern. Throughout our paper, we use the terms “public capital” and “infrastructure” interchangeably.

There are three main difficulties in our research. (1) The first is the lack of an explicit criterion to evaluate the appropriate level of infrastructure capital. (2) Second, there is no official data of capital stock for empirical research. (3) Third, there exist a series of econometric difficulties to measure the output effects of public capital. We have to overcome all these obstacles. We hereby briefly discuss these three issues.

There are a lot of growth models incorporating public capital. However the connection between theoretical model and empirical work is often weak. The theoretical works have already studied many interactive aspects of public infrastructure, such as network effect, congestion, capital adjustment cost, investment efficiency, endogenous depreciation, capital quality and composition. But the empirical studies still prefer measuring the output effects of public capital and then advocating that public capital should be invested more (less) when its estimated productivity is high (low), or when it can (cannot) positively affect economic growth. A more cogitative consideration as

---

3It is even possible to write a literature review on literature reviews for infrastructure topic. Appendix A.2 lists some recent works for reference.
in some empirical works explicitly concerns the tradeoffs among investing in public capital and other sectors, which is essentially a benefit-cost analysis. Since the direct cost of raising funds for infrastructure is difficult to be measured, the most obvious opportunity cost of infrastructure – investing in private capital – is more often compared with. A lot of empirical works cannot proceed farther beyond reporting the output elasticities of infrastructure and non-infrastructure capitals, often because many of them use physical index of infrastructure which is incomparable to currency measure of non-infrastructure. Some literature argues to invest more infrastructure because the estimated elasticity is postive. This is questionable since it is the marginal product rather than elasticity who really matters. A few of researches, including Peterson (1994), Canning (1999) and Shi and Huang (2014), explicitly compare the marginal products of public and private capitals. This public-private capital tradeoff approach typically resorts to the formula $F_k = F_g$ by implicitly assuming that optimality holds when the marginal product of public capital $F_g$ equates that of private capital $F_k$. However, as we will show later the static equality $F_k = F_g$ only holds in macroeconomic dynamics abstract from a lot of elements such as capital depreciation, infrastructure congestion, and investment efficiency. On the other hand, while theoretical models produce fruitful insights onto the real world, we currently still lack a sufficiently simple, without resorting to estimate many parameters, but powerful criterion to evaluate the desirability and tradeoff of infrastructure stock. The misused formula $F_k = F_g$ is so intuitive, at least in a micro-static partial equilibrium. Will a macroeconomic model provide a different equation, or just add something on it? We will show in a Ramsey model that it is particularly not fair to ignore the depreciation of capital, which is concerned by all empirical works constructing capital stock data, when we derive the simple theoretical criterion.

The explicit dataset for aggregate capital stock in developing countries is scarce. This is also true for China. Zhang, Wu and Zhang (2004) and Zhang (2008) estimate China’s provincial gross capital stock during 1952-2000 and 1952-2004, respectively, by Perpetual Inventory Method (PIM). Most recent researches estimating capital stock in China use methods analogous to their works. For our research topic, we need go further to estimate both infrastructure and non-infrastructure capital stocks separately. Jin (2012), and Shi and Huang (2014) recently estimate China’s infrastructure capital stocks at provincial level. However, their works are not without shortcomings and we have to construct the capital stock series from the first step, while the data we can rely on is quite limited.

Following the seminal work of Aschauer (1989), a great amount of empirical researches attempted to quantify the output effects of public capital. However this is econometrically difficult. The spurious regression and reverse causality issues are two of the most severe obstacles but un-

\[4\] Recently, cointegrated panel (e.g. Canning and Pedroni, 2008) or vector error correction model (e.g. Balázs Égert, Tomasz Koźniuk and Douglas Sutherland, 2009) is employed to escape from calculating marginal products. By construction, the estimated long-run effect coefficient of infrastructure capital variable should be (smaller/larger than) zero if the infrastructure is (over-/under-) optimally invested. But in this paper we do not rely on these methods because we do not want the explicit economic dynamics of variables to be hidden by the numerical equations.
fortunately we do not have a perfect way to deal with them. Moreover, since we tend to study the
country China which has very large inter-district heterogeneities and speedy institutional and
economic transitions, we have more difficulties to select a sufficiently robust regression method. We
select the semiparametric varying smooth coefficient model to allow for the inter-district hetero-
genity of estimated parameters, while maintaining the variation meaningful in economics.

The analyses in our paper have three main implications. Firstly, the theoretical model of in-
frastucture investment shows the depreciation rate gap of different capitals should be considered
in determining the optimal allocation between capital types. Secondly, the econometric estimation
indicates a space of investing more infrastructure in most Chinese provinces, even though the in-
frastucture stock has already increased grossly in recent decades. As a byproduct of our regression,
thirdly, the relatively higher ratio of transportation and telecommunication infrastructure in the
eastern provinces can be justified in terms of higher marginal product of capital.

The rest of the paper proceeds as follows. Section 2 discusses the dilemma of using theoretical
models to derive the optimality condition for infrastructure policy analysis, and presents our theo-
retical finding of a simple criterion mainly in a first-best case. Section 3 describes the details of our
semiparametric varying coefficient econometric model. The issues regarding our data and capital
stock estimates are in Section 4. Section 5 displays our regression results and the corresponding
policy implications. Conclusion is given in Section 6.

2 Theoretical Model

In this section, we manage to use endogenous growth model with public capital to derive convenient
optimalization criterion for infrastructure investment evaluation. The desirable criterion should be
writable in simple formula(s) with two crucial features: (1) it should base on only aggregate variables
and parameters that are easily measurable; (2) it should be robust to varieties of model setup and
economic environment. The first feature is indispensable because a reliable microeconomic dataset
covering the most Chinese provinces and industries is scarce, which makes agent-level theoretical
criterion inapplicable. The second feature is necessary because a well-defined economic model for
a growing so fast transitional country like China is lacking.

2.1 Two approaches: a dilemma

The literature mostly uses the framework of Ramsey–Cass–Koopmans model, either with the decen-
tralized economy solution or the social planner problem. The pros and cons of both decentralized
economy analysis and central planner solution are so substantial that we in fact bog down a dilemma
of model selection.

In a decentralized economy, the government is limited and can only sets several policy variables
like tax rate and public expenditure. Given these policies, the households and firms behave au-
tonomously and are not controlled directly by the government. This kind of “second-best” situation is clearly more realistic and prevalent in literature. However, this second-best modeling can hardly be used in Chinese context because of three major disadvantages. (a) The associated analyses work well only when the economy is in or close to the steady state – for instance, we can describe the economy’s dynamics in (and only in) the neighborhood of steady state using a linear differential equation system. But we know little about the steady state of China’s economic development. (b) In order to obtain closed form solutions, the model has to impose specific forms of utility function, production function, and/or some strong assumptions such as constant return to scale in production. These assumptions might not suit China. (c) It is required to measure some key parameters (e.g., the preference parameter of elasticity of substitution in consumption), which are unavailable empirically.

The solution for social planner problem, in which the versatile benevolent government optimizes all variables, supplies a “first-best” case. This is clearly not realistic however useful in our policy analysis. (a) As shown below, a clear and simple optimal relation between marginal products of infrastructure and non-infrastructure capitals can be derived for the first-best situation. Thus we only need to measure a small amount of variables, which greatly facilitates our econometric regression. (b) It is no longer necessary to limit the economy to be close to the steady state. Instead, we are able to analyze the whole transition dynamics. This is especially powerful in Chinese context. Therefore, we make our theoretical model stick to the first-best situation with some loss of modeling reality. In the last part of this section, we give two decentralized economy examples to demonstrate that the central planner solution indeed provides a good benchmark.

2.2 Theoretical model: a first-best case

We construct a social planner problem to derive the optimality conditions, in a first-optimal situation, which generate a convenient equation regarding the tradeoff between private capital and public capital investment. Our social planner model incorporates several key elements developed in recent literature, including the congestion cost of public capital (Turnovsky, 1997; Fisher and Turnovsky, 1998), capital adjustment cost of private and public investment, as well as maintenance in public capital (Kalaitzidakis and Kalyvitis, 2004; Dioikitopoulos and Kalyvitis, 2008).\(^5\) We assume there are two types of agents in a closed economy – households as consumer-producers, and the government who taxes and provides public service. There is no financial sector in this economy. Thus the allocation of private investment and public investment is conducted through taxation.\(^6\)

\(^5\) We do not consider the probable implementation lag of government investment (Leeper, Walker and Yang, 2010), because it can be shown that as long as the implementation delay for building public capital and impatience of household are not too severe, the effects of implementation lag are almost negligible.

\(^6\) Absence of financial friction disables us to model the interaction of private and public investment via financial channels, which is important if we concern the crowding-in or crowding-out effects of public investment. The Section 5.2.1 briefly discusses some relevant issues.
Model setup

The social planner maximizes the social welfare, as the weighted sum of households’ utilities, in the form

\[ \int_0^\infty e^{-\rho t} U(C, L, C_g) dt \]

where \( \rho \) is time preference rate, \( C \) is aggregate private consumption, \( L \) is aggregate labor, \( C_g \) is the public consumption good provided by the government and benefits households directly. Here, following the tradition of growth model we abstract from the time index of each variable. Households are consumer-producers, who accumulate private capital and produce by a concave production function satisfying the Inada conditions. The production function is

\[ Y = F(K, K_g, L) = \tilde{F}(K, K_g, L) \]

where \( K \) is aggregate private capital stock and \( K_g \) is the service provided to each household by aggregate public capital \( K_g \). We set \( K_g = \psi K_g \) where \( \psi = N^{-\xi} \) is assumed a fixed parameter indicating the congestion effect in public capital. Here \( N \) is the amount of production units in the economy and \( 0 \leq \xi \leq 1 \) is congestion parameter. \( K_g \) becomes a pure public good when \( \xi = 0 \) and works like a private good when \( \xi = 1 \). Note that we directly model the aggregate economy as a whole. In fact in such a social planner problem if we as usual first write down the representative agent’s equations, and then aggregate to obtain the economy-wide ones, we will just get to the aggregate model here.

Households allocate after-tax income between consumption and private capital investment. Excluding fiscal risk issue, the government is assumed to always keep a balanced budget in every period. The stock of private capital \( K \) and public capital \( K_g \) evolve according to the transition equations

\[ \dot{K} = (1 - \phi_k)(1 - \tau)Y - S - C - \delta_k K \]

\[ \dot{K}_g = (1 - \phi_g)(1 - \mu_1 - \mu_2)(\tau Y + S) - \delta_g \left( \frac{\mu_1(\tau Y + S)}{Y} \right) K_g \]

where \( \phi_k \) and \( \phi_g \) are the capital adjustment costs of private and public investment, respectively. In order to keep model tractable, we simply assume \( \phi_k \) and \( \phi_g \) are constant parameters.\(^7\) Obviously, \( \phi_k \) and \( \phi_g \) can also be used to indicate the degree of utilization inefficiency of capital investment, which might be important in countries without well established institutions. Moreover, the existence of \( \phi_k \) and \( \phi_g \) indicates that public capital and private capital stocks are generally not one-to-one convertible. Variable \( \tau \) gives the rate of income tax and \( S \) is the amount of lump-sum tax.

\(^7\) The end of Section 2.2.3 presents some justifications for assuming \( \phi_k \) and \( \phi_g \) constant.
the government’s tax income is $\tau Y + S$. Variables $\mu_1$ and $\mu_2$ tell the proportion of tax used for maintenance of public capital and providing public consumption goods $C_g$, respectively. $\delta_k$ is the fixed depreciation rate of $K$. $\delta_g$ is the depreciation rate of $K_g$, which is a function of ratio of government spending on maintenance of public capital $\mu_1(\tau Y + S)$ to economic scale $Y$. We restrict that the government always hold budget balanced. Thus, clearly the flow of public consumption goods $C_g$ is

$$C_g = \mu_2(\tau Y + S)$$

Since $K^*_g = \psi K_g$, the equation of $\dot{K}_g$ can be equivalently written in terms of $K^*_g$ as

$$\dot{K}^*_g = \psi(1 - \phi_g)(1 - \mu_1 - \mu_2)(\tau Y + S) - \delta_g \left( \mu_1(\tau Y + S) \right) Y K^*_g$$

**Equilibrium optimality conditions**

To get the first order optimality conditions, we have the usual (present value) Hamilton function

$$\mathcal{H} = e^{-\rho t}U(C, L, C_g) + \lambda_1 \{(1 - \phi_k)(1 - \tau)Y - S - C\} - \delta_k K$$

$$+ \lambda_2 \left\{ \psi(1 - \phi_g)(1 - \mu_1 - \mu_2)(\tau Y + S) - \delta_g \left( \mu_1(\tau Y + S) \right) Y \right\} K_g^*$$

$$+ \lambda_3 \left\{ \mu_2(\tau Y + S) - C_g \right\}$$

in which the control variables are $C, L, C_g, \tau, \mu_1, \mu_2, S$ and state variables are $K$ and $K^*_g$ (or equivalently $K_g$). The combination of first order conditions $\frac{\partial \mathcal{H}}{\partial C} = 0$, $\frac{\partial \mathcal{H}}{\partial L} = 0$, $\frac{\partial \mathcal{H}}{\partial C_g} = 0$, $\frac{\partial \mathcal{H}}{\partial \tau} = 0$, $\frac{\partial \mathcal{H}}{\partial \mu_1} = 0$, $\frac{\partial \mathcal{H}}{\partial \mu_2} = 0$, $\frac{\partial \mathcal{H}}{\partial S} = 0$, $\frac{\partial \mathcal{H}}{\partial K^*_g} = -\dot{\lambda}_1$, $\frac{\partial \mathcal{H}}{\partial K^*_g} = -\dot{\lambda}_2$ (with the proper transversality conditions) can give us several important relations with clear economic meanings:

$$U_c = U_{cg}$$

$$-U_l = (1 - m)U_c F_l$$

$$\lambda_1(1 - \phi_k) = \lambda_2 \psi(1 - \phi_g) = \lambda_3$$

$$(1 - \phi_k)F_k = (1 - \phi_g)F_g + \frac{1}{1 - m}(\delta_k - \delta_g(m))$$

---

8If consumption tax or government bond is additionally introduced into the model, (just as expected in a social planner economy) basic conclusions would not be changed at all.
where \( m = \frac{\mu_1 (rY + S)}{Y} \) is the relative expenditure of public capital maintenance, \( U_c, U_{cg}, U_l \) are the marginal utilities of \( C, C_g, L \), and \( F_l, F_k, F_g \) are the marginal products of \( L, K, K_g \). Naturally the congestion effect on public capital does not matter in any of these equations since the social planner can fully handle this externality. Among the above four equations, what we want to focus on is the last one. This is a simple criterion telling how to make tradeoff between infrastructure capital and non-infrastructure capital. At optimality, the marginal products of infrastructure and non-infrastructure capitals should not be equal generally. Instead, the relationship depends on the investment efficiency and the corresponding depreciation rates.

If we restrict the simplest case that \( \phi_k = \phi_g = \mu_1 = 0 \) i.e. there is no capital adjustment cost and no need for maintenance spending (and hence \( \delta_g \) is fixed), we would have

\[
F_k = F_g + (\delta_k - \delta_g)
\]

This equation can be equivalently written as \( (F_g - \delta_g) = (F_k - \delta_k) \) which might be interpreted as the equality of return to capitals (net of depreciation) of infrastructure and non-infrastructure in accounting sense. However we decide to not use this interpretation since we are not sure how large the return to capital deviates from marginal product in reality. Since the marginal products \( F_g \) and \( F_k \) are with respect to GDP in which tax is included, it is better to call them “social rate of return” rather than just “rate of return to capital” which is often referred to (after-tax) market return of capital investment. It is notable that Canning and Bennathan (2000) essentially obtain a criterion as the same as \( (F_g - \delta_g) = (F_k - \delta_k) \). But they mainly derive it from an accounting perspective and explain marginal product minus depreciation rate as internal rate of return to the investment project. They omit the possibility of unequal depreciation rates and set \( \delta_k = \delta_g \).

### 2.2.1 Interpretation of the simple formula

From an angle of production maximization, the spirit of this simple formula \( (F_g - \delta_g) = (F_k - \delta_k) \) is intrinsically as the same as \( F_k = F_g \). That is, at optimality the society should allocate capital resource in a way that each marginal unit of capital should have the same contribution everywhere. The difference is that the equation \( (F_g - \delta_g) = (F_k - \delta_k) \), unlike \( F_k = F_g \), considers the dynamic effects of private and public investment that, in order to sustain economic growth the capital decays in the following periods should be compensated. Hence the capital sector which depreciates faster should be invested less until its productivity is large enough to offset this disadvantage. The formula is seemingly so ordinary, or even trivial that it could be derived from many very simple macro-and microeconomic models. But we would like to underline two of its merits. (1) First, as we have already seen it serves as the base to incorporate more elements (e.g. institutional reform to improve capital investment efficiency, government’s budget allocation between public capital maintenance and investing new capital) if the associated data is available for empirical research. (2) Second, it is a robust benchmark formula for decentralized economy equilibrium. (We will discuss this second
point soon.)

The policy implication of the above simple equation is very clear: if the marginal product of infrastructure net of depreciation \((F_g - \delta_g)\) is larger than that of non-infrastructure capital \((F_k - \delta_k)\), it is beneficial to increase the investment of public capital; conversely, when \((F_g - \delta_g)\) is below \((F_k - \delta_k)\) it is better to allocate more resource to non-infrastructure sector.

In some early literature, either the depreciation of capital was not considered (e.g. Barro, 1990; Futagami, Morita and Shibata, 1993; Turnovsky, 1997; Aschauer, 2000; or by modeling public services as flow) or it was assumed that there is no difference between \(\delta_k\) and \(\delta_g\) (e.g. Bajo-Rubio and Díaz-Roldán, 2005). Accordingly the condition \(F_k = F_g\) was stated. Our result shows that, even in a most tractable social planner problem the relation \(F_k = F_g\) does not hold in general, as long as we consider a little bit more realistic economic environment. Hence, in empirical work before we compare the marginal products of public and private capitals from aggregate production function, we should care that between \(F_k\) and \(F_g\) there should exist a wedge which depends on the parameters (and probably policy variables). Because of the data limit, we cannot deal with the capital adjustment cost and maintenance spending in the following empirical analysis. So from now on we focus on the case of \(F_k = F_g + (\delta_k - \delta_g)\).

2.2.2 The importance of depreciation difference

There is no doubt that if the depreciation difference \((\delta_k - \delta_g)\) is very close to zero, we can regard \(F_k = F_g\) as a good approximate criterion indicating a good match of private and public capital. But at least in the context of China, the magnitude of depreciation difference is not negligible.

The difference of depreciation rates is the result of distinct physical composition of infrastructure and non-infrastructure capital. The infrastructure mostly consists of structure, which depreciates at a relatively low rate. In contrast, the equipment, machine and some other materials with faster depreciation rates made up of a large proportion of the non-infrastructure capital. Bai, Hsieh and Qian (2006) assume a depreciation rate of 8% for structures and 24% for machinery in China. Zhang, Wu and Zhang (2004) assume that the depreciation rate is 6.9% for building, 14.9% for equipment, and 12.1% for other capitals. Let us assume that infrastructure is made up completely of structures. And non-infrastructure consists of half structures and half machines. Using the depreciation rates from Zhang, Wu and Zhang (2004), the depreciation difference \((\delta_k - \delta_g)\) would be 4%. In Shi and Huang (2014) (they actually use the value from Bai, Hsieh and Qian, 2006), Huang and Shi (2014), the depreciation difference \((\delta_k - \delta_g)\) is 8% and 7% respectively. From literature, e.g. Kamps (2006), Arslanalp et al. (2010), Arestoff and Hurlin (2006), we can also make an international comparison of depreciation rates and the difference \((\delta_k, \delta_g, \delta_k - \delta_g)\). Table 1 lists the capital depreciation rate values in some empirical literature.
Table 1: Values of Capital Depreciation Rate in Some Empirical Literature

<table>
<thead>
<tr>
<th>Literature</th>
<th>$\delta_k$</th>
<th>$\delta_g$</th>
<th>$\delta_k - \delta_g$</th>
<th>Sample Country and Period</th>
</tr>
</thead>
<tbody>
<tr>
<td>based on</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Zhang, Wu and Zhang (2004)</td>
<td>10.9%</td>
<td>6.9%</td>
<td>4%</td>
<td>China, 1952-2000</td>
</tr>
<tr>
<td>based on</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Bai, Hsieh and Qian (2006)</td>
<td>16%</td>
<td>8%</td>
<td>8%</td>
<td>China, 1978-2005</td>
</tr>
<tr>
<td>Huang and Shi (2014)</td>
<td>14%</td>
<td>7%</td>
<td>7%</td>
<td>China, 1995-2011</td>
</tr>
<tr>
<td>Kamps (2006)</td>
<td>8.5%</td>
<td>4%</td>
<td>4.5%</td>
<td>22 OECD countries, 2001</td>
</tr>
<tr>
<td>Arslanalp et al. (2010)</td>
<td>8.5%</td>
<td>4%</td>
<td>4.5%</td>
<td>13 middle-income non-OECD countries, 2001</td>
</tr>
<tr>
<td>Arestoff and Hurlin (2006)</td>
<td>7%</td>
<td>3.25%</td>
<td>3.75%</td>
<td>26 developing countries, 1998</td>
</tr>
</tbody>
</table>

Note: “based on” means the paper does not directly provide the estimated values of $\delta_k$ and $\delta_g$. We calculate them by assuming infrastructure (non-infrastructure) is made up completely of buildings (half buildings and half machines).

We learn two points from these researches. One is that infrastructure has a significantly lower depreciation rate than private capital. At the second point, we find the magnitude of $(\delta_k - \delta_g)$ is comparable to that of $\delta_g$. In literature the typical estimated magnitude of $F_k$ or $F_g$ roughly ranges from 0.2 to 0.4.\(^9\) Thus, the depreciation difference counts for 10%-40% of estimated marginal product of capital. Obviously neglecting the depreciation difference could significantly misguide the policy making of resource allocation between private and public investment, especially when $F_k$ and $F_g$ are close. (Apparently, if the capital adjustment cost is substantial or investment efficiency is low, the importance of the depreciation rate difference even increases.)

We can use a simple Cobb–Douglas production function to illustrate the importance of $(\delta_k - \delta_g)$. Details of this excise are in Appendix B. Figure 11 in the appendix displays the capital resource misallocation in one single period if a depreciation gap $(\delta_k - \delta_g)=0.06$ is not taken into account. The consequent capital misallocation is visibly serious. By setting the depreciation difference at 0.04, 0.06, 0.08 respectively and calculating the allocation of $K_g$ and $K$ based on equation $F_k = F_g$ and equation $F_k = F_g + (\delta_k - \delta_g)$, it is found that if the social planner incorrectly uses the equation $F_k = F_g$, on average the resource as large as 4.79%, 7.36% and 10.1% of stock $K$ would be misallocated from public capital into private sector. Even worse, in a long-run perspective, the neglecting of depreciation rate difference causes a severe deviation from the optimal convergence path toward steady state. Clearly, the undesirable consequence is a grave welfare loss. In our policy experiment demonstrated in Figure 12 of Appendix B, when the economy starts with half steady state capital stocks with depreciation rates $(\delta_k, \delta_g, \delta_k - \delta_g)=(0.12, 0.06, 0.06)$, a welfare loss

\(^9\)The estimated marginal product of capital varies depending on the method and data used. Here we list some examples of the estimated capital returns in China. (Although the capital return does not necessarily equate the marginal product of capital in aggregate production sense, it should be an indicator of the size.) The pre-tax return on capital is roughly 21% at 2005 in Bai, Hsieh and Qian (2006) and 18% at 2013 in Thomas (2013). The median value of marginal revenue product of capital is 32% at 2007 in Ding, Guariglia and Knight (2012).
as large as 5.54% permanent consumption reduction would occur. The loss is 2.40% and 10.3% for $(\delta_k, \delta_g, \delta_k - \delta_g) = (0.11, 0.07, 0.04)$ and $(0.13, 0.05, 0.08)$, respectively. The numerical instances provide important lessons for developing countries. Especially for low-income countries far behind the steady state and with high depreciation rates, if the planners do not consider the depreciation difference seriously, the associated welfare loss might even climb to 20%, 30% or more.

2.2.3 The special case of constant capital investment efficiency

It is widely believed that, e.g. Pritchett (2000), in developing countries the coefficients of capital investment (in)efficiency $\phi_k$ and $\phi_g$ are significantly nonzero. However we can show in 3 steps that when the coefficients $\phi_k$ and $\phi_g$ are constant, in terms of the classical econometric results there is actually no difference between adjusting the capital stock estimates by investment efficiency or not in order to apply the simple criterion $F_k = F_g + (\delta_k - \delta_g)$. That is, for policy analysis we can equivalently use unadjusted capital stock ($K_{g,t, K_t}$) with formula $F_k = F_g + (\delta_k - \delta_g)$ or adjusted capital stock ($K_{g,t, K_t}^*$) with formula $(1 - \phi_k)F_k^* = (1 - \phi_g)F_g^* + (\delta_k - \delta_g)$.

A normal capital accumulation process is written as $K_{t+1} = (1 - \delta_t)K_t + f(I_t)$ where $f(I_t) \leq I_t$ represents the efficiency (or capital adjustment cost) of investment to generate new capital. In a usual Perpetual Inventory Method (PIM) case, we do not adjust the capital stock by investment efficiency. We have $f(I_t) = I_t$ and hence $K_{t+1} = (1 - \delta_t)K_t + I_t$. If we leave from the PIM but assume that the capital investment efficiency is constant, we get $f(I_t) = \alpha I_t$ where $0 < \alpha \leq 1$ is the constant efficiency parameter. We use the symbol $K_t^*$ to denote the estimated capital stock after adjusting by investment efficiency. Hence we have $K_{t+1}^* = (1 - \delta_t)K_t^* + \alpha I_t$.

In the 1st step, we derive the parameter estimates in usual econometric models. Assuming that $K_0^* = \alpha K_0$ (this assumption is natural), we then have $K_t^* = \alpha K_t$ in all periods. Or, we write it in log-form: $k_t^* = \ln \alpha + k_t$. If we write it in first-differenced form, that is $\Delta k_t^* = \Delta k_t$. In a usual aggregate production function regression, we have $y_t = \text{constant} + \beta_g k_{g,t} + \beta_k k_t + \beta_l l_t$ and $y_t = \text{constant}^* + \beta_g^* k_{g,t}^* + \beta_k^* k_t^* + \beta_l^* l_t^* = \text{constant}^{**} + \beta_g^* k_{g,t}^* + \beta_k^* k_t + \beta_l^* l_t$, respectively. Clearly, we will get $\beta_g = \beta_g^*$, $\beta_k = \beta_k^*$ and $\beta_l = \beta_l^*$. If we regress using the first-differenced variables, we have $\Delta y_t = \text{constant} + \beta_g \Delta k_{g,t} + \beta_k \Delta k_t + \beta_l \Delta l_t$ and $\Delta y_t = \text{constant}^* + \beta_g^* \Delta k_{g,t}^* + \beta_k^* \Delta k_t + \beta_l^* \Delta l_t$. Clearly, we will also get $\beta_g = \beta_g^*$, $\beta_k = \beta_k^*$ and $\beta_l = \beta_l^*$.

In the 2nd step, we calculate the marginal products. We have $F_k^* = \beta_k^* \frac{\dot{y}_t}{K_t^*} = \frac{1}{\alpha_k} \beta_k \frac{\dot{y}_t}{K_t}$ and likewise $F_g^* = \frac{1}{\alpha_g} F_g$. In the 3rd step, we apply our calculated marginal products into our simple criterion $(1 - \phi_k)F_k = (1 - \phi_g)F_g + (\delta_k - \delta_g)$. There is no doubt that $\alpha_k = 1 - \phi_k$ and $\alpha_g = 1 - \phi_g$ here. Then, we have the criterion $F_k = F_g + (\delta_k - \delta_g)$ in PIM case (in which $\alpha_k = \alpha_g = 1$ means $\phi_k = \phi_g = 0$) and $\alpha_k F_k^* = \alpha_g F_g^* + (\delta_k - \delta_g)$ in efficiency-adjusted case. Recall that the 2nd step tells that $\alpha_k F_k^* = F_k$ and $\alpha_g F_g^* = F_g$. Therefore, when $\alpha_k$ and $\alpha_g$ are constant, using unadjusted capital stock ($K_{g,t, K_t}$) and efficiency-adjusted capital stock data ($K_{g,t, K_t}^*$) are
in fact equivalent.

Literature shows that the (nearly) constant capital investment efficiency is likely in practice, at least in some developing countries. Arestoff and Hurlin (2006), Hurlin and Arestoff (2010) investigate the efficiency of public capital investment in Colombia and Mexico. They find that the new capital created can be well approximated by a simple linear function of investment: \( f(I_t) = \alpha I_t \) where the efficiency parameter \( \alpha \) is constant and below 0.5. This is also confirmed by Gupta et al. (2014), from which we find the average “normalized Public Investment Management Indices (PIMI)” for 40 low income and 31 middle income countries are quite similar in 1990s and in 2000s. The possibility of constant capital investment efficiency enhances our confidence of applying our simple criterion in policy analysis, particularly when the information about the exact efficiency is unavailable.

2.3 Two illustrative decentralized economy models

We already showed that in centrally planned economy the social welfare maximization requires \( F_k = F_g + (\delta_k - \delta_g) \). A further question is: does this criterion change in a decentralized economy? Unfortunately a closed form solution for the corresponding criterion is generally infeasible, especially when \( \delta_g = \delta_k = 0 \) does not hold, unless a set of strict assumptions are imposed into the decentralized model. On the other hand, (as mentioned at Section 2.1) several disadvantages of decentralized economy modeling tend to make the effort of finding decentralized solution meaningless in Chinese context. Therefore, in this section we merely use two artificially simplified models to illustrate the form of criterion in decentralized economy.

Using the two illustrative models, we want to show that \( F_k = F_g + (\delta_k - \delta_g) \) can be a convenient good benchmark for empirical policy analysis. We focus the macroeconomic dynamics along the balanced growth path (BGP) (which indeed exists in our simple examples) in which the consumption, capital stock and output grow at the same rate \( \gamma = \dot{C}/C = \dot{K}/K = \dot{K}/K_g = \dot{Y}/Y \).\(^{10}\) We investigate one example of welfare-maximizing, and another of growth-maximizing tax policy.

2.3.1 Example of welfare-maximizing tax policy

We look at a situation under welfare-maximizing tax policy through the len of a model with basic setup from Fisher and Turnovsky (1998). The model is actually a simplified decentralized counterpart of our socially planned economy model in Section 2.2. This model does not incorporate capital investment efficiency and public capital maintenance, since these two features complicate the calculation while merely providing essentially analogous insights. The representative agent

\(^{10}\)The growth rate \( \gamma \) can be zero or strictly positive, depending on the specific model functional form and parameterization. But we would not distinguish these two cases intentionally, since the value of \( \gamma \) is not really important for our purpose in theoretical analysis.
maximizes the life-time utility
\[
\int_0^\infty e^{-\rho t} u(c, l) dt
\]
subject to the flow budget constraint
\[
\dot{k} = (1 - \tau) f(k, K_s, l) - s - c - \delta k k
\]
where \(\tau\) is the distortionary income tax rate and \(s\) the amount of lump-sum tax. Following a set of literature (e.g. Turnovsky, 1997; Fisher and Turnovsky, 1998; Gómez, 2004; Dioikitopoulos and Kalyvitis, 2008) we assume the public service \(K_s = K_g(\frac{k}{K})^\xi\). The individual production function is \(y = f(k, K_g, l) = f(k, K_g(\frac{k}{K})^\xi, l)\). With homogeneous consumer-producers, in equilibrium the aggregate and individual private capital stocks are related by \(K = Nk\). With \(\psi = N^{-\xi}\) we get the aggregate production \(Y = Ny = N f(k, K_g, l) = N f(\frac{k}{N}, K_g, \frac{k}{N}) = F(K, K_g, L) = F(K, \psi K_g, L)\). The properties of the assumed individual and aggregate production functions imply the following relationships:

\[
F_k = f_k; \quad F_{g^*} = N f_{g^*}; \quad F_g = \psi F_{g^*}; \quad F_l = f_l
\]

where \(F_k, F_{g^*}, F_g, F_l\) are the marginal products of \(Y\) with respect to the aggregate variables \(K, K_g, K_g, L\), respectively. And \(f_k, f_{g^*}, f_g, f_l\) are the marginal products of \(y\) associated with \(k, K_g^*, K_g, l\). The government finances the public investment by either distortionary or lump-sum tax. Considering the optimization behavior of agents in a decentralized economy, the government intends to maximize the social welfare as the sum of all consumers’ utilities by setting the policy instrument – taxation. Government’s budget constraint is

\[
\dot{K}_g = \tau F(K, \psi K_g, L) + S - \delta g K_g
\]

Household’s optimization solves the Hamilton function

\[
H = e^{-\rho t} u(c, l) + \lambda \left[ (1 - \tau) f(k, K_g(\frac{k}{K})^\xi, l) - s - c - \delta k k \right]
\]

taking the policy variables \(\tau, s\) and aggregate variables \(K, K_g\) as given. Two of the first order conditions are

\[
e^{-\rho t} u_c = \lambda
\]

\[
(1 - \tau) \left[ f_k + \xi \left( \frac{k}{K} \right)^{\xi-1} \frac{K_g}{K} f_{g^*} \right] - \delta k = - \frac{\dot{\lambda}}{\lambda}
\]

Plugging the relationships \(F_k = f_k, F_{g^*} = N f_{g^*}\) and \(F_g = \psi F_{g^*}\), we get an equation in aggregate
variables:

\[(1 - \tau) \left[ F_k + \xi \frac{K_g^\alpha}{K^\alpha} F_g \right] - \delta_k = -\frac{\lambda}{\lambda} \]

The government flow budget constraint is just another way of writing the transition equation for public capital:

\[\frac{\dot{K}_g}{K_g} = \frac{\tau Y + S}{K_g} - \delta_g\]

Noting that in equilibrium \(\dot{C} = \dot{c}\), obviously imposing several often used assumptions would facilitate us to obtain some simple equations in the macroeconomic dynamic equilibrium.

[Ass. 1] We do not consider the variation of labor input by setting fixed \(l = \bar{l}\). [Ass. 2] We assume consumer has logarithm utility function \(u(c) = \ln c\). These result in \(\dot{C} = \dot{c}\), obviously imposing several often used assumptions would facilitate us to obtain some simple equations in the macroeconomic dynamic equilibrium.

[Ass. 3] We restrict that government only levies income tax \(\tau Y\) but no lump-sum tax. [Ass. 4] A Cobb-Douglas individual production function \(f(k, K_g) = k^{\alpha_k} K_g^{\alpha_g}\) is imposed. Therefore, the aggregate production is \(F(K, K_g) = Nf(k, K_g) = N^{(1 - \alpha_k)} \xi \alpha_g K^\alpha_k (K_g)^{\alpha_g}\) which gives \(F_k = \alpha_k \frac{Y}{K}\) and \(F_g = \alpha_g \frac{Y}{K_g}\). By combining all these four assumptions (i.e. fixed labor input, log-utility, no lump-sum tax, and Cobb-Douglas production function) and noting that \(\dot{C} = \dot{K}_g\) along the BGP, we immediately reach an equality \((1 - \tau) \left[ F_k + \xi \frac{\alpha_g}{\alpha_k} F_k \right] - \delta_k - \rho = \tau \alpha_g F_g - \delta_g\). Rearranging it produces the simple formula:

\[F_k = F_g + (\delta_k - \delta_g) + \left\{ \rho + \left( \frac{\tau}{\alpha_g} - 1 \right) F_g + \left[ \tau - (1 - \tau) \xi \frac{\alpha_g}{\alpha_k} \right] F_k \right\}\]

which can be further simplified if the tax rate \(\tau\) is known.

It is well known that (e.g. Barro, 1990) under the key assumptions of full congestion \(\xi = 1\) and \(\alpha_k + \alpha_g = 1\), the welfare-maximizing tax rate is \(\tau^\ast = \alpha_g\). In this case, we readily get

\[F_k = F_g + (\delta_k - \delta_g) + \rho\]

As discussed in Appendix C.1, in fact both the property of setup \(K_g^\ast = K_g(\frac{k}{K})^\xi\) and empirical evidence guarantee the reasonability to restrict \(\xi\) to be sufficiently close to value 1. Since, fixing other parameters, \(\tau\) would be a continuous function of \(\xi\), the simple formula \(F_k = F_g + (\delta_k - \delta_g) + \rho\) would be a good approximation for situation of \(\xi\) around 1.11 In the next step we further simplify the model to exemplify what could happen when \(\xi < 1\).

Under partial congestion, the public capital is no longer shared completely privately and hence the government is able to use a tax rate \(\tau < \alpha_g\) to finance public service. In order to obtain

---

11Typically the value of time preference rate \(\rho\) is quite small, such as 0.02 in Barro (1999). For example, with logarithm utility we have the relationship between \(\rho\) and steady state real interest rate \(\bar{r}\) that \(\bar{r} = \rho + \bar{g}r\) where \(\bar{g}r\) is the steady state economic growth rate. Let us assume \(\bar{r} = 4\%\) and \(\bar{g}r = 2\%\). We accordingly have \(\rho = 2\%\). The Section 2 of Weitzman (2010) gives some normative justifications for setting \(\rho = 0\).
a clear solution, we impose one additional assumption onto the previous four. [Ass. 5] \( \alpha_k + \alpha_g = 1 \) in the production function. (Hereby, after adding these five assumptions our model is analogous to that in Dioikitopoulos and Kalyvitis, 2008). Equalizing the growth rate of aggregate consumption \( \gamma_c \) in decentralized economy and \( \gamma_{SP}^c \) in social planner problem, a necessary condition of welfare-maximizing income tax is attained: \( \tau^* = \frac{\alpha_k \xi}{\alpha_g \xi + \alpha_k} + \frac{\alpha_k}{\alpha_g \xi + \alpha_k} \). As long as \( \xi \) is sufficiently close to 1 (e.g. \( \xi = 0.95 \)), the term \( \rho \) is already large enough to offset the negative term \( -\frac{\alpha_k (1-\xi)}{\alpha_g \xi + \alpha_k} F_g \). Accordingly we see again that the criterion \( F_k = F_g + (\delta_k - \delta_g) + \frac{\alpha_k (1-\xi)}{\alpha_g \xi + \alpha_k} F_g \) derived from the central planner solution serves as a good benchmark. (Admittedly, if we suppose that \( \xi \) is indeed not enough close to 1 by some reasons, we might get \( F_k = F_g + (\delta_k - \delta_g) + NT \) with \( NT \) negative. Anyhow, we find the existence of a wedge between \( F_k \) and \( F_g \) disavows the relationship of \( F_k = F_g \).)

We already see that the degree of congestion impacts on the magnitude of wedge \( F_k - F_g \), notwithstanding the impact is small when \( \xi \) is close to 1. In different places and for different public capital types, the degree of congestion could be dissimilar. Thus the empirical work can be strengthened if we are able to measure the congestion parameter and determine the exact size of optimal wedge. Howbeit this is a difficult work and is important for our empirical research topic only when \( \xi \) is far from 1 which case is believed by us to be not. Moreover, it will be promising if we can analyze a more elaborate congestion model beyond setting \( K_s^g = K_g (\frac{k}{K})^\xi \). However, this likely renders the wedge to depend on the calibration of much more parameters and make empirical research messy. Also, it is highly suspect whether a closed form solution is available (while noting that we need five strict assumptions to get the abovementioned simple formula for \( \xi < 1 \)). Therefore we leave these aspects for work in the future.

### 2.3.2 Example of growth-maximizing tax policy

Economic growth is usually the main measurable merit of the government’s economic intervention. And in China, the GDP growth rate is one of the crucial indicators determining the political promotion chance of local officials. Hence it is useful to analyze the situation with growth-maximizing tax policy. The intuition to answer what would happen is simple. (As usual, we suppose that the government in a decentralized economy aims to maximize the economic growth and needs to levy distortionary tax to finance its policy.) The distortions associated with taxation discourage the accumulation of private capital. As a result the public to private capital ratio \( z^{max} \) for growth maximization in decentralized economy is higher than (or in some special case, equal to) the ratio \( z^{SP} \) in the first-best social planner situation. Here \( z = K_g/K \). Note that the first-best ratio \( z^{SP} \) is implicitly restricted by the criterion \( F_k = F_g + (\delta_k - \delta_g) \) with usual concave production function. Thus \( z^{max} > z^{SP} \) combining with the properties of concave production function implies that the corresponding growth maximization criterion in a decentralized economy would be \( F_k = F_g + (\delta_k - \delta_g) + PT \) where \( PT \) is some positive term(s).
Following the argument in Dioikitous and Kalyvitis (2008), using the model in Section 2.3.1 with the abovementioned five assumptions it can be shown that $\tau^{max} = \alpha_g$ and

$$F_k = F_g + (\delta_k - \delta_g) + [\rho + \alpha_g(1 - \xi)F_k]$$

accordingly. Clearly the term $\rho + \alpha_g(1 - \xi)F_k$ is positive. In fact, as in the welfare-maximizing case, an explicit simple criterion under growth-maximizing tax policy is (perhaps more) seldom available. But after we check several models it seems that the formula $F_k = F_g + (\delta_k - \delta_g)$ really serves well as the baseline. Here we would like to present another straightforward model for illustrative purpose. A generalized research on growth-maximizing criterion might be done in the future.

The model is revised from Kamps (2005). We additionally introduce the depreciations $\delta_g$ and $\delta_k$ into the model. This model intentionally introduces government debt because this makes the derivation of criterion very easy. The representative household maximizes the life-time utility

$$\int_0^\infty e^{-\rho t} \frac{c^{1-\sigma}}{1-\sigma} dt$$

subjekt to the budget constraint

$$\dot{k} + \dot{d} + c = (1 - \tau)y + rd - \delta_k k$$

with individual production function $y = k^{\alpha_k}k_g^{\alpha_g}$ with $\alpha_k + \alpha_g = 1$. This model is abstract from the labor input choice and assumes full congestion of public capital. The household takes the tax rate $\tau$, market interest rate $r$ as given. Since government raises risk-free public debt $d$ at fair interest rate, household is willing to lend any amount of $d$ (subject to resource constraint, of course,) to government. Hence $d$ is actually decided by government, not household. So household’s first order optimality conditions can be derived from the Hamilton function

$$\mathcal{H} = e^{-\rho t} \frac{c^{1-\sigma}}{1-\sigma} + \lambda \{(1 - \tau)y + rd - \delta_k k - c\}$$

Combining $\frac{\partial \mathcal{H}}{\partial c} = 0$ and $\frac{\partial \mathcal{H}}{\partial k} = -\dot{\lambda}$ generates $(1 - \tau)f_k - \delta_k = \sigma \frac{\dot{c}}{c} - \rho$. Here $f_k$ is the marginal product of $y$ with respect to $k$. Along the balanced growth path (BGP) (which indeed exists in this present simple model), the consumption, capital stock and output grow at the same rate $\gamma = \frac{\dot{c}}{c} = \frac{\dot{k}}{k} = \frac{k}{k_g} = \frac{\dot{y}}{y}$. Thus

$$\gamma = \frac{1}{\sigma}[(1 - \tau)f_k - \delta_k - \rho]$$

In the competitive equilibrium, it must be that

$$r = (1 - \tau)f_k - \delta_k$$
(If \( r < (1 - \tau) f_k - \delta_k \), household would not lend to government. If \( r > (1 - \tau) f_k - \delta_k \), government has no incentive to pay such a high interest rate since the interest payment comes from distortionary tax revenue.)

Government’s flow budget constraint is

\[
\dot{D} = rD + \dot{K}_g - \tau Y + \delta_g K_g
\]

It is assumed that the tax \( \tau Y \) in each period is completely used to pay government debt interest payment \( rD \) and compensate the public capital depreciation \( \delta_g K_g \). i.e. \( \tau Y = rD + \delta_g K_g \). Hence, the net accumulation of public capital \( \dot{K}_g \) is fully financed by \( \dot{D} \) i.e. \( \dot{D} = \dot{K}_g \). This means \( D = K_g \).

The government is assumed to maintain a particular public to private capital ratio \( z \) and maximize economic growth \( \gamma \). Tax and debt are used to serve this purpose. It is obvious that with identical agents under full congestion of public capital, we have \( Y = Ny = K^\alpha_K K^\alpha_g \) and thus \( F_k = f_k \). The equality \( \tau Y = rD + \delta_g K_g \) and \( D = K_g \) give

\[
\tau = \frac{(r + \delta_g)K_g}{Y} = (r + \delta_g)z^\alpha_k
\]

and

\[
r = \tau z^{-\alpha_k} - \delta_g = \frac{\tau}{\alpha_g} F_g - \delta_g
\]

Now we nearly find the aimed criterion. On the one side, we get \( r = \frac{\tau}{\alpha_g} F_g - \delta_g \) from government budget constraint. On the other side, we have \( r = (1 - \tau) f_k - \delta_k = (1 - \tau) F_k - \delta_k \) by household’s optimization. Thus clearly

\[
(1 - \tau)F_k - \delta_k = \frac{\tau}{\alpha_g} F_g - \delta_g
\]

But in this equation the tax rate \( \tau \) is still unknown. In this illustrative example we focus on the case of maximized growth rate \( \gamma^{max} \) along the BGP. Appendix C.2 documents the steps to find the value of growth maximizing \( \tau^{max} \) which is \( \alpha_g \).

Now we ultimately get the optimality condition assuming \( \alpha_k + \alpha_g = 1 \) in a decentralized economy: \( (1 - \alpha_g)F_k - \delta_k = F_g - \delta_g \) i.e.

\[
F_k = F_g + (\delta_k - \delta_g) + \alpha_g F_k
\]

Comparing this with the first-best case criterion \( F_k = F_g + (\delta_k - \delta_g) \), we find the depreciation wedge is increased in the decentralized economy. This is consistent with our intuition \( F_k = F_g + (\delta_k - \delta_g) + PT \) with \( PT \) positive. This model presented above is extremely simplified and only for illustrative purpose. Especially, the constant return to scale assumption \( \alpha_k + \alpha_g = 1 \) is very helpful to get the simple criterion. In a more elaborate and realistic model, it is in general very difficult to get the explicit expression of the \( PT \) term. We might, again, resort to simple intuition to get some
insights.

We consider two possible ways to extend the model – incorporating (i) the congestion of public capital i.e. the negative externality of private capital accumulation, and (ii) maintenance expenditure to reduce depreciation of public capital. In the decentralized economy, the growth-maximizing tax rate will be positively related to the degree of public service congestion in the economy because of the desirability of tax to mitigate the intensity of private capital’s negative externality. The growth-maximizing tax rate will be higher in the existence of public capital maintenance, which needs to be financed to sustain the high usage or lower the depreciation of public capital. So in both cases the growth maximization tax rate is likely higher. In consequence the ratio $z_{max}$ is probably larger; and the depreciation rate wedge between $F_k$ and $F_g$ might increase. Admittedly, the intuition could be not correct. In the model of Dioikitopoulos and Kalyvitis (2008) in which congestion and public capital maintenance are present together, they find the interaction of them might make the overall effect of congestion on the growth-maximizing tax rate ambiguous, depending on the initialization of congestion level and structural parameters.

Although a detailed theoretical analysis based on a generalized decentralized economy model would be beneficial, that is far beyond the scope of our current paper. Now we summarize our finding from theoretical model briefly: the optimization in a centralized or decentralized economy would always require a wedge between $F_k$ and $F_g$ as long as the condition $\delta_g = \delta_k = 0$ does not hold; and the formula $F_k = F_g + (\delta_k - \delta_g)$ is found to be a good benchmark criterion for policy analysis.

3 Empirical Model

There are basically two approaches to combine theory and reality and evaluate the desirability of current public capital stock level. The first approach estimates the productivities or benefits of public capital and private capital, and then compares their relative efficiencies. The second approach directly compares the real level or ratio of public capital with the supposed growth-maximization or welfare-maximization value from calibrated theoretical growth model (e.g. Kamps, 2005). We choose the former approach. Regardless of some drawbacks of the aggregate production function approach (discussed in e.g. Haughwout, 1998, 2002), the estimation of aggregate production function has become the dominant method. Our empirical model also relies on it. Since the empirical part of our paper would be analogous to the paper by Shi and Huang (2014) (henceforth SH for short), we now move to a short discussion on their paper and then lay out our econometric model.

3.1 A short discussion on Shi and Huang (2014) paper

SH paper uses a semiparametric aggregate production function to estimate the marginal products of infrastructure and non-infrastructure capitals in 28 Chinese provinces. Based on their estimation
result, they propose that after 2008 most western (eastern and central) provinces exhibited an oversupply (undersupply) of infrastructure relative to private capital since the marginal product of infrastructure in the region is lower. According to our finding from theoretical model, their policy suggestion would need more inspection because they do not consider the wedge of depreciation rate. However the main merits of their work are obvious. First, they recalculate the provincial capital stock series in China by especially noting that the price of infrastructure increased faster than that of private capital. Also, they depreciate the two types of capitals by different rates. Thus, their estimated capital stock data is perhaps more reliable than many previous works. Secondly, they use recently developed semiparametric varying coefficient method to estimate the output elasticities of capitals. By allowing the coefficient to vary with economic circumstance, they are able to better model the nonlinear interactions among public capital, private capital and production. Thirdly, they are very serious to deal with several severe problems prevalent in aggregate production regression, especially the issues of spurious correlation and reverse causality.

3.2 Empirical model: semiparametric varying coefficient model

3.2.1 Regression model

A lot of econometric challenges stand in front of us when we try to identify the marginal products of aggregate production factors. Now we start to discuss these econometric difficulties one by one and write down our regression model accordingly. The organization of this section is analogous to that in SH since our framework is admittedly revised from theirs. In each model equation below we simply save the subscript index for district in each variable, as it is apparent that we are talking about a panel data set.

(a) Misspecification of functional form

The Cobb-Douglas or translog production function was often resorted to in literature. But it is considerably suspect whether these functional forms correctly specify the nonlinearity between production input and output. Nonparametric or semiparametric models can help us go beyond these imposed function forms. Theoretically, by some proper nonparametric techniques we can directly estimate the aggregate production function of the form:

\[
\ln Y_t = f(\ln K_{g,t}, \ln K_t, \ln L_t) + \ln A_t
\]

where \( K_{g,t}, K_t \) and \( L_t \) is the stock of infrastructure, non-infrastructure capital and labor amount, respectively, and \( Y_t \) is the output and \( A_t \) the TFP. Howbeit a pure nonparametric estimation can hardly be interpreted in economic sense. For the topic of production function, a semiparametric varying coefficient model is especially powerful since it allows the output elasticities of input factor to vary with economic environment, while keeping the ways of variation nonarbitrary.
Classical microeconomic theory shows that the marginal product of one production factor could be influenced by other input factors. Concerning the private and public capital investment at macroeconomic level, it is natural to believe that a good match of public and non-public capitals is crucial for high productivities of both capitals. Hence, we set our model in the way that allows the elasticities of capital (and labor) to change over time, as a function of infrastructure to non-infrastructure capital stock ratio \( z_t = \frac{K_{g,t}}{K_t} \). For example the output elasticity of infrastructure capital can be written as \( \beta_{g,t} = \beta_g(z_t) \). Of course we might also consider the coefficients to vary with respect to other variables. In the estimation robustness check part of Section 5.3, we study other possibilities.

SH impose that the return to scale of inputs is 1 and write the production function in per capita form without argument \( L_t \). This constant to scale assumption, which is widely used in literature estimating the aggregate production function, however is probably not true especially when the elasticity parameters are allowed to vary over time. Hence it is very tempting to not impose this assumption and explicitly include \( L_t \) in our model. In fact, a set of literature sticking to constant return to scale get the estimated coefficient on labor input larger than 1, implying a negative output elasticity of private capital which is questionable (e.g. Kamps, 2006; Sturm and de Haan, 1995).

But we still, as that in SH, restrict the model in per capita form without estimating the coefficient on \( L_t \) because of two reasons. (1) Firstly, unlike at the firm-level production the labor in one province is not really a variable input here. Because of no data on the working hour of Chinese workers, the constructed variable \( L_t \) mostly reflects the demographic information rather than dynamic production input. (2) Secondly, we find that (log-) output \( \ln Y_t \) and capital stocks \( \ln K_{g,t} \) and \( \ln K_t \) are clearly at least \( I(1) \) but \( \ln L_t \) could be not, as documented by the panel unit root test results in Table 8 at Appendix I.1. Directly using population as \( L_t \) is clearly not suitable since the labor proportion in population is substantially heterogeneous across provinces. If we use the amount of labor defined as people aged 15-64 to represent \( L_t \), we will find \( \ln L_t \) in some provinces are nearly \( I(0) \) because the population slightly increases while labor to population ratio slightly decreases. This makes the econometric model questionable. If we (ultimately) use \( L_t \) of skill-adjusted labor amount (which is discussed in detail in Section 4.2.2), it is found that variation of \( \ln L_t \) could be almost constant in some provinces. This renders efficiently estimating parameters difficult. Therefore, as a result of these two abovementioned reasons we consider the aggregate production function in intensive form:

\[
\ln y_t = f(\ln k_{g,t}, \ln k_t) + \ln A_t
\]

where the variables in lower case letter are variables per skill-adjusted labor. We would like to underline again that we use variables per capita because of the data and technique difficulties regarding \( L_t \), not because we believe the production exhibits constant returns to scale (though these two motivations just bring about numerically indistinguishable model setup). We think
avoiding the explicit inclusion of $L_t$ in the model is helpful for better estimation.

(b) **Spurious correlation**

Along with the high speed development in China, the variables of output and capital stock are definitely nonstationary. Thus the problem of spurious correlation arises. In literature, there are three major ways to deal with spurious regression issue.

(i) **Detrending**  There are several detrending methods including first (or higher order) differencing, linear detrending (e.g. Hassler, 1999), stochastic detrending (e.g. Ferson, Sarkissian and Simin, 2010), HP filter and so on. A comparison of detrending methods can be found in Burnside (1998) and Canova (1998a,b). SH use the first differencing. While first differencing removes the spurious correlation, it has defects. It “destroys any long run relationship that may exist amongst the levels of variables of interest” (Abdih and Joutz, 2008), but this does not matter in our model as we focus on the marginal effect within one period rather than the “long run relationship”. In the context of allowing parameters varying over time, the first differencing method possibly makes estimated parameter values uncertainly volatile, depending on the uncontrollable time points based on which the time series is split up. Appendix D discusses this point in detail and shows why this issue is not a severe concern in our research. On the other hand, the other detrending methods do not work since they demand relatively strong assumption underlying the data generating process which we lack here. (Actually we tried the linear detrending, stochastic detrending and HP filter but did not obtain any meaningful result.) Thereby, we have no choice but the first differencing.

(ii) **Cointegration**  The cointegration method is popular. Abdih and Joutz (2008) give an example of using cointegration analysis to investigate the impact of public capital on American private sector output. But we do not employ cointegration technique because (1) we do not find an easy way to incorporate varying coefficient model into cointegration analysis; (2) more importantly, for our data sample the relevant variables are integrated of different orders as documented in Table 8 of Appendix I.1 – this feature violates the precondition of cointegration analysis.

(iii) **Error correction model**  Error correction model (ECM) can estimate both the short term effects and long term effects of explanatory variables. However, the typical ECM suits the case that dependent variable gradually returns to its long run equilibrium after a change in independent variables. But in our aggregate production function situation, we actually assume the dependent variable of output reach the new equilibrium level within one period. (We have indeed tried to estimate our model in ECM but gotten no reasonable result.) Therefore, we resort to the first differencing method and get:

$$\Delta \ln y_t = \Delta f(\ln k_{g,t}, \ln k_t) + \Delta \ln A_t$$
where symbol $\Delta$ refers to the first differencing operator. Next we express the equation using $\Delta \ln k_{g,t}$ and $\Delta \ln k_t$. Mean value theorem tells that $g(x_2) - g(x_1) = \frac{\partial g}{\partial x}(cx_1 + (1 - c)x_2) \cdot (x_2 - x_1)$ where $\frac{\partial g}{\partial x}$ is the gradient, $x$ could be a vector and $c \in (0, 1)$. Thus we have $\Delta f(\ln k_{g,t}, \ln k_t) = \beta_{g,t} \Delta \ln k_{g,t} + \beta_{k,t} \Delta \ln k_t$ where $\beta_{g,t}$ and $\beta_{k,t}$ measure the marginal output contributions of infrastructure and non-infrastructure capital at one specific time point between period $t$ and $t - 1$. As mentioned previously, it is not unreasonable to assume the output elasticities of capitals vary along with the ratio of $z_t = K_{g,t}/K_t$. Built on this assumption, the model now becomes:

$$\Delta \ln y_t = \beta_g(z_t) \Delta \ln k_{g,t} + \beta_k(z_t) \Delta \ln k_t + \Delta \ln A_t$$

(c) Multicollinearity

We find that the correlation between two type capital stocks is very high (correlation coefficient $> 0.9$ for level variables). This raises the problem of multicollinearity in regression. Imposing constant return to scale in inputs, or run cross-section regression with fixed parameters can mitigate this problem however is not suitable for us. Alternatively, we can use the idea of principal component regression if the level variables are used for regression. But this is already not needed after we taking the first differencing – for first differenced variables of capital stocks the correlation coefficient $< 0.5$ in our sample.

(d) Reverse causality

Reverse causality issue arises because an increase of capital stock could be just a result of capital accumulation in an environment of output increase. In econometrics, this mainly turns out to be the endogeneity problem of capital stock variables. Reverse causality problem is often mitigated by IV or GMM regression. But as (1) IV or GMM method cannot be incorporated with varying coefficient model smoothly and (2) our sample’s short time dimension critically hinders their powers, we deal with reverse causality from three other aspects. We add time fixed effect dummies, proxy variable for TFP shock, and rely on 3-year moving average value of variables rather than annual data.

The TFP term can be written as $\Delta \ln A_t = \sigma_t + \Delta \varepsilon_t$ where $\sigma_t$ is the time fixed effect common to all provinces and $\varepsilon_t$ is the residual which is the TFP shock. Reverse causality arises when the term $\ln k_{g,t}$ or $\ln k_t$ is correlated to $\varepsilon_t$. This is quite possibly the case since the government (or firm, household) is able to adjust the factor input when they perceive (anticipate or observe) the TFP shock. For instance, the government can invest more infrastructure as discretionary policy to stimulate the economy when negative TFP shock comes. The existence of reverse causality makes the estimation biased.

The TFP shock $\Delta \varepsilon_t$ can be divided into two components $\Delta \varepsilon_t = \Delta \omega_t + \Delta \mu_t$ where $\omega_t$ is perceived by the agents and hence correlated to $\ln k_{g,t}$ and $\ln k_t$, while $\mu_t$ is the uncorrelated error term. So it is important to find a valid proxy variable for $\omega_t$. SH propose to use the approach from Ackerberg,
Caves and Frazer (2006) (ACF) to address the reverse causality problem. But, because of the assumptions of ACF do not hold in our context of aggregate production function, ACF approach is in fact not suitable for us. Appendix E states why we think SH misuse the ACF approach. However the underlying idea of ACF approach – using proxy variables to control for unobservables when estimating productivity – indeed works. As pointed out in SH, two conditions for proxy variable should hold: (i) the proxy is not an input factor for the output, (ii) it is monotonic with $\omega_t$. We use the HP filtered consumption to GDP ratio to proxy the conceived TFP shocks. The reason is that: based on economics theory, the private consumption should be smooth over time, which indicates the fluctuations in consumption ratio should be the result of some “shocks”.

After we estimate the model, we can obtain the estimated value of $\omega_t$. Then we can test the monotonicity assumption between proxy variable and the perceived TFP shock. Appendix G gives us the result. We find that the monotonicity indeed well holds. We furthermore also calculate the correlation between (first differenced) $\hat{\omega}_t$ and the HP-filtered $\ln y_t$ which is 0.38. Since we basically should expect that the HP-filtered $\ln y_t$ is an indicator of business cycle which is quite positively correlated to the TFP shock, the value of 0.38 tells that our selection of proxy variable is useful.

We assume that $\Delta \text{proxy}_t = \phi(\Delta \omega_t)$ where $\phi(\cdot)$ is a monotonic function of $\Delta \omega_t$. Then we have $\Delta \omega_t = \phi^{-1}(\Delta \text{proxy}_t)$. For simplicity, we use the semiparametric varying coefficient form to approximate the function $\phi^{-1}(\cdot)$. We write as:

$$\Delta \omega_t = \phi^{-1}(\Delta \text{proxy}_t) = \varphi_1(z_t) \Delta \text{proxy}_t = \varphi_1(z_t) \Delta \text{proxy}_t$$

(In the robustness check in Section 5.3 we also check the case of using third order polynomial to proxy $\Delta \omega_t$.) Thus we now have:

$$\Delta \ln y_t = \beta_g(z_t) \Delta \ln k_{t,g} + \beta_k(z_t) \Delta \ln k_t + \varphi_1(z_t) \Delta \text{proxy}_t + \sigma_t + \Delta \mu_t$$

Furthermore, we rely on the 3-year geometric mean values of level variables ($Y_t, K_{t,g}, K_t, L_t$) rather than the original values to run the semiparametric regression. The motivation is simple. First, even though the investment efficiency of capital is 100%, it takes time for the capitals (especially the infrastructure capital) to really work after investment. Since we do not know how much proportion of the capital measured by monetary value is currently working as production factor, a several-year moving average of capital stock is a reasonable approximation to the stock of “effective” capital. Second, by smoothing the data we are able to cancel out some TFP shocks and help mitigate endogeneity problem. Using several year average value of variables in growth regression is not rare. For instance, Devarajan, Swaroop and Zou (1996) and Gupta et al. (2014) all use 5-year average growth rate.
(e) Spillover effects from neighboring regions (and additional control variable)

We follow SH’s setting of spatial spillover effects of infrastructure and non-infrastructure capital but simply assume that there is not spatial autocorrelation of shocks. In addition, while the residential capital is excluded from our constructed infrastructure and non-infrastructure capital stock (see Section 4.2.1 for details), we consider the fact that in most provinces the residential house investment counts for a large part (roughly 10%) of GDP. Then, ignoring the house investment perhaps induces bias in estimating other coefficients. So we add the variable of private residential investment $house_t$ in the regression. Now we get our final complete regression model:

$$\Delta \ln y_t = \beta_g(z_t)\Delta \ln k_{g,t} + \beta_k(z_t)\Delta \ln k_t + \varphi_1(z_t)\Delta proxy_t + \varphi_2(z_t)\Delta house_t$$

$$+ \theta_g * W * \Delta \ln \tilde{k}_{g,t} + \theta_k * W * \Delta \ln \tilde{k}_t + \sigma_t + \Delta \mu_t$$

where $\tilde{k}_{g,t}$ and $\tilde{k}_t$ denote the vector of capital stocks in other provinces, and $W$ is a known spatial weight matrix. We use the relative economic size of one province among all the adjacent provinces as weight, with the geometric mean of 1993-2013 real GDP as indicator of economic size. No geographically adjacent provinces only have zero weights. Table 2 gives the definitions of all variables in our econometric model. Data sources of the variables are discussed in Section 4.

**Table 2: Definitions of Variables in Econometric Model**

<table>
<thead>
<tr>
<th>Variable</th>
<th>Definition</th>
</tr>
</thead>
<tbody>
<tr>
<td>$y_t$</td>
<td>GDP per skill-adjusted labor</td>
</tr>
<tr>
<td>$z_t$</td>
<td>ratio of infrastructure to non-infrastructure capital stock</td>
</tr>
<tr>
<td>$k_{g,t}$</td>
<td>infrastructure capital stock per skill-adjusted labor</td>
</tr>
<tr>
<td>$k_t$</td>
<td>non-infrastructure capital stock per skill-adjusted labor</td>
</tr>
<tr>
<td>proxy$_t$</td>
<td>proxy variable for TFP shock; we use the HP filtered consumption to GDP ratio</td>
</tr>
<tr>
<td>house$_t$</td>
<td>private residential investment per skill-adjusted labor</td>
</tr>
<tr>
<td>$W$</td>
<td>spatial weight matrix; we use the relative economic size of one province among all the adjacent provinces as weight</td>
</tr>
<tr>
<td>$\tilde{k}_{g,t}$</td>
<td>vector of infrastructure capital stocks in other provinces</td>
</tr>
<tr>
<td>$\tilde{k}_t$</td>
<td>vector of non-infrastructure capital stocks in other provinces</td>
</tr>
<tr>
<td>$\sigma_t$</td>
<td>time fixed effect common to all provinces</td>
</tr>
<tr>
<td>$\mu_t$</td>
<td>error term</td>
</tr>
</tbody>
</table>

Note: We use 3-year moving average values of $(Y_t, K_{g,t}, K_t, L_t)$ to construct $(y_t, k_{g,t}, k_t, house_t, \tilde{k}_{g,t}, \tilde{k}_t)$, as mentioned in Section 3.2.1(d).

3.2.2 Estimation

The varying coefficient model can be estimated by the local constant least-squares method as described in Li and Racine (2007). For a regression model $Y_i = X_i'\beta(Z_i) + \varepsilon_i$ the general form of
the estimator is:

$$\hat{\beta}(z) = \left[ \sum_{j=1}^{n} X_j X_j' K \left( \frac{Z_j - z}{h} \right) \right]^{-1} \sum_{j=1}^{n} X_j Y_j K \left( \frac{Z_j - z}{h} \right)$$

We choose the Gaussian kernel function and bandwidth selected by least-squares cross-validation. The coefficients for time fixed effects and spillover effects can be estimated by the three steps procedure suggested in Li and Racine (2007) for the “partially linear varying coefficient model”. We use the residual based wild bootstrap method suggested by Mammen (1993) to estimate standard errors of estimated coefficients.

4 Data

4.1 Estimating capital stock by Perpetual Inventory Method

Without a dataset for the capital stocks, as a tradition, the stocks of both infrastructure and non-infrastructure capitals can be estimated by the Perpetual Inventory Method (PIM). Thus, we have

$$K_t = (1 - \delta_{t-1}) K_{t-1} + I_{t-1}$$

where $K_t$ denotes the capital stock at the beginning of period $t$. It is important to explicitly distinguish the two timing notations of $K_t$ in literature where $K_t$ could be defined as the stock at the beginning or end of period $t$. Unexpectedly, we find some literature confuses this and is actually in mistake. PIM requires the data of 4 elements: initial (base period) capital stock, capital investment time series, capital price, and capital depreciation profile. The capital investment and capital price data are documented in China Statistical Yearbooks. We select the year 1993 as the base year to calculate the capital stock at constant price. Starting from an initialized stock at the beginning of 1993, the capital stock series in subsequent years can be calculated from the investment data. The initial capital stock in 1993 can be estimated in several ways. We calculate by using the capital investment growth rate and depreciation rate in the following years, assuming they were the same in previous years. In formula, we have $K_{1993} = I_{1993}/(gr + \delta)$ where $K_{1993}$ is the capital stock in 1993, $I_{1993}$ is the capital investment in 1993, $gr$ is the assumed growth rate of capital investment before 1993 (for which we use the country-wide average 1993-1997 growth rate 24% for infrastructure and 11% for non-infrastructure), and $\delta$ is the corresponding depreciation rate.

4.2 Data source and sample

We rely on the official data supplied by National Bureau of Statistics (NBS) of China. The data on the capital, GDP and price level are from China Statistical Yearbooks 1994-2014. The data on
the population and labor are from *China Population Statistics Yearbooks 1994-2006* and *China Population and Employment Statistics Yearbooks 2007-2014*. Most of the data in *China Statistical Yearbooks* are also readily downloadable from NBS’ website.

### 4.2.1 Capital

The infrastructure includes 4 industrial sectors: (i) production and supply of electricity, gas and water, (ii) transport, storage and post, (iii) information, transmission, computer service and software, (iv) management of water conservancy, environment, and public facilities. Since the residential capital and investment in human capital (including investment in education and public health) are quite different, we follow SH and rule out them from both infrastructure and non-infrastructure. Thus, the non-infrastructure capital is calculated by total physical capital minus the abovementioned three types of capitals.

In *China Statistical Yearbooks*, there are two types of capital investment series that can be used to construct capital stock data. One is “*Gross Fixed Capital Formation*”, and another one is “*Total Investment in Fixed Assets*”. Xu (2010, 2013) tells the main differences between these two statistics calculated by NBS. Usually in literature the former is used to estimate the total capital stock for all industries as a whole (e.g. Zhang, Wu and Zhang, 2004 for China, Kamps, 2006 for 22 OECD countries). However, the data of “*Gross Fixed Capital Formation*” for each specific industry in China, which we need, is unavailable. Therefore we have to use the data of “*Total Investment in Fixed Assets*”. Figure 2 shows the average ratio of “*Gross Fixed Capital Formation*” to “*Total Investment in Fixed Assets*” over all industries for eastern, central and western areas during year 1993-2013. Roughly speaking, during the two decades the ratio was not too far from 1. A simple average ratio over all provinces over the 21 years is 0.9920. But interestingly, after 2003 this ratio significantly decreased from a level of slightly over 1 to around 0.8. We suppose that the decreasing trend applies to both infrastructure and non-infrastructure industries.

Recall that we have a varying coefficient econometric model $\Delta \ln y_t = \beta_g(z_t)\Delta \ln k_{g,t} + \beta_k(z_t)\Delta \ln k_t + OTs$ where $OTs$ refers to some “other terms”. Suppose currently that $\Delta \ln k_{g,t}$ and $\Delta \ln k_t$ are constructed by “*Total Investment in Fixed Assets*”. If we use “*Gross Fixed Capital Formation*” to derive capital stocks $\ln k^*_{g,t}$ and $\ln k^*_t$, we should generally have $\Delta \ln k^*_{g,t} < \Delta \ln k_{g,t}$ and $\Delta \ln k^*_t < \Delta \ln k_t$ after 2003 which indicates that the estimated marginal product of capital could be larger than that estimated by using “*Total Investment in Fixed Assets*” data. Unfortunately we cannot know how large the difference exactly is because of no data. In our application, we

---

\[\text{If the concept of “infrastructure” is more widely defined as in much literature, the “infrastructure” in this present paper is just “economic infrastructure” which is “a direct input to economic activity”. And the fixed capital for providing social services (such as school, hospital and perhaps also residential house) should be sorted into another part of the concept – “social infrastructure”. Moreover, not all capitals in the 4 infrastructural industries necessarily work as the “infrastructure”. But the data at hand does not allow us to distinguish the different capital components within the industries. Hence, it is important to take into account the possibly distinct definitions of “infrastructure” when comparing our results with other papers.}\]
estimate the capital stocks by using an artificial investment series. This series is obtained by “Total Investment in Fixed Assets” within infrastructure or non-infrastructure sector multiplying by “total capital formation ratio”. The “total capital formation ratio” is just the total “Gross Fixed Capital Formation” in all industries divided by the “Total Investment in Fixed Assets” in the whole economy. Thus Figure 2 in fact displays the “total capital formation ratio” in the three regions.

Figure 2: 

Public investment efficiency is also an important issue for estimating public capital stock. There is no doubt that not all public investment creates public capital. Arestoff and Hurlin (2006), Gupta et al. (2014), Hurlin and Arestoff (2010) and Pritchett (1996) estimate that only roughly “half of the money invested in investment projects will have a positive impact on the public capital stocks in the developing countries”. However we do not know the public investment efficiency in Chinese provinces. Fortunately in Section 2.2.2 we show that in the special case of constant capital investment efficiency, which is in reality likely especially when we only focus on a period of one decade, the unknown investment efficiency parameter would not greatly change our regression result.

NBS reports the price indices for investment in (i) “Construction and Installation”, (ii) “Purchase of Equipment and Instruments”, (iii) “Others”, and calculates the “Price Index for Investment in Fixed Assets” as a weighted average of these three components. But we cannot directly use these price indices because the exact components of infrastructure and non-infrastructure are unknown. Since infrastructure capital mostly consists of structure and buildings, we set the price of infrastructure equal to the “Price Index for Investment in Construction and Installation” reported by NBS. And we assume that “Price Index for Investment in Fixed Assets” is equal to a weighted average of infrastructure and non-infrastructure prices, taking the corresponding current price investment volumes as weights. Then we can calculate the price index for non-infrastructure. We find that in almost all provinces the price of infrastructure inflates faster than non-infrastructure. Figure
3 exhibits the ratios of infrastructure and non-infrastructure capital prices in eastern, central and western districts during the sample period. We know from the graph that the divergence of the ratios in three regions mostly happened in 1990s. In the 21st century, the ratios evolved in a similar way.

![Graph showing the evolution of average Pg/Pk ratios for Eastern (red), Central (black) and Western (blue) provinces, 1993-2013](image)

**Figure 3:** Average ratios of infrastructure and non-infrastructure capital prices for eastern (red), central (black) and western (blue) provinces, 1993-2013

The most debatable variable is the depreciation rate of capital. Wu (2009) briefly investigates the literature and finds that “different rates of depreciation have been used, ranging from 3.6 to 17.0 per cent”. After comparing the depreciation rates used in previous literature, we set the depreciation rates of public and private capitals as constant at 0.06 and 0.12, respectively. We think these rates are reasonable because of two arguments. (i) With these rates, the combination of infrastructure and non-infrastructure capital would give a depreciation rate of total physical capital close to 0.10 which is widely used for Chinese researches (e.g. 0.092 in Jin, 2012; 0.096 in Zhang, Wu and Zhang, 2004). SH use the depreciation rates of 0.08 and 0.16. But we think those rates are probably too high, especially for the non-infrastructure capital. We expect that the infrastructure consists mostly of structure. For the structure, Bai, Hsieh and Qian (2006), Zhang, Wu and Zhang (2004) set the depreciation rate of 8%, 6.9%, respectively. In contrast, the depreciation rates of all capital as a whole in Xu, Zhou and Yuan (2007) and Wu (2009) are 3% and 4.2% respectively, but these values are quite probably too low considering the very high development of Chinese economy. (ii) Then, we have our value of \((\delta_k - \delta_g)\) at 0.06 which is comparable to the size of \(\delta_g\). We discussed in Section 2.2.2 that, both from China-specific and international literature, it is commonly accepted that the values of \((\delta_k - \delta_g)\) and \(\delta_g\) are close. We also alter our depreciation rates and redo our empirical procedure, and the (not much different) results are reported in the robustness check in Section 5.3.
4.2.2 Labor

Three different variables are usually used as aggregate labor supply – population, labor amount, and skill-adjusted labor amount. We follow Barro and Lee (2010) to calculate the skill-adjusted labor amount. The formula is 

\[ L_t = \text{LAB}_t \times \text{ADJ}_t = (\text{POP}_t \times \text{LABRATIO}_t) \times \text{ADJ}_t \]

where \( L_t \) is the amount of skill-adjusted labor, \( \text{LAB}_t \) is raw labor amount, \( \text{POP}_t \) is population, \( \text{LABRATIO}_t \) is the ratio of labor in population which is represented by fraction of people aged 15-64. \( \text{ADJ}_t \) is the coefficient used to adjust the labor according to the skill level. For \( \text{POP}_t \) we do not use the population data reported in China Statistical Yearbooks directly. Instead, we derive it by dividing nominal GDP by nominal GDP per capita as suggested in Li and Gibson (2013). The data of \( \text{LABRATIO}_t \) is easily calculated from China Population Statistics Yearbooks and China Population and Employment Statistics Yearbooks. Since there are several unusual jumps in the \( \text{LABRATIO}_t \) time series, which suddenly return to the original trend in the following years and are obviously the result of sample error, we smooth the series by HP filter. We do not multiply the labor amount by employment rate because the reported employment rate is often not regarded reliable.

It is assumed that \( \text{ADJ}_t = \varphi(s_t) \) where \( s_t \) is a proxy for human capital. As in Barro and Lee (2010), we use years of schooling of the worker \( s_t \) to proxy human capital, and we further assume that \( \varphi(s_t) = e^{\theta s_t} \). The parameter \( \theta \) can be calculated from \( RS = LS \times \theta \) where \( RS \) is the marginal rate-of-return to schooling and \( LS \) is share of labor in total output.

Chen and Hamori (2009), Zhu (2011) report the RS of 7%-8%, 9%-10% in China in recent years. This magnitude is also supported by a series of researches such as Ren and Miller (2012), Zhang et al. (2005). Hence, we set \( RS=0.075 \). Admittedly, some researchers (e.g. Li, Liu and Zhang (2012)) report a lower level of \( RS \) around or even below 5%. We check the case of lower \( RS \) in our robustness check section.

We also need to set the value of labor share parameter \( LS \). It is widely found that the labor share in China declined in recent decade, and is around 40%. Chi and Qian (2013), Karabarbounis and Neiman (2014), Qian and Zhu (2012) among others discuss the associated evidences. Following their findings, we set \( LS=0.4 \). Therefore we have \( \theta=RS/LS=0.1875 \). Of course it can be argued that the \( \theta \) could be heterogeneous for different provinces, but unfortunately we have no data about that and can only set it country-widely constant.

Figure 4 shows the average value of labor-skill-adjustment coefficients \( \text{LABRATIO}_t \times \text{ADJ}_t \) and skill-adjustment coefficients \( \text{ADJ}_t \) for eastern, central and western provinces. We see that the labor markets in the three districts are obviously different. We see that both the labor-skill-adjustment and skill-adjustment coefficients in eastern provinces are highest. This is a result of both highest labor proportion and most schooling years. The western provinces, with the lowest level of labor ratios and schooling, have smallest adjustment coefficients.
4.2.3 Sample

We classify the 31 provinces in mainland China into eastern, central and western parts, by following the official classification of the NBS. Therefore we have 11 eastern provinces: Beijing, Fujian, Guangdong, Hainan, Hebei, Jiangsu, Liaoning, Shandong, Shanghai, Tianjin, Zhejiang; 8 central: Anhui, Heilongjiang, Henan, Hubei, Hunan, Jiangxi, Jilin, Shanxi; and 12 western: Chongqing, Gansu, Guangxi, Guizhou, Inner Mongolia, Ningxia, Qinghai, Shaanxi, Sichuan, Tibet, Xinjiang and Yunan. We find that Tibet is indeed an outlier since it (i) has very high infrastructure ratio, (ii) has a great proportion of data missing, and (iii) is with a very small economy size. Hence we do not include Tibet in our regression. So we have a panel of 30 Chinese provinces. The data for Chongqing before 1997 is calculated from Sichuan data.

The NBS changed the statistical system in 2003, which makes the data before 2003 actually uncomparable. On the other hand, as we start our estimation of capital stocks from 1993, we think that the influence of estimate errors for base year capital stock should sufficiently vanish 10 years later i.e. 2003. Hence we focus on the year 2003-2013 data. In the robustness check section, we also look at the 1998-2013 and 2000-2013 data.

4.3 Estimated capital stocks

Employing the method and data abovementioned, we are able to estimate the infrastructure and non-infrastructure capital stocks for 31 provinces in mainland China.

According to Gupta et al. (2014)’s estimation for the public capital stocks in a group of countries, the mean (non-efficiency-adjusted) public capital stock in middle-income countries during 2000-2009 is 93.2% of GDP. This number is not very far from our estimation. Figure 5 and Figure 6 show
the average $K/Y$, $K_g/Y$ and $K_g/K$ ratios for eastern, central, western regions and the country as a whole during 1993-2013. We know from the graph that $K/Y$ rates were relatively stable before 2004 and obviously rose after then. $K_g/Y$ ratios smoothly increased in three regions during the whole sample period, and the gaps among three regions aggravated in the 21st century. $K_g/K$ ratios climbed up in the first decade and went down after the peaks around 2006.

We see that for $K/Y$ and $K_g/Y$ ratios, there are breaks around year 2003. The breaks are partly the consequences of NBS’ statistical system reform at that year. So it is not very reasonable to pool pre-2003 and post-2003 data in a regression analysis without robust way to deal with the break.

Figure 5: Average $K/Y$ and $K_g/Y$ ratios for eastern, central and western provinces, 1993-2013
Figure 6: Average $K_g/K$ ratios for eastern, central and western provinces, 1993-2013

Figure 14 in Appendix F additionally shows the evolution of $K$, $K_g$, $Y$ shares (as the fraction of country-wide sum) of the three districts. It is obvious that the output shares are stable for all the three regions, while the shares of infrastructure and non-infrastructure capital stocks demonstrate some variations especially after 2003. At the beginning of the sample period, $K_g$ and $Y$ shares were very close to each other in all the three areas. Then except in the central region, $K_g$ and $Y$ deviate from each other gradually. In 2013, the share of $K_g$ was already nearly 10% higher (lower) than the share of $Y$ in western (eastern) zone. One may guess this fact implies that the infrastructure in western provinces is less productive since an increase of $K_g$’s proportion does not go with an increase of $Y$’s. Howbeit, we in fact observe that the share of non-infrastructure $K$ increased as well at a comparable pace. Hence, we need a rigorous comparison about the relative contributions of $K$ and $K_g$ to output. The next section presents our regression results.

5 Regression Result

5.1 Output elasticity

We summarize our estimation of coefficients in Table 3. For comparison, we also report the results for pooled OLS estimation, one-way fixed effects model and median values of estimation for varying coefficient model without spillover and fixed effects. For the complete varying coefficient model with spillover and fixed effects, we report the values of 5%, 25%, 50%, 75% and 95% quantile
over the entire data panel.\textsuperscript{13} The significance level of estimated coefficients is evaluated by the corresponding estimation standard errors which can be obtained by residual based wild bootstrap method, analogous to that in Li, Hsiao and Zinn (2003), Khan and Salim (2015). Result for a monotonicity test for proxy variable with respect to perceived TFP shock can be found in Appendix G.

Table 3: *The Estimated Output Elasticities of Infrastructure and Non-Infrastructure Capitals*

<table>
<thead>
<tr>
<th>Coef.</th>
<th>Pooled OLS</th>
<th>One-Way Fixed Effects Model</th>
<th>without Spillover and Fixed Effects</th>
<th>with Spillover and Fixed Effects</th>
<th>Semiparametric Varying Coefficient Model</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Median</td>
<td>Median</td>
<td>5% Quantile</td>
<td>25% Quantile</td>
<td>75% Quantile</td>
</tr>
<tr>
<td>$\beta_g$</td>
<td>0.192**</td>
<td>0.159**</td>
<td>0.199**</td>
<td>0.173**</td>
<td>0.116**</td>
</tr>
<tr>
<td></td>
<td>[0.026]</td>
<td>[0.025]</td>
<td>[0.041]</td>
<td>[0.026]</td>
<td>[0.041]</td>
</tr>
<tr>
<td>$\beta_k$</td>
<td>0.182**</td>
<td>0.221**</td>
<td>0.154**</td>
<td>0.178**</td>
<td>0.161**</td>
</tr>
<tr>
<td></td>
<td>[0.020]</td>
<td>[0.020]</td>
<td>[0.022]</td>
<td>[0.022]</td>
<td>[0.045]</td>
</tr>
<tr>
<td>$\varphi_1$</td>
<td>-0.254**</td>
<td>-0.135</td>
<td>-0.344*</td>
<td>-0.311**</td>
<td>-0.253**</td>
</tr>
<tr>
<td></td>
<td>[0.065]</td>
<td>[0.069]</td>
<td>[0.099]</td>
<td>[0.084]</td>
<td>[0.085]</td>
</tr>
<tr>
<td>$\varphi_2$</td>
<td>0.069**</td>
<td>0.070**</td>
<td>0.069**</td>
<td>0.069**</td>
<td>0.044**</td>
</tr>
<tr>
<td></td>
<td>[0.011]</td>
<td>[0.011]</td>
<td>[0.019]</td>
<td>[0.019]</td>
<td>[0.022]</td>
</tr>
<tr>
<td>$\theta_g$</td>
<td>0.117*</td>
<td>-0.158*</td>
<td>No</td>
<td>0.083</td>
<td>No</td>
</tr>
<tr>
<td></td>
<td>[0.056]</td>
<td>[0.071]</td>
<td>[0.058]</td>
<td>[0.058]</td>
<td>[0.058]</td>
</tr>
<tr>
<td>$\theta_k$</td>
<td>-0.143*</td>
<td>0.087</td>
<td>No</td>
<td>-0.107*</td>
<td>No</td>
</tr>
<tr>
<td></td>
<td>[0.052]</td>
<td>[0.061]</td>
<td>[0.051]</td>
<td>[0.051]</td>
<td>[0.058]</td>
</tr>
<tr>
<td>$\sigma_t$</td>
<td>-</td>
<td>Yes</td>
<td>No</td>
<td>No</td>
<td>Yes</td>
</tr>
<tr>
<td>$R^2$</td>
<td>0.624</td>
<td>0.670</td>
<td>0.691</td>
<td>0.696</td>
<td>0.727</td>
</tr>
</tbody>
</table>

Note: * $p<0.05$, ** $p<0.01$; (bootstrapped) standard errors in brackets.

The estimated coefficients are located within the range in line with literature. The coefficients for infrastructure, non-infrastructure capital and residential investment are significant in all cases. The coefficient of proxy variable is sometimes not significant. But since the proxy is just an auxiliary to help mitigate reverse causality, the degree of its significance does not matter much. The spillover effect is generally not very significant. This is not surprising since we are working on the data of Chinese provinces with usually very wide area, in which most capitals are located far from the provincial boundary and thus have little connection to other provinces.

We check the cross-section correlation, auto-correlation (by Box-Pierce and Ljung-Box tests),

\textsuperscript{13}In Table 3 the quantile of each coefficient is reported separately. Thus, for instance, a data point with $\beta_g$ at 25% quantile is not necessarily with $\beta_k$ at 25% quantile. That is also the case in Table 4.
normality (by Shapiro-Wilk test) and heteroscedasticity of the residuals. The residuals typically well behave and should not bias our estimation. The $R^2$ value for our semiparametric model is not very high (but already significantly higher than OLS and one-way fixed effects model). In fact we find that if we arbitrarily add the amount of dummies (especially for the provincial fixed effects) or put more variables in the part $z_t$, the value could be increased to 0.8 or even 0.9. But we do not do that to increase $R^2$ value since (i) there is no definite economic sense underlying, (ii) after that either key coefficients are mostly not significant or even worse the whole estimate result becomes disordered and uninterpretable. In Section 5.3 robustness check we discuss some of the situations.

5.2 Marginal product

With the estimated output elasticity, ratio of capital to output and price level, we can calculate the marginal products of infrastructure and non-infrastructure capitals. Table 4 shows the results.

Table 4: The Estimated Marginal Products of Capitals

<table>
<thead>
<tr>
<th>Variable</th>
<th>Estimated Value</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>5%</td>
</tr>
<tr>
<td></td>
<td>Quantile</td>
</tr>
<tr>
<td>$F_g$</td>
<td>0.113</td>
</tr>
<tr>
<td>$F_k$</td>
<td>0.124</td>
</tr>
<tr>
<td>$(F_g+F_k)$</td>
<td>0.128</td>
</tr>
<tr>
<td>$F_g/F_k$</td>
<td>0.627</td>
</tr>
<tr>
<td>MPdiff</td>
<td>-0.035</td>
</tr>
</tbody>
</table>

Note: $F_{(g+k)}$ is the marginal product of capital as a mix of public and private capital with current capital ratio maintained. And the MPdiff is $(F_g - \delta_g) - (F_k - \delta_k)$.

Averagely (here we actually refer to the “median” value when using the term “average”) speaking, the marginal product of infrastructure capital is very close to that of non-infrastructure, which turns out to a median value of $F_g/F_k$ around one. (We find that this point is robust to some variations of our model setup or sample size.) However, the difference of $F_g$ and $F_k$ net of depreciation is positive averagely indicating an underinvestment of public capital.

Figure 7 shows the marginal products of infrastructure and non-infrastructure capitals, as well as the difference of marginal products net of depreciation rates, for 30 provinces during year 2004-2013 (where 2003 data is implicitly exploited in first-differenced model). And Figure 8 demonstrates the average marginal products for eastern, central and western (excluding Yunnan) provinces, taking the capital investment volumes at each year as weights. These two figures enable us to obtain the main finding of our paper. We exclude Yunnan from Figure 8 because the estimation for this province is not strongly reliable (for which a short discussion is below).
Figure 7: The estimated marginal products of infrastructure and non-infrastructure capitals in 30 provinces, 2004-2013

Note: Blue line is $F_g$; red line is $F_k$; green line is $(F_g - \delta_g) - (F_k - \delta_k)$. E is eastern province; C is central province; W is western province.
In most eastern provinces the marginal product of infrastructure capital net of depreciation is clearly above that of non-infrastructure, which is a signal of infrastructure lack. Exceptions are Fujian, some years for Beijing, Guangdong, Hainan and Shanghai, where the dynamic marginal product difference is negative. However, since the negative value still close to the zero line, we can regard these provinces as holding a roughly balanced allocation of infrastructure and non-infrastructure capitals. For the two provinces of Tianjin and Zhejiang, even though the $F_g$ is very close to $F_k$ the infrastructure is still slightly underinvested because of the existence of depreciation rate difference. The red line in Figure 8 tells us that the average marginal product difference in eastern provinces is always higher than 0.1 over 2004-2013. So, a clear underinvestment of infrastructure exists in the east. Howbeit the great inter-province heterogeneity should not be neglected. For instance, in Jiangsu and Shandong $F_k$ is about half of $F_g$ while in Fujian and Guangdong $F_k$ is always larger than $F_g$.

Some researches (such as SH) argue that there is already too much infrastructure in western provinces. However, our estimation demonstrates no evidence of significant overinvestment of infrastructure in most western provinces. For Chongqing, Gansu, Shaanxi, and Xinjiang, $(F_g - \delta_g) - (F_k - \delta_k)$ is close to 0 for the whole period. That is to say, a good match of infrastructure and non-infrastructure capital was roughly maintained. In contrast, for several provinces like Guangxi, Inner Mongolia, Ningxia, Sichuan, the infrastructure was apparently underinvested in the early years when the $K_g/Y$ ratio was low. And the value of $(F_g - \delta_g) - (F_k - \delta_k)$ converged to 0 after 2010. Yunnan province is an outlier in the semiparametric regression because its $z_t$ value was much higher than all other provinces (except Tibet which is not in the sample). The kernel function works not well for the sparsely and isolatedly located sample points. Consequently the result for Yunnan provides actually little information. If we set interval [-0.05, 0.05] of $(F_g - \delta_g) - (F_k - \delta_k)$
as the indicator for a balanced match of infrastructure and non-infrastructure capitals\textsuperscript{14}, we find from Figure 8 that the western region average value was located in this interval after 2010. Thus we think that the current infrastructure investment policy in the west does not result in an overstock of infrastructure.

The situation of central provinces is more or less a mix of situations of eastern and western ones, and we do not analyze in detail here. We see from Figure 8 that the average marginal product difference in central region was closer to that of the west in early years, and then converged to the eastern provinces gradually.

5.2.1 Some supplementary supports for our results

A potential criticism on our inference of underinvestment of infrastructure by comparing the marginal products of capitals is that, we fail to fully consider the dynamic interaction of infrastructure and non-infrastructure investment. Public investment influences the private investment mostly through two channels: first, the public investment could alter the productivities of relevant industries and thus change the profitability of private investment; second, since public investment also needs funds, it usually directly changes the financial conditions of private investment either by taxation or competing in capital market. In the previous part of this paper we have already considered the marginal products of capitals and tax. But we assume the absence of financial friction. Within this setup, an increase of one unit of money in public investment accompanies the decrease of the exact amount of resource available for private investment (or consumption, of course). But in reality when government competes with the private sector in the capital market and tightens the budget of private investment, public investment tends to crowd out private investment, as long as the improvement of productivity induced by more public capital stock is not large enough to offset the deterioration of financial conditions. In China, more than 80% of infrastructure investment is from governments or state-owned firms. And it is widely observed that banks in China rank the government-relevant investment as priority when making loan decisions. So, it is serious to ask whether infrastructure investment crowds out non-infrastructure investment in China? If the answer is true, our previous conclusion would be suspect.

However, the empirical evidences support us. (i) The strongest support comes from the results of SVAR analyses which directly find a net crowding-in effect of public infrastructure investment.

\textsuperscript{14}Because of the inter-province heterogeneities and the time series properties of the data, it is actually difficult to have a robust way to test the significance of difference between \((F_g - \delta_g) - (F_k - \delta_k)\) and 0. The standard deviation of region-weighted-average (with all provinces but Yunnan) marginal product difference \((F_g - \delta_g) - (F_k - \delta_k)\) is 0.035. The corresponding mean value is 0.082. Hence, assuming the marginal product difference is normally distributed, the level of 0.03 (roughly equates the mean minus 1.5 standard deviation, for one-sided test) is the benchmark of “significant difference” at approximately 95% probability. Considering the possible errors from econometric estimation, we extend the benchmark level to 0.05 (somehow arbitrarily, nevertheless).

It is also possible to test the significance for each province. The standard deviation of \((F_g - \delta_g) - (F_k - \delta_k)\) over 290 (also without Yunnan) sample points is 0.079. And the mean value is 0.072. But it is seemingly meaningless to pool all provinces and assume they are from the same sample. We hence do not report statistics taking individual province separately.
Using the country-wide annual data of 1980-2011, Xu and Yan (2014) suggest that government investment in public goods and infrastructure in China crowds in private investment significantly. Guo, Liu and Ma (2015) have similar finding based on an annual county-level dataset for approximately 1800 Chinese counties over 2001-2009. Using quarterly data during 1995Q1–2009Q2, Hur, Mallick and Park (2014) find an increase of private investment both in impact and long-run as a response to expansionary government expenditure shocks in China. (ii) The indirect evidences concern a significant productivity enhancement in private sector by investing infrastructure. For instance, Zhang, Wang and Chen (2013) based an intertemporal dynamic computable general equilibrium (CGE) model, show that higher infrastructure investment (financed by either foreign borrowing or production tax) substantially raises productivity in all sectors and income in all household categories. Wu (2008), using stochastic frontier approach, finds that capital efficiency (for all industries as a whole) is affected positively by the level of infrastructure development. (iii) We could also note the fact that the infrastructure investmen in many less developed Chinese western provinces, where the infrastructure investment rate is relatively high while the private industries are less flourishing compared to the eastern counterparts, is largely from the fiscal transfers by central government (with money mostly originates from the eastern regions). This kind of infrastructure investment should have no direct influence on the financial condition of, and thus should not crowd out local private investment.

Another doubt on our finding perhaps comes from the impression of low efficiency of state-owned enterprises (SOEs) in China. In recent years, more than 80% of investment of infrastructure industries comes from governments or state-owned firms. It is not unreasonable to guess that the relatively less efficiency of SOEs possibly results in a low return of infrastructure investment, and thus contradicts our previous finds. But in contrast, we actually also find firm-level evidence on our side. Ding, Guariglia and Knight (2012) investigate 2000-2007 data\footnote{Using the data for more recent years could provide stronger support. But unfortunately, as far as we know the corresponding dataset does not provide new data after 2009. In the dataset, only two (i.e. electronic and transportation) of our previously defined infrastructure industries are documented. So we do not discuss another two (i.e. telecommunication and water) infrastructural industries.} of over 100,000 firms. Amongst all 10 industries, they find the 2 infrastructural sectors – electronic equipment industry and transport equipment industry – have the highest values of investment rate and “very high” return of capital (measured by both average and marginal revenue product of capital). Although these two industries are mainly invested by governments or SOEs, the inter-industrial efficiency difference crucially outweighs the average efficiency gap between SOEs and other firms . (Admittedly, the efficiency disparity across industries might be partially a result of market distortion – some economists warn that the government intervention maintains the production cost (accounting rate of return) of several industries at an artificially low (high) level. But inspecting this issue in detail is far beyond the scope of our paper. Anyhow, at least we see that infrastructure investment indeed owns a high marginal return of capital.)
5.3 Robustness analysis

Nonparametric model is usually very flexible. We check the robustness of estimated coefficients in different model setups and by different samples, and find no crucial change of the result in most cases.

(a) Without proxy and house investment variables

Column (1), (2), (3) in Table 5 show the regression result of excluding both of proxy and house investment from the explanatory variables, or either of them. The results are similar to our baseline regression.

(b) Proxy variables

Previously we only use the first order polynomial of proxy variable \( \phi_1(z_t) \Delta proxy_t \) to get \( \Delta \omega_t \). Column (4) in Table 5 presents the result of using the third order polynomial \( \phi_1(z_t) \Delta proxy_t + \phi_2(z_t) (\Delta proxy_t)^2 + \phi_3(z_t) (\Delta proxy_t)^3 \), whose difference is negligible.

(c) More \( z_t \) variables

In our baseline model, we assume the estimated coefficients \( \beta_t \)'s are a function of variable \( z_t \). It is possible that they are functions of several variables. Then we consider to add more variables in the part of \( z_t \). We try to change the model and assume \( \beta_t \)'s are functions of \( (z_t, r_t, (K_{g,t} + K_t)/Y_t) \). Here \( r_t = K_{g1,t}/K_{g2,t} \) is the degree of match between transportation and telecommunication infrastructure, and energy and water infrastructure. The reason why we divide infrastructure into these two groups is clear in Section 5.4. And, \( (K_{g,t} + K_t)/Y_t \) is the ratio of total (infrastructure plus non-infrastructure) capital to GDP. We report the results in column (5) and (6) of Table 5. Adding more \( z_t \) variables greatly improves the \( R^2 \) value but introduces a severe problem. Since we only have 300 observations (10 years times 30 provinces), the sample points are sparsely distributed in the two dimensional space of \( (z_t, r_t) \) and three dimensional space of \( (z_t, r_t, (K_{g,t} + K_t)/Y_t) \). This severely damages the power of our kernel function and results in very volatile estimates.

(d) Location fixed effects dummies

It is possible that the location can have some fixed effects on the aggregate production function. Even though first differencing should eliminate the section fixed effects, there are possibly some location features that vary constantly and slightly in each year and still appear in our first differenced equation. We try to add the dummy variables for location, either for the 30 provinces or 8 economic regions (Northeast, North Coastal, Eastern Coastal, South Coastal, the Middle Yellow River, Middle Reaches of the Yangtze River, Southwest, and Big Northwest regions) according the
NBS. But we find that the results are quite implausible. The location fixed effects are very large and count for a substantial part (for example 25%) of $\Delta \ln y_t$. So we rule out the models with the location fixed effects dummies and do not report the results.

Table 5: The Robustness Check for Estimated Coefficients in Semiparametric Varying Coefficient Model, Different Models

<table>
<thead>
<tr>
<th>Coefficient</th>
<th>(0) baseline model</th>
<th>(1) without pr.&amp;hous.</th>
<th>(2) without proxy</th>
<th>(3) without house</th>
<th>proxy</th>
<th>$z + r$</th>
<th>$z + r + \frac{K_g + K}{Y}$</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\beta_g$</td>
<td>0.147</td>
<td>0.157</td>
<td>0.144</td>
<td>0.156</td>
<td>0.150</td>
<td>0.138</td>
<td>0.119</td>
</tr>
<tr>
<td>$\beta_k$</td>
<td>0.220</td>
<td>0.240</td>
<td>0.223</td>
<td>0.239</td>
<td>0.223</td>
<td>0.231</td>
<td>0.247</td>
</tr>
<tr>
<td>$\varphi_1$</td>
<td>-0.221</td>
<td>-</td>
<td>-</td>
<td>-0.253</td>
<td>-0.187</td>
<td>-0.238</td>
<td>-0.086</td>
</tr>
<tr>
<td>$\varphi_2$</td>
<td>0.075</td>
<td>-</td>
<td>0.078</td>
<td>-</td>
<td>0.075</td>
<td>0.053</td>
<td>0.049</td>
</tr>
<tr>
<td>$\theta_g$</td>
<td>-0.161</td>
<td>-0.249</td>
<td>-0.164</td>
<td>-0.235</td>
<td>-0.163</td>
<td>-0.225</td>
<td>-0.408</td>
</tr>
<tr>
<td>$\theta_k$</td>
<td>0.088</td>
<td>0.150</td>
<td>0.084</td>
<td>0.145</td>
<td>0.098</td>
<td>0.135</td>
<td>0.203</td>
</tr>
<tr>
<td>$\sigma_t$</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
</tr>
<tr>
<td>$R^2$</td>
<td>0.727</td>
<td>0.657</td>
<td>0.728</td>
<td>0.666</td>
<td>0.647</td>
<td>0.832</td>
<td>0.935</td>
</tr>
</tbody>
</table>

Note: all the reported values for varying coefficients are in the median.

(e) Alternatively estimated capital stock

There are several alternative ways to estimate the capital stock, depending the assumptions underlying the estimation. (i) First, we might change the depreciation rate. We check the results of setting $(\delta_g, \delta_k) = (0.07, 0.12), (0.08, 0.11), (0.08, 0.14)$. The column (1), (2), (3) of Table 6 display the results. (ii) We might define the non-infrastructure as including the residential capital, compared to our baseline case without residential capital. The estimation result is just analogous. (iii) We also change the base year by which we calculate the constant price capital stock. We choose the year to 2003, 2013 and find no much difference.

(f) Sample size

We change our sample period to year 1998-2013 or 2000-2013. The changes are not really meaningful because of the uncomparability of pre- and post-2003 data as a consequence of statistical system change. Anyhow, we report the result in column (4) and (5) in Table 6.
Table 6: The Robustness Check for Estimated Coefficients in Semiparametric Varying Coefficient Model, Different Samples

<table>
<thead>
<tr>
<th>Coefficient</th>
<th>Estimated (Median) Value</th>
<th>(0)</th>
<th>(1)</th>
<th>(2)</th>
<th>(3)</th>
<th>(4)</th>
<th>(5)</th>
</tr>
</thead>
<tbody>
<tr>
<td>baseline</td>
<td>δg=0.07 δk=0.12</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>model</td>
<td>δg=0.08 δk=0.11</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>δg</td>
<td>0.07 0.08 0.08 0.08 0.08 1998-2000-</td>
<td>0.104 0.109</td>
<td>0.135 0.138</td>
<td>0.219 0.218</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>βg</td>
<td>0.147 0.141 0.136 0.135 0.104 0.109</td>
<td>0.147</td>
<td>0.141</td>
<td>0.136</td>
<td>0.135</td>
<td>0.104</td>
<td>0.109</td>
</tr>
<tr>
<td>βk</td>
<td>0.220 0.222 0.229 0.219 0.181 0.188</td>
<td>0.220</td>
<td>0.222</td>
<td>0.229</td>
<td>0.219</td>
<td>0.181</td>
<td>0.188</td>
</tr>
<tr>
<td>ϕ1</td>
<td>-0.221 -0.222 -0.225 -0.213 -0.003 -0.003</td>
<td>-0.221</td>
<td>-0.222</td>
<td>-0.225</td>
<td>-0.213</td>
<td>-0.003</td>
<td>-0.003</td>
</tr>
<tr>
<td>ϕ2</td>
<td>0.075 0.076 0.076 0.074 0.118 0.102</td>
<td>0.075</td>
<td>0.076</td>
<td>0.076</td>
<td>0.074</td>
<td>0.118</td>
<td>0.102</td>
</tr>
<tr>
<td>ϑg</td>
<td>-0.161 -0.149 -0.133 -0.146 -0.077 -0.144</td>
<td>-0.161</td>
<td>-0.149</td>
<td>-0.133</td>
<td>-0.146</td>
<td>-0.077</td>
<td>-0.144</td>
</tr>
<tr>
<td>ϑk</td>
<td>0.088 0.080 0.066 0.081 0.018 0.072</td>
<td>0.088</td>
<td>0.080</td>
<td>0.066</td>
<td>0.081</td>
<td>0.018</td>
<td>0.072</td>
</tr>
<tr>
<td>σt</td>
<td>Yes Yes Yes Yes Yes Yes</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
</tr>
<tr>
<td>R²</td>
<td>0.727 0.727 0.727 0.724 0.631 0.657</td>
<td>0.727</td>
<td>0.727</td>
<td>0.727</td>
<td>0.724</td>
<td>0.631</td>
<td>0.657</td>
</tr>
</tbody>
</table>

Note: all the reported values for varying coefficients are in the median.

5.4 A possible justification for current high ratio of transportation and telecommunication infrastructure in eastern provinces

5.4.1 Dividing different types of infrastructure into two groups

The capital can be measured in physical units or monetary units. When measuring monetarily, it is usual to not distinguish different capital types. However, the heterogeneity among distinct capital types, especially of the infrastructure, might be quite large. Without serious inspection the estimation from a mixture of infrastructure types could misguide economic policy. Considering this, we use the principal component analysis (PCA) to investigate the physical stocks of infrastructure in Chinese provinces. We take 5 different types of physical infrastructure indicators: \( x_1 \) is Electricity Power Generation (100 million \( kwh \)/100 thousand people); \( x_2 \) is Proportion of Internet Users; \( x_3 \) is Length of Highways (10000 \( km \)/10000 \( km^2 \) land area); \( x_4 \) is Length of Railways in Operation (10000 \( km \)/10000 \( km^2 \) land area); \( x_5 \) is Total Amount of Water Supply (100 million \( m^3 \)/million people). We choose them to represent diversified industries: energy, telecommunication, transportation, and sanitation. The basic PCA result is in Table 7. The data is during 2004-2012 (for reason of data availability) and from NBS website except the data of land area. We obtain the land area data from the website of Chinese central government. By including 27 provinces, the sample excludes 4 provinces of Inner Mongolia, Tibet, Qinghai and Xinjiang because their unusually low population density easily makes them outliers. The PCA result is similar if we change the sample period or add some more physical stock indicators into analysis.
Table 7: Eigenvectors for Principal Components in PCA

<table>
<thead>
<tr>
<th>Variable</th>
<th>Prin1</th>
<th>Prin2</th>
<th>Prin3</th>
<th>Prin4</th>
<th>Prin5</th>
</tr>
</thead>
<tbody>
<tr>
<td>x1</td>
<td>0.194</td>
<td>-0.684</td>
<td>0.630</td>
<td>0.302</td>
<td>0.074</td>
</tr>
<tr>
<td>x2</td>
<td>0.550</td>
<td>-0.170</td>
<td>0.026</td>
<td>-0.684</td>
<td>-0.448</td>
</tr>
<tr>
<td>x3</td>
<td>0.578</td>
<td>0.038</td>
<td>-0.196</td>
<td>-0.069</td>
<td>0.788</td>
</tr>
<tr>
<td>x4</td>
<td>0.546</td>
<td>0.162</td>
<td>-0.260</td>
<td>0.660</td>
<td>-0.415</td>
</tr>
<tr>
<td>x5</td>
<td>-0.168</td>
<td>-0.689</td>
<td>-0.704</td>
<td>0.020</td>
<td>-0.017</td>
</tr>
</tbody>
</table>

Cumulative Proportion 47.1% 75.8% 86.0% 93.9% 100%

The first 2 principal components can already explain most of the variance in 5 infrastructure types. We see that the PC1 majorly combines \(x_2\), \(x_3\) and \(x_4\) which are internet, highway and railway. PC2 majorly consists of \(x_1\) and \(x_5\) of electricity and water. The PCA indicates it is reasonable to split infrastructures into two types – \(K_{g1,t}\) for transportation and telecommunication, and \(K_{g2,t}\) for energy and water. We then calculate the ratio of two types of infrastructure \(r_t = K_{g1,t}/K_{g2,t}\), and plot it (at y-axis) vs. infrastructure and non-infrastructure capital ratio \(z_t = K_{g,t}/K_t\) (at x-axis) for each province during year 2003-2013 in Figure 9. In the figure each separate line indicates one province, with the name of the province at the year 2003 sample point.

![Evolution of z, r of Eastern, Central and Western Provinces, 2003-2013](image)

*Note: red lines for eastern, black lines for central and blue lines for western provinces.*

**Figure 9: Evolution of \((z, r)\) of eastern, central and western provinces, 2003-2013**

It is apparent that the eastern and western provinces hold distinct features in terms of \(z_t\) and \(r_t\).
the central provinces seem to be a mixture of eastern and western ones. Particularly the eastern ones have significantly lower infrastructure capital ratios while holding a clearly larger proportion of type 1 infrastructure. A calculation of historical investment flow data finds that, during 1993-2013 and 2003-2013 the annual average proportions of type 1 infrastructure investment (as percentage of total local infrastructure investment, at 1993 constant price) in eastern area are, respectively, 4.8% and 5.7% larger than that in western part. Compared to the central provinces, the numbers are 5.3% and 2.3%. The analyses in the following Section 5.4.2 will show that the distinguishing of two infrastructure types makes sense.

5.4.2 Linear interpolant fitted contour map

We use linear interpolant method to fit the surface and display the contour map for estimated $F_{(g+k)}$ with respect to $z_{1,t} = K_{g1,t}/K_t$ and $z_{2,t} = K_{g2,t}/K_t$. Figure 10 shows the situation. Recalling that the eastern provinces have a higher ratio of $z_{1,t}$ and lower $z_{2,t}$, as demonstrated by using $r_t = K_{g1,t}/K_{g2,t}$ in Figure 9, the contour map provides a possible justification for current high ratio of transportation and telecommunication infrastructure in eastern provinces.

For the eastern provinces, the data points close to the lower-right boundary in Figure 10 ex-
hibit higher marginal product of total capital $F_{(g+k)}$. This means that, keeping the ratio of total infrastructure $z_t$ unaltered, a higher proportion of transportation and telecommunication type composition helps eastern provinces to improve the productivity of capital.

For the western provinces, there is no clear answer about which mixture of two infrastructure types (transportation and telecommunication vs. energy and water) is desirable. Some data points near the coordinate $(z_1, z_2) = (0.5, 0.4)$ have high $F_{(g+k)}$ value, but the sample points are too sparse and provide little information.

The correlation coefficients between $\ln K_{g1,t}$ and $\ln K_{g2,t}$ is close to 1. And even for the first differenced variables $\Delta \ln K_{g1,t}$ and $\Delta \ln K_{g2,t}$, the correlation is much larger than 0.5. This very huge multicollinearity prevents us to definitely measure the productivity of two infrastructure types in our regression model. Anyhow, the above contour map supplies a possible justification for current high ratio of transportation and telecommunication infrastructure in eastern provinces.

6 Conclusion

The analyses in our paper have three main implications. Firstly, the theoretical model of infrastructure investment shows the depreciation rate gap of different capitals should be considered in determining the optimal allocation between capital types. The depreciation rate is widely ignored in literature. But it can be crucial if the marginal products of capitals are low or close to each other. While we have concentrated on the depreciation difference, our simple growth model furthermore also shows the importance of some other aspects such as the investment efficiency (if it is not constant). This confirms the effort of estimating efficiency-adjusted public capital stocks across countries by Gupta et al. (2014). All in all, more elements of the economy and further analysis can build on our simple model and investigate the wedge between marginal products of public capital and private capital.

Secondly, focusing on the objective of regional economic growth, the econometric estimation indicates a space of investing more infrastructure in most Chinese provinces, even though the infrastructure stock has already increased grossly in recent decades. This proposes one path for structural reform of improving capital product and towards internal rebalancing. When recently the people found difficult to maintain the economic growth rate above 7%, some began to reflect whether something is wrong in government-led high investment rate. Howbeit, our empirical finding supports the idea that the key issue facing China is the distribution of capital and how it is used (Purdy and Qiu, 2014). Freestone and Horton (2014) show using a global computable general equilibrium (CGE) model that even assuming in the future 10 years the investment ratio decrease by 10%, if the rise in productivity can be sufficiently large, the growth rate of real GDP can still be around 6%. We think a more efficient allocation between infrastructure and non-infrastructure investment would greatly facilitate the favored improvement of productivity.
As a byproduct of our regression, thirdly, the relatively higher ratio of transportation and telecommunication infrastructure in the eastern provinces can be justified in terms of higher marginal product of capital. This highlights the importance of component of capital investment. Intuitively the sectors that use infrastructure more intensively would benefit more from it. Based on pre-2006 data, Chen and Yao (2011) find that Chinese government infrastructural investment encourages the secondary sector while negatively affecting another two sectors. As an unpleasant result, household consumption declines. This is consistent with the finding of Lee, Syed and Liu (2013) that investments in agriculture and services seem to be superior as a result of their more direct impact on household incomes. Because of data limit, we lack a detailed analysis on how different components of infrastructure can benefit different regions in China. But some general claims can be posted: the investment should satisfy the need of local residents and industries; and since the demand is visibly strong, investing in rural infrastructure would be quite helpful.

Our paper has several main shortcomings. The first is about the short period of available time series sample. Since the NBS reformed its statistical system in 2003, we currently only can work on the consistent data series of about one decade. Considering that we intend to study the 30 (or 31) provinces in mainland China, the sample period is clearly short. However this data limit seems insuperable. What we can do in the future is inspecting the currently available data seriously and improving the utilization of it.

Secondly, in empirical part we did not take the quality of infrastructure into account and assume that each unit of money spent is homogeneous. This concern is related to infrastructure’s investment efficiency as well. It is widely argued that in less developed areas the money is more probably wasted because of worse management or other institutional reasons. If furthermore, the quality and investment efficiency are not constant and in reality vary over time, neglecting this might bring about significant errors. Some authors, such as Démurger (2001) and Francois and Manchin (2013), use the (constructed) physic indicators of capital stocks to control for infrastructure quality and category. But this method is more or less arbitrary, because there exists no general rule for doing this. Or it is possible to use some quality or efficiency index to adjust the raw capital stocks data, as done by Gupta et al. (2014) for 71 countries (but without China). This approach is very promising for our future research.

Thirdly, we failed to consider the region-specific rates of depreciation. The depreciation rate is crucial for us, both in the empirical estimation of capital stocks and the theoretical criterion of evaluating the tradeoff between infrastructure and non-infrastructure capitals. Wu (2009) estimates the different depreciation rates in 31 provinces for 3 sectors (agriculture, manufacturing, and services). But his result is not really helpful for us because we divide the production into two sectors – infrastructure and non-infrastructure. Jia and Zhang (2014), and Wu (2009) find that in

\[16\] This does not contradict our basic conclusion that more infrastructure is still good because, after a more efficient allocation of capital the enhancement of future output might be large enough to offset the temporary decline of consumption, and then improve the life-time household utility.
general the depreciation rate is “high in the more developed regions and low in the less developed regions”. For us, it is important to investigate the depreciation differences \((\delta_k - \delta_g)\) across different provinces. We should do that if we can find a satisfactory way with sufficient data. Also, in the long run we might also need to consider the time-varying depreciation rates. But now it does not really matter for our research since we only focus on a period of just one decade.

Another limitation of our paper is regarding the basic focus of the paper. Our theoretical model and aggregate production function econometric framework only consider the spending aspect of public investment policy, rather than the financing procedure. In fact a shortage of infrastructure does not necessarily imply a need of immediate investment. For instance, a radical development of infrastructure might bring about grave fiscal risk of local governments, which outweighs the production benefit from increased public capital. Moreover, our aggregate production framework essentially takes growth as objective. Many other impacts of public capital construction, such as poverty reduction, regional development convergence, and employment, are crucial as well. Consequently, the policy makers should always investigate the miscellaneous aspects of infrastructure before deciding where and when to invest.
References


Wu, Yanrui. 2009. “China’s Capital Stock by Sector and by Region.” *University of Western Australia, Business School, Discussion Paper 09.02*.


Appendix

Appendix A. Literature reviews on some relevant topics

A.1. Typical arguments for overinvestment (or no overinvestment) of public capital or gross capital as a whole in China

The typical arguments can be grouped into 7 types. (1) The first is using neoclassical aggregate indicator such as capital-to-output ratio as in Gros (2015), difference between Gross Capital Income/GDP and Gross Investment/GDP ratios as in He, Zhang and Shek (2007), or (GDP-Investment)/GDP ratio as in Lee, Syed and Liu (2013). But this approach depends on many strong assumptions and is perhaps not applicable in a transitional economy like China. (2) The second is to predict some “reasonable” investment rate by estimated coefficients from regression based on the data of “comparable” countries, as done in Lee, Syed and Liu (2012). But ignoring the heterogeneity of Chinese economy can make the estimated “good” investment rate misleading. (3) The third approach is to measure the crowding-in and crowding-out effects of public investment on private investment, as in Cavallo and Daude (2011), Xu and Yan (2014). This approach provides some hints whether the overinvestment exists. But not only the amount but also the productivity of investment matter for economic growth. Thus, only the investigation for crowding-out effect of public investment cannot generate clear answer for our question. (4) The fourth calculates the
rates of return in industries, as in Ding, Guariglia and Knight (2012). This is probably not really useful since the rate of return of a certain industry in the market often differs from its social rate of return, especially for public-goods-relevant industries. (5) Another approach employs frontier production function techniques such as data envelopment analysis. This approach, including Qin and Song (2009), Wu (2008) etc., generally finds that the investment efficiency in western and inland provinces is lower than that in eastern coastal areas. But the strict assumptions underlying this approach limit its reliability for policy analysis. And the estimated efficiencies for different capitals are not comparable. (6) Another type of argument directly compares the domestic physical availability of facilities (e.g. the transportation network density, proportion of population accessible to the Internet) with the international counterparts. This, while supplying information about the future development potential of capital in the country, does not take account of the contemporary interaction with other economic sectors. (7) The last one is using some micro-level evidences or impressions, such as the phenomenon of “ghost towns” in some regions or the high-speed train without many passengers. This is often discussed in nonacademic media and influential on shaping public awareness. Howbeit, the fact that some places are rich of investment can occur at the same time while investment is sparse elsewhere, even within the same city or same town. From an aggregate perspective we need some “average” conclusion across different areas and industries.

Ding, Guariglia and Knight (2012) and Lee, Syed and Liu (2012, 2013) present brief literature reviews on the macro- and micro-data based researches for China grounded on (some of) the 7 approaches aforementioned.

A.2. Some recent literature reviews for infrastructure topic

The theoretical researches usually focus on two areas: optimal financing method (e.g. taxation) for infrastructure and the optimal supply (e.g. expenditure) of infrastructure. For the expenditure topic, the literature mostly uses the framework of Ramsey–Cass–Koopmans model, either with the decentralized economy solution or the social planner problem. The frequently cited work of Barro (1990) devised a fundamental endogenous growth model with public capital. Irmen and Kuehnel (2009) and Dissou and Didic (2013) provide helpful literature reviews for the theoretical development since 1990s. In these two review papers, different growth models from literature are incorporated into one consistent framework which evolves according to the change of underlying assumptions. But unfortunately, in order to obtain explicit expression of solution to offer comprehensive perspective, the model has to be subject to some specific function forms and loss of generality. In the domain of empirical researches four approaches are usually relied on: estimating the production function, estimating the cost function, using vector autoregression model, or working on cross-section regression. Romp and de Haan (2007) is often cited because of their comprehensive review on the empirical researches. Dissou and Didic (2013) as well as Duran-Fernandez and Santos (2014) also give the discussions on the empirical literature. So far, most empirical studies
have provided evidence for the importance of public capital on economic activity. However much more effort of offering refined data and robust methodology is called for, since the documented magnitude and characteristics of public capital’s impact vary greatly and depend critically on the data, econometric model and technique used. Estache (2007) is a policy-oriented review article focusing on infrastructure in developing countries. That paper examines a series of important topics including the linkage between infrastructure and growth, effects of infrastructure on poverty reduction, fiscal cost of public capital financing, influence of corruption and so forth, which supplies hints for inspiring research orientations.

Appendix B. Illustrating the importance of depreciation difference by example of a simple Cobb-Douglas production function

We assume a Cobb-Douglas aggregate production function $Y = K^{\alpha_k} K_g^{\alpha_g} L^{1-\alpha_k-\alpha_g}$. We set fixed $L = 1$, and $1 - \alpha_k - \alpha_g = 0.4$ since the estimated labor share in China is roughly 40%. Hence, the production function can be simply written as $Y = F(K, K_g) = K^{\alpha_k} K_g^{\alpha_g}$ with $\alpha_k + \alpha_g = 0.6$.

B.1. Capital resource misallocation in one period

First, we look at a one-period example to see how capital resource would be allocated. Suppose that at the beginning of the period, the economy as a whole is endowed with 1 unit of resource which can be allocated between private and public capital i.e. $K + K_g = 1$. Obviously, using $F_k = F_g$ will

![Figure 11: Capital resource misallocation in one period, with depreciation difference $(\delta_k - \delta_g) = 0.06$](image)

First, we look at a one-period example to see how capital resource would be allocated. Suppose that at the beginning of the period, the economy as a whole is endowed with 1 unit of resource which can be allocated between private and public capital i.e. $K + K_g = 1$. Obviously, using $F_k = F_g$ will
generate \( \frac{K_g}{K} = \frac{\alpha_g}{\alpha_k} \). Also we numerically find the solution \((K, K_g)\) for \( F_k = F_g + (\delta_k - \delta_g) \) satisfying the resource constraint. Figure 11 displays the results in the case of \((\delta_k - \delta_g)=0.06\) for parameter \( \alpha_k \in [0.2, 0.45] \). Particularly we find misusing the formula \( F_k = F_g \) results in averagely 7.36% misallocation of \( K \) capital stock. The misallocation would be 4.79% and 10.1% if the depreciation gap is 0.04 and 0.08.

**B.2. Long-run dynamics and welfare loss**

Now we turn to investigate the long-run dynamics of convergence toward the steady state. Since if we use criterion \( F_k = F_g \) the balanced growth path (BGP) is not sustainable under \((\delta_k - \delta_g) \neq 0\), we look at the situation before the economy reaches its BGP. Specifically, we let the economy start at \((0.5\bar{K}, 0.5\bar{K}_g)\) where \( \bar{K} \) and \( \bar{K}_g \) are the steady state capital stocks. Just for illustrative convenience, we currently use a discrete-time version of endogenous growth model in which welfare is \( \sum_{t=0}^{\infty} \beta^t U(C_t) \). We set quarterly discount rate \( \beta = 0.99 \) and log-utility \( U(C_t) = \ln(C_t) \). The capital stocks are accumulated according to \( K_{t+1} = \omega_k Y_t + (1 - \delta_k) K_t \) and \( K_{g,t+1} = \omega_g Y_t + (1 - \delta_g) K_{g,t} \). Here \( \omega_k,t \) and \( \omega_g,t \) are the fractions of output used for private and public capital accumulation, respectively. Definitely, the proportion for consumption is \( \omega_c,t = 1 - \omega_k,t - \omega_g,t \) i.e. \( C_t = \omega_c,t Y_t \).

It is not difficult to calculate that the welfare maximizing steady state private capital stock is

\[
K^* = \left[ \frac{(1+\theta)\delta_k}{(\alpha_k+\alpha_g)\Theta^{\alpha_g}} \right]^{\frac{1}{\alpha_k+\alpha_g-1}}
\]

where \( \Theta = \frac{\alpha_k}{\alpha_k + \frac{1}{\delta_k} - (1 - \delta_k)} \) and \( \theta = \frac{\delta}{\delta_k} \Theta \). Accordingly, \( K_g^* = \Theta K^* \) and \( Y^* = \Theta^{\alpha_g} (K^*)^{\alpha_k+\alpha_g} \). The steady state fraction \( \omega_k = \delta_k \Theta^{-\alpha_g} (K^*)^{1-\alpha_k-\alpha_g} \) and \( \omega_g = \theta \omega_k \).

Now we would like to see how the economy gradually converges to the steady state from the start point \((0.5\bar{K}, 0.5\bar{K}_g)\). We consider three types of policy rules for resource allocation: (i) always stick to \( \omega_k,t = \omega_k \) and \( \omega_g,t = \omega_g \); (ii) ensure \( F_{k,t+1} = F_{g,t+1} + (\delta_k - \delta_g) \); (iii) target \( F_{k,t+1} = F_{g,t+1} \). In order to maintain the comparability of consumption process, we let \( \omega_{c,t} = \omega_c \) under all three scenarios. We plot the first 80 periods (i.e. 20 years) of the economic evolution in Figure 12. In the graph, we set \( \alpha_k = 0.33 \) and \( \alpha_g = 0.27 \) and the variables \((Y, C, K, K_g)\) are expressed in terms of the proportion of the steady state values.
From the figure, we see the policy (i) and (ii) are almost overlapped while policy (iii) results in a severe deviation from the optimal convergence path. As expected, the policy (iii) results in a large welfare loss. We set different values of \((\delta_k, \delta_g, \delta_k - \delta_g)\) at \((0.11, 0.07, 0.04)\), \((0.12, 0.06, 0.06)\), or \((0.13, 0.05, 0.08)\). Under these parameterizations, the welfare loss of policy (iii) is equivalent to 2.40%, 5.54%, 10.3% permanent consumption reduction, respectively. And the welfare difference between rule (i) and (ii) are completely negligible.

Appendix C. Supplement for Section 2.3

C.1. Argument for (nearly) full congestion \((\xi \text{ close to 1})\) setup of public capital

(1) Property of of setup \(K_g^s = K_g^{(\frac{k}{\bar{K}})}\xi\)

Following a set of literature we assume the public service \(K_g^s = K_g^{(\frac{k}{\bar{K}})}\xi\). For simplicity, we focus on the Cobb-Douglas individual production function \(y = k^{\alpha_k}(K_g^s)^{\alpha_g}\). In equilibrium, the aggregate and individual private capital stocks are related by \(K = Nk\). Thus we have the aggregate production \(Y = Ny = N^{(1-\alpha_k)-\xi\alpha_g}K^{\alpha_k}K_g^{\alpha_g}\). As an example we restrict that \(\alpha_k + \alpha_g = 1\). Then we get \(Y = K^{\alpha_k}K_g^{\alpha_g}N^{1-\xi}\alpha_g\) and accordingly \(\frac{Y}{K} = N^{(1-\xi)}\alpha_g \left(\frac{K_g}{K}\right)^{\alpha_k}\). Usually in real economy,
the value of $N$ (i.e. the amount of economic units e.g. population) is very large. Hence the term $N^{1-\xi}\alpha_g$ is large when $\xi$ is not sufficiently close to 1. On the other hand, we know if we count $Y$, $K$ and $K_g$ in the real world by monetary value, we should have $Y = O(1)$ and $K_g = O(1)$. This fact contradicts the situation that $\frac{Y}{K} = N^{1-\xi}\alpha_g \left(\frac{K_g}{K}\right)^{\alpha_k}$ when $N^{1-\xi}\alpha_g$ is very large (compared to value 1). In other words, the setup of $K_g^{s} = K_g\left(\frac{k}{N}\right)^{\xi}$ naturally only allows the parameterization that $\xi$ is sufficiently close to 1 i.e. (nearly) full congestion.

In order to eliminate the scale effect related to large $N$, Dioikitopoulos and Kalyvitis (2008) set up that $y = k^{\alpha_k} \left(\frac{K_g}{N^{1-\xi}}\right)^{\alpha_g}$. In this case, we get a very convenient aggregate production function $Y = K^{\alpha_k} K_g^{\alpha_g}$. But now the individual production itself is problematic. In equilibrium of $K = Nk$, we have $y = k^{\alpha_k} \left(\frac{K_g N^{1-\xi}}{K^{1-\xi}}\right)^{\alpha_g} = k^{\alpha_k} \left(\frac{K_g}{N}\right)^{\alpha_g}$. The aggregate public capital stock is equally divided by all $N$ agents and each agent only takes the $\frac{K_g}{N}$ piece of government capital into account. This turns out to be the full congestion circumstance regardless of the value of $\xi$, which denies our attempt of using parameter $0 \leq \xi < 1$ to model non-full congestion situation. So, whether eliminating the scale effect or not, we see that letting $\xi$ to be close to 1 is the reasonable setup of congestion model.

(2) Empirical evidence

Full congestion of public capital means that the public service enjoyed by a production unit is proportional to its relative economic scale among all firms. In order to ensure the level of public service available to the individual firm to remain fixed, when individual private capital stock $k$ is unchanged the aggregate public capital $K_g$ should increase in proportion to the aggregate private capital stock $K$. Those infrastructures such as transportation, energy used directly in private production and input and output distribution may have this full congestion characteristics. Also, in population dense area such as big cities the high degree of public service congestion is often observed. Craig (1987) and Edwards (1990) discuss some evidences that the local public goods could be quite congested. Recently, Breunig and Rocaboy (2008) use French data to estimate the effect of municipality size on per capita public expenditure. They find the very large population is associated with a severe congestion of public goods. This finding would be quite probably applicable to China.

Furthermore, the local governments often pay much attention and is willing to incline public policy toward the large firms located in the domain because of their crucial roles of taxpayer, GDP creator and employer. In some developing countries without transparent politics, the economic scale might be one-to-one or even one-to-more convertible to political influence. This generates economic rent and enables the public capital congestion parameter $\xi$ close or even over 1.

C.2. Derivation of $\tau_{max} = \alpha_g$ in illustrative decentralized economy growth-maximizing tax policy example

Plugging $\tau = (r + \delta_g)z^{\alpha_k}$ into $r = (1-\tau)F_k - \delta_k = (1-\tau)\alpha_k z^{\alpha_g} - \delta_k$, after arrangement we express $r$ in terms of $z$: $r = \frac{\alpha_k z^{\alpha_g} - \alpha_k \delta_k z - \delta_k}{1 + \alpha_k z}$. Since we already have $\gamma = \frac{1}{\sigma}(1 - \tau)F_k - \delta_k - \rho = \frac{1}{\sigma}(r - \rho)$,
we now get γ as a function of z: \( \gamma = \frac{1}{\sigma} \left[ \frac{\alpha_k z^{\alpha_g - \alpha_k \delta_k z - \delta_k}}{1 + \alpha_k z} - \rho \right] \). Along the BGP, γ is maximized with respect to z. Thus \( z^{\text{max}} \) is the solution to \( \frac{\partial \gamma}{\partial z} = 0 \) which is equivalent to

\[
\alpha_k^2 z^{\text{max}} - (\delta_k - \delta_g)(z^{\text{max}})^{\alpha_k} - \alpha_g = 0
\]

This equation has no general closed form solution. Howbeit it is sufficient to help us obtain growth maximization tax rate \( \tau^{\text{max}} \).

We can write \( \tau \) in terms of \( z \). Plugging \( r = \frac{\alpha_k z^{\alpha_g - \alpha_k \delta_k z - \delta_k}}{1 + \alpha_k z} \) into \( \tau = (r + \delta_g) z^{\alpha_k} \) we have \( \tau = \frac{\alpha_k z^{\alpha_k} - (\delta_k - \delta_g) z^{\alpha_k}}{1 + \alpha_k z} \). Rearranging this generates

\[
(1 - \tau)\alpha_k z - (\delta_k - \delta_g) z^{\alpha_k} - \tau = 0
\]

This apparently means that \( \tau^{\text{max}} = \alpha_g \) gives the desirable \( z^{\text{max}} \).

**Appendix D. Issue of first differencing to convert data stationary in a varying coefficient framework**

Using first differencing to convert data stationary possibly makes the estimated coefficients uncertainly volatile, depending on the uncontrollable time points based on which the time series is split up. Without rigorous proof, we use a simple example to show the situation when we estimate the (varying) slope parameter \( \beta_t \) from model \( \Delta y_t = \beta_t \Delta x_t + \epsilon_t \). Figure 13 shows a stylized example of real data \( y_t = f(x_t) \) and observed data \( y_t = f(x_t) + \epsilon_t \), and the consequences of taking first difference on the data. Here the real data at one time point is an accumulation of the data in the past periods: \( y_t = y_0 + \Delta y_1 + \Delta y_2 + \ldots + \Delta y_s \). What we are interested in is each segment \( \Delta y_s \). For instance if we are in the case 1, obviously the estimated slope \( \hat{\beta}_t \) would significantly deviate from real slope \( \beta_t \). In contrast, in the case 2 the estimated is very close to the real one. Unfortunately the “exactness” of estimation depends on the points based on which the curve is split up, which are in fact uncertain and uncontrollable. Generally, there are two ways to mitigate that problem. (1) We can use a relatively large sample to cancel out potential influences of uncertain segmentations. In our paper we use a panel data set of 300 samples. This sample size should be already large enough. (2) As shown in the case 3 of the figure, if we use a relatively longer segment, the chance is high that the estimated \( \hat{\beta}_t \) is close to real parameter \( \beta_t \). In Section 3.2.1(d) we state that we would rely on the 3-year average values of variables to run the regression. That actually increases the length of segment. (To see this, we just need to notice that \( \Delta \tilde{x}_s = \tilde{x}_s - \tilde{x}_{s-1} = \frac{1}{3} \sum_{t=s-1}^{s+1} x_t - \frac{1}{3} \sum_{t=s-2}^{s} x_t = \frac{1}{3} (x_{s+1} - x_{s-2}) \) where \( \tilde{x}_s \) refers to the 3-year average value.) In a nutshell, as a whole we think the problem of making the estimated coefficients uncertainly volatile does not matter much in our research.
Appendix E. Misusage of ACF approach in Shi and Huang (2014)

First let us recall how SH use the ACF approach (by using the notation in SH paper and suppressing the first difference operator without loss of generality). It is assumed that the TFP shock ε_{st} can be decomposed into two components: ε_{st} = ω_{st} + µ_{st} where ω_{st} is the part of the TFP shock anticipated in advance (by capital investors but unobservable to econometrician) which affects ln k_{i,st} and ln k_{p,st}, and µ_{st} is the part of the TFP shock unanticipated in advance which is uncorrelated with ln k_{i,st} and ln k_{p,st}. Then SH argues that the observable private consumption per working resident ln c_{st} increases monotonically with the variable ω_{st}. Thus for given (ln k_{i,st}, ln k_{p,st}), ln c_{st} is a monotonic function of ω_{st}: ln c_{st} = φ(ω_{st}, ln k_{i,st}, ln k_{p,st}) which gives ω_{st} = φ^{-1}(ln c_{st}, ln k_{i,st}, ln k_{p,st}). Assuming that ω_{st} follows a first-order Markov process, ω_{st} can be rewritten as ω_{st} = ρ(ω_{st−1}) = ρ [φ^{-1}(ln c_{st−1}, ln k_{i,st−1}, ln k_{p,st−1})] = φ(ln c_{st−1}, ln k_{i,st−1}, ln k_{p,st−1}) where the function ρ governs the first-order Markov process of ω_{st} and function φ is the composition of ρ and φ^{-1}. Ultimately in the regression model the coefficients for ln k_{i,st} and ln k_{p,st} can be identified because the correlated term ω_{st} is proxied by φ(ln c_{st−1}, ln k_{i,st−1}, ln k_{p,st−1}).

Below we show that SH’s application of ACF approach is quite implausible from both the aspects regarding “timing” and “Markov” assumptions (which are both crucial in ACF approach). The original idea of ACF method refers to Ackerberg, Caves and Frazer (2006). Ornaghi and Van Beveren (2011) compares several relevant approaches, including ACF, of using observable...
proxy variables to control for unobservables when estimating productivity.

E.1. Regarding “timing” assumption

In SH, \( \omega_{st} \) is the part of the TFP shock anticipated in advance. Let us use another notation to indicate the nature of “anticipation” and write it as \( E(\omega_{st}|I_{t-1}) \) where \( I_{t-1} \) is the information set at the start of period \( t \). We use \( p_{st} \) to refer to the selected proxy variable which is \( \ln c_{st} \) in SH. Under these notations we have \( p_{st} = \phi_t(E(\omega_{st}|I_{t-1}), k_{i,st}, k_{p,st}) \). Without loss of generality we assume the private capital stock \( k_{p,st} \) has the timing analogous to infrastructure stock \( k_{i,st} \). In period \( t \), the infrastructure and non-infrastructure stock are not predetermined and correlated to the anticipated TFP shock \( E(\omega_{st}|I_{t-1}) \). Without loss of generality we can express that \( k_{i,st} = g_t(E(\omega_{st}|I_{t-1}), Plan_{i,st}) \) and \( k_{p,st} = g_p(E(\omega_{st}|I_{t-1}), Plan_{p,st}) \) where \( Plan \) is some predetermined investment and production plans. In this way we actually have \( p_{st} = \phi_t(E(\omega_{st}|I_{t-1}), k_{i,st}, k_{p,st}) \) and based on historical data of \( p_{st}, k_{i,st}, k_{p,st} \) we can never know \textit{a priori} whether the proxy \( p_{st} \) is really monotonic to \( E(\omega_{st}|I_{t-1}) \). Since we do not have the knowledge about \( g_t(\cdot) \) and \( g_p(\cdot) \), we cannot infer the details of \( Plan_{i,st} \) and \( Plan_{p,st} \) and hence cannot test \textit{a posteriori} whether \( p_{st} \) is a monotonic function of \( E(\omega_{st}|I_{t-1}) \).

Therefore, in SH we in fact can never assure that the selected proxy \( \ln c_{st} \) is monotonic to \( \omega_{st} \).

E.2. Regarding “Markov” assumption

Even though we assume that \( p_{st} \) is a monotonic function of \( E(\omega_{st}|I_{t-1}) \) and hence \( E(\omega_{st}|I_{t-1}) = \phi_t^{-1}(p_{st}, k_{i,st}, k_{p,st}) \) holds, we still have the following problems. Under SH’s assumption that \( \omega_{st} = \rho(\omega_{st-1}) \) we have \( E(\omega_{st}|I_{t-1}) = \rho[E(\omega_{st-1}|I_{t-2})] \) in our notation. This means the agents only form \( E(\cdot|I_t) \) based on \( E(\cdot|I_{t-1}) \) i.e. \( E(\cdot|I_t) = \rho[E(\cdot|I_{t-1})] \) and never adjust the expectation based on the realization of innovation in TFP observed in the last period. This is clearly not a reasonable assumption.

Hence, it can hardly be true that \( \omega_{st} = \rho(\omega_{st-1}) \) as in SH.

E.3. Regarding “timing” assumption

Regardless of the reasonableness of \( \omega_{st} = \rho(\omega_{st-1}) \) assumption, we now check the timing of the variables’ sequence in SH which is:

| Period \( t - 1 \) | \( \Rightarrow y_{st-1} \) | \( \Rightarrow E(\omega_{st}|I_{t-1}) \) | \( \Rightarrow (k_{i,st}, k_{p,st}) \) | \( \Rightarrow p_{st} \) | \( \Rightarrow \omega_{st} \) | \( \Rightarrow y_{st} \) | \( \Rightarrow E(\omega_{st+1}|I_t) \) | \( \Rightarrow \omega_{st+1} \) | Period \( t + 1 \) |
|-------------------|----------------|------------------|----------------|----------------|----------------|----------------|----------------|----------------|----------------|

Under SH’s timing assumption, the TFP shock \( \omega_{st} \) realizes at last, after all other variables except output (otherwise the proxy variable \( \ln c_{st} \) should be a function of realized \( \omega_{st} \) rather than the anticipated \( E(\omega_{st}|I_{t-1}) \)). This timing structure is obviously implausible.
E.4. Regarding “Markov” assumption

Perhaps when SH express “... \( \varpi_{st} \) is the part of the TFP shock anticipated in advance”, they actually want to refer to \( \omega_{st} \) rather than \( E(\omega_{st}|I_{t-1}) \). However, using \( \omega_{st} \) does not improve the plausibility of SH’s application of the ACF approach. Generally, in macroeconomic models with the assumption that TFP shock \( \omega_{st} \) follows a first-order Markov process, we cannot write \( \omega_{st} = \rho(\omega_{st-1}) \).

In contrast, the true formulation of the process is \( \omega_{st} = E(\omega_{st}|I_{t-1}) + \xi_{st} = \tilde{\rho}(\omega_{st-1}) + \xi_{st} \) (by noting that with Markov property the information set \( I_{t-1} \) is \( \omega_{st-1} \)) where \( \xi_{st} \) is the innovation to TFP, which is uncorrelated to \((k_{i,st}, k_{p,st}, p_{st})\).

In order to make SH’s argument working, it is required that \( \omega_{st} = \rho(\omega_{st-1}) \) for some certain function \( \rho(\cdot) \). But undoubtedly \( \omega_{st} \neq \tilde{\rho}(\omega_{st-1}) \) unless \( \xi_{st} = 0 \forall t \) (or likewise, \( \xi_{st} \) is a known non-random function of some known variables) i.e. the whole TFP process is not stochastic. This case is definitely not realistic. Even though we accept that the TFP process can be wholly deterministic, we can never accept what is implied from this determinism. In the deterministic case that \( \xi_{st} = 0 \forall t \), we have \( \omega_{st} = E(\omega_{st}|I_{t-1}) = \rho(\omega_{st-1}) = \rho[\rho(\omega_{st-2})] = ... = \rho^t(\omega_0) \). Thus, (since the form of function \( \rho(\cdot) \) can be estimated by some, for example nonparametric, method) the whole path of TFP shock can be known as long as the initial value \( \omega_0 \) or the value \( \omega_{st-n} \) in a specific period is exactly measured. In this nonrandom circumstance, there is no reason why the econometrician cannot measure \( \omega_0 \) or \( \omega_{st-n} \) and then get \( \omega_{st} \) directly (while the government and private sector know it and adjust capital investment as a response).

Thereupon using the realized TFP shock \( \omega_{st} \) instead of \( E(\omega_{st}|I_{t-1}) \) in SH’s argument does not help it work.

E.5. Short summary

Although we have not strictly proven the infeasibility of the application of ACF approach in SH, we have shown logically that it stood on a set of unconvincing assumptions that should be ruled out. The key problem is that: in order to mitigate the endogeneity bias and estimate the coefficients for \((\ln k_{i,st}, \ln k_{p,st})\) directly, SH attempt to introduce the last period variables \((\ln c_{st-1}, \ln k_{i,st-1}, \ln k_{p,st-1})\) to replace \( \varpi_{st} \) by assuming \( \varpi_{st} = \rho(\varpi_{st-1}) \) – but \( \varpi_{st} = \rho(\varpi_{st-1}) \) does not depict the Markov process of TFP shock in a correct manner. (Even though assuming that SH’s approach is without problem, we will face a difficulty to explain the regression result since the time \( t-1 \) capitals appear in the time \( t \) regression model. Thus, by construction the capitals have one period lag impacts on production which make the definition of marginal product ambiguous.)

In fact, if we tightly follow ACF’s original two-stage regression procedure (rather than SH’s revised approach), it is still possible to get a relatively good estimator. However, we do not do that because the intrinsic characteristics of aggregate production function on which we discuss below.

ACF (and several other relevant) approach was devised initially for firm-level analysis. Since in firm-level production the inputs are often lumpy, it is much easier to find clear timing structure.
and select suitable proxy. But in aggregate economy things differ. (i) At country or province-level the input, output, investment and consumption happen every day. While GDP measures a flow during a period interval, aggregate capitals which vary frequently are only available in stock data at specific time points. So, using the start of year capital stock $K_t$ might underestimate the (average) input really utilized while using the end of year capital stock $K_{t+1}$ (which is used in SH) might overestimate the real (average) input. A weighted average of $K_t$ and $K_{t+1}$ perhaps approximates the amount of input well but we do not know the proper weight. (ii) More importantly, the timing assumption of ACF approach requires that the selected proxy variable should occur after the capital inputs were determined. There exist no such an aggregate variable at all -- the capital inputs actually vary all the year round. Thus, applying AFC method restricts us to think the world in an unjustified way: aggregate capital stock is determined in one specific moment and the proxy variable is produced after that. (iii) Furthermore, (perhaps most crucially) we select to use first differenced data to avoid spurious regression. The timing structure is disordered after first differencing. Considering the reluctance of employing the method in aggregate production model, we in practice do not fully rely on proxy variable to control for unobservable TFP shock. We instead take a 3-year moving average to $Y_t, K_t, K_{g,t}, L_t$ supposing that this is a reasonable way to approximate the “effective” variables and meanwhile cancel out some TFP shocks. We also use one proxy variable in addition. Details of our empirical approach are in Section 3.

Appendix F. Evolution of $K$, $K_g$ and $Y$ shares (as fraction of country-wide sum) of eastern, central and western provinces, 1993-2013

Figure 14: $K_g, K$ and $Y$ shares of eastern, central and western provinces, 1993-2013
A brief discussion of Figure 14 is at the end of Section 4.3.

**Appendix G. Monotonicity test of proxy variable and perceived TFP shock**

Figure 15 tests the assumed monotonicity between (first differenced) proxy variable and perceived TFP shock. The blue sample points plot estimated $\hat{\Delta} \omega = \hat{\varphi}(z_t) \Delta proxy_t$ vs. $\Delta proxy_t$. The black points are the corresponding quadratic approximations. Clearly there is a monotonic positive correlation between proxy variable and estimated TFP shock. Thus the use of proxy variable shows its merit to control for TFP shock.

Moreover, the correlation between our estimated first differenced perceived TFP shock $\hat{\omega}_t$ and HP-filtered $\ln y_t$ is 0.355 for pooled data and 0.38 for 30 provinces average. This also indicates that the proxy variable is useful to partly demonstrate the shock to the economy.

**Appendix H. A discussion of magnitude of estimated marginal product of capital**

*To be finished.*

We find our estimated $F_k$ mostly lies in the interval of [0.1, 0.3], lower than many other results in literature which often lie in [0.2, 0.4]. Howbeit we do not regard this as a signal of mistake in our estimation. There are five possible explanations for our relatively low estimation results.

1. Land and natural resources
In fact a great deal of values of land and natural resources were usually not considered. Caselli and Feyrer (2007) argue that if we take the land and natural resources into account, the estimated marginal product of capital would be stably close to 0.1 around the world.

(2) Distortion in Chinese economy

Most literature actually estimates the marginal return, rather than the marginal product of capital in China. It is found that the income share of capital in China is significantly higher than most countries. It is possible that the distortions in Chinese economy result in $r > F_k$.

(3) Double counts of firm-level production when the GDP is calculated

(4) TFP drives how much $Y$ in China?

(5) We use first differenced version of production function, which only measures the production in a specific interval of $K$.

Appendix I. Panel unit root and Granger causality tests for regression variables

I.1. Panel unit root test

Table 8 presents the panel unit root test result for variables of output, capital stock and skill-adjusted labor amount. Since we only have sample with a relatively large amount of cross-sections and short time period where heterogenous business cycle would be important, we include an individual intercept and trend in test equation. Lag length is selected automatically based on Schwarz information criterion. We report the Breitung $t$-, IPS (Im, Pesaran and Shin) $W$-, ADF-Fisher $\chi^2$- and PP-Fisher $\chi^2$-Statistic in the table. Intentionally, the LLC (Levin, Lin and Chu) and Hadri test are not relied on because they are not informative for our sample, as LLC almost always rejects the unit root and Hadri always rejects the stationarity null hypothesis for all variables.
Table 8: Panel Unit Root Tests for Regression Variables

<table>
<thead>
<tr>
<th>Variable</th>
<th>Test Statistics</th>
<th>Whether It is Stationary?</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Breitung</td>
<td>IPS</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Levels</td>
<td></td>
<td></td>
</tr>
<tr>
<td>ln Y</td>
<td>3.658</td>
<td>3.238</td>
</tr>
<tr>
<td>ln y</td>
<td>-4.667**</td>
<td>-0.136</td>
</tr>
<tr>
<td>ln K</td>
<td>2.918</td>
<td>0.554</td>
</tr>
<tr>
<td>ln k</td>
<td>3.351</td>
<td>-1.021</td>
</tr>
<tr>
<td>ln L</td>
<td>6.921</td>
<td>2.228</td>
</tr>
<tr>
<td>ln L</td>
<td>0.505</td>
<td>-1.022</td>
</tr>
<tr>
<td>First Differences</td>
<td></td>
<td></td>
</tr>
<tr>
<td>∆ ln Y</td>
<td>1.989</td>
<td>-0.718</td>
</tr>
<tr>
<td>∆ ln y</td>
<td>-2.166*</td>
<td>-2.301*</td>
</tr>
<tr>
<td>∆ ln K</td>
<td>2.287</td>
<td>-0.417</td>
</tr>
<tr>
<td>∆ ln k</td>
<td>-0.979</td>
<td>-1.503*</td>
</tr>
<tr>
<td>∆ ln K</td>
<td>5.205</td>
<td>0.258</td>
</tr>
<tr>
<td>∆ ln k</td>
<td>2.339</td>
<td>-1.515*</td>
</tr>
<tr>
<td>∆ ln L</td>
<td>-2.580**</td>
<td>-2.735**</td>
</tr>
<tr>
<td>Second Differences</td>
<td></td>
<td></td>
</tr>
<tr>
<td>∆(∆ ln Y)</td>
<td>-7.150**</td>
<td>-0.955</td>
</tr>
<tr>
<td>∆(∆ ln y)</td>
<td>-2.757**</td>
<td>-2.641**</td>
</tr>
<tr>
<td>∆(∆ ln K)</td>
<td>-4.491**</td>
<td>-0.577</td>
</tr>
<tr>
<td>∆(∆ ln k)</td>
<td>-3.082**</td>
<td>-2.863**</td>
</tr>
<tr>
<td>∆(∆ ln K)</td>
<td>-2.099*</td>
<td>-1.361*</td>
</tr>
<tr>
<td>∆(∆ ln k)</td>
<td>-2.591**</td>
<td>-4.222**</td>
</tr>
<tr>
<td>∆(∆ ln L)</td>
<td>-1.966*</td>
<td>-3.088**</td>
</tr>
</tbody>
</table>

Note: (i) */**/*** indicate the rejection of the null hypothesis of unit root (Breitung, IPS, Fisher-ADF, Fisher-PP) at the 10%/5%/1% significance level, respectively. (ii) In column “based on 4 tests”, “Mixed” refers to mixed evidence in the case that only 2 out of 4 tests reject unit root hypothesis. (iii) Estimations are conducted by Eviews 8.0.

Furthermore, we report whether the variables are stationary based on the test statistics at 5% and 10% significance level. Considering that sometimes different tests indicate distinct consequences, the column “based on 4 tests” refers to the judgement depending on whether at least 3 tests reject/unreject unit root hypothesis. We also report the judgement using the Fisher-ADF statistic in the column “based on Fisher-ADF”, as this test is most often used in literature.

A careful reading of Table 8 finds that the variables of output, capital stock and labor are integrated of different orders. Especially for the aggregate (non-per capita) capital stocks ln \( K_g \) and ln \( K \), stationarity demands second order differences. Thus we can hardly implement panel
cointegration tests (and not to mention cointegrating regression such as Fully Modified OLS or Dynamic OLS) on these variables, as long as we want to reserve economic meaning of regression equation. Next we turn to the possible Granger causality test.

I.2. Granger causality test

The first differenced series of per capita variables $\Delta \ln y$, $\Delta \ln k_g$ and $\Delta \ln k$ are stationary. We check the stacked test (assuming common coefficients) and Dumitrescu Hurlin test (assuming individual coefficients) and report results as below in Table 9. Since AIC (Akaike information criterion), SC (Schwarz information criterion) and LR (likelihood ratio) test statistic indicate different lag length selection which varies between 1 and 4, we document the outcomes with lag 1-4 for Granger causality test. Because of small sample, we are only able to report the Dumitrescu Hurlin panel causality test with lag 1.

<table>
<thead>
<tr>
<th>Null Hypothesis:</th>
<th>Dumitrescu Hurlin Panel Causality Test</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>F-Statistic</td>
</tr>
<tr>
<td>$\Delta \ln y$ does not Granger Cause $\Delta \ln k_g$</td>
<td>105.786**</td>
</tr>
<tr>
<td>$\Delta \ln k_g$ does not Granger Cause $\Delta \ln y$</td>
<td>21.196**</td>
</tr>
<tr>
<td>$\Delta \ln y$ does not Granger Cause $\Delta \ln k$</td>
<td>160.301**</td>
</tr>
<tr>
<td>$\Delta \ln k$ does not Granger Cause $\Delta \ln y$</td>
<td>7.219**</td>
</tr>
<tr>
<td>$\Delta \ln k_g$ does not Granger Cause $\Delta \ln k$</td>
<td>0.563</td>
</tr>
<tr>
<td>$\Delta \ln k$ does not Granger Cause $\Delta \ln k_g$</td>
<td>0.132</td>
</tr>
</tbody>
</table>

Note: (i) */**/*** indicate the rejection of the null hypothesis at the 10%/5%/1% significance level, respectively. (ii) To be precise, the null hypothesis of Dumitrescu Hurlin Panel Causality Test is “X does not homogeneously cause Y” i.e. “there is no causal relationship for all the cross-units of the panel”; the alternative hypothesis is “there is a causal relationship from X to Y at least for one cross-unit”. (iii) Estimations are conducted by Eviews 8.0.

Although as well known that Granger causality does not necessarily imply explicit causality in economic sense. We can, more or less, obtain some information. (i) First, we see from the table that the growth of output always increases the stock of capitals. This is not surprising. On the other hand, this indicates the importance of reverse causality effect when we intend to analyze the output effect of capital stocks. (2) While the growth of public capital $k_g$ and private capital $k$ potentially enhance GDP increase, the subtle interaction between $\Delta \ln k_g$ and $\Delta \ln k$ should be noticed. We think the influence might happen in the way that the amount of one production input changes the productivity of another production factor.